PUBLIC HEALTH BENEFITS AND HARMS OF BREAST CANCER SCREENING

Microsimulations informing screening recommendations

JEROEN JOS VAN DEN BROEK

Public Health Benefits and Harms of Breast Cancer Screening

Microsimulations informing screening recommendations

Jeroen Jos van den Broek

Public Health Benefits and Harms of Breast Cancer Screening:
Microsimulations informing screening recommendations

Doctoral thesis, Erasmus University Rotterdam, the Netherlands

ISBN 978-94-6361-305-7
Lay-out and printing by Optima Grafische Communicatie

Financial support for printing this thesis was kindly provided by the Department of Public Health of the Erasmus Medical Center

Public Health Benefits and Harms of Breast Cancer Screening

Microsimulations informing screening recommendations

Voor- en nadelen van borstkankerscreening voor de volksgezondheid

Microsimulaties ter ondersteuning van screeningsaanbevelingen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. R.C.M.E. Engels
en volgens het besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
dinsdag 17 september 2019 om 15.30 uur

door

Jeroen Jos van den Broek
geboren te Eindhoven

Erasmus University Rotterdam

PROMOTIECOMMISSIE

Promotor
Prof. dr. H.J. de Koning
Overige leden
Prof. dr. P.J.E. Bindels
Prof. dr. R.M. Pijnappel
Prof. dr. M.A. Joore
Copromotor
Dr. N.T. van Ravesteyn

CONTENTS

Chapter 1	Introduction	7
Part One: Breast cancer microsimulation: model, methods, comparison, and validation.		
Chapter 2	Simulating the impact of risk based screening and treatment on breast cancer outcomes with MISCAN-Fadia.	29
Chapter 3	Modeling ductal carcinoma in situ (DCIS) – an overview of CISNET model approaches.	51
Chapter 4	Comparing CISNET Breast Cancer Incidence and Mortality Predictions to Observed Clinical Trial Results of Mammography Screening from Ages 40 to 49.	71
Chapter 5	Comparing CISNET Breast Cancer Models Using the Maximum Clinical Incidence Reduction Methodology.	91
Part Two: Quantifying the harms and benefits of age-based breast cancer screening in the United States		
Chapter 6	Association of Screening and Treatment With Breast Cancer Mortality by Molecular Subtype in US Women	117
Chapter 7	Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies.	165
Chapter 8	Radiation-Induced Breast Cancer Incidence and Mortality From Digital Mammography Screening: A Modeling Study.	191
Part Three: Projecting the harms and benefits of risk-based breast cancer screening in the United States		
Chapter 9	Tailoring Breast Cancer Screening Intervals by Breast Density and Risk for Women Aged 50 Years or Older: Collaborative Modeling of Screening Outcomes.	227
Chapter 10	Personalizing Breast Cancer Screening Based on Family History & Polygenic Risk.	255
Chapter 11	Discussion	283
	Conclusions	299
	Summary (EN)	303
	Samenvatting (NL)	309
	Curriculum Vitae	315
	List of Publications	316
	Phd Portfolio	318
	Acknowledgements	320

Chapter 1

Introduction

1. What is breast cancer
2. Etiology and risk-factors of breast cancer
3. Breast cancer epidemiology
4. Primary prevention of breast cancer
5. Secondary prevention of breast cancer
6. Breast cancer treatment
7. Evidence on breast cancer screening
8. Current breast cancer screening guidelines
9. Moving towards risk-based breast cancer screening
10. The use of models next to randomized controlled trials
11. Research questions and thesis outline

1. WHAT IS BREAST CANCER

Cancer is the uncontrolled growth of cells into a malignant tumor. Breast cancer usually begins in the lobules, ducts, or connective tissue of the breast. The lobules are the glands that produce milk in nursing women. The ducts are thin tubes that drain milk from the lobules to the nipple. The connective tissue, consisting of fibrous and fatty tissue holds everything together. Most breast cancers begin in the ducts called ductal carcinoma in situ (DCIS) or, less common, in the lobules (lobular carcinoma in situ). Non-invasive cancers are confined to the milk ducts or lobules in the breast and do not evade into normal tissues. The non-invasive cancers may be pre-cancer and are sometimes called stage-0 breast cancer. Breast cancers become invasive when they grow into healthy tissue and can eventually spread outside the breast (metastasize) to other parts in the body through blood vessels and lymph vessels. Breast cancer diagnosed at an early stage when it has not spread, is more likely to be treated successfully. Vice versa, women's chances of surviving breast cancer are much lower when the cancer has spread throughout the body and effective treatment becomes increasingly difficult.(1)

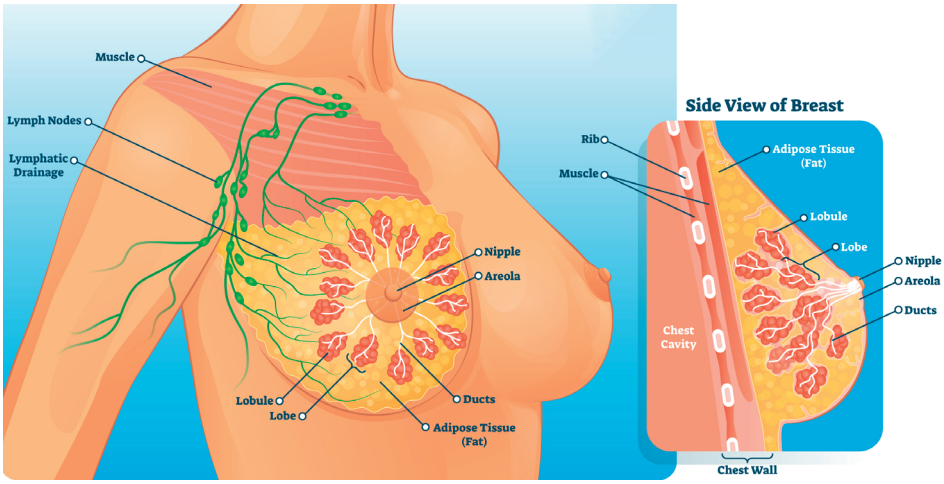


Figure 1 Anatomy of the female breast

Breast cancer staging

Breast cancer staging is used by doctors, hospitals, and others to characterize breast cancer upon diagnosis. Staging describes where the cancer is present in the body in relation to the primary tumor and in particular whether, and to what extent the cancer has spread. Staging is useful for guiding the treatment strategy and assessing the prognosis of the cancer. A widely used staging system for cancer is the tumor, node, metastasis (TNM) system.(2) The T refers to the size of the primary tumor from which the cancer

originates. The number of nearby lymph nodes involved is indicated with N. The M refers to metastasis of cancer and indicates whether the cancer has spread from the primary tumor to other parts in the body.

A similar staging system used by the Surveillance, Epidemiology, and End Results (SEER) program is the local-regional-distant system. In situ; abnormal cells, which may be a precursor of cancer, are present but have not spread to nearby tissue. Localized; cancer is present, but only in the organ where it started. Regional; the cancer has spread to nearby lymph nodes or organs. Distant; the cancer has spread from the place of the primary tumor to distant parts of the body.

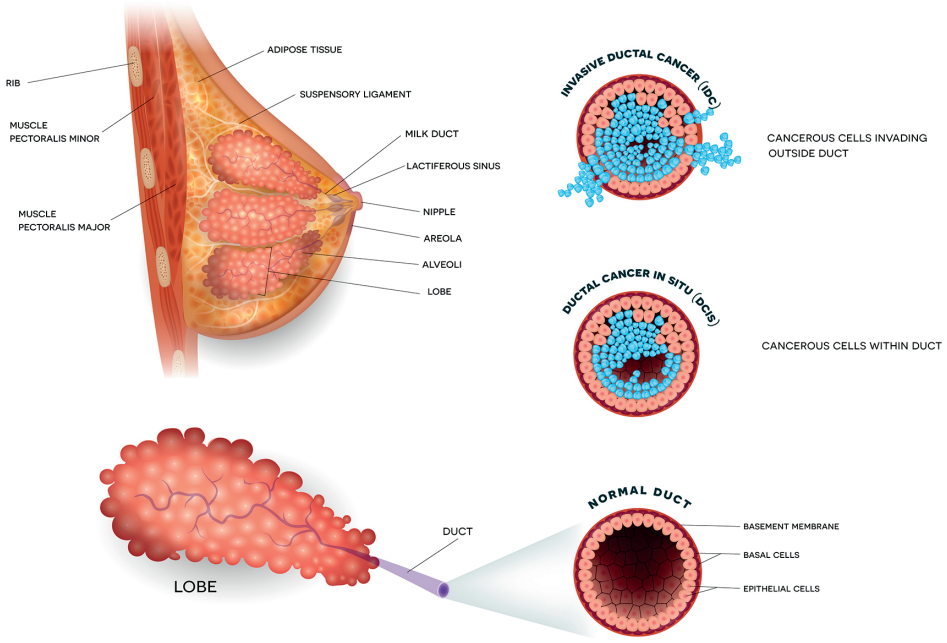


Figure 2 Ductal Carcinoma in Situ (DCIS) – non-invaiseve or pre-invasive breast cancer.

2. ETIOLOGY AND RISK-FACTORS OF BREAST CANCER

Research has identified hormonal, lifestyle, environmental and genetic factors that may increase the risk of developing breast cancer. (3) Breast cancer is likely caused by a complex interaction of genetic makeup and environment. While there are known risk factors, many women who develop breast cancer have no evident risk factors other than being women and in the age range of 50-74 when breast cancer incidence is the highest. As women get older, there are more opportunities for genetic damage in the breast and

Table 1 Overview of major and minor risk-factors of breast cancer.(3)

Breast cancer risk factors	Relative risk	Reference population
Personal information		
Age	20-30	Breast cancer at age 20 vs. at age 70
Body Mass Index	2	Obesity (BMI>30) vs. no obesity
Alcohol consumption	1.28	4 glasses containing alcohol vs. none
Breast density	4-6	Extremely dense vs. fatty breast
Hormonal / reproductive risk factors		
Age of first menarche	1.5	Before age 10 vs. after age 16
Age of menopause	2	After age 55 vs. before age 40
Age of first live birth	3	After age 35 vs. before age 19
Breast feeding	0.8	More than 4 years vs. No breast feeding
Use of hormonal replacement therapy	2	10 years usage vs. never
Family history of breast cancer		
First degree family history of breast cancer	3.6	2 first degree with breast cancer vs. none
Second degree family history of breast cancer	1.5	Second degree with breast cancer vs. none
Age of breast cancer onset	3	Onset before age 50 in sister vs. none
Ovarian cancer	1.5	Ovarian cancer in family vs. none
Personal history with breast cancer		
Atypical ductal hyperplasia	4	Ductal hyperplasia vs. no hyperplasia
Previous breast biopsy	2	No previous breast biopsy
Lobular carcinoma in situ (LCIS)	4	LCIS vs. no LCIS
Genetic breast cancer risk		
Single Nucleotide Polymorphisms	10	Top 1% vs. bottom 1% based on 77 SNPs
Mutations in BRCA 1/2	15	Mutation in BRCA genes vs. no mutation

the entire body. At the same time, the human body becomes less capable of repairing genetic damage that may cause cancer.

A previous breast biopsy, dense breasts, and a positive family history of breast cancer are strong risk factors for breast cancer. Inherited cases of breast cancer are often associated with mutations in genes BRCA1, BRCA2, ATM, CHEK2, and PALB2 which are known to increase breast cancer risk by a large factor.(4) Minor risk factors include reproductive factors such as low parity, and young age at first menarche which expose women to female hormones estrogen and progesterone that are linked to breast cancer onset and growth.(3) Breast cancer single nucleotide polymorphisms (SNPs) are common variations in the DNA sequence associated with small increases or decreases in breast cancer risk.(5) Polygenic risk combines information from multiple SNPs and could potentially achieve a degree of risk discrimination useful for population screening and be suitable to stratify risk in women of all ages.(6) Several other risk factors are related to personal behaviors, such as lack of exercise, alcohol consumption, smoking, and an unhealthy diet. While

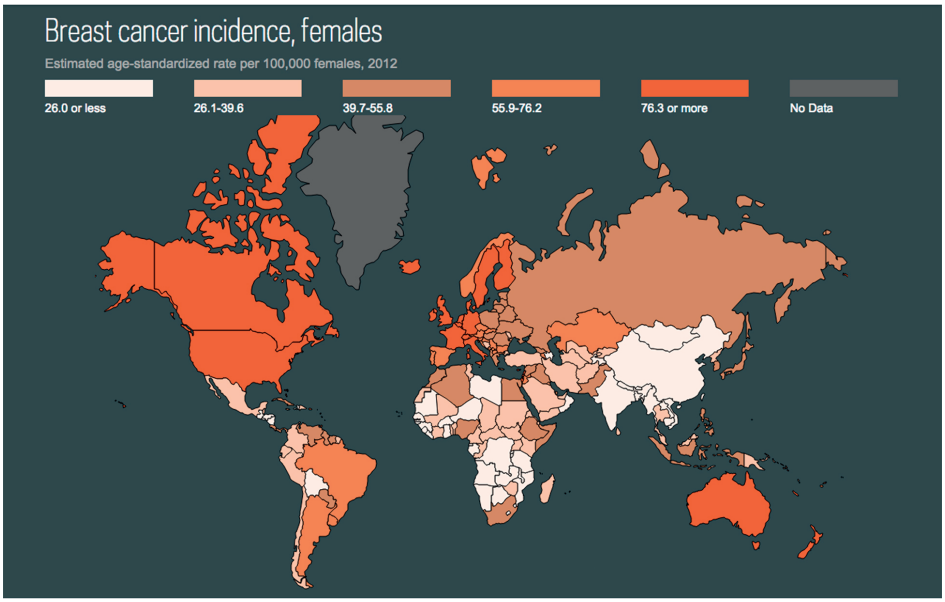


Figure 3 Worldwilde female breast cancer incidence in 2012. All incidence rates are age-standardized to the 1960 world population. Source: Ferlay J. Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No 11.

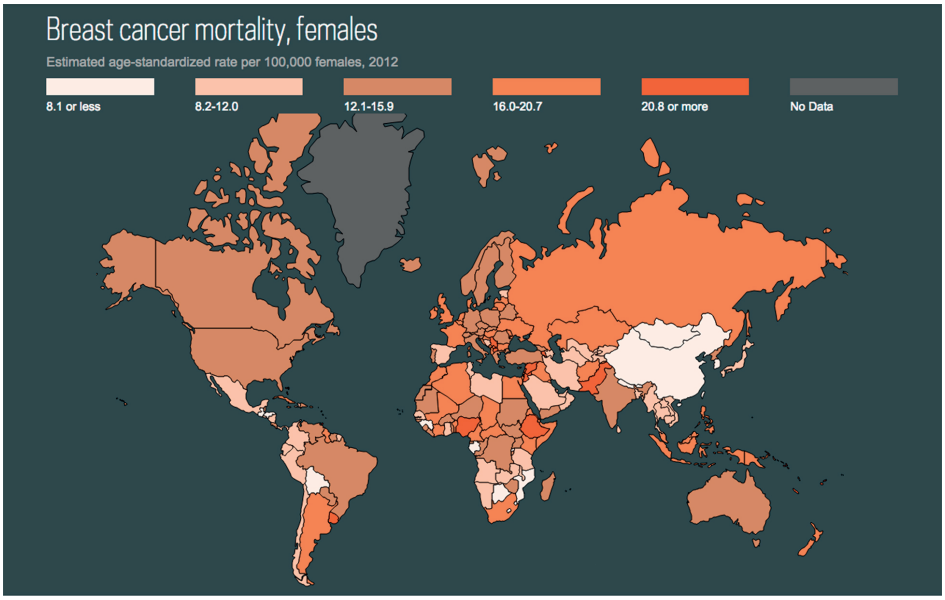


Figure 4 Worldwilde female breast cancer mortality in 2012. All mortality rates are age-standardized to the 1960 world population. Source: Ferlay J. Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No 11.

many known factors increase the risk of developing breast cancer, a large part of breast cancers are due to random, unpredictable, mistakes in DNA copying which is essential for cell division and life itself.

3. BREAST CANCER EPIDEMIOLOGY

Breast cancer incidence worldwide

Breast cancer is a major health problem with an estimated 2.1 million new cases and 0.63 million breast cancer deaths worldwide in 2018. (Figure 1, 2) In many developed countries around 1 in 8 (13%) women are diagnosed with breast cancer in their lifetime. (7)

Age-specific breast cancer incidence

Breast cancer correlates strongly with age regardless of race or ethnicity. At age 50 around which most women start screening, 200 cases per 100,000 women are observed in the United States. The peak in incidence lies between ages 70 and 80. This age-specific pattern is seen in most western countries.

Breast cancer incidence over time

Invasive breast cancer incidence has seen a sharp increase in the United States up to the year 2000. (Figure 5) After 2000, there was a drop in incidence up to 2003 which was followed by a period of relatively stable incidence levels. These changes over time have been attributed to the increase in use- and performance of mammography, changes in hormone use after 2000, risk factor prevalence, and differential birth cohort effects. The use of Hormone Replacement Therapy (HRT) was reduced in 2000-2003 as it became apparent at the time that it was associated with increased risk of breast cancer.(8, 9) This led to a decrease in breast cancer incidence up to 2003.

Breast cancer mortality

Breast cancer mortality was relatively stable up to the mid-nineties of the 20th century and gradually declined thereafter. The decrease in breast cancer mortality has been attributed to the increase in screening, better access to healthcare, and advances in breast cancer.(10, 11)

Breast cancer survival

Survival rates are an estimate of the percentage of patients who survive for a given period of time after a cancer diagnosis. Relative breast cancer survival compares survival among women with breast cancer to women of the same age without breast cancer.

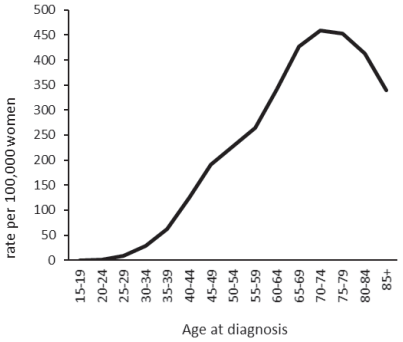


Figure 5 U.S. BC incidence by age, 2011-2015.

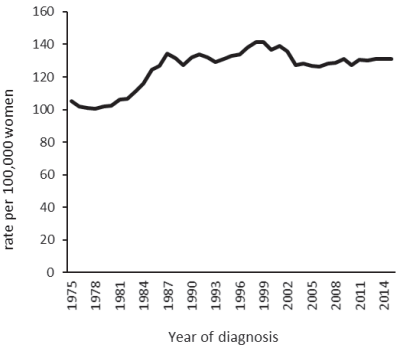


Figure 6 U.S. BC incidence - age-adjusted

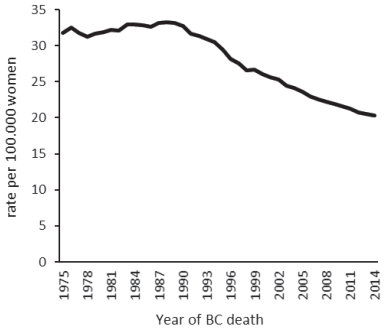


Figure 7 U.S. BC mortality, age-standardized '75-'13

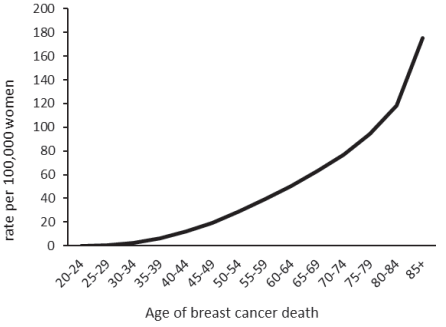


Figure 8 U.S. age-specific BC mortality '11-'15

Based on data over a 14-year period from 2000 to 2014, the 10-year survival rate for U.S. women diagnosed with breast cancer was 83.3% and varied strongly by stage at diagnosis. (Figure 8)

4. PRIMARY PREVENTION OF BREAST CANCER

Primary prevention aims to prevent disease before it begins. This is typically done by changing unhealthy behavior or prevent exposure to hazardous chemicals or situations. In breast cancer, the modifiable risk factors include postmenopausal obesity, alcohol consumption, physical inactivity, and exposure to radiation. A healthy bodyweight, balanced diet and regular physical activity reduce breast cancer risk and improve general health as well. A balanced diet is one that consists of sufficient fruit, fibers, vegetables, healthy fats, proteins and preferably no or little red or processed meat and added salt. In a proper diet the total caloric intake should maintain a healthy body mass index to prevent obesity. Physical activity should ideally be at least 30 minutes of walking, biking

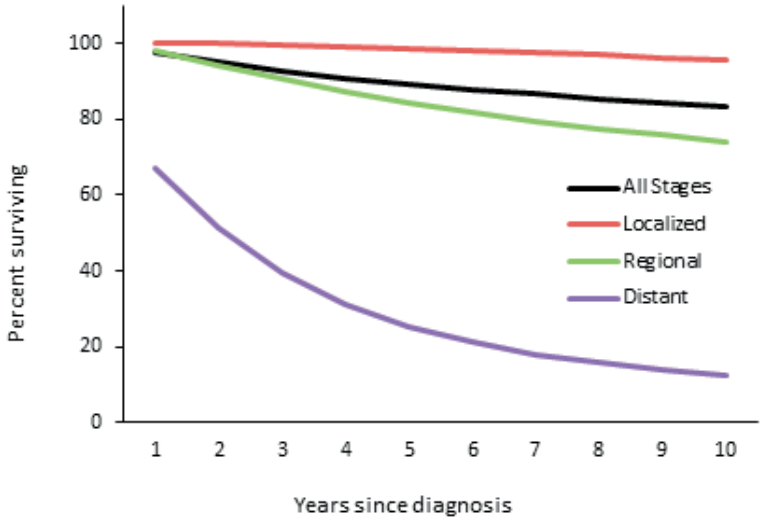


Figure 9 U.S. Survival Rates by Time Since Breast Cancer Diagnosis, 2000-2014.

or other sports according to the world cancer research fund. Further, primary prevention among high risk women may entail the use of medications that modulate estrogen receptors such as tamoxifen and raloxifene.

5. SECONDARY PREVENTION OF BREAST CANCER

Screening aims at finding breast cancer in early stages of the disease when tumors are less likely to have spread in the body. Screening can find in healthy, asymptomatic women in multiple different ways. For example, breast self-examination is a screening technique which allows women to examine their breast tissue at home for any physical or visual changes. More modern screening techniques include the use of digital mammography, ultrasound, magnetic resonance imaging (MRI), or Tomosynthesis. Mammography is an X-ray image taken of the breasts called a mammogram which has relatively high sensitivity and specificity. (12) Mammograms and other medical imaging techniques, allow radiologists to look for changes in breast tissue that could be pre-cursor, or early stage breast cancer.

Benefits of screening

True positives screening outcomes correctly identify abnormalities in the breast as cancer. True negatives correctly provide reassurance when no cancer is present in the breast. Chances of successful treatment and survival are higher for breast cancer diagnosed at

an early (localized) stage. Screening increases the number of early stage breast cancer and thereby improves breast cancer survival of the majority of screen-detected cancers. Next to life years gained, averting breast cancer deaths is an important goal of screening. In the absence of screening, more cancers are diagnosed at a more advanced stage of breast cancer. Consequently, more advanced treatment is necessary and if the cancer is lethal, life years are lost or quality of life is significantly reduced. Overall, regular screening at the population level provides large benefits for a small number of women, and harms among the majority of women who undergo screening but never develop breast cancer.

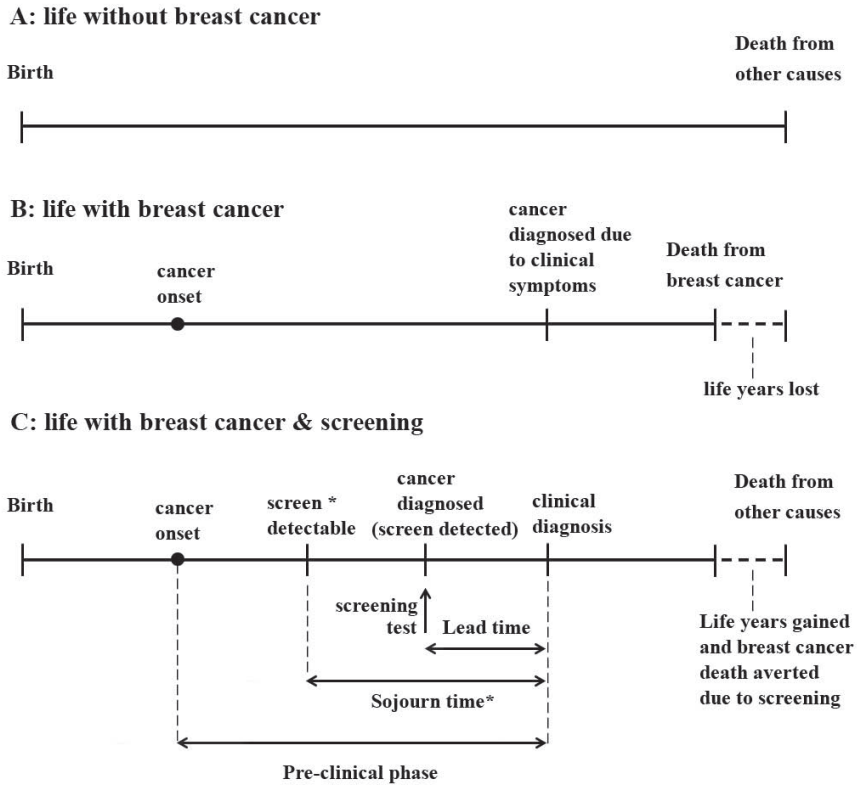


Figure 10 Three possible life-history scenarios. A: women without breast cancer, B:women with breast cancer who are not screened, C: women with breast cancer who are screened. In scenario C, the pre-clinical phase is the period of time between tumor inception and clinical diagnosis in the absence of screening. The *sojourn time* for a screening test, e.g., mammography, is the period of time within the pre-clinical phase that a cancer can be screen-detectable; this period can also be termed the *pre-clinical screen-detectable phase*. The point when the cancer is detected by screening depends on when the screening test is performed and the sensitivity of the screening test. The period before the sojourn time represents a period in which the tumor is present but undetectable by mammography. Should the sensitivity of mammography improve, or new types of screening tests evolve, the point of screen-detectability would be closer to tumor inception.

harms of screening

On mammograms, tissue may show up that looks like breast cancer, but may in fact be benign (non-cancerous) tissue. If the abnormalities are flagged as breast cancer and additional imaging shows that there is no cancer, this is called a false positive screening that may cause unnecessary anxiety and distress. One other important harm of breast cancer screening is over diagnosis. Overdiagnosis is the diagnosis of breast cancer by screening that would never have caused symptoms and be diagnosed in the absence of screening in a woman's lifetime. Besides false positives and overdiagnoses, false negative screening outcomes can also be harmful. False negatives may provide a sense of false reassurance while in fact cancer is growing in the breast. Lastly, regular screening increased the overall exposure to ionizing radiation and could lead to radiation-induced breast cancer in some cases.

Quality of life

Through screening, cancer diagnoses are advanced in time and in the majority of cases treatment can be less invasive and still be curative. In general, this results in a better quality of life for women who are diagnosed with breast cancer. For the majority of women who will never be diagnosed with breast cancer, mammography screening involves planning, travel, and waiting time. Before the actual mammogram, women may feel anxious or worry about the possible abnormal outcomes of the screening. Undergoing screening means that women have to undress from the waist up and may feel pain, pressure and discomfort in their breasts from the mammogram. After the examination, it takes some time before women are notified about the outcomes of the screening. This waiting period could be experienced as uncertain and stressful, but may be worth the reassurance, be it early diagnosis of breast cancer. Because women differ in their willingness to accept the harms of screening for potential benefits, a personal consideration is advised before attending screening.

6. BREAST CANCER TREATMENT

The majority of breast cancers will eventually metastasize without treatment. To prevent breast cancer death after diagnosis, the tumor is surgically removed and the patient usually receives adjuvant treatment to help decrease the risk of breast cancer recurring. Effective adjuvant treatments are commonly called systemic treatment and include: radiation, chemotherapy, and hormone therapy. There are additional supplemental treatments which might increase the effectiveness of these three treatments, but chemical, radiation, and hormonal treatments are the first ones considered to successfully treat breast cancer.(13)

If breast cancer is contained in the breast regions, localized treatment is considered. To help prevent local recurrence, a surgeon will try to remove the tumor, possibly with surrounding tissue, and treat the patient with radiation. The molecular nature of the tumor may also determine whether chemo- and/or hormonal therapy is used. Systemic treatment comes into play when breast cancer has spread or metastasized to the lymph nodes. In this stage of breast cancer, surgery alone is not curative anymore and systemic therapies are considered. Neoadjuvant breast cancer treatment is applied before surgical intervention aiming to stop the cancer growth and shrink the tumor size before surgical intervention.(14)

In the past, radical mastectomy of the breast was much more common. This involved surgery to remove the entire breast including the axillary lymph nodes and chest wall. Today, this medical procedure is less common and lumpectomy, i.e., breast conserving surgery, is more common. Lumpectomy aims to remove the cancer while preserving as much of the normal breast as possible.

7. EVIDENCE ON BREAST CANCER SCREENING

Large randomized trials have been introduced in 1960's and '70's and conducted throughout to the early 2000's. These include the New York Health Insurance Plan (HIP) (15), Malmö I and II (16), Swedish two county trial(17), Canada I and II (18), Göteborg (19), Stockholm(20), and the UK age trial(21). These trials compared breast cancer incidence and mortality among women invited to screening to women not invited to screening. While most studies found a reduction in breast cancer mortality from screening, controversy about the harms of breast cancer screening remains. In 2013, an independent panel extensively reviewed published work about the evidence on breast cancer screening to reach conclusions about the benefits and harms.(22) They found that 43 breast cancer deaths are prevented and 129 cases are overdiagnosed per 10,000 women screened triennially for 20 years from age 50 onwards in the UK.

In 2014, the International Agency for Research on Cancer (IARC) convened 29 independent experts from 16 countries to review the scientific evidence of various methods of screening for breast cancer.(23) The IARC concludes that women in the age range of 50 to 69 invited to mammography screening have a 23% breast cancer mortality reduction. Older women, in age ranges 70-74 also observed a substantial reduction in risk of breast cancer death. The reduction in risk of breast cancer death in studies among women aged 40 to 49 was less pronounced. Estimates of the cumulative risk of false positive results differ between organized programs and opportunistic screening. The cumulative risk of having at least one false-positive is about 20% for a woman who had 10 screens between the ages of 50 and 70 years. Overdiagnosis was estimated to be in

the range of %1 to 10% of all breast cancer diagnoses, with point estimate of 6.5% based on data from European studies that adjusted for both lead time and trends in incidence between screened and unscreened women.

8. CURRENT BREAST CANCER SCREENING GUIDELINES

Breast cancer screening guidelines recommending who should undergo screening, how often and at what ages vary within and among developed countries. The United States Preventive Services Task Force (USPSTF) 2016 guidelines recommend that women aged 50 to 74 years of age be screened with digital mammography every two years. According to the USPSTF, screening before age 50 is an individual decision women should make including their values about the (possible) harms and benefits of screening and attitude towards breast cancer risk.(24)

The American Cancer Society (ACS) recommends that women between ages 40 and 45 should have the choice to be screened based on their own considerations. Women between ages 45 and 54 are recommended to undergo annual mammography, followed by biennial screening between ages 55 and 74.(25) The International Agency for Research on Cancer (IARC), part of the World Health Organization (WHO), recommends women aged 50 to 69 to be screened and is next to the USPSTF one of the least intensive screening guidelines.(23) Overall, these guidelines agree that women aged 50 to 69 should be screened and vary to some extent in screening initiation and stopping age and screening interval.

9. MOVING TOWARDS RISK-BASED BREAST CANCER SCREENING

Historically, breast cancer screening guidelines have been age-based even though we know that at any given age there is variability in breast cancer risk due to earlier mentioned risk factors. By better understanding which women are at increased or decreased breast cancer risk, risk stratification can target screening to those who are most likely to benefit from different screening strategies than currently recommended. This could individualize breast cancer care and potentially reduce the population-level harms of screening and increase the benefits. Projections for groups of women differing in risk due to family history, breast density, polygenic risk, and other risk factors have been made under various screening and treatment interventions by breast cancer simulation models in the chapters of this thesis.

10. THE USE OF MODELS NEXT TO RANDOMIZED CONTROLLED TRIALS

Randomized clinical trials (RCT) are considered the gold standard to assess the effectiveness of breast cancer screening and treatment interventions. However, there are several reasons why modeling is essential to complement and extend the evidence from randomized trials. First, RCTs to assess screening and treatment interventions with cause of death as primary outcome are time consuming and relatively expensive to set up. Second, lifetime follow-up is difficult logistically as participants may move abroad, are lost to follow-up, or decide to stop their participation. Consequently, the long-term benefits and harms of medical interventions such as screening are difficult to assess. Third, trials are usually set up to evaluate a limited number of interventions. In screening this would be different starting ages, intervals, and treatment regimens. Fourth, in RCTs ethical concerns have to be taken into account. If routine screening of healthy women is part of usual practice, it could be unethical to include a non-screening (control) group in the trial that is at increased risk of late stage cancer. Finally, trials usually provide outcomes in a specific setting, for a specific group of people in a certain region with screening and treatment methods available at that time. We know screening and treatments methods have improved since the large mammography trials and are likely to have a different impact on breast cancer detection and breast cancer mortality. Simulation models can synthesize data on breast cancer epidemiology, population demographics, screening accuracy, and treatment effectiveness from different sources and produce outcomes for multiple screening and treatment strategies among varying risk groups.

Microsimulation model MISCAN-Fadia

In this thesis, MISCAN-Fadia which is an acronym for Microsimulation Screening Analysis – Fatal Diameter is used to make predictions about breast cancer incidence and mortality following from varying screening and treatment strategies, Chapter 2 of this thesis (26). The model simulates individual life histories from birth to death, with and without breast cancer, in the presence and in the absence of screening and treatment. Life histories are simulated according to discrete events such as birth, tumor inception, metastasis, and death from breast cancer or death from other causes. The model consists of four main components: demography, natural history of breast cancer, screening, and treatment. The impact of screening on the natural history of breast cancer is assessed by simulating continuous tumor growth and the "fatal diameter" concept. This concept implies that tumors diagnosed at a size that is between the screen detection threshold and the fatal diameter are cured, while tumors diagnosed at a diameter larger than the fatal tumor diameter metastasize and lead to breast cancer death.

Collaborative modeling

Erasmus Medical Center part of a collaborative modeling initiative called the cancer intervention and surveillance modeling network (CISNET). We use statistical modeling to improve understanding of cancer control interventions in prevention, screening, and treatment and their effects on population trends in incidence and mortality. Models are used to guide public health research and priorities, and they can aid in the development of optimal cancer control strategies. Collaborative modeling can enhance the rigor of modeling research using multiple independent models to answer the same research question. Conclusions supported by multiple independently developed models provide greater credibility than conclusions obtained from a single model.

11. RESEARCH QUESTIONS AND THESIS OUTLINE

This thesis consists of three main parts: 1. Breast cancer micro-simulation modeling, 2.Quantification of current breast cancer screening practice among average-risk women in the United States. 3. Outcome projections of risk-based screening strategies. This thesis concludes with a discussion of the work in this thesis in relation to the field of breast cancer screening.

PART 1: BREAST CANCER MICROSIMULATION: MODEL, METHODS, COMPARISON, AND VALIDATION

***Research question 1*: How can model description, comparison, and validation contribute to a better understanding of model predictions?**

Chapter 2 provides an overview of the past, current and future applications of breast cancer simulation model MISCAN-FADIA. In chapter 3, different approaches to modeling the natural history ductal carcinoma in situ are compared. Chapter 4 presents an external validation and comparison of CISNET models´ breast cancer incidence and mortality predictions to the observed clinical trial outcomes. Chapter 5 investigates the impact of model structure and model assumptions about tumor onset and progression on predictions of breast cancer incidence and mortality.

PART 2: QUANTIFYING THE HARMS AND BENEFITS OF AGE-BASED BREAST CANCER SCREENING IN THE UNITED STATES.

***Research question 2*: What are the benefits and harms of current age-based breast cancer screening in the United States?**

In chapter 6, the contributions associated with screening and treatment to breast cancer mortality reductions by molecular subtype-specific breast cancer are evaluated. In chapter 7, six simulation models use U.S. national data on incidence, digital mammography performance, treatment effects, and other-cause mortality to evaluate screening outcomes among average risk women. In chapter 8, we estimated the distributions of radiation-induced breast cancer incidence and mortality from digital mammography screening while considering exposure from screening and diagnostic mammography and dose variation among women.

PART 3: PROJECTING THE HARMS AND BENEFITS OF RISK-BASED BREAST CANCER SCREENING IN THE UNITED STATES

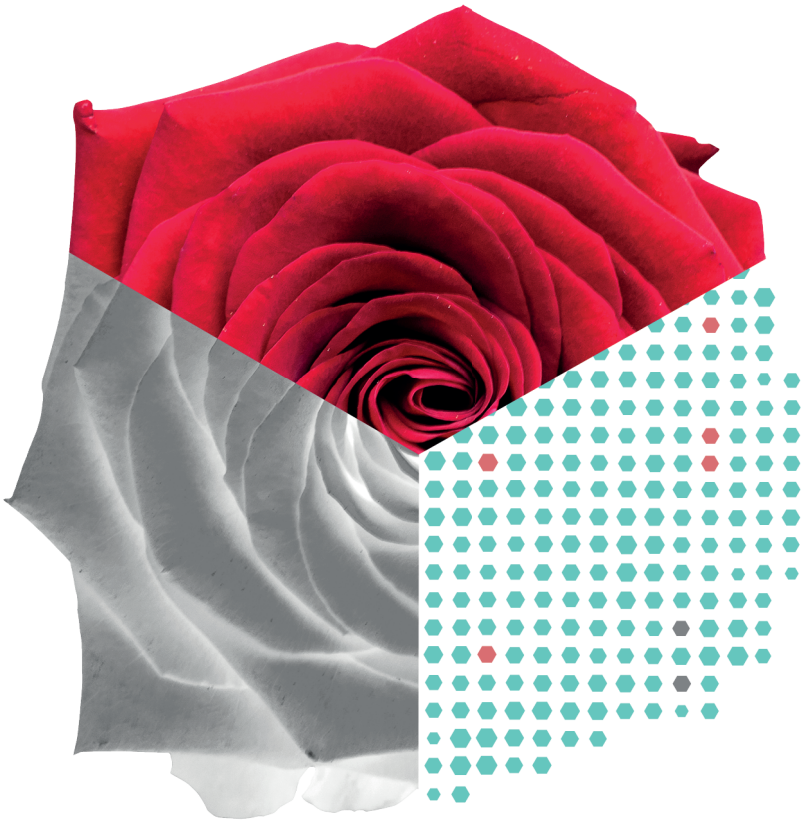
***Research question 3*: To what extent can risk-based breast cancer screening improve the harm-benefit ratio of current age-based screening guidelines?**

In chapter 9, we estimated the outcomes for various screening strategies in the U.S. tailored to women aged 50 years or older with various combinations of breast density and relative risk. Chapter 10 assessed screening approaches using first-degree family history (FH) and polygenic risk scores (PRS) to identify women for risk-based screening.

REFERENCES

1. Guo F, Kuo YF, Shih YCT, Giordano SH, Berenson AB. Trends in breast cancer mortality by stage at diagnosis among young women in the United States. Cancer. 2018;124(17):3500-9.
2. Tio TL. The TNM staging system. Gastrointest Endosc. 1996;43(2 Pt 2):S19-24.
3. Evans DG, Howell A. Can the breast screening appointment be used to provide risk assessment and prevention advice? Breast Cancer Res. 2015;17:84.
4. Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, et al. Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med. 2015;372(23):2243-57.
5. Cuzick J, Brentnall A, Dowsett M. SNPs for breast cancer risk assessment. Oncotarget. 2017;8(59):99211-2.
6. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 2015;107(5).
7. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359-86.
8. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA. 2002;288(3):321-33.
9. Ravdin PM, Cronin KA, Howlader N, Berg CD, Chlebowski RT, Feuer EJ, et al. The decrease in breast-cancer incidence in 2003 in the United States. N Engl J Med. 2007;356(16):1670-4.
10. Jatoi I, Miller AB. Why is breast-cancer mortality declining? Lancet Oncol. 2003;4(4):251-4.
11. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005;353(17):1784-92.
12. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology. 2008;246(2):376-83.
13. Early Breast Cancer Trialists' Collaborative G, Peto R, Davies C, Godwin J, Gray R, Pan HC, 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.
14. Steenbruggen TG, van Ramshorst MS, Kok M, Linn SC, Smorenburg CH, Sonke GS. Neoadjuvant Therapy for Breast Cancer: Established Concepts and Emerging Strategies. Drugs. 2017;77(12):1313-36.
15. Shapiro S. Periodic screening for breast cancer: the HIP Randomized Controlled Trial. Health Insurance Plan. J Natl Cancer Inst Monogr. 1997(22):27-30.
16. Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. BMJ. 1988;297(6654):943-8.
17. Tabar L, Vitak B, Chen TH, Yen AM, Cohen A, Tot T, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology. 2011;260(3):658-63.

18. Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. BMJ. 2014;348:g366.
19. Bjurstam N, Bjorneld L, Warwick J, Sala E, Duffy SW, Nystrom L, et al. The Gothenburg Breast Screening Trial. Cancer. 2003;97(10):2387-96.
20. Frisell J, Lidbrink E, Hellstrom L, Rutqvist LE. Followup after 11 years--update of mortality results in the Stockholm mammographic screening trial. Breast Cancer Res Treat. 1997;45(3):263-70.
21. Moss SM, Wale C, Smith R, Evans A, Cuckle H, Duffy SW. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. Lancet Oncol. 2015;16(9):1123-32.
22. Independent U. K. Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet. 2012;380(9855):1778-86.
23. Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, et al. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med. 2015;372(24):2353-8.
24. Siu AL, On behalf of the U. S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2016;164(4):279-96.
25. Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. JAMA. 2015;314(15):1599-614.
26. van den Broek JJ, van Ravesteyn NT, Heijnsdijk EA, de Koning HJ. Simulating the Impact of Risk-Based Screening and Treatment on Breast Cancer Outcomes with MISCAN-Fadia. Med Decis Making. 2018;38(1_suppl):54S-65S.



PART ONE: Breast cancer microsimulation: model, methods, comparison, and validation

Chapter 2

Simulating the impact of risk based screening and treatment on breast cancer outcomes with MISCAN-Fadia.

Jeroen J. van den Broek, Nicolien T. van Ravesteyn, Eveline A. Heijnsdijk, Harry J. de Koning.

Medical Decision Making. 2018;38(1S):54–65.

ABSTRACT

The MISCAN-Fadia microsimulation model uses continuous tumor growth to simulate the natural history of breast cancer and has been used extensively to estimate the impact of screening and adjuvant treatment on breast cancer incidence and mortality trends. The model simulates individual life histories from birth to death, with and without breast cancer, in the presence and in the absence of screening and treatment. Life histories are simulated according to discrete events such as birth, tumor inception, the tumor's clinical diagnosis diameter in the absence of screening, and death from breast cancer or death from other causes. MISCAN-Fadia consists of four main components: demography, natural history of breast cancer, screening, and treatment. Screening impact on the natural history of breast cancer is assessed by simulating continuous tumor growth and the "fatal diameter" concept. This concept implies that tumors diagnosed at a size that is between the screen detection threshold and the fatal diameter are cured, while tumors diagnosed at a diameter larger than the fatal tumor diameter metastasize and lead to breast cancer death. MISCAN-Fadia has been extended by including a different natural history for molecular subtypes based on a tumor's estrogen receptor (ER) status and human epidermal growth factor receptor 2 (HER-2) status. In addition, personalized screening strategies that target women based on their risk such as breast density have been incorporated into the model. This personalized approach to screening will continue to develop in light of potential polygenic risk stratification possibilities and new screening modalities.

INTRODUCTION

Randomized trials are considered the gold standard to assess the efficacy of cancer screening interventions. However, ethical concerns, participants lost to follow-up, feasibility issues regarding the number of evaluated screening strategies, and limited quantification abilities of the harms of screening such as overdiagnosis, emphasize the need for ways to complement randomized trials. The breast cancer models of the Cancer Intervention and Surveillance Modeling Network (CISNET) simulate the effects of screening and treatment for lifetime follow up, with varying compliance rates, for an unlimited number of screening strategies, and thereby extrapolate the findings from randomized trials.

MISCAN-Fadia, acronym for Micro Simulation Screening Analysis – Fatal Diameter, has been part of CISNET since its start in 2000, usually referred to as Model E (i.e., Erasmus Medical Center). Before the development of MISCAN-Fadia, a microsimulation model with discrete tumor progression was developed at Erasmus already in the 1980's to evaluate the effects of breast cancer screening in the Netherlands [1]. However, compared to observed stage distribution data, the model over-estimated the number of early-stage cancers diagnosed at subsequent screens. Sensitivity analysis of screening sensitivity did not lead to better estimates [2]. Moreover, it was difficult to explore different natural history assumptions because tumor progression was directly linked to discrete stages. MISCAN-Fadia, with continuous tumor growth, was initiated to overcome this rigid property. This model was developed with the intent of creating a more biologically oriented breast cancer model to evaluate the impact of screening and treatment on breast cancer incidence and mortality. Since tumor size is measurable and tumor growth is continuous, these properties form the biological approach to modeling the natural history of breast cancer. In the model, a distinction is made between tumor biology (growth function) and other model variables that are more likely to vary by calendar year and possibly differ between geographical areas such as access to screening facilities, screening equipment and consequently screening test sensitivity, clinical diagnosis in the absence of screening due to fewer breast self-examinations and less public awareness of breast cancer risk. Sensitivity of a screening test is translated into a diameter size at which tumors become screen detectable. In MISCAN-Fadia, ductal carcinoma in situ (DCIS) as well as invasive tumors are simulated. Tumor properties like exponential growth rate, clinical diagnosis diameter, minimal diameter for screen detection and fatal diameter are drawn from probability distributions to account for variability between tumors. The fatal diameter concept implies that available treatment only cures tumors that are diagnosed at a smaller diameter than the tumor's fatal diameter. Available treatment options are not sufficient for tumors diagnosed past their fatal diameter and these tumors will cause breast cancer death.

Disease processes such as the moment of onset of breast cancer and progression or regression of DCIS and breast cancer are unobservable in reality. These are nonetheless important determinants that influence the balance of harms and benefits of screening and treatment. Modeling allows us to explore the effect of changing one of these unobservable factors on modelled outcomes such as breast cancer incidence and mortality. Likewise, it is possible to study the effect of changing tumor onset and tumor growth while keeping all other parameters unchanged to gain insight into the natural history of breast cancer and its interaction with cancer control interventions. To quantify the harms and benefits of different screening and treatment strategies, the model simulates the same female population twice. First, a population is simulated in the absence of screening, and second, in the presence of screening. Key outcomes such as the number of breast cancers, the number breast cancer deaths and over diagnosed breast cancers can be calculated for lifetime follow-up for any possible screening strategy.

Population demography, natural history of breast cancer, screening and treatment are the four main parts of the model. All model inputs and model parameters belong to one of these components and are either calibrated to data from trials or are based on empirical research [3-5]. This paper presents the current model status and in particular the progress and extensions with respect to the first model paper [6], as well as the latest model applications that explore the possibilities of risk-based breast cancer screening.

METHODS

Discrete event-driven microsimulation

Discrete event simulation implies that the model moves from the time of one event (e.g., birth) to the next event (e.g., tumor onset). The events in a woman's lifetime are discrete and mutually exclusive. Microsimulation modeling entails simulation of independent life histories that can be aggregated to estimate the effects of screening and treatment at the population level. Life histories are simulated according to discrete events such as birth, a possible tumor inception, the diameter of the tumor when it would be clinically diagnosed in the absence of screening, a date of death from other causes, or, for woman with breast cancer, a date of breast cancer death. Events that affect the natural history of breast cancer, such as screening and treatment, are tied to the tumor's continually growing diameter (i.e., screen detection of the tumor may take place from a certain tumor size and treatment may treat tumors successfully up to a certain tumor size). Each woman is simulated from birth and followed until death and time plays an essential role in the order of events in a woman's life.

Parallel universe approach

In randomized controlled trials, randomization of participants is a key step to reduce the chance of systematic differences between study participants in the intervention and control group. In MISCAN-Fadia, this is imitated by simulating the same female population twice. First, the population is simulated in a no screening world, then, the identical population is simulated again and subjected to screening to evaluate the effects of screening and treatment on incidence and mortality. In microsimulation modeling this approach is often referred to as a parallel universe structure. Usually, populations of tens of millions of women are simulated with a model runtime of approximately fifteen minutes.

Breast cancer onset

The risk of developing breast cancer increases as women get older, while at the same time breast cancer risk may differ by birth cohort [7, 8]. Therefore, breast cancer onset in Model E is mainly driven by an age risk factor combined with a birth cohort risk factor to account for variations in the prevalence of risk factors that are related to birth cohort. The model uses as input breast cancer incidence (invasive and DCIS) in the absence of screening to derive breast cancer onset probabilities that vary by age and cohort. Considering breast cancer incidence in the absence of screening has not been available at the population level in the U.S. since routine mammography screening started in the 1980's, most CISNET breast models used the breast cancer incidence in the absence of screening derived by Holford et al. [9]. Currently in Model E, the breast cancer onset parameters are calibrated to the U.S. incidence in the absence of screening that was derived and estimated by Gangnon et al. who extended the work by Holford by disentangling breast cancer incidence by cohort- and age-related factors, and the impact of mammography screening dissemination in the U.S.. [10].

The continuous tumor growth natural history model

Among women who develop breast cancer, the natural history of the disease is simulated as a continuously growing tumor. At tumor inception, the tumor's diameter is 0.1 millimeter and based on the time it takes for the tumor to double in size, (i.e., the tumor volume doubling time) it grows exponentially. The DCIS model was originally based on the DCIS model of the Erasmus MISCAN breast model [11]. Once a breast lesion emerges from normal breast tissue, a woman is in the pre-clinical undetectable DCIS phase (Figure 1). The two possible transitions from there are either: pre-clinical screen-detectable DCIS, the state that all CISNET breast models that include DCIS have in common [12], or pre-clinical invasive breast cancer. From the pre-clinical screen-detectable state three different transitions are possible; regression to a breast cancer-free life, progression to pre-clinical invasive breast cancer, or progression to the clinical DCIS state. The duration (years) in each DCIS state is assumed to be exponentially distributed and transitions

between DCIS states happen at exponential rates. These transition rates were estimated using SEER American Joint Committee on Cancer (AJCC) data on stage distributions and age-specific DCIS and invasive incidence rates between 1975 and 1999 [3].

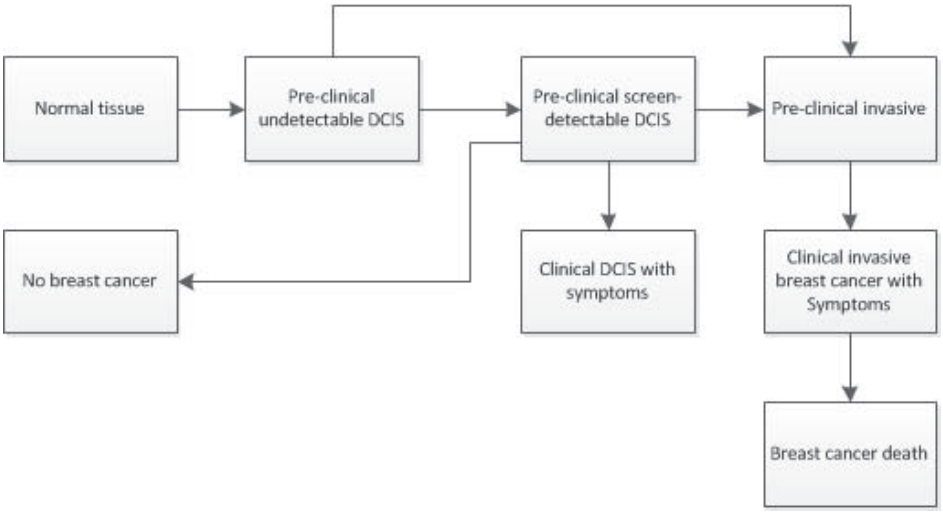


Figure 1 The Ductal Carcinoma in Situ sub-model in MISCAN-Fadia.
Once a breast lesion emerges from normal breast tissue, a woman is in the pre-clinical undetectable DCIS phase. The two possible transitions from there are either: pre-clinical screen detectable DCIS or pre-clinical invasive breast cancer. From the pre-clinical screen detectable DCIS phase the tumor may regress and the woman will end up in the 'No Breast Cancer' pool. However, from the pre-clinical screen detectable DCIS phase the tumor may also progress to pre-clinical invasive breast cancer or the tumor may cause clinical symptoms and a DCIS case will be diagnosed as a result of clinical symptoms. If a tumor is in the pre-clinical invasive breast cancer state, the cancer may be screen detected or cause clinical symptoms that lead to a clinical breast cancer diagnosis. Depending on the moment of diagnosis and the type of treatment a women may cure or die from breast cancer.

The tumor diameter at which available treatment options no longer result in cure is the fatal disease diameter and reflects the spread of breast cancer, i.e., distant metastasis. If the disease is fatal at the moment of diagnosis (i.e., the tumor diameter at diagnosis is larger than the tumor's fatal diameter), the time until death from breast cancer is determined by a draw from the survival distribution at the moment the disease became fatal (Figure 2). Tumors that are diagnosed at a smaller diameter than their fatal diameter are surgically removed, possibly radiated and adjuvant treatment ensures the woman will not die of breast cancer. Each tumor is unique and has different diameter sizes for: clinical diagnosis, screen detectability and metastasis (fatal diameter). As listed under 'the life course of a tumor', these tumor properties are governed by probability distributions to bring about variation between tumors.

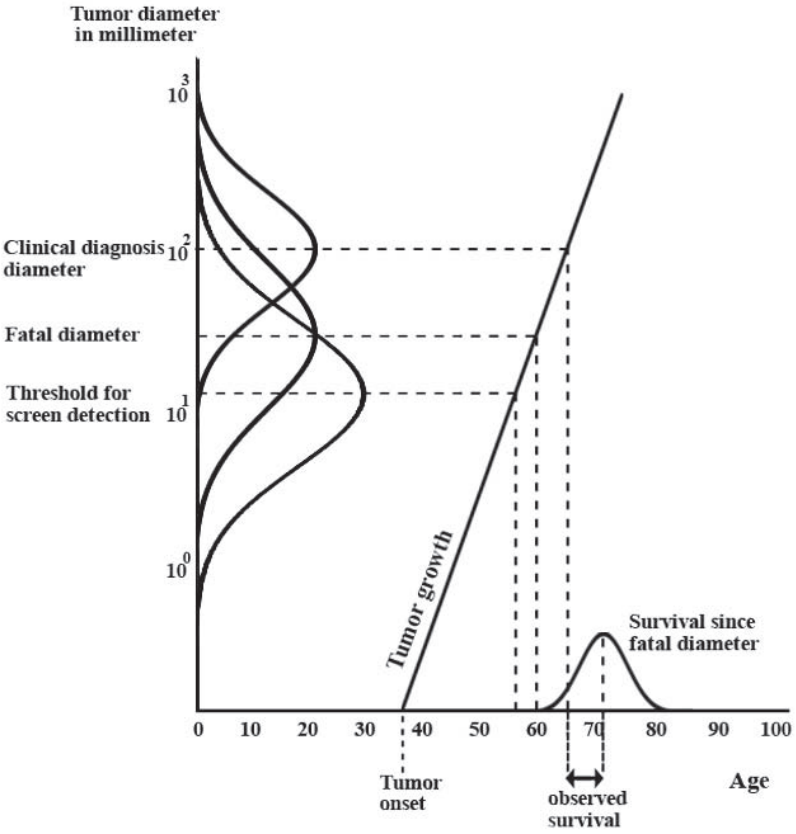


Figure 2 The MISCAN-Fadia breast cancer natural history model.
When a breast tumor is initiated in a simulated woman, values of six tumor characteristics are generated: growth rate of the tumor, the tumor's fatal diameter that represents distant metastasis, survival time after reaching the fatal diameter, screen detectability diameter (threshold), and the clinical diagnosis diameter. The distribution curves on the y-axis demonstrate the probabilistic nature of the simulations and the variation between the screen-detection, fatal and clinical diagnosis diameter of tumors. The growth rate of the tumor determines the times since its initiation at which the tumor reaches the screen detectability diameter, the clinical diagnosis diameter, and the fatal diameter. If in the absence of screening the clinical diagnosis diameter is larger than the fatal diameter, the woman will die of breast cancer and the observed survival time is given as depicted in Figure 2. A woman will be cured if the breast cancer is detected, either clinically or through screening, before the fatal diameter is reached. Treatment (not shown in Figure 2) is modeled as a shift in fatal diameter and may affect survival and in the best scenario cause of death.

Our natural history approach makes a distinction between tumor biology (i.e., growth rate of the tumor) and variables that are more likely to change over time, by age, or differ by geographical region. The advantage of this approach is that it readily lends itself to define separate distributions for different parameters based on risk groups and molecular tumor subtypes for example [13, 14]. As such, adapting the model to simulate subgroups

of more aggressive and faster growing tumors (e.g., ER/HER2 molecular subtypes of breast cancer) was done by changing the growth rate of tumors while keeping other tumor aspects such as the clinical diagnosis diameter and tumor diameter threshold for screen detectability unchanged.

The life course of a tumor is described by

1. Tumor growth rate ~ Log Normal (μ_1,σ_1)
2. Fatal diameter of the tumor ~ Weibull (λ_1,κ_1)
3. Survival time after reaching fatal diameter ~ Log Normal (μ_2,σ_2)
4. Screen detectable (threshold) tumor diameter ~ Weibull (λ_2,κ_2)
5. Clinical diagnosis diameter of the tumor ~ Log Normal (μ_3,σ_3)
6. Clinical diagnosis of the tumor caused by distant metastasis. This is modeled as a constant fraction of the survival after reaching the tumor's fatal diameter.
7. Correlation between tumor growth rate and the tumor's clinical diagnosis diameter: ρ_1
8. e.g., fast growing tumors are diagnosed at larger diameters.
9. Correlation between tumor growth rate and survival time after reaching the tumor's fatal diameter: ρ_2 e.g., fast growing tumors have a shorter survival.
10. Correlation between tumor diameter at clinical diagnoses and survival time after reaching the tumor's fatal diameter: ρ_3
11. e.g., tumors with a large size at clinical diagnosis have a shorter survival.
12. The tumor diameter at which N1 lymph node disease becomes detectable ~ Weibull (λ_3,κ_3)
13. Difference in tumor size at which N1 and N2 lymph node disease become detectable.

When a breast tumor is initiated in a simulated woman, values of the six (1-6) tumor variables are generated. For each simulated tumor, the clinical diagnosis diameter is determined by the smallest tumor diameter of either the diameter at clinical diagnosis or the diameter at clinical diagnosis because of fatal metastases. After tumor initiation, the growth rate of the tumor determines the times at which the tumor reaches the threshold diameter for detectability by screening, the clinical diagnosis diameter, and the fatal diameter. If the tumor diameter at diagnosis is larger than the fatal diameter, then the survival time after reaching the fatal diameter will give the time at which a woman will die of breast cancer. On the other hand, if a tumor is detected, either clinically or through screening, before the fatal diameter is reached, the woman will be cured of cancer and die of other causes. A graphical representation of how the natural history of breast cancer is modeled in MISCAN-Fadia is provided in Figure 2. In MISCAN-Fadia, initially, Weibull distributions were assumed for all variables. However, when it became apparent that correlations had to be assumed, the more convenient multivariate lognormal distribution

was used for three correlated variables. The main reason was to get a better fit on the data of the base-case analysis.

For the CISNET breast "Base Case" analysis [15, 16], the maximum likelihood estimates of MISCAN-Fadia for the natural history parameters were initially based on detailed data from the Swedish Two County Study [4, 5]. These included estimates for tumor growth, tumor fatal diameter, survival duration since fatal diameter, clinical diagnosis diameter, and screen detectability diameter. The tumor size distribution and number of screen detected cancers and interval cancers per screening round were simulated and compared to the findings of the trial. A detailed description and estimation of these natural history parameters can be found elsewhere [6]. Since the base case analysis, the natural history parameters such as tumor growth rate, tumor fatal diameter, survival duration after reaching the fatal diameter, and the threshold for screen detection have been re-estimated for the simulation of various breast cancer molecular subtype combinations of ER and HER2. [13, 14]

Population Demographics

MISCAN-Fadia can simulate one specific birth cohort, or, to account for varying demographic characteristics, a dynamic population consisting of multiple birth cohorts can be simulated. Certain birth cohorts may be assigned a different relative risk of developing breast cancer when cohort effects are present in the population. Nevertheless, each birth cohort is assigned an all-cause mortality table from which breast cancer as cause of death is removed. These mortality tables determine the date of non-breast cancer related death. A woman dies either from breast cancer or from other causes, whichever comes first. MISCAN-Fadia uses population parameters such as the number of birth cohorts and the proportion of each birth cohort in the overall U.S. population. These model inputs, as well as the other cause mortality tables are common CISNET model inputs [3].

Screening and screen detection

Characteristics of organized screening programs, such as screening ages, intervals, screening modality, and attendance by first and subsequent screens can be inserted directly into the model. The mammography screening dissemination that reflects the historic opportunistic screening patterns observed in the U.S. can also be simulated [17, 18]. Parameters to simulate screen detection, such as the sensitivity of the screening test, are translated into a diameter size at which tumors become screen detectable. By means of model calibration of tumor size distributions to observed tumor size distributions, the model estimates the screen detection (threshold) parameter. By varying of only the screen detection parameters, the model finds the parameter values that resemble the best match between the simulated data and observed data.

If a woman is screened after a tumor onset, but before the threshold tumor diameter of screen-detectability, the result of the screening test is false negative. If that woman would be screened when the tumor diameter is larger than the tumor's screen-detectability diameter, the result of the screening test is true positive. This structure for screen detection implies that no false positives are registered as direct output from the model. The number of false positive mammograms is calculated based on the total number of mammograms performed in the model and the observed false positive rates. Screening sensitivity differences between screening modalities, as well as improvements in screening performance are modeled as a shift in the threshold diameter for screen-detectability. The advent of digital mammography between 2000 and 2010 has been incorporated into the model by calibrating the threshold to digital mammography data [19].

Overdiagnosis is defined as screen-detected DCIS or invasive breast cancer that would not have been diagnosed in a woman's life in the absence of screening. The parallel universe approach; simulating the same population of women twice, implies that the women in the screened population are exactly the same women as in the unscreened population. This allows for exact quantification of overdiagnosis due to screening because of the lifetime follow-up of all women.

Breast cancer staging

In MISCAN-Fadia, the severity of breast cancer is described by the diameter of the primary tumor and the extent to which the cancer has spread to lymph nodes or distant organs. This corresponds to the Tumor Node Metastasis (TNM) staging system that was developed and is maintained by the AJCC union that classifies tumors based on the size of the primary tumor (T), the nearby lymph nodes that are involved (N), and the spread of cancer as distant metastasis (M). To get to a stage at diagnosis, MISCAN-Fadia links tumor diameter to staging by including 3 parameters. First, continuous growth of the tumor diameter; the main concept of the natural history model, covers the T part of the staging system by the unique size of the tumor at diagnosis. Second, the lymph node status of tumors is covered by the inclusion of two parameters; N1: the size of the tumor that reflects the spread to 1-3 nearby lymph nodes, N2: the size of the tumor that corresponds to the diameter at which breast cancer has spread to 4 to 9 lymph nodes. This is modeled as a fixed diameter size larger than N1. Third, metastasis of the primary tumor is modeled and covered by the unique fatal diameter of each tumor. The values of N1 and N2 were calibrated to SEER data on stage at diagnosis of cancers diagnosed between 1975 and 2000 as part of the base-case analysis[6]. The definition of the AJCC staging system determines how cancers are staged at diagnosis; all DCIS diagnoses are staged as 0. Tumors smaller than 2 cm that have not spread to any nearby lymph nodes are staged as 1, tumors that are between 2 and 5 cm at diagnosis that have not spread to nearby lymph nodes are staged as 2a, and so on.

Adjuvant treatment

The benefit of adjuvant treatment is modeled as a shift in the fatal diameter. For each adjuvant treatment an age-specific cure proportion is estimated using the common CISNET model inputs [3] based on treatment effectiveness data from the meta-analyses by the Early Breast Cancer Trialists' Collaborative Group (EBCTCG) [20, 21]. The cure proportions are translated into tumor diameters so that more effective treatment can cure a larger tumor. Women diagnosed at a tumor diameter greater than the tumor's fatal diameter, benefit from adjuvant treatment by a shift to a larger fatal disease diameter. If the new fatal diameter is larger than the diameter at diagnosis, the treatment results in cure and ultimately death from other causes. However, if the new fatal diameter is still smaller than the diameter at diagnosis, surgery and radiation combined with adjuvant treatment will not results in cure and the tumor will eventually cause breast cancer death. The dissemination of adjuvant treatment is modeled as the probability of being treated with a certain type of treatment (e.g. chemotherapy, tamoxifen) given stage at diagnosis, calendar year, age at diagnosis, ER and HER2 status.

Parameter estimation

Parameter estimates are obtained by optimizing the goodness of fit between simulated data and observed data. The stochastic nature of the model output and duration of the model runs make the process of finding solid parameter estimates time-consuming. For selected starting values of the parameters, one microsimulation run will produce, for instance, age-specific breast cancer incidence trends over time, and compare it to the observed breast cancer incidence levels. Maximum likelihood estimates of the model parameters are obtained by repeated evaluation of the simulated breast cancer incidence for different sets of parameter values. Parameters are estimated by minimizing the sum of squared differences between observed and simulated data. This weighted sum measures the goodness of fit of the simulation results and is defined as a chi-squared distributed statistic. [22]. Minimization of the goodness of fit statistic leads to the optimal parameters, but requires frequent, and time-consuming evaluations of the objective function. We used the Nelder and Mead Simplex (NMSM) algorithm [23], which has the advantage that it only uses the value of the objective function, i.e., the goodness of fit of the model, to find the minimum. In the NMSM approach, each step in the optimization algorithms is based on output from previous simulation runs in which large numbers of life histories have been simulated, and it performs quite well in locating the optimum.

Extensive model calibration for the CISNET base case analysis provided parameter estimates that resulted in a close match between the simulated U.S. incidence and mortality over time and the observed trends in incidence and mortality from 1975 to 2000 [16]. These parameter estimates from the base case analysis were only re-calibrated for a limited number of parameters at a time and within logical parameter bounds (e.g., new

screening modalities with higher sensitivity of screening correspond to, and resulted in, a smaller threshold diameter for screen-detectability).

Validation

Establishing the degree to which MISCAN-Fadia is an accurate representation of the real world, is validation. Five types of validation [24] are addressed: face validity, internal validity, cross validity, external validity, and predictive validity. Face validity means the model makes sense at face value. MISCAN-Fadia's structure with a biological entry of continuous tumor growth makes sense at face value. The model structure and data sources used as input lead to credible results that show no logical contradictions such as screening resulting in the diagnosis of more late stage tumors, or decreasing risk of developing breast cancer as women get older. Internal consistency, or verification, examines the mathematical calculations performed and its consistency with what could be expected based on the model's specification. MISCAN-Fadia, programmed in Delphi, is a microsimulation model in which disease processes are mainly driven by clearly specified probability distributions that are widely used in modern programming software packages. Results of mathematical calculations for published parameter values can easily be verified when using these probability distributions.

Cross-validity covers the aspect of comparing model results to the results of other modeling groups. As MISCAN-Fadia has been part of CISNET since the start of its collaboration, this form of validation of the model has been done extensively [15, 25, 26]. External validity is the comparison of model outcomes to observed data that was not used for calibration and development of the model. MISCAN-Fadia is currently part of an independent external validation exercise wherefore we validated the results of five CISNET breast cancer models against the UK Age trial [27]. In the past, we conducted a dependent model validation against the UK Breast Screening Frequency trial [28]. UK specific breast cancer incidence and life tables were used, and the threshold diameter as well as the diameter of clinical diagnosis were re-estimated based on the trial's data. The model accurately reproduced the cumulative incidence in the intervention and control groups. Also, the percentage of screen detected and clinically diagnosed breast cancers were similar to the observed percentages in both groups, as were the number of breast cancer deaths [29]. Predictive validation is done by making model predictions for future outcomes of, for example, patterns in incidence and mortality. MISCAN-Fadia has made predictions about future trends in incidence and mortality [30], but it still remains to be seen how these predictions unfold.

Model input and output of MISCAN-Fadia

Differences in patterns of breast cancer incidence and mortality can often be traced back to different screening and treatment regimens, adherence patterns, and different

underlying risks. To simulate the harms and benefits of screening and treatment at the population level, the model requires data for the four major model components: population demographics, natural history of breast cancer, screening and treatment. A list of inputs of MISCAN-Fadia is provided and described as common CISNET model inputs [3].

The outcomes listed in Table 1 can be produced for any screening scenario with different start and stop ages of screening, screening frequency and screening modality. In addition to different screening strategies, the model output can also be broken down by: calendar year, age group, and by tumor size or breast cancer stage such as AJCC. By assigning health utilities to specific health states and unit costs to specific events, total costs and Quality Adjusted Life Years (QALYs) can be calculated. Consequently cost-effectiveness analyses can be performed [31]. In addition, radiation-induced breast cancers and breast cancer deaths can be calculated using model output together with radiation dose [32].

Table 1 Model output MISCAN-Fadia model

	Output description
1	Invasive Breast cancer cases diagnosed clinically
2	Invasive Breast cancer cases diagnosed by screening
3	DCIS cases diagnosed clinically
4	DCIS cases diagnosed by screening
5	Life years in the absence of screening
6	Life years in the presence of screening
7	DCIS over diagnosed cases (in the presence of screening)
8	Invasive over diagnosed cases (in the presence of screening)
9	Breast cancer deaths in the absence of screening
10	Breast cancer deaths in the presence of screening
11	Deaths from other causes in the absence of screening
12	Deaths from other causes in the presence of screening
13	Number of mammograms
14	Number of cancers diagnosed in AJCC stage I, II, III, IV
15	Number of cancers diagnosed in SEER stage local, regional, distant
16	Number of cancers diagnosed by tumor size 0-20mm, 20-50mm, 50+ mm
17	Number of cancers treated with adjuvant treatment
18	Intervals between events, e.g., lead time (time between screen detection and diagnosis in the absence of screening), survival (time between diagnosis and death)

Extensions and applications of the model

Targeting screening to women with the highest potential benefit and lowest potential harm can improve the overall balance between benefits and harms in the population. In recent years, we explored the effects of obesity and race on U.S. breast cancer mortality

[30, 33] as well as the cost effectiveness of ultrasonography screening [31]. In the past years, we also examined the contributions of screening and treatment to reduction in molecular subtype specific breast cancer mortality by evaluating different screening scenarios, including risk-based screening strategies. We present some examples of the model adaptations that formed the basis of these collaborative modeling studies.

Personalizing screening

To evaluate screening outcomes while taking into account advances in mammography and treatment of breast cancer, several screening strategies were modeled differing by age at which screening starts and screening interval. Biennial screening from age 50 to 74 years avoided a median of 7 breast cancer deaths per 1,000 women screened compared to no screening and is generally considered to have a favorable balance between benefits and harms. More intensive screening leads to more benefits (breast cancer deaths averted), but also to more harms (false-positives and over diagnosis). For example, annual screening from age 40 to 74 years avoided an additional 3 deaths, but yielded 1988 more false-positive results and 11 more over diagnosed cases per 1000 women screened [26]. Women aged 40 with a two-fold risk (compared to average risk) can expect the same balance of benefits and harms as average-risk women who receive biennial screening starting from age 50 [25].

Breast density and breast cancer

Breast density has been proposed to personalize mammography screening. Dense breast tissue is prevalent and associated with a higher risk of developing breast cancer [34]. Moreover, since breast density is relatively easy to measure on a mammogram, it can be used for risk stratification. Some studies have found that tumors in dense breasts (categorized as BI-RADS 3 and 4) may progress more rapidly than those in fatty breasts, categorized as BI-RADS 1 and 2 [35]. Based on this, breast density could be taken into account when personalizing a woman's screening frequency. Breast density does not only affect risk of developing breast cancer, it also affects screening test sensitivity as dense breast tissue is comprised of less fat and more connective breast tissue which appears white on a mammogram. Moreover, cancer appears white on a mammogram and is therefore easier overlooked by radiologists, resulting in a lower screening test sensitivity.

Breast density in MISCAN-Fadia

Breast density has been incorporated into MISCAN-Fadia to assess the effects of personalized screening; breast density was assumed to influence the sensitivity of the screening test (threshold diameter) as well as the onset of breast cancer. We also incorporated the decrease in breast density as women age because mammographic density decreases

after the menopause when ovarian function declines. When modeling both risk and density, we found that average-risk women (low breast density) undergoing triennial screening and higher-risk women (high breast density) receiving annual screening will maintain a similar or better balance of benefits and harms compared to biennially screening average-risk women [36].

Simulating molecular subtypes of breast cancer

It has been widely acknowledged that breast cancer is a heterogeneous disease and more knowledge is emerging on distinct molecular subtypes. Combinations of Estrogen Receptor (ER) and Human Epidermal Growth Factor Receptor 2 (HER-2) status have different tumor growth and are associated with different treatment responses that have been found to be important in targeting the treatment of breast cancer. To understand the relative contributions of screening and treatment to U.S. breast cancer mortality, first the major subtype combinations of ER positive and ER negative have been included in MISCAN-Fadia. Across CISNET models we found greater absolute breast cancer mortality declines in ER-positive cancers than among ER-negative cancers. The relative contribution of adjuvant treatment vs screening to breast cancer mortality reductions was higher for ER-positive cases; for ER-negative cases, the relative contributions were similar [13]. We have recently also included HER-2 in the model [14], as well as the treatment Trastuzumab (Herceptin) that is an antibody that interferes with the HER2 receptor.

Future directions of MISCAN-Fadia

Risk based screening based on genetic risk profile

Genomic discoveries of genes associated with breast cancer risk may have the potential to personalize screening based on a woman's genetic risk profile. It is one of our primary goals in the upcoming years to continue our research on estimating the population impact of using polygenic risk to tailor screening strategies. A growing group of single nucleotide polymorphisms (SNPs) are discovered that are associated with an elevated risk for breast cancer [37]. Individual SNPs identify a small increase in risk, however, multiple SNPs combined together can be translated into a polygenic risk score to stratify women based on their polygenic risk. We divide the population into risk groups based on observed polygenic risk score distributions. For each risk group, the models simulate routine digital mammography screening strategies by varying starting and stopping ages of screening and screening frequency. To warrant a more intense screening scenario for high risk groups and a less intense screening strategy for low risk groups, we compare the benefits and harms of the different screening strategies. The polygenic risk distribution in the U.S. female population determines how many women are eligible for each selected screening strategy and what the overall harms and benefits of polygenic risk-based screening will be.

A simplified analysis of using polygenic risk to inform screening strategies, can be performed by dividing the population into three (low, median, high) risk groups with varying prevalence (Figure 3). Targeted screening based on polygenic risk leads to a redistribution of benefits and harms. A more in-depth analysis will be performed in the near future within CISNET. MISCAN-Fadia will be used to quantify the benefits such as the breast cancer deaths averted, quality-adjusted life years saved, breast cancer mortality reduction, costs, and harms such as the false positive mammograms, over diagnosed cases, unnecessary biopsies, false negatives.

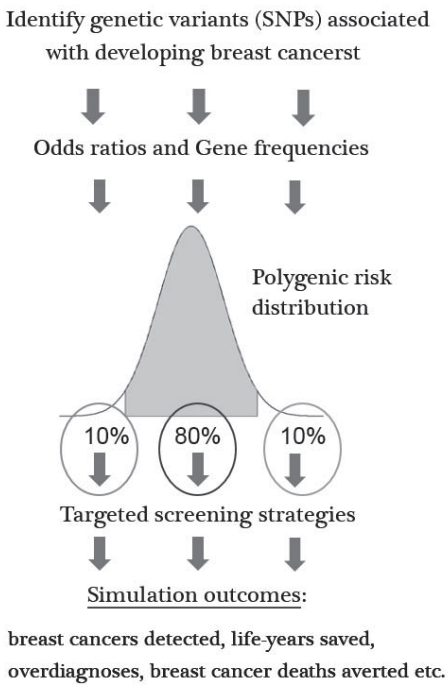


Figure 3 Simulating a personalized approach to breast cancer screening based on genetic risk profile.
Genetic variants for breast cancer have different risk alleles. Multiple single nucleotide polymorphisms (SNPs) combined together can be translated into a polygenic risk score to stratify women based on their polygenic risk. In Figure 3, a simplified analysis of the potential population impact of using polygenic risk to inform screening strategies is demonstrated by dividing the population into three (low, median, high) risk groups with varying prevalence. In this simplified example 10% of the population has a low risk of developing breast cancer, 80% an average risk, and 10% a high risk. More frequent screening could be offered to the high risk group and less frequent screening (compared to average risk group) could be offered to the low risk group. With more risk groups, or even a continuous risk distribution we could potentially optimize the tailoring of screening strategies based on polygenic risk which would lead to a redistribution of benefits and harms compared to current practice. A more in depth analysis will be performed in the near future within CISNET.

Strategies to reduce overtreatment of DCIS

While early detection of breast cancer and consequently less invasive treatment are often mentioned as benefits of screening, overtreatment of DCIS lesions that otherwise would not have clinically surfaced without screening is an increasing harm of screening since DCIS rates have increased dramatically over the last 30 years. Studies have shown that an increase in breast cancer mortality reduction due to screening comes with a substantial increased number of over diagnosed DCIS cases [11, 38]. MISCAN-Fadia will be extended to investigate if, how, and to what extent the harms of screening and treatment

of DCIS can be reduced. By simulating 'watchful waiting' strategies and exploring risk factors for progression to invasive breast cancer such as cytological grade, ER status, age at diagnosis and ethnicity, MISCAN-Fadia will be used to assess how different screening strategies and treatment routines may affect incidence and mortality for varying progression and regression rates of DCIS.

CONCLUSION

Trends in breast cancer incidence and mortality depend on many factors related to the biology and natural history of breast cancer. As tumor size is observable at diagnosis and tumors are considered to grow in continuous time rather than discrete time, these two aspects form MISCAN-Fadia's biological entry to modeling the effects of screening and treatment on breast cancer incidence and mortality. The advantage of this biologically oriented approach is that it allows for simple hypothesis testing because the core biological mechanisms are separated from cancer control interventions. Changes or improvements in screening and treatment that may vary by age, or over time, can be implemented directly and be dealt with without changing breast cancer onset or tumor growth parameters. On the other hand, simulating less or even more aggressive tumor subtypes with a different growth function is also possible. Moreover, correlations that were added to the base case model in order to get a good overall fit with observed data, were plausible, and with a biological reasoning, intuitive to understand. In particular, one may expect faster growing tumors to be diagnosed at larger tumor diameters, and faster growing tumors to have a shorter survival as well as a larger clinical diagnosis diameter.

However, MISCAN-Fadia also has limitations and makes use of simplifying assumptions. We model only one tumor per woman while it may be possible that breast cancer develops in both breasts independently or at the same time, although such cancer development is not prevalent. Also, recurrence of breast cancer is not simulated in our model. We do not model specific factors associated with an elevated risk for breast cancer such as reproductive history, alcohol use, hormone therapy use or familial risk. These different risk groups are assumed to be captured by the distribution we simulate tumors from. The spread between slower and faster growing tumors with unique tumor characteristics is assumed to capture the entire population risk profile.

Future development of the model will focus on evaluating the impact of using polygenic risk to inform screening strategies, evaluating the clinical management of screen-detected DCIS, and incorporating alternative and emerging screening modalities such as breast MRI and tomosynthesis.

REFERENCES

[1] Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed. 1985; 20(1):79-93.

[2] Boer R, de Koning H, van Oortmarssen G, Warmerdam P, van der Maas P. Stage distribution at first and repeat examinations in breast cancer screening. J Med Screen. 1999; 6(3):132-8.

[3] Mandelblatt JS, Near AM, Miglioretti DL, Munoz D, Sprague BL, Trentham-Dietz A, et al. Common Model Inputs used in CISNET Collaborative Breast Cancer Modeling Medical Decision Making. 2016; submitted.

[4] Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am. 1992; 30(1):187-210.

[5] Tabar L, Fagerberg G, Day NE, Duffy SW. The Swedish two-county trial of mammographic screening for breast cancer: recent results on mortality and tumor characteristics. Pathol Biol (Paris). 1992; 39(9):846.

[6] Tan SY, van Oortmarssen GJ, de Koning HJ, Boer R, Habbema JD. The MISCAN-Fadia continuous tumor growth model for breast cancer. J Natl Cancer Inst Monogr. 2006; (36):56-65.

[7] Clemmesen J. Carcinoma of the breast; results from statistical research. Br J Radiol. 1948; 21(252):583-90.

[8] Weedon-Fekjaer H, Bakken K, Vatten LJ, Tretli S. Understanding recent trends in incidence of invasive breast cancer in Norway: age-period-cohort analysis based on registry data on mammography screening and hormone treatment use. BMJ. 2012; 344:e299.

[9] Holford TR, Cronin KA, Mariotto AB, Feuer EJ. Changing patterns in breast cancer incidence trends. J Natl Cancer Inst Monogr. 2006; (36):19-25.

[10] Gangnon RE, Sprague BL, Stout NK, Alagoz O, Weedon-Fekjaer H, Holford TR, et al. The contribution of mammography screening to breast cancer incidence trends in the United States: an updated age-period-cohort model. Cancer Epidemiol Biomarkers Prev. 2015; 24(6):905-12.

[11] de Gelder R, Fracheboud J, Heijnsdijk EA, den Heeten G, Verbeek AL, Broeders MJ, et al. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. Prev Med. 2011; 53(3):134-40.

[12] van Ravesteyn NT, van den Broek JJ, Li X, Weedon-Fekjar H, Schechter CB, Alagoz O, et al. Modeling ductal carcinoma in situ (DCIS) – an overview of CISNET model approaches Medical Decision Making. 2016; submitted.

[13] Munoz D, Near AM, van Ravesteyn NT, Lee SJ, Schechter CB, Alagoz O, et al. Effects of screening and systemic adjuvant therapy on ER-specific US breast cancer mortality. J Natl Cancer Inst. 2014; 106(11).

[14] Plevritis S, Munoz D, Kurian A, Alagoz O, Near AM, Stout NK, et al. Contributions of screening and systemic therapy to molecular subtype specific U.S. breast cancer mortality from 2000 to 2010. 2016; submitted.

[15] Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005; 353(17):1784-92.

[16] Feuer EJ. Modeling the impact of adjuvant therapy and screening mammography on U.S. breast cancer mortality

between 1975 and 2000: introduction to the problem. J Natl Cancer Inst Monogr. 2006; (36):2-6.

[17] Cronin KA, Mariotto AB, Clarke LD, Feuer EJ. Additional common inputs for analyzing impact of adjuvant therapy and mammography on U.S. mortality. J Natl Cancer Inst Monogr. 2006; (36):26-9.

[18] Cronin KA, Yu B, Krapcho M, Miglioretti DL, Fay MP, Izmirlian G, et al. Modeling the dissemination of mammography in the United States. Cancer Causes Control. 2005; 16(6):701-12.

[19] Stout NK, Lee SJ, Schechter CB, Kerlikowske K, Alagoz O, Berry D, et al. Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography. J Natl Cancer Inst. 2014; 106(6):dju092.

[20] Early Breast Cancer Trialists' Collaborative G, Peto R, Davies C, Godwin J, Gray R, Pan HC, 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.

[21] Yin W, Jiang Y, Shen Z, Shao Z, Lu J. Trastuzumab in the adjuvant treatment of HER2-positive early breast cancer patients: a meta-analysis of published randomized controlled trials. PLoS One. 2011; 6(6):e21030.

[22] Barton RR, Ivey JS. Nelder-Mead simplex modifications for simulation optimization. Management Science. 1996; 42:954–73.

[23] Tan SYGL, Van Oortmarssen GJ, Piersma N. Estimating Parameters of a Microsimulation Model for Breast Cancer Screening Using the Score Function Method. . Annals of Operations Research. 2003; 119(1):43-61.

[24] Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Med Decis Making. 2012; 32(5):733-43.

[25] van Ravesteyn NT, Miglioretti DL, Stout NK, Lee SJ, Schechter CB, Buist DS, et al. Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: a comparative modeling study of risk. Ann Intern Med. 2012; 156(9):609-17.

[26] Mandelblatt JS, Stout NK, Schechter CB, van den Broek JJ, Miglioretti DL, Krapcho M, et al. Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies. Ann Intern Med. 2016; 164(4):215-25.

[27] Moss SM, Wale C, Smith R, Evans A, Cuckle H, Duffy SW. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. Lancet Oncol. 2015; 16(9):1123-32.

[28] Breast Screening Frequency Trial G. The frequency of breast cancer screening: results from the UKCCCR Randomised Trial. United Kingdom Co-ordinating Committee on Cancer Research. Eur J Cancer. 2002; 38(11):1458-64.

[29] van Ravesteyn NT, Heijnsdijk EA, Draisma G, de Koning HJ. Prediction of higher mortality reduction for the UK Breast Screening Frequency Trial: a model-based approach on screening intervals. Br J Cancer. 2011; 105(7):1082-8.

[30] Mandelblatt J, van Ravesteyn N, Schechter C, Chang Y, Huang AT, Near AM, et al. Which strategies reduce breast cancer mortality most? Collaborative modeling of optimal screening, treatment, and obesity prevention. Cancer. 2013; 119(14):2541-8.

[31] Sprague BL, Stout NK, Schechter C, van Ravesteyn NT, Cevik M, Alagoz O, et al.

Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. Ann Intern Med. 2015; 162(3):157-66.

[32] Miglioretti DL, Lange J, van den Broek JJ, Lee CI, van Ravesteyn NT, Ritley D, et al. Radiation-Induced Breast Cancer Incidence and Mortality From Digital Mammography Screening: A Modeling Study. Ann Intern Med. 2016; 164(4):205-14.

[33] van Ravesteyn NT, Schechter CB, Near AM, Heijnsdijk EA, Stoto MA, Draisma G, et al. Race-specific impact of natural history, mammography screening, and adjuvant treatment on breast cancer mortality rates in the United States. Cancer Epidemiol Biomarkers Prev. 2011; 20(1):112-22.

[34] McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2006; 15(6):1159-69.

[35] Aiello EJ, Buist DS, White E, Porter PL. Association between mammographic breast density and breast cancer tumor characteristics. Cancer Epidemiol Biomarkers Prev. 2005; 14(3):662-8.

[36] Trentham-Dietz A, Kerlikowske K, Stout NK, Miglioretti DL, Schechter CB, Ergun MA, et al. Tailoring breast cancer screening intervals by breast density and risk for women 50 and older: Collaborative modeling of screening outcomes. 2016; submitted.

[37] Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 2015; 107(5).

[38] van Luijt PA, Heijnsdijk EA, Fracheboud J, Overbeek LI, Broeders MJ, Wesseling J, et al. The distribution of ductal carcinoma in situ (DCIS) grade in 4232 women and its impact on overdiagnosis in breast cancer screening. Breast Cancer Res. 2016; 18(1):47.

Chapter 3

Modeling ductal carcinoma in situ (DCIS) – an overview of CISNET model approaches.

Nicolien T. van Ravesteyn, Jeroen J. van den Broek, Xiaoxue Li, Harald Weedon-Fekjær, Clyde B. Schechter, Oguzhan Alagoz, Xuelin Huang, Don Weaver, Elizabeth S. Burnside, Rinaa S. Punglia, Harry J. de Koning, Sandra J. Lee.

Medical Decision Making. 2018;38:126S-39S.

ABSTRACT

Ductal carcinoma in situ (DCIS) can be a precursor of invasive breast cancer. Since the advent of screening mammography in the 1980's, the incidence of DCIS has increased dramatically. The value of screen detection and treatment of DCIS is a matter of controversy, since it is unclear to what extent detection and treatment of DCIS prevents invasive disease and reduces breast cancer mortality. The aim of this paper is to provide an overview of existing Cancer Intervention and Surveillance Modelling Network (CISNET) modeling approaches for the natural history of DCIS, and to compare these to other modeling approaches reported in the literature. Five of the six CISNET models currently include DCIS. Most models assume that some, but not all, lesions progress to invasive cancer. The natural history of DCIS cannot be directly observed and the CISNET models differ in their assumptions and in the data sources used to estimate the DCIS model parameters. These model differences translate into variation in outcomes such as the amount of overdiagnosis of DCIS with estimates ranging from 34%-72% for biennial screening from age 50-74 years. The other models described in the literature also report a large range in outcomes with progression rates varying from 20%-91%. In the future, DCIS data by grade from active surveillance trials, development of predictive markers of progression probability, and evidence from other screening modalities, such as tomosynthesis, may be utilized to inform and improve the models' representation of DCIS and might lead to convergence of the model estimates. Until then, the CISNET model results consistently show a considerable amount of overdiagnosis of DCIS, supporting the safety and value of observational trials for low-risk DCIS.

Key Words: Cancer simulation, breast cancer epidemiology, simulation models, ductal carcinoma in situ

INTRODUCTION

Ductal carcinoma in situ (DCIS) represents a spectrum of abnormal cells confined to the breast duct and is a risk factor for invasive breast cancer development [1]. Before the introduction of mammography screening, DCIS was not often diagnosed. Since the advent of screening mammography in the 1980s, the incidence of DCIS has increased dramatically. In the United States, the incidence of DCIS increased from 5.8 per 100,000 women in 1975 to 68.9 per 100,000 women in 2010 [2-4]. By the year 2020, more than one million US women are expected to be living with and have been treated for a DCIS diagnosis [1].

The etiology of DCIS is presumably heterogeneous and its natural history is poorly understood as onset, progression and regression rates are not directly observable. Some DCIS lesions likely represent a precursor to subsequent invasive breast cancer, but DCIS may also remain indolent for sufficiently long that a woman dies of other causes [5-7]. The proportion of untreated DCIS that will progress to invasive breast cancer is unknown [1], and therefore, the impact of detecting and treating DCIS, particularly for any given woman, is unclear. Treating some DCIS lesions will probably prevent invasive disease, and consequently might reduce breast cancer mortality, thus can be considered a benefit. Other lesions might remain indolent in the absence of treatment with only harms related to their treatment (representing overdiagnosis and overtreatment). Since we do not know which and how many DCIS lesions will progress, the value of screen detection and treatment of DCIS remains unknown and is a matter of considerable controversy.

Despite the uncertainty around the natural history of DCIS, some predictors for progression have been identified. For example, younger age at diagnosis and black ethnicity are associated with higher breast cancer-specific mortality among patients with DCIS [8, 9]. Other identified factors for progression include estrogen receptor (ER) negative status, larger DCIS tumor size, and comedonecrosis [9]. In addition, DCIS progression to invasive breast cancer can be predicted by cytologic grade [5, 7, 9]. Pathologists use three grading categories: corresponding to well (grade 1), moderately (grade 2), and poorly (grade 3) differentiated DCIS [10], also referred to as "low grade", "intermediate grade", and "high grade", respectively. Grade has been found to be associated with recurrence [11, 12] and the survival benefit of surgical treatment has been found to be lower for low-grade DCIS than that for intermediate or high-grade DCIS [13]. Furthermore, the DCIS Score, based on Oncotype DX, has been found to be associated with recurrence of DCIS (either as DCIS or invasive breast cancer) [14].

These identified prognostic factors for recurrence may enable physicians to tailor treatment strategies. Specifically, recommending treatment that is less aggressive would be appropriate for DCIS that has a low risk for future recurrence, and predictors such as age, ER status, and/or grade might be used to identify low-risk lesions. Thus, understanding the

natural history of DCIS and its recurrence and progression predictors to guide treatment strategies is important for both clinical and public health decisions. However, investigating the natural history of DCIS is difficult as ideal high-quality data is lacking, given that progression paths are not directly observable. In addition, data are also limited because survival for women diagnosed with DCIS is very high and a trial would need to enroll very large numbers of women and follow them for a lifetime to be adequately powered to detect an impact of screening and treatment on mortality or other endpoints. Moreover, the natural history of DCIS is difficult to study because the standard of care is immediate treatment following diagnosis. In these instances (comparative) modeling can be useful, for example to provide a range of plausible DCIS progression and regression rates by evaluating what set of assumptions about these rates best fit the existing observable data. In addition, in natural history models, the difference in risk of progression based on age, grade and ER status can be included by allowing varying transition rates for these factors, which has already been done in a well-established microsimulation model to include grade [15].

Furthermore, within the Cancer Intervention and Surveillance Modelling Network (CISNET) comparative modeling work has been done. Previously, three CISNET models estimated the amount of DCIS overdiagnosis in women age 74 and older. The results indicated that at older ages harms began to outweigh benefits, largely as a consequence of the increasing amount of overdiagnosis of DCIS at older ages [16], which is partly due to the higher death rate from competing causes with aging. Together, these modeling papers, on one hand highlight the uncertainty regarding the natural history of DCIS, but also show the potential value of modeling in providing information where results are consistent.

The aim of this paper is to provide an overview of the ways CISNET models simulate the natural history of DCIS, illustrate how different assumptions affect results, to compare the CISNET models to other models described in the literature, and to highlight developments that might lead to model improvements or refinements.

CISNET models

CISNET DCIS models – model overview

CISNET is a consortium of National Cancer Institute (NCI)-sponsored investigators who use statistical modeling to improve our understanding of cancer control interventions in prevention, screening, and treatment and their effects on population trends in incidence and mortality. The CISNET breast models have been described in detail previously and recently updated descriptions have been given [17-22]. Briefly, the models are designed to match breast cancer incidence and mortality rates observed in the US. Four models are micro-simulation models (models developed by Erasmus MC, University Medical Center Rotterdam, model E; Georgetown University Medical Center, and Albert Einstein

College of Medicine, model G-E; MD Anderson Cancer Center, model M; and University of Wisconsin, Madison and Harvard Medical School, model W), one model uses an analytic approach (model developed by Dana-Farber Cancer Institute, model D), and the remaining model is a hybrid Monte Carlo simulation (model developed by Stanford University, model S). The micro-simulation models include natural history components that approximate tumor progression in size and stage (https://resources.cisnet.cancer.gov/registry/site-summary/breast/). Five of the six CISNET models currently include DCIS (all except model S). Most models assume that some, but not all, lesions progress to invasive cancer, for example by including three different types of preclinical DCIS: DCIS that progresses to invasive disease during the preclinical phase, progressive DCIS that is diagnosed clinically, and DCIS that does not progress (and might regress). However, the models differ in natural history of DCIS (Table 1) and model structure (see Figure 1), with different pathways for the progression and regression of DCIS and breast cancer. For example, invasive cancer can either develop through pre-clinical screen-detectable DCIS (Figure 1C), or also develop directly from pre-clinical DCIS that is not detectable at screening (Figure 1A and 1B). In the models, DCIS can regress from pre-clinical screen-detectable DCIS to pre-clinical undetectable DCIS (Figure 1A) or to an absorbing 'no breast cancer' state and disappear ("cease to exist") (Figure 1B and 1C). One model (model W) allows regression of pre-clinical DCIS as well as invasive disease (Figure 1D). Although the regression of breast cancer, especially invasive disease, is controversial, there is some evidence supporting the possibility of regressing tumors, including epidemiologic evidence [23] and a case report on regression of breast on imaging [24].

Most of the CISNET models have used data from the Surveillance, Epidemiology, and End Results (SEER) Program [25], typically age-specific incidence over time, combined with data from other sources (Wisconsin cancer registry for model W, Dutch data for model E) to estimate DCIS parameters, although one model used data from another source to develop their model (Norwegian data for model D) [26]. All CISNET models include a certain probability for mammography to detect DCIS at screening (Table 2). Specifically models D and GE use the same detection mechanism for DCIS as for invasive disease by including a sensitivity of screening. Model W uses the detection probability as a function of tumor size and because in situ lesions are small the likelihood of detecting DCIS is lower than that for detecting invasive breast cancer. Model E includes two separate detection mechanisms; DCIS detection is modeled by including a sensitivity, whereas screen-detection of invasive disease is modeled by a threshold diameter. Thus, in some models the sensitivity of a screening test differs for DCIS and invasive cancer.

CISNET models – analysis

The CISNET models were recently applied to evaluate screening outcomes of various screening strategies differing by age at which screening starts (40, 45, or 50 years) and

screening interval (annual, biennial) for the US female population [27]. We assessed the results of those prior analyses by focusing on the (as yet unpublished) model-specific rates of DCIS detection and overdiagnosis of the five CISNET models that include DCIS [28]. Overdiagnosis was defined as the detection of tumors that would not have been detected in a woman's lifetime in the absence of screening. We estimated the detection and overdiagnosis rate per 1000 women screened followed from age 40 over their lifetimes. In addition, the percentage overdiagnosis was calculated by dividing the rate of overdiagnosed DCIS by the rate of detected DCIS. We focus on four screening scenarios: biennial screening from 50-74 years (base), more frequent screening (annual screening from age 50-74 years; A50-74), an earlier starting age (biennial screening from age 40-74 years; B40-74), and later stopping age (biennial screening from age 50-84 years; B50-84).

Table 1 Natural history of DCIS in the CISNET models.

Model	in situ or DCIS?*	Do all tumors start as in situ?	Progression/regression	Model structure
D	DCIS only	Yes, but some DCIS is not screen-detectable and assumed to progress to invasive directly	DCIS progress to clinical DCIS or invasive breast cancer at exponential rates with mean sojourn time of 1.5-3 years; DCIS may also go back to a state in which it is undetectable [19]	Figure 1A
E	All in situ	Yes	DCIS progress to clinical or invasive breast cancer at an exponential rate with age and calendar year dependent sojourn times; DCIS may also regress [22]	Figure 1B
GE	DCIS only	Yes	DCIS progress to clinical or invasive breast cancer at an exponential rate with mean sojourn time of 2.97 years; DCIS may also regress [21]	Figure 1C
M	Model M is not a natural history model. It does not specify how tumors grow. It is an empirical model to describe screening, incidence, treatment and mortality. Under different screening scenarios, different stage distribution tables obtained from observed data [28] are used to assign tumor stages: DCIS, stages I, II, III or IV. DCIS patients are assumed to have the same survival as normal population, given age and birth year, no matter what treatments they receive.[18]			
W	All in situ. Model W also separated in situ into DCIS and non-DCIS in situ	Yes	All tumors, including DCIS, progress according to a Gompertz-type growth function, where the growth parameter is a random variable distributed with Gamma. Small size defines in situ. All tumors grow until they reach a maximum size. All tumors progress although a subset with "limited malignant potential" (LMP) stop at early invasive. LMPs comprise approximately 30-50% of all onset tumors [17]	Figure 1D

Model D: Dana-Farber Cancer Institute, Boston, Massachusetts. **Model E**: Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands**. Model GE**: Georgetown University Medical Center, Washington, DC, and Albert Einstein College of Medicine, Bronx, New York. **Model M**: MD Anderson Cancer Center, Houston, Texas. **Model W**: University of Wisconsin, Madison, Wisconsin, and Harvard Medical School, Boston, Massachusetts.

* in situ: DCIS and lobular carcinoma in situ (LCIS)

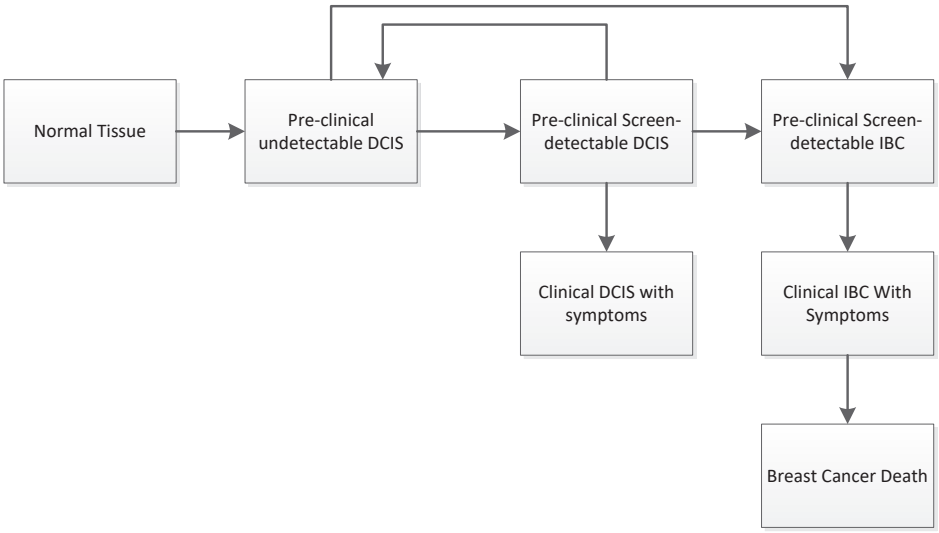


Figure 1A Model D

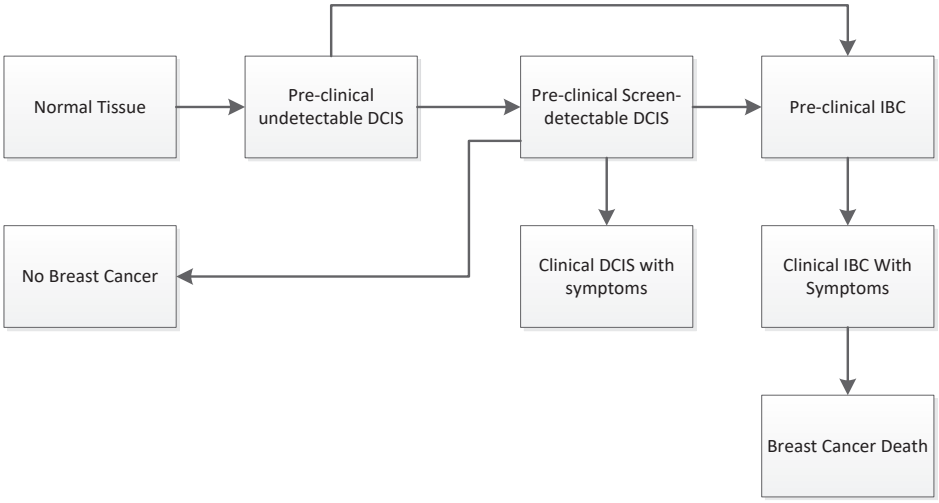


Figure 1B Model E

Figure 1 Schematic overview of models for the natural history of DCIS and invasive breast cancer. Invasive cancer can either develop through pre-clinical screening detectable DCIS (Figure 1C), or also develop directly from pre-clinical DCIS not detectable at screening (Figure 1A, 1B and 1D). Models include progression from preclinical screen-detectable DCIS to either clinical DCIS or preclinical invasive disease (Figure 1A, 1B, 1C, 1D), regression from preclinical DCIS to normal tissue (Figure 1D), to pre-clinical undetectable DCIS (Figure 1A), or to a 'no breast cancer' (absorbing) state in which women are no longer at risk for developing DCIS or invasive breast cancer (Figure 1B and 1C). Regression from invasive disease is also possible (Figure 1D).

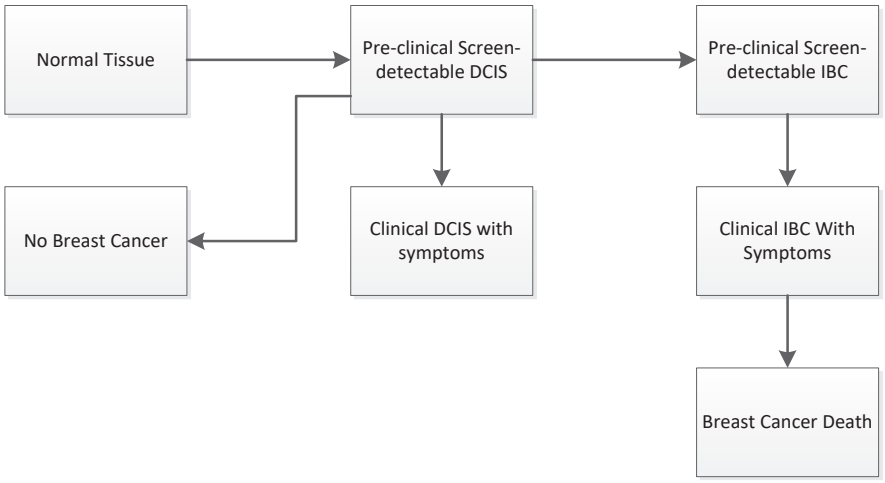


Figure 1C Model GE

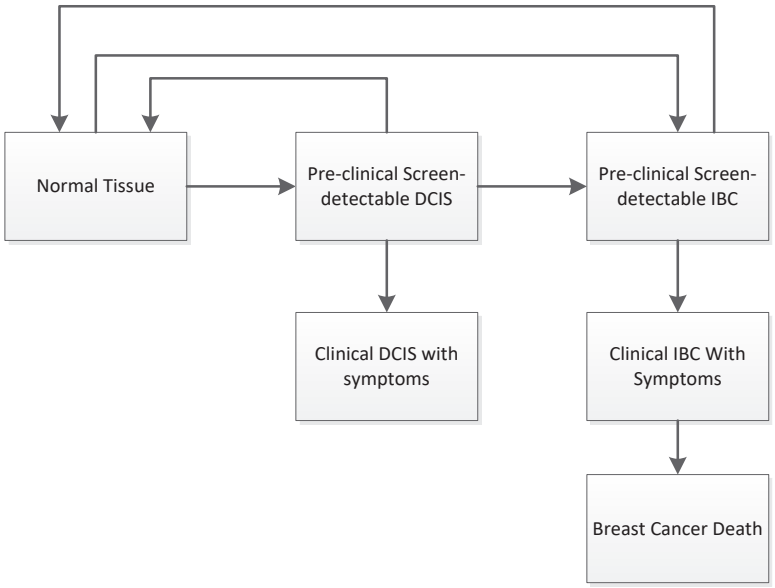


Figure 1D Model W

Figure 1 Schematic overview of models for the natural history of DCIS and invasive breast cancer. Invasive cancer can either develop through pre-clinical screening detectable DCIS (Figure 1C), or also develop directly from pre-clinical DCIS not detectable at screening (Figure 1A, 1B and 1D). Models include progression from preclinical screen-detectable DCIS to either clinical DCIS or preclinical invasive disease (Figure 1A, 1B, 1C, 1D), regression from preclinical DCIS to normal tissue (Figure 1D), to pre-clinical undetectable DCIS (Figure 1A), or to a 'no breast cancer' (absorbing) state in which women are no longer at risk for developing DCIS or invasive breast cancer (Figure 1B and 1C). Regression from invasive disease is also possible (Figure 1D).

Table 2 Detection mechanism of DCIS in the CISNET models.

Model	Clinical detection mechanism	Screen detection mechanism	Detection mechanism DCIS vs. invasive cancer
D	Some DCIS progress to clinical DCIS with symptoms - this rate matches age-specific incidence rate of DCIS in pre-screening era	Sensitivity varying by screening modality, age, calendar year	Same mechanism for DCIS and invasive cancer by test sensitivity
E	Some DCIS progress to clinical DCIS with symptoms - this rate matches age-specific incidence rate of DCIS in pre-screening era	Sensitivity varying by calendar year	DCIS is detected by test sensitivity; invasive disease is detected using a threshold diameter
GE	Progressive DCIS are clinically detected the same as more advanced lesions. Non-progressive DCIS are NEVER clinically detected.	Sensitivity varying by screening modality, age, calendar year	Same mechanism for DCIS and invasive cancer by test sensitivity
M	Model M makes no explicit mechanism assumptions regarding DCIS detection.		
W	Some DCIS are clinically diagnosed similarly as more advanced lesions. Clinical detection probability is an increasing function of tumor size and varies by age and calendar year. Clinical detection probabilities are in general smaller than screen detection probabilities; therefore a tumor is less likely to be detected via clinical surfacing than by screening.	Sensitivity varying by is tumor size, age, calendar year	Detection probability is an increasing function of tumor size, thus because in situ are small by definition, likelihood of detection of DCIS is less than that for invasive cancer

Model D: Dana-Farber Cancer Institute, Boston, Massachusetts. **Model E**: Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands**. Model GE**: Georgetown University Medical Center, Washington, DC, and Albert Einstein College of Medicine, Bronx, New York. **Model M**: MD Anderson Cancer Center, Houston, Texas. **Model W**: University of Wisconsin-Madison, Madison, Wisconsin, and Harvard Medical School, Boston, Massachusetts.

CISNET models – results and implications

For biennial screening between age 50 and 74 years, the five models that include DCIS predict that 154.4 women (median; range across five models 137.4 – 158.5; Table 3) are diagnosed with breast cancer per 1000 women followed from age 40 over their lifetimes. Of these women, 26.7 (25.8 – 32.3) are diagnosed with DCIS and 128.2 (110.7 – 131.8) with invasive disease. Of the women diagnosed with DCIS, 15.6 (9.0-18.8) are overdiagnosed, representing 51.3% (33.7%-71.8%) of the detected DCIS (Table 3). In contrast, for invasive disease, the models estimate that of the 128.2 (110.7-131.8) breast cancers detected, 3.3 (1.8-15.4) are overdiagnosed, corresponding to 2.6% (1.5%-12.0%; Table 3). This means that 2.6% (1.5-12.0%) of the invasive breast cancers that are detected would not have been detected in the absence of screening and are overdiagnosed. There is no direct connection between the amount of overdiagnosis of DCIS and overdiagnosis of invasive disease in the models. For example, one model predicts relatively low overdiagnosis percentages for DCIS as well as invasive breast cancer (model GE), whereas another model predicts relatively high percentages for both (model M). In contrast, there are also models that have modest estimates of DCIS overdiagnosis combined with relatively high estimates of invasive disease overdiagnosis (model W) or the other way around (model E).

Table 3 Detection and overdiagnosis of DCIS and invasive disease across the CISNET models for biennial screening from age 50-74 years.

Model	DCIS dx per 1000	DCIS overdx per 1000	%overdx DCIS	invasive dx per 1000	invasive overdx per 1000	%overdx invasive	total dx per 1000	overdx per 1000	%overdx (DCIS + invasive)
D	30.2	15.5	51.3%	128.3	3.3	2.6%	158.5	18.8	11.9%
E	25.8	16.1	62.4%	131.8	2.0	1.5%	157.6	18.1	11.5%
GE	26.7	9.0	33.7%	110.7	1.8	1.6%	137.4	10.8	7.9%
M	26.2	18.8	71.8%	128.2	15.4	12.0%	154.4	34.2	22.2%
W	32.3	15.6	48.3%	114.8	9.9	8.6%	147.1	25.5	17.3%
Median	*26.7*	*15.6*	*51.3%*	*128.2*	*3.3*	*2.6%*	*154.4*	*18.8*	*11.9%*

Model D: Dana-Farber Cancer Institute, Boston, Massachusetts. **Model E**: Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands**. Model GE**: Georgetown University Medical Center, Washington, DC, and Albert Einstein College of Medicine, Bronx, New York. **Model M**: MD Anderson Cancer Center, Houston, Texas. **Model W**: University of Wisconsin-Madison, Madison, Wisconsin, and Harvard Medical School, Boston, Massachusetts.

When annual screening from age 50-74 years is simulated, the models estimate 0.1-14.0 additional cases of DCIS being detected of which 0.1-13.7 are overdiagnosed (Table 4). Also, the models differ for the source for additional DCIS cases. For Models D, M, the increase in detection of DCIS is entirely overdiagnosis, whereas in models E, GE, W it is combination of overdiagnosis and earlier detection of lesions with progressive potential.

In addition, the order of scenarios that have the largest increase in overdiagnosis of DCIS varies across models, as well as the magnitude of the increase. For example, for annual screening the increase in overdiagnosis varies between 0.1 and 13.7 overdiagnosed DCIS cases across models. Some models estimate the largest change in detection and overdiagnosis when annual screening is considered (models E, M, W), whereas other models predict the largest increase when upper age of screening is extended to age 84 (models D and GE).

For the biennial screening scenario from age 50-74 years, the highest percentage of overdiagnosis of DCIS and invasive breast cancer was estimated by model M followed by W. This can be explained by the modeling choice of model M to assume a rather stable trend in breast cancer incidence (background trend) over time and, therefore, assign more of the increase to overdiagnosis than other CISNET models. Model W assumes that some invasive disease is non-progressive, and consequently, has a higher estimate for overdiagnosis than the other three models, especially for invasive disease.

For the other scenarios, annual screening from age 50-74 years, biennial screening from age 40-74 years, and biennial screening from age 50-84 years, there are two clusters of models: models D and M assign the increase in detection of DCIS when screening more intensively entirely to overdiagnosis. For model M that is again related to the stable background trend and for model D, the screen detectable period for DCIS is relatively

Table 4 Changes in DCIS detection and overdiagnosis of DCIS when moving from biennial 50-74 years to other screening scenarios.

Model	change in DCIS detection			change in DCIS overdiagnosis			change in DCIS overdx / change in DCIS detection		
	A50-74	B40-74	B50-84	A50-74	B40-74	B50-84	A50-74	B40-74	B50-84
D	0.1	0.0	2.8	0.1	0.1	2.8	100%	N/A	100%
E	8.5	4.8	5.6	6.7	3.3	5.2	79%	69%	93%
GE	3.2	3.6	6.3	0.4	1.2	3.0	13%	33%	48%
M	13.6	5.0	5.5	13.7	5.1	5.6	101%	102%	102%
W	14.0	2.4	9.7	7.1	1.5	-1.1	51%	63%	-11%

A50-74: **annual** screening from age 50-74 years.
B40-74: biennial screening from age **40**-74 years.
B50-84: biennial screening from age 50-**84** years.
Model D: Dana-Farber Cancer Institute, Boston, Massachusetts. **Model E**: Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands**. Model GE**: Georgetown University Medical Center, Washington, DC, and Albert Einstein College of Medicine, Bronx, New York. **Model M**: MD Anderson Cancer Center, Houston, Texas. **Model W**: University of Wisconsin-Madison, Madison, Wisconsin, and Harvard Medical School, Boston, Massachusetts.

short. The other three models (models E, GE, and W) only assign a proportion of the increase to overdiagnosis and a proportion to earlier diagnosis. Models E and GE assign most of the increase to overdiagnosis when moving to older ages and a smaller percentage when moving to younger ages.

Literature

Description of other DCIS models in the literature

To improve the understanding of the natural history of DCIS, we conducted a literature search to identify DCIS models that have been described in the literature. We searched PubMed and JSTOR for "DCIS natural history modeling" and "DCIS progression", and selected the articles that focus on the estimation of key DCIS natural history parameters, such as mean sojourn time for screen-detectable pre-clinical DCIS, and percent of DCIS cases that progress to either invasive cancer, clinical DCIS, or potentially regress. We identified 10 relevant studies, of which nine include DCIS natural history modeling (Table 5). Among them, four studies use Markov models [29-32] and five use simulation models [15, 33-36], with parameters estimated with either maximum likelihood, Bayesian Gibbs sampling or least square methods, and varying assumptions about DCIS natural history pathways. Seven studies assumed that all invasive breast cancers progress through a pre-clinical in situ or DCIS state that can be detected at screening [15, 29, 32-34, 36], whereas the other two studies assumed that some DCIS or in situ lesions first become visible on mammograms as small invasive tumors [30, 35]. DCIS or in situ is assumed to have both progressive and non-progressive paths in eight studies [15, 29-34, 36], with one study also including non-progressive invasive cancers [36].

Table 5 Overview of studies on modeling DCIS.

1st Author (Year), Journal	Paper title	Approaches/Models for DCIS natural history
Yen (2003), Eur J Cancer. [32]	Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening	Markov model
Ozanne (2011), Breast Cancer Res Treat. [35]	Characterizing the impact of 25 years of DCIS treatment	Simulation model
de Gelder (2011), Epi Rev. [33]	Interpreting overdiagnosis estimates in population-based mammography screening	Simulation model
Gunsoy (2012), Breast Cancer Res. [29]	Modeling the overdiagnosis of breast cancer due to mammography screening in women aged 40-49 in the United Kingdom	Markov model
Tan (2013), Br J Cancer. [31]	Quantifying the natural history of breast cancer	Markov model (Bayesian)
Ryser (2016), J Natl Cancer Inst. [30]	Outcomes of Active Surveillance for DCIS: A Computational Risk Analysis	Markov model
Duffy (2016), Lancet Oncol. [37]	Screen detection of ductal carcinoma in situ and subsequent incidence of invasive interval breast cancers: a retrospective population-based study	Poisson regression
de Koning (2006), Breast Cancer Res. [34]	Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based on observed screen and clinical data	Simulation model
Seigneurin (2011), BMJ. [36]	Overdiagnosis from non-progressive cancer detected by screening mammography: stochastic simulation study with calibration to population based registry data	Simulation model (Bayesian)
van Luijt (2016), Breast Cancer Res. [15]	The distribution of ductal carcinoma in situ (DCIS) grade in 4232 women and its impact on overdiagnosis in breast cancer screening	Simulation model

These 10 studies used various data sources including different combinations of: i) data aggregated from population registries [15, 30, 35, 36], ii) observed national screening service program data [32, 33, 37], iii) detailed data from randomized screening trials [29, 31, 32, 34] and iv) estimates made from previously reported studies including studies of DCIS first overlooked at mammography [30, 36]. Generally, more detailed screening data makes it possible to deduce more realistic natural history models, fitting the model using data from different screening rounds and screening histories [29, 32]. In addition to the different data sources, three studies include all in situ lesions [29, 31, 36], while seven others only include DCIS [15, 30, 32-35, 37].

Data sources	Natural History assumptions
Swedish two county trial, service screening programs from UK, US Netherlands, and Australia	Healthy cases can progress to pre-clinical screen detectable progressive or non-progressive DCIS; progressive DCIS progress to invasive breast cancer; non-progressive DCIS regress to a separate state where no tumor is apparent.
US SEER (1975-2005) incidence	The percentage of the DCIS lesions that are assumed to progress to invasive breast cancer varies between 0% and 100%. The initial assumption that DCIS is a short-term obligate precursor of invasive cancer must be reevaluated based on the results.
Dutch population data from public screening program	Healthy cases can progress to pre-clinical screen detectable DCIS or invasive breast cancers; pre-clinical screen detectable DCIS can regress, progress to clinical DCIS, or progress to invasive breast cancer.
UK Age trial	Healthy cases can progress to pre-clinical screen detectable progressive in-situ or non-progressive in-situ; progressive in situ progress to invasive breast cancers
Swedish randomized trials	Healthy cases can progress to pre-clinical screen detectable progressive DCIS or non-progressive DCIS; progressive DCIS progress to invasive breast cancer.
US SEER (1999-2011) for cumulative mortality estimates and natural history model summarized from a variety of studies	Healthy cases can progress to the pre-clinical screen detectable progressive DCIS or non-progressive DCIS; progressive DCIS progress to localized invasive breast cancer.
UK National Health Service Breast Screening Program (NHSBSP)	Not specified.
Dutch pilot studies in Utrecht & Nijmegen; EORTC	Healthy cases can progress to pre-clinical screen detectable DCIS; pre-clinical screen detectable DCIS cases can progress to clinical DCIS or invasive breast cancer.
Isere, France incidence rates of breast cancer and DCIS (1991-2006) with some screening information	Healthy cases can progress to in situ; in situ cases can be non-progressive, progressive to clinical, and progressive to invasive; invasive cancer can also be non-progressive or progressive.
Nationwide network and registry of histopathology and cytopathology in the Netherlands (PALGA) data	Healthy cases can progress to different grades of DCIS; lower grade DCIS can progress to higher grade DCIS and vice versa; each grade of DCIS can progress to invasive cancer that are charaterized by tumor stage.

Parameters in the literature useful for DCIS modeling

The estimated proportion of DCIS progressing to invasive cancer varies widely in the literature (Table 5), mainly due to the available data, study-specific model assumptions, and different model structures. When all invasive breast cancer is assumed to go through a pre-clinical screen detectable DCIS state, the estimated progression rate of DCIS to invasive varies from 61% to 91% [15, 29, 31-34, 36]. When this assumption is not made, the estimated progression rate from DCIS to invasive varies from 20% to 24.4% [30, 35]. Some studies report a large proportion of progressive DCIS [31, 33, 34, 36], while other studies report that most DCIS cases do not progress to invasive cancer [30, 35]. When the

1st Author (Year), Journal	All invasive cancers progress through screening detectable DCIS?	Screening detectable DCIS might regress to a non-detectable stage	Regression
Yen (2003), Eur J Cancer. [32]	Yes	Yes	37% (19%-46%) at 1st screening; 4% (3%-21%) at 2nd screening
Ozanne (2011), Breast Cancer Res Treat. [35]	No	Not specified	Not specified
de Gelder (2011), Epi Rev. [33]	Yes	Yes	11% of DCIS regress
Gunsoy (2012), Breast Cancer Res. [29]	Yes	No	Not specified
Tan (2013), Br J Cancer. [31]	Yes	Yes	Not specified
Ryser (2016), J Natl Cancer Inst. [30]	No	Yes	Not specified
Duffy (2016), Lancet Oncol. [37]	Not specified	Not specified	Not specified
de Koning (2006), Breast Cancer Res. [34]	No	Yes	Not specified
Seigneurin (2011), BMJ. [36]	Yes	Yes	6% non-progressive in situ (95%CI 0%--17%)
van Luijt (2016), Breast Cancer Res. [15]	Yes	Yes	4% low, 2% intermediate, and 1% for high grade DCIS

Note: ranges present values estimated from different studies or data sources unless otherwise specified.

proportion of progressive DCIS is reported by screening round, the subsequent screening rounds often reported smaller proportions of progressive DCIS [29, 32] compared to initial screening, as cases with a long sojourn time were diagnosed in earlier screening exams. High-grade DCIS cases have a larger proportion progressing to invasive than low-grade DCIS cases [15].

As for the mean sojourn time, when all invasive cancer are assumed to be screen detectable at a pre-clinical DCIS stage, the estimated mean sojourn time for progressive DCIS cases in the pre-clinical screen-detectable DCIS state are usually short varying from 1 month to 5 years [29, 31, 32, 34, 35]. On the other hand, the sojourn time estimates are much longer if it is assumed that only a small fraction of invasive cancers comes from pre-clinical screen-detectable DCIS [30]. The estimated mean sojourn time in pre-clinical screen-detectable DCIS state for DCIS cases that progress to clinical DCIS or regress is in typically longer than the mean sojourn time of DCIS cases that progress to invasive cancer [29, 32].

Progression	Mean sojourn time	Mammography sensitivities to detect DCIS/in situ
To invasive: 100-%non-progression	for non-progression: 30y (6y-37y), for progression to invasive: 3mo (2mo-5mo)	Not specified
To invasive: 20% of progression rate matches SEER data best	Not specified	Not specified
To clinical DCIS: 28% ; To invasive: 61%	2.6y	for DCIS: 72%
To invasive: 45% (95%CI: 23%-75%) at 1st screen, 60% (95%CI: 40%-78%) at incidence screen	for pre-clinical non-progressive DCIS to clinical DCIS: 1.3y (95%CI: 0.4y-3.4y), for pre-clinical progressive DCIS: 0.11y (95%CI: 0.05y-0.19y).	for in situ: 82% (95%CI: 43%-99%)
91%(95%CI: 85%-97%) aggressive	for aggressive DCIS to invasive 0.5mo (95%CI: 0-1mo)	for DCIS: 88% (95%CI: 83%-92%)
24.4% (11.3%-67%)	for progressive DCIS to localized invasive (did not specify whether to pre-clinical or clinical invasive): 9.8y (6.4y-13.5y)	for MRI: 84% (77%-100%); for mammography: 40% (33%-50%)
1 invasive interval cancer case is estimated to be avoided per 5 DCIS cases	Not quantified, but short	Not specified
To either invasive or clinical : 90%	Dutch pilot study suggests 2.8y with 99% sensitivity. Nijmegen data suggests 2.5y. EORTC trial suggests 5y with 40% sensitivity.	
To invasive: 91% (95%CI: 84%-97%)	Not specified	Not specified
To invasive: 16% low, 31% intermediate, 53% for high grade DCIS	Not specified	Not specified

The mammography sensitivity for DCIS varies from 40% to 99% [29, 31, 33, 34]. The mean sojourn time for progressive DCIS in the pre-clinical screen detectable DCIS state tends to be smaller when mammography sensitivity is high. These variations reveal the uncertainty regarding the natural history of DCIS, highlighting the need and potential directions of CISNET modeling.

DISCUSSION

While the CISNET models have generated relatively similar results and conclusions in most other respects, DCIS detection rates and overdiagnosis reveal more variation in results, with predicted DCIS incidence ranging from 25.8 – 32.3 per 1000 women age 40 followed over their lifetimes, and estimates of DCIS overdiagnosis ranging from 34%-72% for biennial screening from age 50 to 74 years. The large difference in the

predicted amount of overdiagnosis of DCIS between models likely reflects the continued uncertainty about DCIS natural history, in particular the progression rates, which is also reflected in the results from other models described in the literature with reported progression rates varying from 20% to 91%.

In the literature outside of CISNET, several approaches have been proposed to model DCIS. The variations in model structure, assumptions and results make it challenging to deduce good overall estimates of key natural history parameters. Given the uncertainties in the DCIS models, a realistic approach to DCIS modeling is to adopt several plausible sets of model parameters and to evaluate a range of outcomes generated from the models. The CISNET models are well-suited for this type of analysis. CISNET models have the ability to project long-term implications for DCIS assumptions in terms of breast cancer outcomes such as life expectancy and overdiagnosis, and can thus assess how much early detection impacts breast cancer mortality. Also, moving forward, CISNET models are capable of utilizing multiple models and vary model parameters, to explore the impact of different DCIS assumptions on outcomes more systematically. In addition, both the impact of screening and treatment on DCIS-related outcomes can be systematically reviewed and compared. Although it remains to be seen to what extent these analyses will provide sufficiently accurate and consistent findings to inform clinical practice, the comparative modeling effort of the CISNET models will likely contribute to a greater understanding of DCIS.

Despite the large difference in the predicted amount of overdiagnosis of DCIS between models, all models indicated that the amount of overdiagnosis of DCIS is substantial (i.e., 34%-72% for biennial screening from age 50-74 years), indicating that per 1000 women followed over their lifetimes 9-19 are overdiagnosed with DCIS and the majority of those women will undergo treatment for their non-invasive disease. Almost all women (98%) diagnosed with DCIS undergo a surgical procedure [13, 38] and recent work found an increase in the utilization of mastectomy with reconstruction and contralateral risk-reducing mastectomy over time [39]. There was also an increase in the proportion of women undergoing adjuvant radiation therapy after surgery from 58.5% in 1998-1999 to 70% during 2006-2011 [39].

Modeling estimates might improve and results might converge when new data becomes available. A unique opportunity to improve DCIS natural history modeling comes from trials on active surveillance. Several trials are currently underway to evaluate active surveillance approaches for DCIS. In the UK, the Low Risk DCIS Trial (LORIS), is comparing surgical excision to active surveillance without excision [40, 41]. Similarly, the European Organisation for Research and Treatment of Cancer (EORTC) has started a trial on the management of low-risk DCIS (LORD), which is a randomized, multicenter, non-inferiority trial, between standard therapy approach versus active surveillance [42]. In the US a prospective, randomized trial, Comparing Operative to Medical Endocrine

Therapy for low-risk DCIS (COMET), has recently been funded. Women diagnosed with low-risk DCIS will be randomized to receive either guideline-concordant care of surgical intervention, with or without radiation, or active surveillance of a mammogram every 6 months for 5 years. Patients in both trial arms are free to choose endocrine therapy. Also, in the US, several research networks, called cooperative groups, that conduct cancer clinical research primarily under the sponsorship of the NCI, are presently testing the use of neo-adjuvant hormonal therapy in postmenopausal women with ER-positive DCIS prior to surgery; those with a complete response based on magnetic resonance imaging (MRI) will not receive additional therapy. However, it will take a long time before results are available, e.g., for LORIS initial results are expected in 2020 and for LORD the results are not expected before 2029. When they do become available these data present a unique opportunity to validate models by comparing the model projections to the final trial data.

In the meantime, thus, before final results from these trials become available, the models can be used to evaluate which assumptions affect outcomes most. Also, data from several different sources might be used and combined to compare model outcomes and see what model structure and progression rates fit the data best. For example, data from different screening modalities can inform models, as the ability to detect DCIS varies across modalities. Screening ultrasound is less likely to detect DCIS compared to mammography in the small number of controlled experiments available that make this comparison, because ultrasound is unlikely to detect micro-calcifications. MRI may be more sensitive than mammography [43, 44] by detecting the pathophysiologic properties like basement membrane permeability in DCIS [45] perhaps explaining the tendency of MRI to detect intermediate and high grade DCIS more readily than mammography. By using a particular set of parameters and modelling different screening modalities, it might become possible to narrow down the range of plausible progression parameters. Furthermore, data by ER and grade might be used to refine the models. Subsequently, the updated and refined models can be used to simulate active surveillance strategies and quantify the predicted outcomes for subgroups of women varying by age and with DCIS varying by grade and ER status.

Until then, the model results consistently show a considerable amount of overdiagnosis of DCIS, which increases with more frequent screening. This indicates that women undergoing regular screening with a screen-detected DCIS are quite likely to be overdiagnosed. Thus, given the substantial amount of overdiagnosis estimated by the CISNET models for DCIS in general, the model results support the safety and value of observational trials for low-risk DCIS.

REFERENCES

[1] Allegra CJ, Aberle DR, Ganschow P, Hahn SM, Lee CN, Millon-Underwood S, et al. National Institutes of Health State-of-the-Science Conference statement: Diagnosis and Management of Ductal Carcinoma In Situ September 22-24, 2009. J Natl Cancer Inst. 2010; 102(3):161-9.

[2] Kerlikowske K. Epidemiology of ductal carcinoma in situ. J Natl Cancer Inst Monogr. 2010; 2010(41):139-41.

[3] Virnig BA, Tuttle TM, Shamliyan T, Kane RL. Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes. J Natl Cancer Inst. 2010; 102(3):170-8.

[4] Gangnon RE, Sprague BL, Stout NK, Alagoz O, Weedon-Fekjaer H, Holford TR, et al. The contribution of mammography screening to breast cancer incidence trends in the United States: an updated age-period-cohort model. Cancer Epidemiol Biomarkers Prev. 2015; 24(6):905-12.

[5] Collins LC, Tamimi RM, Baer HJ, Connolly JL, Colditz GA, Schnitt SJ. Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. Cancer. 2005; 103(9):1778-84.

[6] Sanders ME, Schuyler PA, Dupont WD, Page DL. The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. Cancer. 2005; 103(12):2481-4.

[7] Mokbel K, Cutuli B. Heterogeneity of ductal carcinoma in situ and its effects on management. Lancet Oncol. 2006; 7(9):756-65.

[8] Esserman L, Yau C. Rethinking the Standard for Ductal Carcinoma In Situ Treatment. JAMA Oncol. 2015; 1(7):881-3.

[9] Narod SA, Iqbal J, Giannakeas V, Sopik V, Sun P. Breast Cancer Mortality After a Diagnosis of Ductal Carcinoma In Situ. JAMA Oncol. 2015; 1(7):888-96.

[10] Allred DC. Ductal carcinoma in situ: terminology, classification, and natural history. J Natl Cancer Inst Monogr. 2010; 2010(41):134-8.

[11] Silverstein MJ, Lagios MD. Choosing Treatment for Patients With Ductal Carcinoma In Situ: Fine Tuning the University of Southern California/Van Nuys Prognostic Index. JNCI Monographs. 2010; 2010(41):193-6.

[12] Kerlikowske K, Molinaro A, Cha I, Ljung BM, Ernster VL, Stewart K, et al. Characteristics associated with recurrence among women with ductal carcinoma in situ treated by lumpectomy. Journal of the National Cancer Institute. 2003; 95(22):1692-702.

[13] Sagara Y, Mallory MA, Wong S, Aydogan F, DeSantis S, Barry WT, et al. Survival benefit of breast surgery for low-grade ductal carcinoma in situ: a population-based cohort study. JAMA surgery. 2015; 150(8):739-45.

[14] Solin LJ, Gray R, Baehner FL, Butler SM, Hughes LL, Yoshizawa C, et al. A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast. J Natl Cancer Inst. 2013; 105(10):701-10.

[15] van Luijt PA, Heijnsdijk EA, Fracheboud J, Overbeek LI, Broeders MJ, Wesseling J, et al. The distribution of ductal carcinoma in situ (DCIS) grade in 4232 women and its impact on overdiagnosis in breast cancer screening. Breast Cancer Res. 2016; 18(1):47.

[16] van Ravesteyn NT, Stout NK, Schechter CB, Heijnsdijk EA, Alagoz O, Trentham-Dietz A, et al. Benefits and harms of mammography screening after age 74

years: model estimates of overdiagnosis. J Natl Cancer Inst. 2015; 107(7).

[17] Alagoz O, Ergun MA, Cevik M, Sprague BL, Fryback DG, Gangnon R, et al. The University Of Wisconsin Breast Cancer Epidemiology Simulation Model: An Update (Medical Decision Making. 2016; submitted.

[18] Huang X, Li Y, Song J, Berry DA. The MD Anderson CISNET Model for Estimating Benefits of Adjuvant Therapy and Screening Mammography for Breast Cancer: An Update. Medical Decision Making. 2016; submitted.

[19] Lee SJ, Li X, Huang H. Models for Breast Cancer Screening Strategies Updated for Ductal Carcinoma In Situ and Subgroups Medical Decision Making. 2016; submitted.

[20] Mandelblatt JS, Near AM, Miglioretti DL, Munoz D, Sprague BL, Trentham-Dietz A, et al. Common Model Inputs in Collaborative Breast Cancer Medical Decision Making. 2016; submitted.

[21] Schechter CB, Near AM, Jayasekera J, Chang Y, Mandelblatt JS. Structure, Function, and Applications of the Georgetown-Einstein (GE) Breast Cancer Simulation Model Medical Decision Making. 2016; submitted.

[22] van den Broek JJ, van Ravesteyn NT, Heijnsdijk EA, de Koning HJ. Estimating the effects of risk-based screening and adjuvant treatment using the MISCAN-Fadia continuous tumor growth model for breast cancer. Medical Decision Making. 2016; submitted.

[23] Segnan N, Minozzi S, Armaroli P, Cinquini M, Bellisario C, González- Lorenzo M, et al. Epidemiologic evidence of slow growing, nonprogressive or regressive breast cancer: A systematic review. International journal of cancer. 2016; 139(3):554-73.

[24] Burnside ES, Trentham-Dietz A, Kelcz F, Collins J. An example of breast cancer regression on imaging. Radiology case reports. 2006; 1(2):27-37.

[25] SEER*Stat Database: Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases. Linked To County Attributes - Total U.S., 1969-2010 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2012. based on the November 2011 submission.

[26] Larsen IK, Småstuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. European journal of cancer. 2009; 45(7):1218-31.

[27] Mandelblatt JS, Stout NK, Schechter CB, van den Broek JJ, Miglioretti DL, Krapcho M, et al. Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies. Ann Intern Med. 2016; 164(4):215-25.

[28] Mandelblatt JS, Cronin K, de Koning H, Miglioretti DL, Schechter CS, Stout N. Collaborative modeling of US breast cancer screening strategies. 2015.

[29] Gunsoy NB, Garcia-Closas M, Moss SM. Modelling the overdiagnosis of breast cancer due to mammography screening in women aged 40 to 49 in the United Kingdom. Breast Cancer Res. 2012; 14(6):R152.

[30] 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).

[31] Tan KH, Simonella L, Wee HL, Roellin A, Lim YW, Lim WY, et al. Quantifying the natural history of breast cancer. Br J Cancer. 2013; 109(8):2035-43.

[32] Yen MF, Tabar L, Vitak B, Smith RA, Chen HH, Duffy SW. Quantifying the potential problem of overdiagnosis of ductal carci-

noma in situ in breast cancer screening. Eur J Cancer. 2003; 39(12):1746-54.

[33] de Gelder R, Heijnsdijk EA, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiol Rev. 2011; 33(1):111-21.

[34] de Koning HJ, Draisma G, Fracheboud J, de Bruijn A. Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based on observed screen and clinical data. Breast Cancer Res. 2006; 8(1):202.

[35] Ozanne EM, Shieh Y, Barnes J, Bouzan C, Hwang ES, Esserman LJ. Characterizing the impact of 25 years of DCIS treatment. Breast Cancer Res Treat. 2011; 129(1):165-73.

[36] Seigneurin A, Francois O, Labarere J, Oudeville P, Monlong J, Colonna M. Overdiagnosis from non-progressive cancer detected by screening mammography: stochastic simulation study with calibration to population based registry data. BMJ. 2011; 343:d7017.

[37] Duffy SW, Dibden A, Michalopoulos D, Offman J, Parmar D, Jenkins J, et al. Screen detection of ductal carcinoma in situ and subsequent incidence of invasive interval breast cancers: a retrospective population-based study. Lancet Oncol. 2016; 17(1):109-14.

[38] Baxter NN, Virnig BA, Durham SB, Tuttle TM. Trends in the treatment of ductal carcinoma in situ of the breast. J Natl Cancer Inst. 2004; 96(6):443-8.

[39] Shiyanbola OO, Sprague BL, Hampton JM, Dittus K, James TA, Herschorn S, et al. Emerging trends in surgical and adjuvant radiation therapies among women diagnosed with ductal carcinoma in situ. Cancer. 2016; 122(18):2810-8.

[40] Francis A, Fallowfield L, Rea D. The LORIS Trial: Addressing overtreatment of ductal carcinoma in situ. Clin Oncol (R Coll Radiol). 2015; 27(1):6-8.

[41] Francis A, Thomas J, Fallowfield L, Wallis M, Bartlett JM, Brookes C, et al. Addressing overtreatment of screen detected DCIS; the LORIS trial. Eur J Cancer. 2015; 51(16):2296-303.

[42] Elshof LE, Tryfonidis K, Slaets L, van Leeuwen-Stok AE, Skinner VP, Dif N, 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.

[43] Warner E, Causer PA, Wong JW, Wright FC, Jong RA, Hill KA, et al. Improvement in DCIS detection rates by MRI over time in a high-risk breast screening study. The breast journal. 2011; 17(1):9-17.

[44] Lehman CD. Magnetic Resonance Imaging in the Evaluation of Ductal Carcinoma In Situ. JNCI Monographs. 2010; 2010(41):150-1.

[45] Kuhl CK. Why do purely intraductal cancers enhance on breast MR images? Radiology. 2009; 253(2):281-3.

Chapter 4

Comparing CISNET Breast Cancer Incidence and Mortality Predictions to Observed Clinical Trial Results of Mammography Screening from Ages 40 to 49.

Jeroen J. van den Broek, Nicolien T. van Ravesteyn, Jeanne S. Mandelblatt, Hui Huang, Mehmet Ali Ergun, Elizabeth S. Burnside, Helen C. Xu, Yisheng Li, Oguzhan Alagoz, Sandra J. Lee, Natasha K. Stout, Juhee Song, Amy Trentham-Dietz, Sylvia K. Plevritis, Sue M. Moss, Harry J. de Koning.

Medical Decision Making 2018;38:140S-50S

ABSTRACT

Background

The U.K. Age trial compared annual mammography screening of women ages 40 to 49 to no screening and found a statistically significant breast cancer mortality reduction at 10-year follow-up, but not at 17-year follow-up. The objective of this study was to compare the observed Age trial results to the Cancer Intervention and Surveillance Modeling Network (CISNET) breast cancer model predicted results.

Methods

Five established CISNET breast cancer models used data on population demographics, screening attendance, and mammography performance from the Age trial together with extant natural history parameters to project breast cancer incidence and mortality in the control and intervention arm of the trial.

Results

The models closely reproduced the effect of annual screening from ages 40 to 49 on breast cancer incidence. Restricted to breast cancer deaths originating from cancers diagnosed during the intervention phase, the models estimated an average 15% (range across models 13% to 17%) breast cancer mortality reduction at 10-year follow-up compared to 25% (95% CI 3% to 42%) observed in the trial. At 17-year follow-up, the models predicted 13% (range 10% to 17%) reduction in breast cancer mortality compared to the non-significant 12% (95% CI -4% to 26%) in the trial.

Conclusions

Overall, the models captured the observed effect of screening from age 40 to 49 on breast cancer incidence and mortality in the U.K. Age trial, suggesting that the model structures, input parameters, and assumptions about breast cancer natural history are reasonable for estimating the impact of screening on mortality in this age group.

INTRODUCTION

The breast cancer models of the Cancer Intervention and Surveillance Modeling Network (CISNET) synthesize data on breast cancer epidemiology, population demographics, screening accuracy, and treatment to simulate the impact of screening and treatment interventions on breast cancer incidence and mortality. Prior comparative modeling studies, i.e., cross-validations [1], by the CISNET models have illustrated the ability of the models to reproduce the trends in breast cancer incidence and mortality in the United States. [2-4] The models generated similar rankings of the effects of different screening scenarios and the relative impact of screening and treatment on breast cancer mortality. Moreover, the simulation results provided quantitative information about the harms and benefits of various screening strategies not examined in randomized clinical trials, and have been used by policy makers to inform decisions about breast cancer screening guidelines. [3, 5]

The consistency of previous collaborative modeling research provides a level of evidence for cross-validation. However, none of the prior collaborative CISNET research by the Breast Working Group has included external model validation. The International Society for Pharmacoeconomics and Outcomes Research in collaboration with the Society for Medical Decision Making (ISPOR-SMDM) recommends external model validation as part of good modeling practices, where external model validation is defined as, "the comparison of model predictions to observed event data not used in model development"[1]. The purpose of this paper is to conduct an external validation and compare CISNET breast cancer incidence and mortality predictions to observed clinical trial results of mammography screening from ages 40 to 49.

To date, the model parameters were primarily developed based on U.S. data on breast cancer epidemiology, screening, treatment, and population demographics.[6] Outcomes of our simulations indicated that offering screening to women in their fifties results in a more favorable ratio of benefits and harms than offering screening to women in their forties. [3, 7] This difference between the benefits and harms between these age groups, corresponds to the available evidence of screening women aged 50 and older [8] and the uncertainty about screening women in their forties, considering the inconclusive evidence from fewer studies, and the different guidelines for this age group [5, 9, 10]. Given the high prevalence of dense breast tissue, faster growing tumors, and inferior sensitivity of mammography in these younger women [11-13], it is important to validate the models for the effectiveness of screening in the forties. The U.K. 'Age' trial is a well-documented [14-20] trial, investigating the effect of annually screening women from ages 40 to 49 compared to no screening, and provided a unique opportunity to externally validate the CISNET breast cancer models for screening in the forties.

In this study, we present the first external validation performed by the CISNET breast cancer models that use different structures and assumptions about breast cancer natural history to project the impact of screening. We compare breast cancer incidence and mortality predictions to the observed results from the U.K. Age trial. The findings from this study are intended to inform CISNET model users as they can account for this information when considering and interpreting future model outcomes.

METHODS

The U.K. Age trial was the only randomized controlled trial designed specifically to investigate the effect of annual mammography screening from ages 40 to 49. Between October 1990 and September 1997, 160,836 women aged 40-41 were randomly assigned in a ratio of 1 : 2 to either the intervention group or the control group. The 53,883 women in the intervention arm were offered annual screening by mammography, and the 106,953 women in the control arm received usual care (no screening). We collaborated with the Age trial investigators to obtain the observed de-identified data from the trial.

Simulation models

Five CISNET breast cancer models were included in this analysis: Model D (Dana-Farber), Model E (Erasmus), Model M (MD Anderson), Model S (Stanford), and Model W (Wisconsin-Harvard). These models have been developed independently within CISNET over the past 15 years and are described in detail elsewhere [21-25]. Briefly, women are born in a breast cancer-free stage, some women develop a tumor that may progress to a pre-clinical stage where it could be screen-detected in its pre-clinical sojourn time, or be diagnosed with breast cancer due to clinical symptoms. Once diagnosed with breast cancer, women receive age-, stage-, and biomarker-specific treatment. Breast cancer incidence and mortality projections depend on age, start and stopping ages of screening, screening frequency, mammography screening performance, stage at diagnosis, estrogen receptor (ER) and Human Epidermal growth factor Receptor 2 (HER2) status of the tumor, breast cancer treatment, and factors related to the natural history of breast cancer (Tables 1 & 2). However, since the Age trial did not collect HER2 status, the models did not simulate HER2 specific molecular subtypes of breast cancer. The models adopt a 'parallel universe' approach; the same population of women is simulated twice: in one scenario women were invited to annual screening in the forties (intervention group), and in the second scenario women did not receive any screening in the forties (control group).

Table 1 Key differences and similarities between the CISNET breast models.

Model	D	E	M	S	W
Model type	Analytic, Parallel universe	Simulation, Parallel universe	Bayesian, Parallel universe	Simulation, Parallel universe	Simulation, Parallel universe
Natural history modeled as	State-transition	Continuous tumor growth	Bayesian model	Continuous tumor growth	Continuous tumor growth
Tumor inception	Start of the sojourn time	Prior to start of sojourn time	N/A	Prior to start of sojourn time	Start of the sojourn time
DCIS included	Since 2014	Yes	Yes	No	Yes
Tumor ER status	Yes	Yes	Yes	Yes	Yes
Screen detection depends on	Modality, age, density, frequency	Tumor size, modality, age, density, frequency	Modality, age, frequency	Tumor size, ER status, age, hormone repl., frequency	Tumor size, modality, age, density, frequency
Screening benefit	Stage shift	Detection at smaller tumor size	Stage shift, beyond stage shift	Stage shift, smaller tumor size	Younger age, smaller tumor size
Estimation of over diagnosis	Difference screen & no-screen	Difference screen & no-screen	Difference screen & no-screen	Difference screen & no-screen	Difference screen & no-screen
Treatment benefit	Hazard reduction	Cure fraction, larger fatal diameter	Cure fraction, hazard reduction,	Hazard reduction, non-proportional	Cure fraction
Death from breast cancer determined by	Survival from BC < survival other cause mortality	Fatal diameter, survival from BC < survival other cause mortality	Survival from BC < survival other cause mortality	Survival from BC < survival other cause mortality	Survival from BC < survival other cause mortality

Model type

Analytic: Analytical approach to estimate the impact of mammography screening and treatment on incidence and mortality of breast cancer.

Simulation: Stochastic simulation is based on the Monte Carlo method and use of random numbers.

Bayesian: The model does not include a natural history and estimates prior probability distributions for all unknown parameters.

Parallel universe: Screening and treatment is modeled in a parallel universe, implying that the same population is simulated twice: once to determine the impact of breast cancer without screening, and once to determine the impact of breast cancer with screening.

Breast cancer natural history and breast cancer death

ER**:** Onset and progression of breast cancer is different for Estrogen Receptor positive and negative tumors.

Tumor stage transition**:** Tumor progression is modeled as transitions between different stages of breast cancer.

Continuous tumor growth**:** Tumors grow continuously after tumor onset.

Death from breast cancer**:** Once diagnosed with breast cancer, a survival until breast cancer death is competing with the other cause mortality survival. That is, breast cancer death occurs only if the patient does not die from other causes.

Screening & Treatment

Sensitivity**:** Sensitivity can be used directly or indirectly (e.g., when translated to tumor size).

Over diagnosis**:** The detection and diagnosis of a condition that would not go on to cause symptoms or breast cancer death in a woman's lifetime.

Hazard reduction**:** Reduction in breast cancer mortality hazard, calculated by 1 minus the hazard ratio for the different treatment regimes.

Cure fraction**:** If hazard rate reduction is not a model input, it is translated into a cure fraction.

Table 2 Model inputs used for the Age trial simulation:

Model Input	Description	Source
Population demographics		
Birth cohort	Birth years of women participating in the Age trial	Age trial
Life years	Number of life years by trial arm by age	Age trial
Natural history of breast cancer		
Incidence	Control arm incidence (incidence in the absence of screening)	Age trial
Tumor onset	The moment tumors start to grow (tumor inception)	CISNET[1]
Sojourn time	Time between when a cancer is first screen-detectable and cancer diagnosis in the absence of screening.	CISNET[2]
Tumor progression	Tumor growth, tumor progression and regression affect tumor sojourn times and breast cancer survival.	CISNET[3]
Estrogen receptor distribution	Age-specific ER positive and ER negative distributions	U.K.[4]
Breast cancer screening		
Attendance	Adherence to annual screening in the intervention arm	Age trial
Sensitivity	Probability that the screen will be positive among women with breast cancer by age, screening round (first vs. subsequent)	Age trial
Mammography	Two-view mammography for first screens, for all subsequent screens one-view mammography	Age trial
Breast cancer treatment		
Treatment dissemination	Breast cancer treatment by age, stage and ER-status	BASO[5]
Effectiveness	Hazard reduction breast cancer mortality by age and ER-status	EBCTCG[6]
Breast cancer survival		
Survival	Breast cancer survival by age, stage and ER-status	CISNET[7]
Other-cause mortality	Probability of dying from causes other than breast cancer	U.K.[8]

[1-3] Tumor onset, sojourn time and tumor progression are model-specific parameters. These, and other model-specific assumptions about breast cancer natural history are described elsewhere [6, 21-25].
[4] Estrogen receptor status comes from observed U.K. data [26].
[5] The treatment dissemination was derived from BASO reports [26] published by the NHSBSP.
[6] Treatment effectiveness / hazard reduction for breast cancer death was published by the Early Breast Cancer Trialists Collaborative Group (EBCTCG) that included the U.K. trials [27]
[7] Breast cancer survival by age and ER status from the UK is not available for the time period of the trial, the existing survival in the models which is based on U.S. data was used.
[8] Other cause mortality was taken from the Human Mortality Database [30] with breast cancer deaths removed.

As summarized in Table 1, the models differ in the ways they approximate unobservable events in the natural history of breast cancer. In model D, tumors progress via discrete state transitions [23], models E, S and W have continuous tumor growth [21, 22, 25], and model M uses Bayesian simulation [24] and does not have a natural history component. In models D and W, tumors are technically screen-detectable from the moment at tumor inception. Models E and S start simulating tumors at small tumor sizes, prior to the start of the sojourn time, when tumors are not yet screen-detectable by film or digital mam-

mography. Screening benefit in models D and M is modeled as a stage shift to earlier stage breast cancer, with the latter model including an additional benefit of screening beyond stage shift. The benefits of screening in models E, S and W are simulated by the detection of tumors at smaller sizes than at clinical diagnosis in the absence of screening. (Table 1)

Model inputs

The Age trial data that the CISNET models obtained included control arm incidence in the absence of screening, mammography screening performance, screening attendance patterns, and demographic data such as life years and the distribution of birth years of women participating in the trial (Table 2). In the Age trial, data were not collected for breast cancer treatment. To fill this gap we modeled the breast cancer treatment dissemination between 1991 and 2006, the intervention period of the trial, based on reports from the British Association of Surgical Oncology [26]. The effectiveness of breast cancer treatment was taken from analyses by the Early Breast Cancer Trialists' Collaborative Group (EBCTCG) that included trials conducted in the U.K. [27]. Model parameters related to the natural history of breast cancer such as tumor onset and tumor growth were based on the original CISNET parameters and no calibration was performed to the results from the Age trial.

Simulation of the Age trial

The women who participated in the Age trial were born between 1950 and 1957, therefore, we simulated the 1950-1957 birth cohort. In the trial, two thirds of women aged 40 to 41 were randomized to the control group and were not invited to any screening in their forties. The models simulated 2 to 10 million women in each arm of the trial as they were not limited by practical issues concerning invitations and the number of women who can be included in the simulation of the trial. (Table 3) Any unscheduled screening in the control group was primarily a consequence of clinical symptoms and not because of routine screening [17], so we did not model screening contamination in the control group explicitly.

We used the control arm incidence as model input for a baseline projection of breast cancer incidence in the absence of screening. The models then overlaid the screening parameters according to the observed screening attendance patterns of the 53,883 women in the intervention group of the Age trial [18]. The percent uptake of invitations increased by screening round while the absolute number of invitations sent to the women in the trial decreased by almost 50% near the end of the intervention period and consequently the absolute number of women who were screened decreased as well. [18] The models accounted for this by simulating the decrease in the number of women who were screened by age. The first analog mammogram in the trial included two views, and

all subsequent mammograms were single-view, similar to the standard practice in the U.K. at the time of the trial. Screen detection of pre-clinical breast cancer was modeled on the basis of observed sensitivity data published by the trial investigators [16].

The U.K. treatment dissemination developed for this project indicated whether a breast cancer is treated with hormone therapy and/or chemotherapy after surgical removal of the tumor. Overall, ER-positive breast cancers were primarily treated with hormone therapy and ER-negative breast cancers with chemotherapy. Since, the trial did not collect HER2 status, and Trastuzumab (Herceptin) was not yet disseminated in the U.K. at the time of the trial, it was not included in the treatment regimens.

Table 3 Number of women included in the control and intervention group

	Nr. of women in the control arm	**Nr. of women in the intervention arm**
Age trial	106,953	53,883
Model D	N/A*	N/A
Model E	10,000,000	10,000,000
Model M	4,000,000	4,000,000
Model S	5,000,000	5,000,000
Model W	2,000,000	2,000,000

All models simulated at least about 20 times as many women in the control group and 40 times as many women as in the intervention group. The number of women simulated was selected by each model to balance feasibility of simulation time with model output that yields relatively smooth incidence and mortality curves.

*Model D uses entirely analytical formulations to evaluate the impact of screening and treatment on breast cancer incidence and mortality, i.e., the number of women simulated does not apply to Model D.

Analysis

Model predictions were compared to breast cancer incidence and mortality observations from the Age trial by arm without calibrating the natural history parameters of the models to the trial. In addition, we compared the number of mammograms in the intervention group to that of the Age trial to investigate whether any differences in model predictions were related to variations in the number of mammograms.

We compared model outcomes to those from the trial at 10-year and 17-year follow-up, corresponding to the most recent analysis by the Age trial investigators [15]. The trial used 'incidence based mortality' to measure the effect of screening and treatment on breast cancer mortality. This implies, only counting cancer deaths that originated from cancers diagnosed during the intervention phase of the trial (ages 40 to 49). This is necessary because all women from both the intervention and control group 'rolled' into the national U.K. breast cancer screening program at age 50 and were invited to screening once every three years. For example, if at age 54 there would be fewer breast cancer deaths among women randomized to the intervention group than among the women

in the control group, one could conclude that the intervention of annual screening in the forties effectively reduced breast cancer mortality at age 54. However, because all women 'rolled' into the national screening program at age 50, it may be the case that the breast cancer deaths prevented at age 54 were actually from breast cancers diagnosed by screening at age 50 as part of the national program and not by the trial's annual screening intervention in the forties. Therefore, the trial and the models only used breast cancer deaths from cancers diagnosed during the intervention phase to measure the effect of annual screening in the forties on breast cancer mortality.

The confidence intervals associated with the mortality reduction observed in the Age trial at 10-and 17-year follow-up are useful as these are mainly influenced by the finite number of women included in the trial. The CISNET models have not included confidence intervals on their results given the millions of women simulated per trial arm. The model estimates will have a negligible range, given that the model outcomes are based on simulations of millions of women, each with varying combinations of variables constituting the life history, and sampled across the distribution of each variable. However, the model results do have uncertainty due to assumptions about unobservable parameters and structural uncertainties that are addressed. The use of multiple models provides a range of results that captures this structural uncertainty and could be considered to provide information comparable conceptually to a confidence interval.

RESULTS

Breast cancer incidence

The average simulated invasive breast cancer incidence among women aged 40 to 49 in the control arm was 131 per 100,000 women (range across models 124 - 138) compared to 132 observed in the Age trial (Figure 1). The modeled ductal carcinoma in situ (DCIS) incidence was 11 per 100,000 women on average (range across models 7 - 17), and equivalent to the 11 per 100,000 observed in the Age trial.

The average number of mammograms per woman in the intervention arm of the simulated trial was 5.2 (range across models 4.9 – 5.4) compared to 4.84 in the Age trial. Modeled invasive breast cancer incidence in the intervention arm increased by age and was an average of 135 per 100,000 among women aged 40 to 49 (range across models 131 - 141). This is consistent with the pattern for the 139 invasive breast cancers diagnosed per 100,000 women in the trial (Figure 2). DCIS intervention arm incidence varied more across the models (range 18 - 38) and with 27 diagnoses on average, higher than the 21 DCIS diagnoses per 100,000 women in the trial. Models with continuous tumor growth (Models E and W) and models with tumor inception prior to the start of the tumor's sojourn time (Model E) tend to have the highest incidence of screen-detected DCIS.

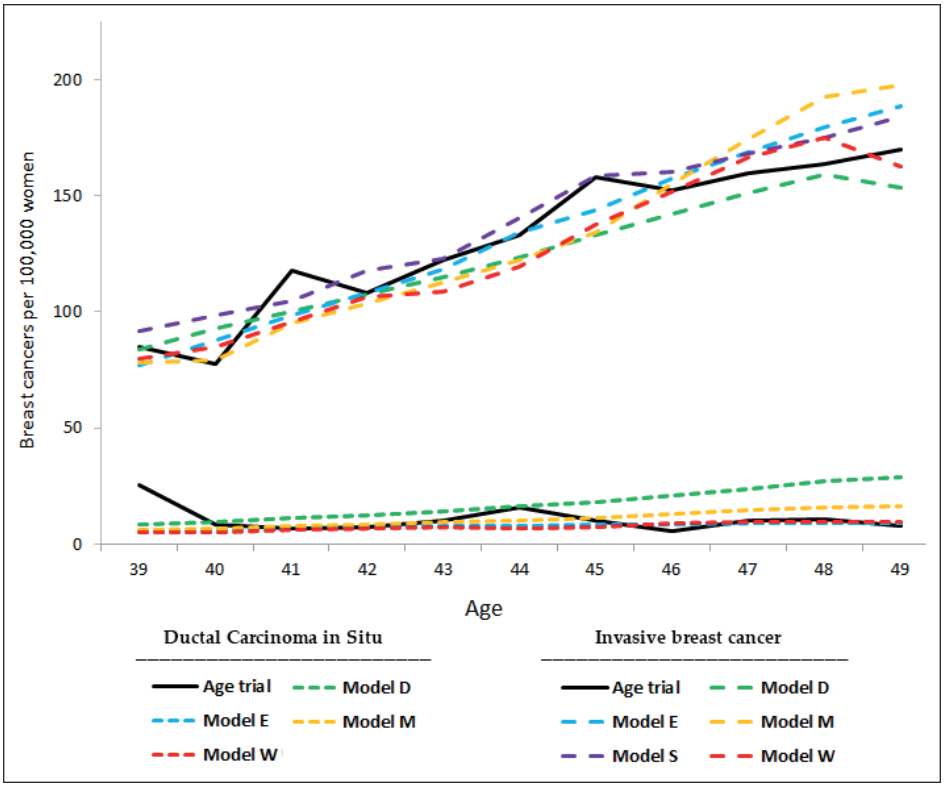


Figure 1 Control group breast cancer incidence (DCIS and invasive separately) per 100,000 women, compared to the Age trial.

Both the model results and the observed Age trial data included a small peak (Figure 3) at age 40 in screen-detected breast cancers due to the detection of (prevalent) cases on the first mammogram, the only two-view mammogram in the trial with better sensitivity than subsequent screens (Table 4). This was the only age during the trial at which the rate of screen detected cancers was higher than the rate of clinically diagnosed cancers in the intervention group. The average rate of screen-detected DCIS and invasive breast cancers in the intervention arm in the age range 40 - 49 was 69 per 100,000 women in the Age trial, compared to the models' average of 75 (range 63 - 89). The rate of clinically diagnosed cases (DCIS and invasive breast cancers) in the intervention arm was 97 in the trial and 93 in the models (range 82 - 99). Regardless of mode of detection, the rate of breast cancers diagnosed in the intervention arm between ages 40 - 49 was 161 per 100,000 women on average (range across models 154 - 169) and similar to 162 in the Age trial.

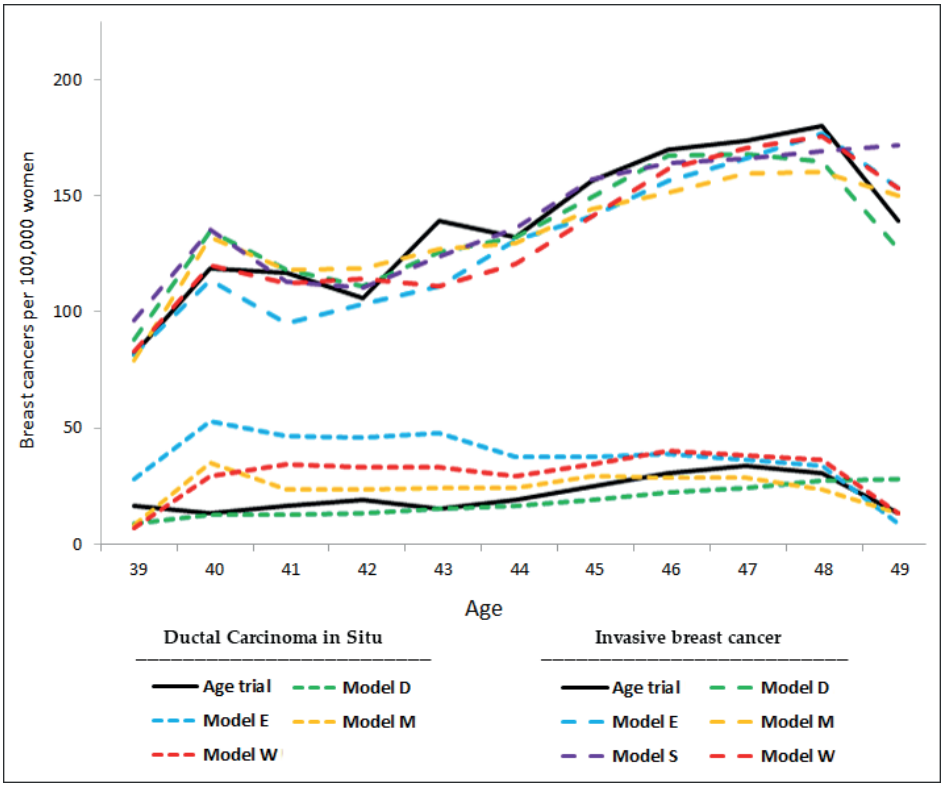


Figure 2 Intervention group breast cancer incidence (DCIS and invasive separate) per 100,000 women, compared to the Age trial.

Breast cancer mortality

Among breast cancers diagnosed between ages 40 to 49, the Age trial found a total of 83 breast cancer deaths in the first 10 years of follow-up in the intervention arm (16 breast cancer deaths per 100,000 women) and 219 breast cancer deaths in the control arm (21 per 100,000 women). At 10-year follow-up, the rate of breast cancer deaths per 100,000 women predicted by the models was 20 on average (range across models 17 to 22) in the intervention arm, and 23 (range across models 20 to 25) in the control arm (Table 5). The number of breast cancer deaths predicted by the different models consistently somewhat higher in both arms than in the trial.

On average, the modeled breast cancer mortality reduction due to screening was 15% (range across models 13% to 17%) at 10-year follow-up vs. 25% (95% CI 3% to 42%) observed in the Age trial. At 17-year follow-up, the models predicted 13% (range across models 10 – 17%) breast cancer mortality reduction when restricted to breast cancer deaths that originated from breast cancers diagnosed during the intervention phase (incidence-based mortality) vs. 12% (95% CI -4% to 26%) observed in the trial (Table 6).

Table 4 Sensitivity of screening in the Age trial and in the models.

	First screen (two view mammography)	Subsequent screens (single view mammography)
Age trial	73.6	55.2
Model D	73.6	55.3
Model E	72.5	55.7
Model M*	-	-
Model S	75.5	59.0
Model W	67.7	59.6

*Model M is a Bayesian without a natural history part and a woman's disease status is unknown. As a result sensitivity is not applicable. Model M simulates screen- and clinically-detected incidences without knowing the true disease status.
Sensitivity of screening and screen detection is modeled differently in various models. In the continuous tumor growth models E, S, and W screen detection of tumors is simulated by transforming sensitivity to a threshold tumor size at which tumors can be screen detected. On the other hand, model D uses sensitivity of screening by simulating a shift to a less-advanced stage of breast cancer.

Table 5 Breast cancer mortality outcomes at 10-years follow-up.

	Mammograms per woman	Breast cancer deaths per 100,000 women		Rate ratio BC deaths	Breast cancer ** mortality reduction
		intervention group	control group		
Age trial	4.84	16	21	0.75	25% (3 to 42%) *
Model average	5.23	19	23	0.85	15.3% [range 13-17%]
Model D	5.30	17	20	0.83	17.0%
Model E	4.90	20	25	0.83	16.9%
Model M	5.43	20	23	0.86	13.6%
Model S	5.29	22	25	0.87	13.2%
Model W	5.23	19	22	0.84	16.0%

* 95% confidence interval in parentheses
** The Age trial measured the effect of annual screening of women aged 40 to 49 on breast cancer mortality. Therefore, the trial and the simulation models excluded breast cancer deaths that occurred in women diagnosed with breast cancer before age 40 and after age 49.

Table 6 Breast cancer mortality outcomes at 17-years follow-up, restricted to breast cancer deaths that stem from cancers diagnosed during the intervention phase.

	Mammograms per woman	Breast cancer deaths per 100,000 women		Rate ratio BC deaths	Breast cancer ** mortality reduction
		intervention group	control group		
Age trial	4.84	19	22	0.88	12% (-4 to 26%) *
Model average	5.23	20	23	0.87	13.2% [range 10 -17%]
Model D	5.30	20	22	0.90	9.7%
Model E	4.90	18	22	0.83	17.1%
Model M	5.43	20	24	0.85	15.2%
Model S	5.29	21	24	0.89	11.0%
Model W	5.23	18	21	0.86	13.7%

* 95% confidence interval in parentheses

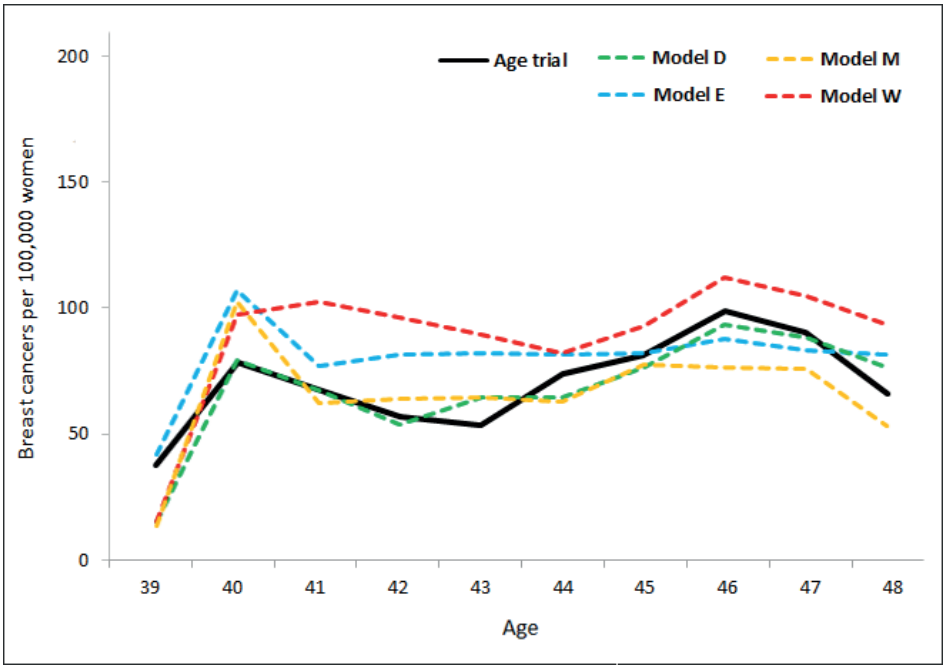


Figure 3 Intervention group (screen detected) breast cancer incidence per 100,000 women. Screening ceased at age 48 in the Age trial.

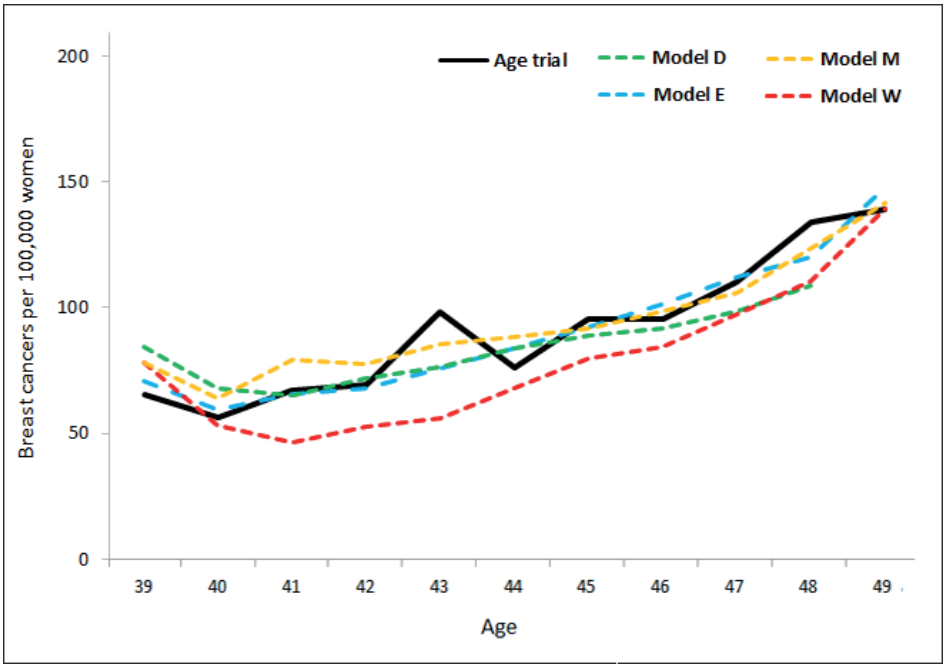


Figure 4 Intervention group (clinically diagnosed) breast cancer incidence per 100,000 women.

The models with either tumor onset at tiny tumor sizes prior to the start of the sojourn time and on average slow tumor progression (Model E), or with tumor cure fractions for treatment benefit (Models E, M and W) maintained their 10-year follow-up breast cancer mortality reduction prediction at 17-year follow-up, whereas mortality reduction in the trial decreased. Similar to the Age trial, the models showed a turning point around age 50 where the increase in the cumulative number of breast cancer deaths averted started to diminish (Figure 5).

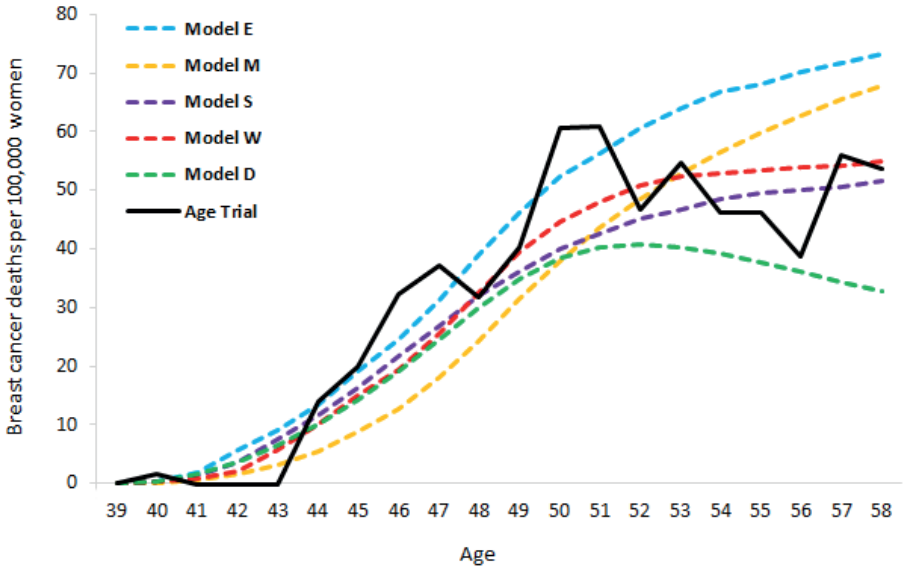


Figure 5 Cumulative breast cancer deaths averted per 100,000 women*. *Cumulative breast cancer deaths averted only using breast cancer deaths from cancers diagnosed in the intervention period per 100.000 women. Calculated by the rate of breast cancer deaths in the control group minus the rate of breast cancer deaths in the intervention group.

DISCUSSION

This is the first collaborative CISNET breast cancer study comparing model predictions to observed clinical trial results not used in the development of any model parameters. The results indicate that all five models estimate the long-term effect of annual screening between the ages of 40 to 49 well within the observed confidence intervals of the U.K. Age trial. The impact of screening on breast cancer mortality was also internally consistent with individual model structures regarding the natural history of breast cancer.

The ISPOR-SMDM Modeling Good Research Practices TaskForce-7 [1] states that predictive and external validation are the strongest forms of model validation as decision-

makers can account for this information when considering model outcomes. In the past, the breast CISNET models have illustrated accurate predictions of molecular-subtype-specific and overall U.S. breast cancer incidence and mortality trends. [3, 4, 28] This study extends these prior cross-validations by independently estimating the observed results from a U.K. randomized controlled trial.

All models reproduced the trend in control group breast cancer incidence from ages 40 to 49, implying that the extant model structures and assumptions about the natural history of breast cancer in the absence of screening are reliable. Despite the intensive (annual) screening intervention, the models predicted more clinically diagnosed than screen-detected breast cancers in the intervention group. This was likely to be explained by the relatively low sensitivity of all subsequent single-view mammograms that followed after the more sensitive prevalent two-view mammogram, and the decrease in the number of women screened by screening round in the trial [18]. Although the models utilized different mechanisms such as a threshold tumor size (Models E, S, and W) or stage shift (Models D and M) to simulate screen detection of pre-clinical breast cancer, they were all able to accurately estimate the impact of screening from ages 40 to 49 on invasive breast cancer incidence.

The effect of screening and treatment on breast cancer mortality was underestimated by all models at 10-year follow-up compared to the reduction observed in the Age trial. Since all models accurately predicted breast cancer incidence, and the fact that the underestimation of the mortality reduction was present across all models, it might be explained by a common model input not related to screening. Specifically, the derived U.K. treatment dissemination may not represent the actual treatment received by women diagnosed with breast cancer in the trial. This is in line with the higher rate of breast cancer deaths predicted by the models in the control arm in the absence of screening.

After 10 years of follow-up, breast cancer mortality reduction observed in the trial decreased and lost significance, whereas most models predicted a fairly constant mortality reduction between 10- and 17-year follow-ups. Previous analysis of the CISNET models [29] illustrated that Model D, with tumor inception at the start of the sojourn time, has fast tumor progression on average, and Model E, with tumor inception prior to the start of the sojourn time, has the slowest tumor progression on average. These individual model structures affect the pattern in breast cancer deaths averted after age 49 when screening ceased, because cancers diagnosed in the control arm caused breast cancer death at a younger age in Model D and at a later age in Model E. Consequently, mortality reduction due to screening was greater at later ages (between 10- and 17-year follow-up) in Model E than in Model D. While the model structure of Model S is similar to that of Model E, Model S does not include DCIS, which implies no possible benefit in terms of mortality reduction from screen-detected DCIS. However, these otherwise screen-detected DCIS cases will likely be diagnosed as local stage small invasive tumors

(size <1 cm.) in Model S with relatively high, and similar survival as DCIS cases. Model W is unique in that it simulates tumors with a limited malignant potential [25]. This may have resulted in a substantial amount of screen-detected tumors that did not cause breast cancer death during the 17-year follow-up. Consequently, Model W's mortality reduction decreased slightly after age 49 despite their high rate of screen-detected cancers in the forties.

In summary, at 10- and 17-year follow-up, the models reproduced the effects of annual screening in the forties on breast cancer mortality well within the trial's confidence intervals [15]. In terms of model validation, it can be questioned what these model outcomes imply, as it is quite common to have relatively wide confidence intervals in randomized trials on cancer screening. The wide confidence intervals in the trial are partly due to the limited number of women included and breast cancer deaths observed in the trial. The models' outcomes may be less sensitive to the number of women that are simulated because they simulated at least 2 million women in each arm of the trial, notwithstanding the fact that the models are ultimately based on observed data as well.

The CISNET breast models used Age trial-specific model inputs and data sources applicable to the U.K., but we can still draw a comparison between the outcomes of this study and published results from a recent collaborative modeling study on screening in the United States [3]. In the U.S. study, we simulated annual screening from age 40 to 74 and compared it to annual screening from age 50 to 74. This implies that the difference in breast cancer deaths averted between these two scenarios over the women's lifetime, is due to the effect of annual screening in the forties. Similar to the results of this analysis, the outcomes indicated that Model M and E avert the most breast cancer deaths from annual screening in the forties followed by Models W, S and D. In other words, the ranking of the models is fairly consistent when applied in another country with different model inputs.

This study presented the first external comparison performed by multiple breast cancer simulation models applied in a different country and setting. A strength of this analysis is that we used detailed observed de-identified trial data as model inputs. Another important strength is that we performed an independent external validation [1] in which no model calibration was performed to ensure credibility of the model outcomes.

Although the CISNET breast models used Age trial-specific model inputs and data sources applicable to the U.K., there were several limitations in this analysis. The trial did not collect data on breast cancer molecular sub-type and treatment, these were estimated based on U.K. data. It is possible that these data underestimated the actual treatment patterns of trial participants. That this is the case is suggested by the fact that all models had estimates for mortality reduction that were consistently lower than the point estimate from the trial. Moreover, when the models simulated the Age trial assuming all women received the most effective therapy available, the average model

estimate was very close to trial result. [3] The lack of precision in being able to model the treatment of women in the Age trial is likely to have contributed more to the differences between model and trial results than the screening and natural history components of the models. Other limits include the fact that the models did not explicitly simulate screening in the control arm because the reported amount of unscheduled screening was low, and primarily due to symptomatic reasons. [17] While this may not affect conclusions of the simulations, it is a limitation.

The quantitative information in this study demonstrated how well the models reproduced the effects of annual screening from ages 40 to 49 on breast cancer incidence and mortality. In the future, the CISNET models could simulate the impact of what would have happened if two-view digital mammography had been used for all screening examinations in the Age trial, simulate the impact of different patterns of screening attendance, provide estimates on overdiagnosis, and estimate the lifetime effects of different screening programs offered to women in their forties. The demonstration that the models can reproduce observed external trial results should increase confidence in models results to inform policy decisions about breast cancer screening.

REFERENCES

[1] Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Med Decis Making. 2012; 32(5):733-43.

[2] Plevritis S, Munoz D, Kurian A, Alagoz O, Near AM, Stout NK, et al. Contributions of screening and systemic therapy to molecular subtype specific U.S. breast cancer mortality from 2000 to 2010. 2016; submitted.

[3] Mandelblatt JS, Stout NK, Schechter CB, van den Broek JJ, Miglioretti DL, Krapcho M, et al. Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies. Ann Intern Med. 2016; 164(4):215-25.

[4] Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005; 353(17):1784-92.

[5] Siu AL, On behalf of the U. S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2016; 164(4):279-96.

[6] Mandelblatt JS, Near AM, Miglioretti DL, Munoz D, Sprague BL, Trentham-Dietz A, et al. Common Model Inputs used in CISNET Collaborative Breast Cancer Modeling Medical Decision Making. 2016; submitted.

[7] van Ravesteyn NT, Miglioretti DL, Stout NK, Lee SJ, Schechter CB, Buist DS, et al. Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: a comparative modeling study of risk. Ann Intern Med. 2012; 156(9):609-17.

[8] Independent U. K. Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet. 2012; 380(9855):1778-86.

[9] American College of O-G. Practice bulletin no. 122: Breast cancer screening. Obstet Gynecol. 2011; 118(2 Pt 1):372-82.

[10] Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. JAMA. 2015; 314(15):1599-614.

[11] Sprague BL, Gangnon RE, Burt V, Trentham-Dietz A, Hampton JM, Wellman RD, et al. Prevalence of mammographically dense breasts in the United States. J Natl Cancer Inst. 2014; 106(10).

[12] Anders CK, Fan C, Parker JS, Carey LA, Blackwell KL, Klauber-DeMore N, et al. Breast carcinomas arising at a young age: unique biology or a surrogate for aggressive intrinsic subtypes? J Clin Oncol. 2011; 29(1):e18-20.

[13] Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med. 2007; 356(3):227-36.

[14] Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L, et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. Lancet. 2006; 368(9552):2053-60.

[15] Moss SM, Wale C, Smith R, Evans A, Cuckle H, Duffy SW. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. Lancet Oncol. 2015; 16(9):1123-32.

[16] Moss S, Thomas I, Evans A, Thomas B, Johns L, Trial Management G. Randomised controlled trial of mam-

mographic screening in women from age 40: results of screening in the first 10 years. Br J Cancer. 2005; 92(5):949-54.

[17] Kingston N, Thomas I, Johns L, Moss S, Trial Management G. Assessing the amount of unscheduled screening ("contamination") in the control arm of the UK "Age" Trial. Cancer Epidemiol Biomarkers Prev. 2010; 19(4):1132-6.

[18] Johns LE, Moss SM, Trial Management G. Randomized controlled trial of mammographic screening from age 40 ('Age' trial): patterns of screening attendance. J Med Screen. 2010; 17(1):37-43.

[19] Johns LE, Moss SM, Age Trial Management G. False-positive results in the randomized controlled trial of mammographic screening from age 40 ("Age" trial). Cancer Epidemiol Biomarkers Prev. 2010; 19(11):2758-64.

[20] Moss S. A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40. Trial Steering Group. J Med Screen. 1999; 6(3):144-8.

[21] van den Broek JJ, van Ravesteyn NT, Heijnsdijk EA, de Koning HJ. Estimating the effects of risk-based screening and adjuvant treatment using the MISCAN-Fadia continuous tumor growth model for breast cancer. Medical Decision Making. 2016; submitted.

[22] Munoz D, Xu C, Plevritis S. A Molecular Subtype-Specific Stochastic Simulation Model of US Breast Cancer Incidence and Mortality Trends from 1975 to 2010 Medical Decision Making. 2016; submitted.

[23] Lee SJ, Li X, Huang H. Assessment of Screening Strategies for Breast Cancer: Model-Based Approaches Medical Decision Making. 2016; submitted.

[24] Huang X, Li Y, Song J, Berry DA. The MD Anderson CISNET Model for Estimating Benefits of Adjuvant Therapy and Screening Mammography for Breast Cancer: An Update. Medical Decision Making. 2016; submitted.

[25] Alagoz O, Ergun MA, Cevik M, Sprague BL, Fryback DG, Gangnon R, et al. The University Of Wisconsin Breast Cancer Epidemiology Simulation Model: An Update (Medical Decision Making. 2016; submitted.

[26] National Health Service Breast Screening Programme & British Association of Surgical Oncology. Audit of screen detected breast cancers. 2000-2007.

[27] Early Breast Cancer Trialists' Collaborative G, Peto R, Davies C, Godwin J, Gray R, Pan HC, 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.

[28] Mandelblatt JS, Cronin KA, Bailey S, Berry DA, Koning H J de, Draisma G, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms Ann Int Med. 2009; 151:738-47.

[29] van den Broek JJ, van Ravesteyn NT, Cevik M, Schechter CB, Lee SJ, Munoz D, et al. Comparing CISNET breast cancer models using Maximum Clinical Incidence Reduction methodology. Medical Decision Making. 2016; submitted.

[30] Barbieri, Magali et al. "Data Resource Profile: The Human Mortality Database (HMD)." International Journal of Epidemiology 44.5 (2015): 1549–1556. PMC. Web. 27 Apr. 2017..

Chapter 5

Comparing CISNET Breast Cancer Models Using the Maximum Clinical Incidence Reduction Methodology.

Jeroen J. van den Broek, Nicolien T. van Ravesteyn, Mucahit Cevik, Clyde B. Schechter, Sandra J. Lee, Hui Huang, Yisheng Li, Diego F. Munoz, Sylvia Plevritis, Jeanne Mandelblatt, Harry J. de Koning, Natasha K. Stout*, Marjolein van Ballegooijen*.

Medical Decision Making 2018;38:112S-25S.

ABSTRACT

Background

Collaborative modeling has been used to estimate the impact of potential cancer screening strategies worldwide. A necessary step in the interpretation of collaborative cancer screening model results is to understand how model structure and model assumptions influence cancer incidence and mortality predictions. In this study we examined the relative contributions of the pre-clinical duration of breast cancer, the sensitivity of screening, and the improvement in prognosis associated with treatment of screen-detected cases to the breast cancer incidence and mortality predictions of five Cancer Intervention and Surveillance Modeling Network (CISNET) models.

Methods

To tease out the impact of model structure and assumptions on model predictions, the Maximum Clinical Incidence Reduction (MCLIR) method compares changes in the number of breast cancers diagnosed due to clinical symptoms and cancer mortality between 4 simplified scenarios: 1) no-screening; 2) one-time perfect screening exam that detects all existing cancers and perfect treatment (i.e., cure) of all screen-detected cancers; 3) one-time digital mammogram and perfect treatment of all screen-detected cancers; and 4) one-time digital mammogram and current guideline-concordant treatment of all screen-detected cancers.

Results

The five models predicted a large range in maximum clinical incidence (19%-71%) and in breast cancer mortality reduction (33%-67%) from a one-time perfect screening test and perfect treatment. In this perfect scenario, the models with assumptions of tumor inception prior to when it is first detectable by mammography predicted substantially higher incidence and mortality reductions than models with assumptions of tumor onset at the start of a cancer's screen-detectable phase. The range across models in breast cancer clinical incidence (11%-24%) and mortality reduction (8%-18%) from a one-time digital mammogram at age 62 with observed sensitivity and current guideline-concordant treatment was considerably smaller than achievable under perfect conditions.

Conclusions

The timing of tumor inception and its effect on the length of the pre-clinical phase of breast cancer had substantial impact on the grouping of the models based on their predictions for clinical incidence and breast cancer mortality reduction. This key finding about the timing of tumor inception will be included in future CISNET breast analyses to enhance model transparency. The MCLIR approach should aid in the interpretation of variations in model results and could be adopted in other disease screening settings to enhance model transparency.

INTRODUCTION

Collaborative modeling can enhance the rigor of modeling research through the use of multiple independent models to answer the same research question. The National Cancer Institute-funded Cancer Intervention and Surveillance Modeling Network (CISNET) was established in 2000 to use collaborative modeling to improve our understanding of the impact of cancer prevention, screening, and treatment dissemination on population trends in cancer incidence and mortality. The CISNET Breast Cancer Working Group includes six modeling teams: Dana-Farber (Model D) [1], Erasmus (Model E) [2], Georgetown-Einstein (Model GE) [3], MD Anderson (Model M) [4], Stanford (Model S) [5], and Wisconsin-Harvard (Model W) [6]. The modeling groups have collaborated to estimate the effects of breast cancer prevention [7], mammography screening [8-11], and systemic adjuvant treatment on trends in breast cancer incidence and mortality [12, 13]. Prior research has also investigated the impact of different screening scenarios on the balance of population-level benefits and harms, and the results have been used by policymakers to inform breast cancer screening guidelines [9, 14, 15].

Each of the models is unique in its structure, assumptions, and methods of synthesizing data. Consequently, they are unique in how they project the impact of screening and treatment on breast cancer incidence and mortality. Results that are similar across multiple models despite differences in assumptions and modeling approach, enhance the credibility of the findings and are more likely to be robust than conclusions obtained from a single model. For instance, in prior analyses, the models all closely estimated observed trends in US breast cancer incidence and mortality and consistently agreed on the ranking of screening scenarios based on several metrics, including mortality reductions. [9, 15]

Despite the consistency of prior conclusions about the effects of screening across the models, there are variations in the magnitude of the effects. [9, 15] For the interpretation of collaborative modeling results, it is important to understand how different model structures and combinations of assumptions contribute to this variation. Detailed model descriptions (Table 1) are informative and contribute to model transparency. However, conveying between-model differences is not always straightforward for reasons related to the nature of modeling disease processes and their interaction with cancer control interventions. In particular, breast cancer modeling involves the representation of unobservable aspects of natural history such as tumor onset and tumor progression upon which interventions (e.g., screening and treatment) are overlaid. To do so, models make assumptions about the timing of tumor inception, tumor progression (e.g., discrete or continuous tumor growth), and progression variability among tumors. These assumptions in conjunction with model structure impact the three key determinants of screening effectiveness: 1) pre-clinical duration of breast cancer, i.e., the time period during which

prevalent undiagnosed cancers could be detected by screening; 2) sensitivity of the screening test, i.e., the likelihood that cancers are detected at screening; and 3) improvement in prognosis from treatment, e.g., whether (earlier) treatment reduces (more) breast cancer mortality.

Given the complexity of interpreting outcomes from multiple models in a collaborative setting, it can be useful to isolate portions of the models to gain greater insight into how model structure and natural history assumptions systematically affect model results. The maximum clinical incidence reduction (MCLIR) method can be used to isolate the effects of tumor onset, tumor progression, screening test sensitivity, and treatment by comparing model results *before* and *after* imposing a one-time screening intervention under varying assumptions about screening performance and treatment effectiveness.

In the absence of screening, breast cancers will only be diagnosed as a result of clinical symptoms, i.e., clinical incidence, which is defined as breast cancers diagnosed due to symptoms. Screening is assumed to detect some of these cancers prior to symptomatic diagnosis, thereby reducing clinical incidence, and possibly cancer mortality. The MCLIR method measures this reduction in breast cancer clinical incidence and mortality. While all models use the same data on screening sensitivity and breast cancer treatment, the implementation of screening and treatment in the models varies as model structures are different. Therefore, differences in *clinical incidence reduction* should reflect model-specific choices in their portrayal of the pre-clinical detectable phase of breast cancer (tumor onset and progression) and mechanisms of screen detection (e.g., how they incorporate sensitivity). Differences in *breast cancer mortality* are expected to capture model-specific assumptions about tumor onset and progression and how the implementation of treatment affects the natural history.

To date, the MCLIR method has been applied to three CISNET colorectal cancer models to clarify the effect of natural history assumptions and model structure on colorectal cancer incidence predictions. [16] In this study, we extended the MCLIR method to understand how differences among the CISNET breast cancer models affect predictions for screening effectiveness by projecting the clinical *incidence and mortality* reductions after a one-time screening exam at age 62 among women without prior screening or a past breast cancer diagnosis. The results are intended to provide a greater understanding of how the CISNET breast models depict unobservable processes, and how those representations may systematically affect conclusions about screening effects on incidence and mortality.

METHODS

This research was approved as exempt by the Georgetown Institutional Review Board based on use of de-identified, publically available data. Five of the six CISNET breast models (those with natural history components) participated in this analysis.

Model Overview

The general model structure of the five models involves the simulation of women who may develop breast cancer in the *absence* or *presence* of screening. In all models, the majority of women live a breast cancer-free life and eventually die of causes other than breast cancer (Figure 1, panel A). For women who develop breast cancer, tumor inception is simulated either prior to (models E and S) or at the start of (models D, GE, and W) the tumor's sojourn time. We define the sojourn time as the portion of time in the pre-clinical phase between when a cancer can be first screen-detectable (e.g., by mammography) and when clinical cancer diagnosis would occur due to symptoms in the absence of screening.[17] Tumor sojourn time is also termed 'pre-clinical screen-detectable phase' (Figure 1).

The point when a tumor becomes screen-detectable may depend on the sensitivity of the screening test, such that more sensitive tests can detect tumors closer to inception, and hence in earlier stages or at smaller tumor sizes. Tumor growth is simulated either continuously (models E, S, and W) or as movement through discrete stages (models D and GE). All models except model S include ductal carcinoma in situ (DCIS). Nonetheless, model S simulates the progression of breast cancers prior to clinical symptoms based on continuous tumor growth of invasive cancer (Table 1). [5]

In the absence of screening, the models assume that some cancers will eventually cause symptoms and be clinically diagnosed (Figure 1, panel B). If a woman participates in screening during the cancer's sojourn time, the cancer may be screen-detected in an earlier stage or at a smaller size than would have occurred with clinical diagnosis in the absence of screening.

The time period between when a cancer is screen-detected and when it would have been clinically diagnosed in the absence of screening is referred to as the *lead-time* (Figure 1, panel C). The lead time is part of the sojourn time, and the duration of the sojourn time is an important unobservable determinant of screening effectiveness because a longer sojourn time implies a longer period during which a screening test can potentially detect cancer. The sojourn time is based on assumptions about tumor inception and tumor growth, and the start of the sojourn time is determined by the sensitivity of the screening test (Figure 1, panel C).

Table 1 Overview of Key Differences and Similarities Between The CISNET Breast Models Structures and Key Model Components

Model	D	E
Model type	Analytic, Parallel universe	Simulation, Parallel universe
Tumor progression modeled as	State-transition	Continuous tumor growth
Incidence in the absence of screening	Age Period Cohort model	Age Period Cohort model
DCIS included	Yes	Yes
ER/HER2 Included	Yes	Yes
Individual risk factors for breast cancer	Breast density	Breast density, obesity
Screen detection conditioned on	Modality, age, density, frequency	Tumor size, modality, age, density, frequency
Implementation of screen benefit	Stage shift	Detection at smaller tumor size
Estimation of over diagnosis	Difference screen & no-screen	Difference screen & no-screen
Implementation treatment benefit	Hazard reduction	Cure fraction, larger fatal diameter
Death from breast cancer determined by	Survival from BC < survival other cause mortality	Fatal diameter, survival from BC < survival OC mortality
SEER data used for calibration	No	Incidence, mortality, stage distribution
Addition based on MCLIR analyses		
Tumor inception point	At the start of pre-clinical screen-detectable phase	Prior to start of pre-clinical screen-detectable phase

* Model S uses background breast cancer incidence derived from the APC framework that explicitly considers the effects of screening and menopausal hormone replacement therapy. [5] Among the other modeling groups breast cancer incidence in the absence of screening is estimated starting from a common APC model. [19, 24]

Explanation of Terms Used in Table 1

Model type

Analytic: Analytical approach to estimate the impact of mammography screening and breast cancer treatment on incidence and mortality of breast cancer.
Simulation: Stochastic simulation is based on the Monte Carlo method and use of random numbers to evaluate the impact of screening on life histories, cancer incidence and mortality.
Parallel universe: Screening and treatment is modeled in a parallel universe, implying that the same population is simulated twice: once to determine the impact of breast cancer without screening, and once to determine the impact of breast cancer with screening.

Natural history and factors affecting breast cancer onset

APC model**:** Breast cancer onset and breast cancer in the absence of screening was derived by Gangnon et al. [23] and is driven by an age-period-cohort model:
Age**:** As women age, their risk of developing breast cancer increases.
Period**:** Onset of breast cancer is different in certain calendar time periods.
Cohort**:** Year of birth influences the risk of developing breast cancer.

GE	S	W
Simulation, Parallel universe	Simulation, Parallel universe	Simulation, Parallel universe
Stage-transition	Continuous tumor growth	Continuous tumor growth
Age Period Cohort model	Age Period Cohort model*	Age Period Cohort model
Yes	No	Yes
Yes	Yes	Yes
Breast density	Hormone replacement	Breast density
Modality, age, density, frequency	Tumor size, ER status, age, hormone repl., frequency	Tumor size, modality, age, density, frequency
Younger age, earlier stage	Stage shift, detect at smaller tumor size	Younger age, detect at smaller tumor size
Difference screen & no-screen	Difference screen & no-screen	Difference screen & no-screen
Hazard reduction, cure fraction	Hazard reduction, non-proportional	Cure fraction
Survival from BC < survival other cause mortality	Survival from BC < survival other cause mortality	Survival from BC < survival other cause mortality
Incidence, stage distribution	Incidence, stage distribution	Incidence, mortality
At the start of the pre-clinical screen-detectable phase	Prior to start of the pre-clinical screen-detectable	At the start of the pre-clinical screen-detectable phase

Breast density**:** Breast density is associated with different levels of risk for developing breast cancer and modifies the operating characteristics of breast cancer screening.
ER/HER2**:** Onset of breast cancer is different for molecular subtypes ER and HER2.
Tumor stage transition**:** Tumor progression is modeled as transitions between discrete stages.
Continuous tumor growth**:** Tumors grow continuously after tumor inception.

Screening mechanism
Sensitivity**:** Sensitivity can be used directly or indirectly (e.g., when translated to tumor size).
Overdiagnosis**:** The detection and diagnosis of a condition that would not go on to cause symptoms or breast cancer death in a woman's lifetime in the absence of screening.
Duration of preclinical detectable phase**:** The period between tumor onset and the start of a cancer's screen-detectable phase.

Treatment effect
Hazard reduction**:** Reduction in breast cancer mortality hazard, derived from the hazard ratio reported for the different treatment regimens [19].
Cure fraction**:** If hazard rate reduction is not a direct model input, it can be translated into a cure fraction to implement breast cancer treatment.
Death from breast cancer**:** Once diagnosed with breast cancer, a survival until breast cancer death is competing with the other cause mortality survival. That is, breast cancer death occurs only if the patient does not die from other causes.

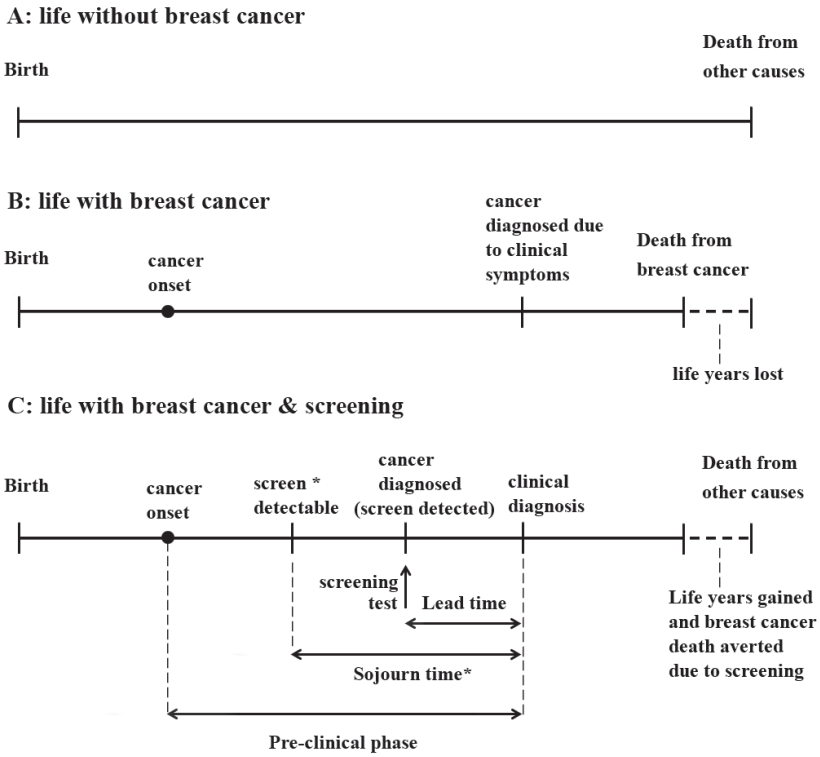


Figure 1 Three versions of a woman's life history. A: without breast cancer, B: with breast cancer and without screening, C: with breast cancer and mammography screening.
In scenario C, the pre-clinical phase is the period of time between tumor inception and clinical diagnosis in the absence of screening. The sojourn time for a screening test, e.g., mammography is the period of time within the pre-clinical phase that a cancer can be screen detectable; this period can also be termed the pre-clinical *screen-detectable* phase. The point when the cancer is detected by screening depends on when the screening test is performed and the sensitivity of the screening test. The period before the sojourn time represents a period in which the tumor is present but undetectable by mammography. Should the sensitivity of mammography improve, or new types of screening tests evolve, the point of screen-detectability would shift to the left and tumors could be detected closer to tumor inception.

Mortality reductions from screening may occur via improvements in survival related to the earlier stage or smaller tumor size at diagnosis of screened vs. unscreened women, given receipt of breast cancer treatment.

MCLIR Analysis

To illustrate the effects of model structure and assumptions about tumor inception, tumor progression, screening test ability to detect tumors, and treatment on breast cancer incidence and mortality predictions, the MCLIR analysis consists of comparisons between four scenarios. Three scenarios involve a one-time screening test at age 62 and the

remaining no-screening scenario serves as a comparator (Table 2). The study population for each scenario is a cohort of average risk women born in 1965, that have never been screened or diagnosed with breast cancer prior to age 62. Age 62 was chosen to illustrate model differences because it is in the middle of the start and stop ages of the generally recommended mammography screening guidelines [14, 18] and there is sufficiently high prevalence of breast cancer at this age to illustrate model differences. While all models have the capacity to include risk factors, to isolate model differences these analyses focused on the average risk population. Women were followed for 15 years (i.e., up to age 77) to capture the immediate and long-term effects of the intervention. Model outcomes were breast cancer clinical incidence and breast cancer mortality by age.

Table 2 Description of Maximum Clinical Incidence Reduction (MCLIR) Method

Scenarios	Scenario Description	Implication	Analyses
No Screening **(Scenario 1)**	No screening: no screening during a woman's lifetime. Diagnosed breast cancers will be treated with current treatment*	All cancers diagnosed in this scenario are diagnosed due to clinical symptoms and will be treated with guideline-concordant treatment.	Comparator to calculate the screening effect in scenarios 2, 3, and 4.
Perfect screening **Perfect treatment** **(Scenario 2)**	A one-time perfect screen with 100% sensitivity$^{\phi}$ at age 62, all screen-detected cancers are treated with perfect treatment$^{\approx}$	All existing cancers at age 62 will be screen-detected and cured by perfect treatment and will not lead to breast cancer death.	Comparison of Scenario 2 to 1 isolates the effect of the pre-clinical detectable duration of breast cancer and provides the tumor progression distribution.
Current sensitivity **Perfect treatment** **(Scenario 3)**	One-time digital mammogram with current sensitivity^ at age 62, all screen-detected cancers are treated with perfect treatment	Some of the existing cancers at age 62 are screen-detected. All screen-detected cancers are cured and will not lead to breast cancer death.	The comparison of scenario 3 to 2 isolates the effect of digital mammography (non-perfect) sensitivity on reductions in clinical incidence and breast cancer mortality.
Current sensitivity **Current Treatment** **(Scenario 4)**	One-time digital mammogram with current sensitivity at age 62, all screen-detected cancers are treated with current treatment	Some of the existing cancers at age 62 are screen-detected. All screen-detected cancers are treated with guideline-concordant treatment and some will not lead to breast cancer death.	Comparison of scenario 4 to 3 isolates the effect of guideline-concordant (imperfect) treatment effectiveness on breast cancer mortality reduction.

* Current treatment: All diagnosed breast cancers receive guideline-concordant breast cancer treatment with observed treatment effectiveness. [19]

$^{\phi}$ Perfect sensitivity: All existing breast cancers are screen-detected at screening (e.g., Sensitivity is 100%).

$^{\approx}$ Perfect treatment: All diagnosed breast cancers are "cured" and women will not die of breast cancer.

^ Current sensitivity: Screening is performed with the observed sensitivity of digital mammography. [19]

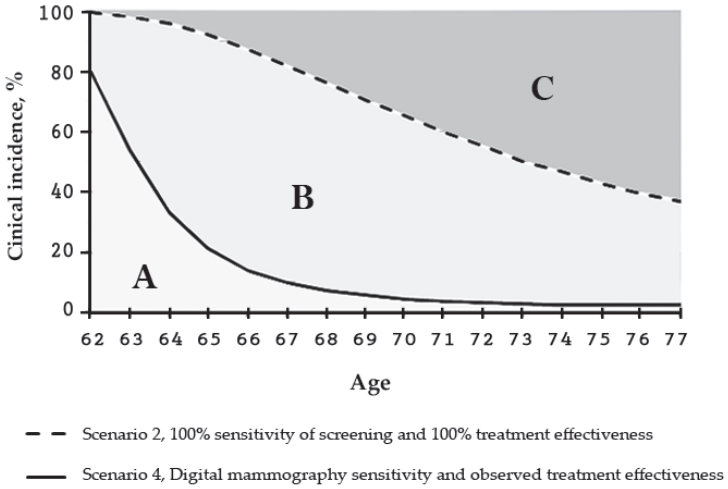


Figure 2 The MCLIR Metrics Explained For Breast Cancer Incidence
Overall Reductions in Breast Cancer Incidence at 15-Year Follow-Up
The light gray area denoted by A is the overall clinical incidence reduction over the 15 years after the digital mammogram at age 62. The area B alone represents the proportion of clinical incidence that could not be reduced because of the non-perfect sensitivity of the digital mammogram. As a digital mammogram does not detect all tumors in existence, the area B provides a measure of the room to further reduce breast cancer clinical incidence if better (more sensitive) screening would become available. The 2 light gray areas combined (A and B) are the maximum clinical incidence reduction from perfect screening. The dark gray area denoted by C, is the proportion of clinical incidence that is not reducible by a perfect screen at age 62 because these clinical cancers had a tumor onset after age 62.
Age-Specific Reductions in Breast Cancer Incidence
Scenario 1, the no-screening scenario, serves as comparator from which the reductions, as measured on the y-axis, are calculated. Scenario 2 (dashed line) is the age-specific percent reduction in clinical incidence from one perfect screening test at age 62 with perfect sensitivity relative to the clinical incidence in the no-screening scenario. Scenario 4 (solid line) is the age-specific percent clinical incidence reduction from one digital mammogram at age 62 relative to the no-screening scenario. Scenario 3 (also solid line) uses sensitivity of current digital mammography and in contrast to scenario 4 has perfect treatment effectiveness which only affects breast cancer mortality, and thus, scenario 3 has the same impact on breast cancer incidence as scenario 4.

MCLIR Scenarios

Scenario 1 is the baseline scenario without screening where *all* breast cancers will be diagnosed due to clinical symptoms. Upon diagnosis, cancers are treated according to current guideline recommended treatment. [19] Guideline concordant treatment roughly implies that, after surgical removal of the tumor, estrogen receptor (ER)-positive breast cancers are primarily treated with hormone therapy and advanced stage ER positive tumors may also receive chemotherapy. ER-negative breast cancers are treated with chemotherapy only. Tumors that are Human epidermal growth factor Receptor 2 (HER2) positive are also treated with Trastuzumab (Herceptin). The effectiveness of breast cancer

treatment was based on the most recent meta-analyses of randomized clinical trials reported by the Early Breast Cancer Trialists' Collaborative Group (EBCTCG). [20] Scenario 1 provides baseline information about the number of cancers that will lead to symptoms and be clinically diagnosed as well as the number of breast cancer deaths occurring in the 15-year follow-up period.

Scenario 2 involves a one-time perfect screening test at age 62 and perfect treatment. The hypothetical perfect screening test assumes that all tumors in existence are screen-detected, including those that may not be detectable by digital mammography. Perfect treatment means treatment is curative and that all women will be cured and will die from other causes than breast cancer. Comparing results from this scenario with the baseline (no-screening) scenario provides the maximum achievable clinical incidence and mortality reduction. It is a measure of the pool of cancers that technically could be screen-detected at age 62 and thus avoid clinical diagnoses of these cancers at a later age when they would cause symptoms. The change in the maximum achievable clinical incidence reduction over time as women age provides insight into the distribution of sojourn times of the existing tumors at age 62, i.e., key determinant 1 of screening effectiveness. The mortality results from this scenario provide information on how many of the breast cancer deaths between ages 62 and 77 stem from cancers that were present at age 62. Relative to the no-screening scenario, it is the maximum achievable mortality reduction from screening and treatment, and the converse (1 minus the maximum mortality reduction) is the portion of unavoidable breast cancer deaths because these cancers had tumor onset after age 62 when the screening test was done (Figure 3). The age-specific maximum achievable mortality reduction after the screen test at age 62 also provides insight into the survival time of pre-clinical cancers in existence at age 62.

Scenario 3 involves a one-time digital mammogram at age 62 with sensitivity based on observed mammography performance in the Breast Cancer Screening Consortium [9, 19] and perfect treatment (i.e., cure) of screen-detected cancers. In this scenario, some of the cancers in existence at age 62 will be missed by screening and this will affect clinical incidence and mortality at later ages. Because scenarios 2 and 3 vary screening performance while holding the treatment effects constant, the comparison of these two scenarios isolates the impact of perfect vs. observed sensitivity of screening on reductions in clinical incidence and breast cancer mortality, i.e., key determinant 2 of screening effectiveness. This comparison also illustrates the room for improvement in terms of fewer clinically diagnosed cases and cancer deaths should the sensitivity of screening would improve (e.g., new radiology technology or circulating tumor DNA detection).

Scenario 4, the realistic scenario, involves a one-time digital mammogram at age 62 and treatment according to current guidelines [19]. Because scenarios 3 and 4 vary treatment effectiveness while holding the sensitivity of screening constant, the comparison of these scenarios isolates the impact of perfect vs. actual treatment effectiveness on breast

cancer mortality, i.e., key determinant 3 of screening effectiveness. This comparison isolates the portion of cancers that, despite earlier detection by screening, will not be cured with current guideline recommended treatment. Also, this provides insight into the room for improvement should breast cancer treatment improve in the future, given current performance of digital mammography.

For ease of comparison and interpretation of outcomes across the four scenarios for five different models, results are reported as percent reductions in clinical incidence and breast cancer mortality relative to each model's clinically diagnosed breast cancers and breast cancer deaths in the absence of screening (Figure 2 & 3).

RESULTS

The results for each scenario for the impact of a one-time screen at age 62 among women with no prior screening or past diagnosis of breast cancer are presented separately for incidence and mortality.

Breast Cancer Incidence

Tumor Onset and Progression

The maximum achievable clinical incidence reduction from a perfect screening test at age 62 (scenario 2) relative to the no-screening scenario (scenario 1) illustrates the impact of natural history assumptions such as tumor onset and tumor progression on screening effectiveness. The maximum clinical incidence reduction ranged from 19% to 71% across the five models with models D, GE, and W grouping towards the lower end of the range and models E and S towards the top of the range (Table 3). This wide variation was the result of differences in the modeling of the timing of tumor inception relative to the start of the sojourn time. For example, Model E's assumption of tumor onset long before the start of the sojourn time led to a large screening effect when the perfect screening test was applied that detects all tumors from their inception even before the pre-clinical screen-detectable phase begins. The majority (71%) of the cancers in this model had an onset prior to age 62 and were therefore screen-detected by a perfect screening test at age 62, avoiding clinical diagnoses at a later age. The remaining (29%) of cancers had an onset after age 62. Model S makes similar assumptions about tumor onset and growth as Model E, and has fairly similar patterns in their results as Model E. In contrast, in Models D, GE, and W, which simulate tumor inception at the start of the pre-clinical screen-detectable phase, only 19% to 27% of cancers were in existence at age 62, leading to a lower maximum clinical incidence reduction from a perfect screening test than Models E and S.

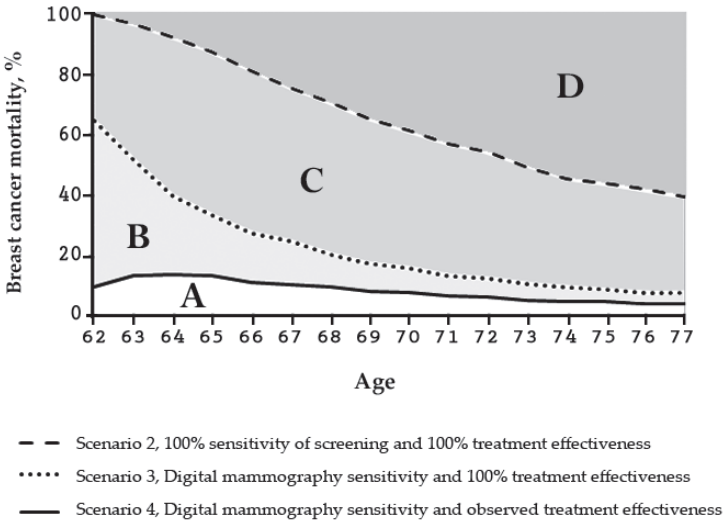


Figure 3 The MCLIR Metrics Explained For Breast Cancer Mortality
Overall Reductions in Breast Cancer Mortality at 15-Year Follow-Up
The light gray area denoted by A is the overall breast cancer mortality reduction over the 15 years after one digital mammogram at age 62 and guideline recommended treatment with observed treatment effectiveness. The area B alone represents the proportion of breast cancer mortality that could not be reduced because of the non-perfect treatment effectiveness of current guideline recommended treatment. Since guideline recommended treatment does not cure all screen-detected cancers, B provides a measure of the room to further reduce breast cancer mortality if better (more effective) treatment would become available.
The area C alone represents the proportion of breast cancer mortality that could not be reduced because of the non-perfect sensitivity of currently available digital mammography. As a digital mammogram does not detect all tumors in existence, B provides a measure of the room to further reduce breast cancer mortality if better (more sensitive) screening would become available. The 3 areas combined (A, B and C) are the maximum mortality reduction from perfect screening and perfect treatment where B + C represent the maximum room to further reduce breast cancer mortality if screening sensitivity and treatment effectiveness would become improve. The dark gray area, denoted by D, is the proportion of breast cancer deaths that is not reducible by a perfect screen at age 62 and perfect treatment because these breast cancer deaths had tumor onset after age 62.
Age-Specific Reductions in Breast Cancer Mortality
Scenario 1, the no-screening scenario, serves as comparator from which the reductions, as measured on the y-axis, are calculated. Scenario 2 (dashed line) is the age-specific percent breast cancer mortality reduction from one perfect screening test with perfect sensitivity and perfect treatment relative to the breast cancer mortality in the no-screening scenario. Scenario 3 (dotted line) is the age-specific percent breast cancer mortality reduction from one digital mammogram at age 62 and perfect treatment relative to the no-screening scenario. Scenario 4 (solid line) is the age-specific percent mortality reduction from one digital mammogram at age 62 and guideline-concordant treatment with observed treatment effectiveness in screen-detected cases relative to the no-screening scenario.

Table 3 Percent Reductions in *Breast Cancer Incidence* after One Mammography Screen at Age 62 over a 15-Year Follow-Up, %

Scenario	Intervention	D	E	GE	S	W
2	100% screening sensitivity and 100% treatment effectiveness (vs. no screening)	23	71	27	43	19
4	Current screening sensitivity and current treatment effectiveness (vs. no screening)	21	15	24	16	11
4 vs. 2	Breast cancer clinical incidence not reduced because of imperfect (current) screening sensitivity	2	56	3	27	8

Scenario 3 is not shown because this scenario has the same screening sensitivity (digital mammography) as scenario 4 and hence has the same clinical incidence reduction as scenario 4.

The shape of the maximum clinical incidence reduction curve provides insight into the variability of tumor growth and disease progression of tumors in existence at age 62 (Figure 4). In models D, GE and W, the age-specific clinical reductions from the perfect screen declined more rapidly in the first five years than in the other two models, indicating the presence of more quickly progressing tumors relative to the other models. The non-steep and slower linear decline of the age-specific maximum clinical incidence reduction in Models E and S is the consequence of greater *variability* in tumor progression and overall *slower* tumor growth among the tumors in existence at age 62 than seen in the other models.

The models have structural differences in the timing of tumor inception relative to the sojourn time *and* they had the same calibration targets (observed trends in U.S. breast cancer incidence and mortality) in their development phase. This explains why Models E and S with tumor inception long before the start of the sojourn time have slower overall tumor progression and Models D, GE, and W with tumor inception at the start of the sojourn time have faster progressing tumors.

Screening Sensitivity

Reductions in clinical incidence based on the observed sensitivity of digital mammography varied less across models than when assuming perfect sensitivity, with ranges of 11% to 24%. Since assumptions about tumor onset and progression differ, how the models arrive at this result differs and is illustrated by comparison to their predictions for maximum reductions achievable (Scenario 3 vs 2). In models D, GE, and W, the differences in clinical incidence reduction were 2%, 3%, and 8%, respectively, and in models E and S these were 56% and 27%. While models E and S have more tumors in existence at age 62, the majority of tumors were in their pre-sojourn period and not yet screen-detectable with

a digital mammogram having actual observed sensitivity. On the other hand, in models D, GE, and W, the majority of tumors in existence at age 62 were in their sojourn period and could be detected by the digital mammogram. Thus, the variations between model clusters E and S vs. D, GE, and W indicate that modeling assumptions about the timing of tumor inception in relation to the implementation of digital mammography can have substantial impact on screen detection and reductions in clinical breast cancer incidence.

Breast Cancer Mortality

Tumor Onset and Progression

Based on one perfect screening test at age 62 and perfect treatment for screen-detected cancers, the maximum reductions in breast cancer mortality relative to the no-screening scenario ranged from 33% to 67% over 15 years of follow-up (Table 4). Similar to the impact of tumor onset on clinical incidence reductions, Models D, GE and W had a higher percent (55% to 67%) of breast cancer deaths stemming from cancers with onset *after* age 62 than Models E and S (33% to 38%). These variations reflect interactions between assumptions about tumor onset and survival times.

Table 4 Percent Reduction in *Breast Cancer Mortality* after One Digital Mammography Screen at Age 62 with 15-Year Follow-Up, %

Scenario	Intervention	D	E	GE	S	W
2	100% screening sensitivity and 100% treatment effectiveness (vs. no screening)	40	67	45	62	33
	% Breast cancer deaths with onset after age 62.*	60	33	55	38	67
3	Current screening sensitivity and 100% treatment effectiveness (vs. no screening)	37	23	40	31	23
3 vs. 2	Breast cancer mortality *not* reduced because of imperfect (current) screening sensitivity	3	44	5	31	10
4	Current screening sensitivity and current treatment effectiveness (vs. no screening)	17	8	17	18	8
4 vs. 3	Breast cancer mortality *not* reduced because of imperfect (current) treatment effectiveness	20	15	23	13	15
4 vs. 2	Breast cancer mortality *not* reduced because of imperfect screening sensitivity *and* imperfect treatment effectiveness	23	59	23	44	25

* The percent breast cancer deaths that stem from cancers with onset after age 62 is given by 100% minus the cancer deaths from cancers with onset before age 62 (Scenario 2).
Scenarios 3 vs. 2, 4 vs. 3, and 4 vs. 2, show the percentage point breast cancer mortality that is not reduced due to imperfect screening sensitivity and/or imperfect treatment effectiveness.

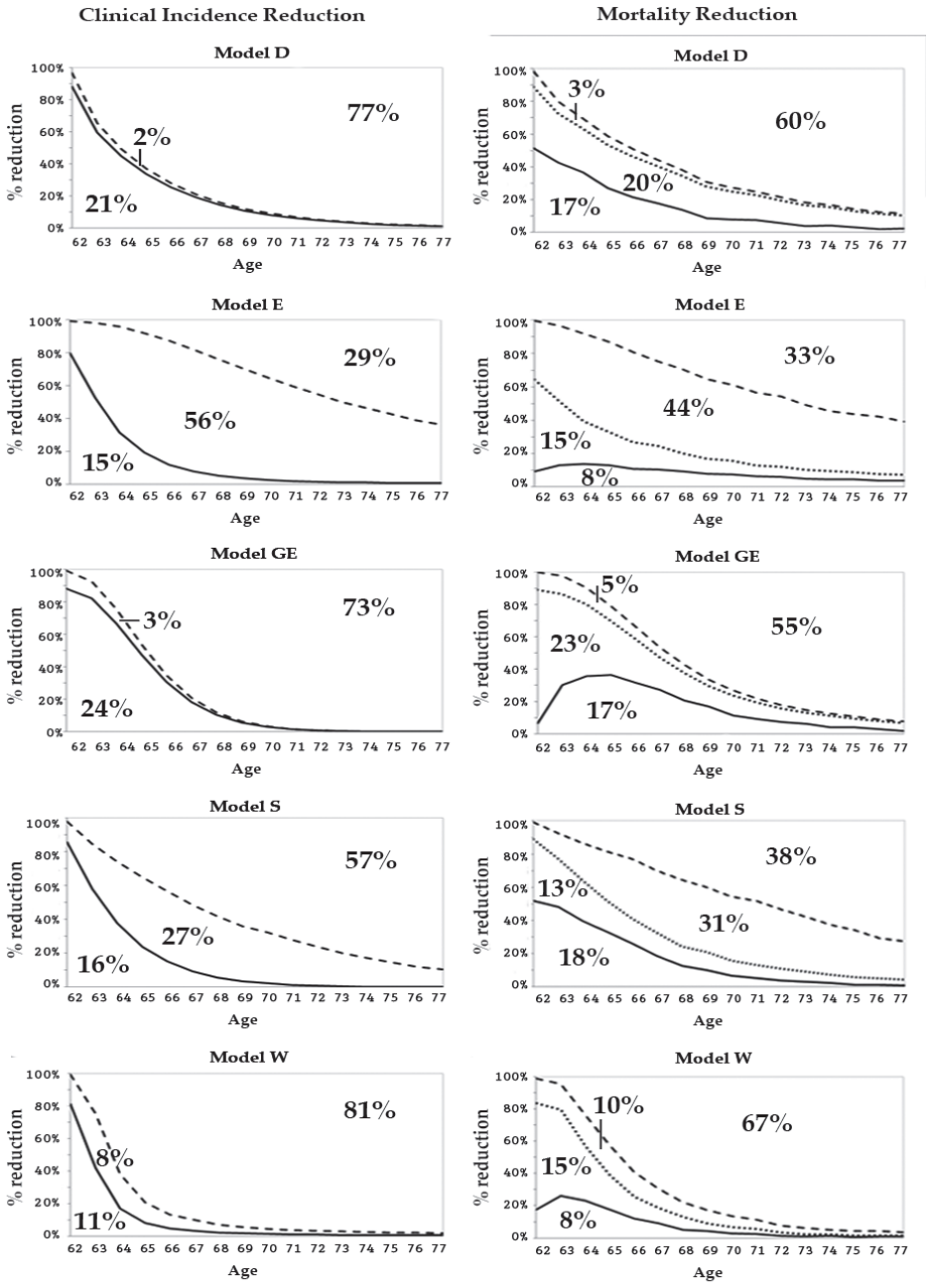

Clinical Incidence Reduction
Mortality Reduction
Model D
Model E
Model GE
Model S
Model W
% reduction
Age
77%
2%
21%
60%
3%
20%
17%
29%
56%
15%
33%
44%
15%
8%
73%
3%
24%
55%
5%
23%
17%
57%
27%
16%
38%
13%
31%
18%
81%
8%
11%
67%
10%
15%
8%
Scenario 2, 100% sensitivity of screening and 100% treatment effectiveness
Scenario 3, Digital mammography sensitivity and 100% treatment effectiveness
Scenario 4, Digital mammography sensitivity and observed treatment effectiveness

Figure 4 Age-Specific Reductions In Breast Cancer Clinical Incidence And Mortality Over 15 Years After A One-Time Screening Test At Age 62 By Model

The percent marks in the panels of Figure 4 represent the cumulative outcomes for the 15-year follow-up period from age 62 to age 77.

The line at the top in the breast cancer incidence panels on the left of Figure 4 is the maximum clinical incidence reduction from a screening test at age 62 with 100% sensitivity and perfect treatment of screen-detected cancers (Scenario 2). Just after the screening test, the reduction in clinical incidence (panels on the left) is highest and decreases by age as it becomes less likely that clinical cancers at later ages were already in existence at age 62 and could have been prevented by a screening test at that age.

The percentages in the left-panel figures represent, for example for Model S: 57% of the cancers that are clinically diagnosed in the absence of screening between ages 62 and 77 have an onset after age 62, this implies that 100-57=43% (Scenario 2, Table 3) of the cancers diagnosed in the absence of screening could be prevented from becoming clinical diagnosis at later ages by a perfect screening test at age 62. The solid line below the dashed line is the clinical incidence reduction from a digital mammography screening test: 16% of clinical diagnoses could be prevented by a one-time digital mammogram at age 62 (Scenario 3, Table 3). This implies that 27% of clinical incidence between ages 62 and 77 was not reduced by the one-time digital mammogram at age 62 (Scenario 3 vs 2).

The dashed line at the top in the breast cancer mortality panels on the right of Figure 4 is the maximum achievable mortality reduction from a screening test with 100% sensitivity combined with perfect treatment in screen-detected cases (Scenario 2). The dotted line below the top line represents the breast cancer mortality reduction over the 15-years after a current digital mammogram at age 62 and perfect treatment in the screen-detected cases (Scenario 3). The solid line at the bottom is the reduction in breast cancer mortality from a current screening test combined with current treatment (Scenario 4).

The percentages in these figures are, for example for Model S: 38% of breast cancer deaths observed in the scenario without screening stem from cancers with onset after age 62 and could therefore not be screen-detected and prevented from breast cancer death by screening at age 62. This implies that 100-38=42% of breast cancer deaths could be reduced by perfect screening and perfect treatment of screen-detected cases (Scenario 2, Table 4). However, 31% of breast cancer deaths are not prevented due to lack of screen-detection if screening is performed with a digital mammogram (Scenario 3 vs 2, Table 4), and 13% of breast cancer deaths is not prevented because current guideline-concordant treatment lacks the effectiveness to cure those screen-detected breast cancers (Scenario 4 vs 3). The 18% mortality reduction follows from current screening and current treatment (Scenario 4).

The steep declines of the maximum mortality reduction curves (Figure 4, right panels) of models D, GE and W reveal that, on average breast cancers in these models have *shorter survival* times and *less variability* in survival times relative to models E and S. These results for average survival times correspond to the findings about tumor progression in the models: the relatively slow tumor progression, based on earlier inception of tumors, in models E and S relate to longer survival times, and the faster tumor progression in Models D, GE and W arising from tumor inception at the beginning of the sojourn period ultimately lead to shorter survival times on average.

Screening sensitivity

Compared to the maximum achievable mortality reduction, a one-time digital mammogram having actual observed sensitivity missed between 3% (Model D) to 44% (Model E) of the avoidable cancer deaths. Overall, the mortality reduction from a one-time digital mammogram at age 62 and perfect treatment relative to no-screening (scenario 3 vs. 1) was 23% to 40% across models (Table 4, Figure 4). The ability to detect lethal tumors by mammography screening was higher among the models (D, GE, and W) with assumptions of tumor onset at the start of the sojourn time than the models (E and S) with tumor onset prior to the start of the sojourn time.

Treatment effectiveness

Assuming observed guideline-concordant treatment effectiveness in screen-detected cancers (scenario 4), the percent breast cancer mortality that was not reduced compared to Scenario 3 with perfect treatment was 13% to 23% (Table 4, Figure 4). The difference between scenario 3 and 4 show that Models E and GE, have a relatively high percentage of cancer deaths that were not averted in the first 3 years after the screen at age 62. This illustrates the substantial portion of cancers screen-detected at a relatively advanced stage that was not curable with current treatment effectiveness. These findings showed that the lethality of the cancers found at screening impacts breast cancer mortality differently over time and in magnitude by model.

Sensitivity and Treatment

The combination of screening with a digital mammogram at age 62 and guideline-concordant treatment with current treatment effectiveness (Scenario 4 vs. Scenario 1) provides insight into how assumptions about currently available screening and treatment interact with breast cancer natural history to affect breast cancer mortality. Models E, W and S grouped towards the lower end and models D and GE towards the higher end of the clinical incidence reductions (Table 3). But for breast cancer mortality slightly different groupings of models were seen: Models D, GE and S predicted 17 to 18% breast cancer mortality reduction relative to the no-screening scenario, whereas models E and W predicted 8% breast cancer mortality reduction (Table 4).

The lower breast cancer mortality reductions predicted by models E and W were primarily the result of a low screen-detection rate of lethal cancers and the lack of improving prognosis with treatment of screen-detected cases: in both models 23% of the cancers destined to cause breast cancer death were screen-detected (Scenario 3), and of those detected only one-third (8 out of 23; Scenario 4 vs. Scenario 3, Table 4) were cured.

Models D and S predicted a similar 17 and 18% mortality reduction as model GE, also due to a combination of relatively high screen-detection and high improvement of prognosis from treatment. However, the shape of the mortality reduction curve of

Model GE, relative to other models, was distinct. The inverted shape of model GE can be explained by the presence of more advanced-stage cancers at screen detection, where breast cancer death could not be avoided.

DISCUSSION

This study is the first to apply the maximum clinical incidence reduction (MCLIR) method to illustrate how model structure and assumptions impact both *clinical incidence and cancer mortality* predictions. To understand variations in model estimates of screening effects, the analysis decomposed the relative contributions of model-specific structures and assumptions regarding the pre-clinical duration of breast cancer, the ability of a screening test to detect cancers, and breast cancer treatment to breast cancer incidence and mortality predictions. The results illustrated that models with similar predictions for screening effectiveness may use differing assumptions about screening, treatment, tumor onset, and tumor progression. Altogether, the key finding was that assumptions about the timing of tumor inception and its effect on the pre-clinical duration of breast cancer had the greatest impact on the model groupings on predicted clinical breast cancer incidence and mortality reductions. As a result of this finding, we now include this model-specific tumor attribute in our CISNET model comparison table (Table1).

The MCLIR scenarios showed that models E and S simulate the longest pre-clinical duration of breast cancer. While this implies a longer period to detect cancers by screening and possibly avert cancer deaths, these models showed the greatest difference in breast cancer mortality reduction between the scenarios with perfect detection to those with (realistic) digital mammography . Again, this was related to those models' assumptions about early tumor onset prior to the start of a cancer's sojourn time. The loss in breast cancer mortality reduction due to digital mammography (imperfect) screening provides information about the further reductions in breast cancer mortality should screening sensitivity improve in the future, given the current state of the models. On the other hand, models D, GE and W had similar and relatively short pre-clinical durations due to their assumptions of tumor inception at the start of the sojourn time and therefore ultimately predicted smaller losses in breast cancer mortality reduction due to digital mammography screening. The effect of guideline-concordant treatment with actual observed treatment effectiveness on breast cancer mortality reduction differed by model structure. In general, greater breast cancer mortality reductions were predicted by models that use a hazard-reduction treatment structure than the models with cure fractions to implement breast cancer treatment. These types of insights from the MCLIR results provide further clarity on the differences and similarities across models and can be used to interpret variations in model outcomes.

The MCLIR analyses also illustrated model variation in the distributions for tumor progression assumed in the models, with models D, GE, and W tending to have faster progressing tumors than models E and S. This knowledge about the models can help interpret model differences in predictions of screening effectiveness by screening frequency. For example, one would expect more cancers to be diagnosed with more frequent screening in models that have relatively faster tumor progression on average and vice versa. This was confirmed in a recent analysis of the impact of screening intervals on breast cancer mortality, with Models D, GE, and W showing greater benefits (breast cancer deaths averted preceded by more cancer diagnoses) from more frequent screening than models E and S. [9]

The MCLIR methodology was first used to evaluate differences in the CISNET colorectal cancer screening models. [16] The colorectal cancer findings indicated that assumptions about the duration between adenoma onset and clinical diagnosis were an important factor in explaining colorectal cancer model differences. The results of this study were similar in demonstrating that models with long pre-clinical durations of breast cancer and relatively low screen detection rates project similar screening effects as models with a shorter pre-clinical durations and higher screen detection rates.

Usually, models are characterized by describing modeling approach, model inputs and assumptions. [19,21,22] In this research, we examined model *outcomes* to drill down to the mechanics of incidence and mortality predictions. There are several caveats that should be considered in evaluating this method. First, the effect of a single screen on breast cancer incidence and mortality is not the same as the effect of routine screening from age 50 to 74. The results in this study are therefore not directly translatable to projections of the effects of a periodic screening program on overall breast cancer incidence and mortality. Second, it was beyond the scope of this paper to perform and evaluate the MCLIR scenarios at different ages or at multiple ages across five different models. Evaluating the MCLIR scenarios at different ages would provide insight into age-specific and between-model differences in tumor inception, progression, and test-characteristics and the impact of these on breast cancer incidence and mortality. Third, the MCLIR methodology employed did not explicitly allow for formal assessments of the factors that account for differences in rates of over-diagnosis. This will be an interesting area for future research and extended use of the MCLIR method.

CISNET collaborative modeling predictions are increasingly used by policy makers to inform screening guidelines [9, 14], evaluate screening and treatment programs [12, 13], and can be used by clinicians to assist in decision-making about breast cancer screening. [23] How different models arrive at their predictions of harms and benefits of screening and treatment may be perceived as opaque due to the complexity of the models. This study complements model descriptions [1-6] by using MCLIR analyses to illustrate and compare which structural differences and natural history assumptions may be important

to consider by policy makers when using collaborative modeling outcomes. The MCLIR approach could be adopted in other comparative modeling research to improve model transparency.

REFERENCES

[1] Lee SJ, Li X, Huang H. Models for Breast Cancer Screening Strategies Updated for Ductal Carcinoma In Situ and Subgroups Medical Decision Making. In Press. 2017.

[2] van den Broek JJ, van Ravesteyn NT, Heijnsdijk EA, de Koning HJ. Estimating the effects of risk-based screening and adjuvant treatment using the MISCAN-Fadia continuous tumor growth model for breast cancer. Medical Decision Making. In Press. 2017.

[3] Schechter CB, Near AM, Jayasekera J, Chang Y, Mandelblatt JS. Structure, Function, and Applications of the Georgetown-Einstein (GE) Breast Cancer Simulation Model Medical Decision Making. In Press. 2017.

[4] Huang X, Li Y, Song J, Berry DA. The MD Anderson CISNET Model for Estimating Benefits of Adjuvant Therapy and Screening Mammography for Breast Cancer: An Update. Medical Decision Making. In Press. 2017.

[5] Munoz D, Xu C, Plevritis S. A Molecular Subtype-Specific Stochastic Simulation Model of US Breast Cancer Incidence and Mortality Trends from 1975 to 2010 Medical Decision Making. In Press. 2017.

[6] Alagoz O, Ergun MA, Cevik M, Sprague BL, Fryback DG, Gangnon R, et al. The University Of Wisconsin Breast Cancer Epidemiology Simulation Model: An Update (Medical Decision Making. In Press. 2017.

[7] Chang Y, Schechter CB, van Ravesteyn NT, Near AM, Heijnsdijk EA, Adams-Campbell L, et al. Collaborative modeling of the impact of obesity on race-specific breast cancer incidence and mortality. Breast Cancer Res Treat. 2012; 136(3):823-35.

[8] Trentham-Dietz A, Kerlikowske K, Stout NK, Miglioretti DL, Schechter CB, Ergun MA, et al. Tailoring Breast Cancer Screening Intervals by Breast Density and Risk for Women Aged 50 Years or Older: Collaborative Modeling of Screening Outcomes. Ann Intern Med. 2016; 165(10):700-12.

[9] Mandelblatt JS, Stout NK, Schechter CB, van den Broek JJ, Miglioretti DL, Krapcho M, et al. Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies. Ann Intern Med. 2016; 164(4):215-25.

[10] Sprague BL, Stout NK, Schechter C, van Ravesteyn NT, Cevik M, Alagoz O, et al. Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. Ann Intern Med. 2015; 162(3):157-66.

[11] Stout NK, Lee SJ, Schechter CB, Kerlikowske K, Alagoz O, Berry D, et al. Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography. J Natl Cancer Inst. 2014; 106(6):dju092.

[12] Munoz D, Near AM, van Ravesteyn NT, Lee SJ, Schechter CB, Alagoz O, et al. Effects of screening and systemic adjuvant therapy on ER-specific US breast cancer mortality. J Natl Cancer Inst. 2014; 106(11).

[13] Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005; 353(17):1784-92.

[14] Siu AL, On behalf of the U. S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2016; 164(4):279-96.

[15] Mandelblatt JS, Cronin KA, Bailey S, Berry DA, Koning H J de, Draisma G, et al. Effects of mammography screening under different screening schedules:

model estimates of potential benefits and harms Ann Int Med. 2009; 151:738-47.

[16] van Ballegooijen M, Rutter CM, Knudsen AB, Zauber AG, Savarino JE, Lansdorp-Vogelaar I, et al. Clarifying differences in natural history between models of screening: the case of colorectal cancer. Med Decis Making. 2011; 31(4):540-9.

[17] Walter SD, Day NE. Estimation of the duration of a pre-clinical disease state using screening data. American Journal of Epidemiology. 1983; (118(6):865-86.).

[18] Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. JAMA. 2015; 314(15):1599-614.

[19] Mandelblatt JS, Near AM, Miglioretti DL, Munoz D, Sprague BL, Trentham-Dietz A, et al. Common Model Inputs used in CISNET Collaborative Breast Cancer Modeling Medical Decision Making. In Press. 2017.

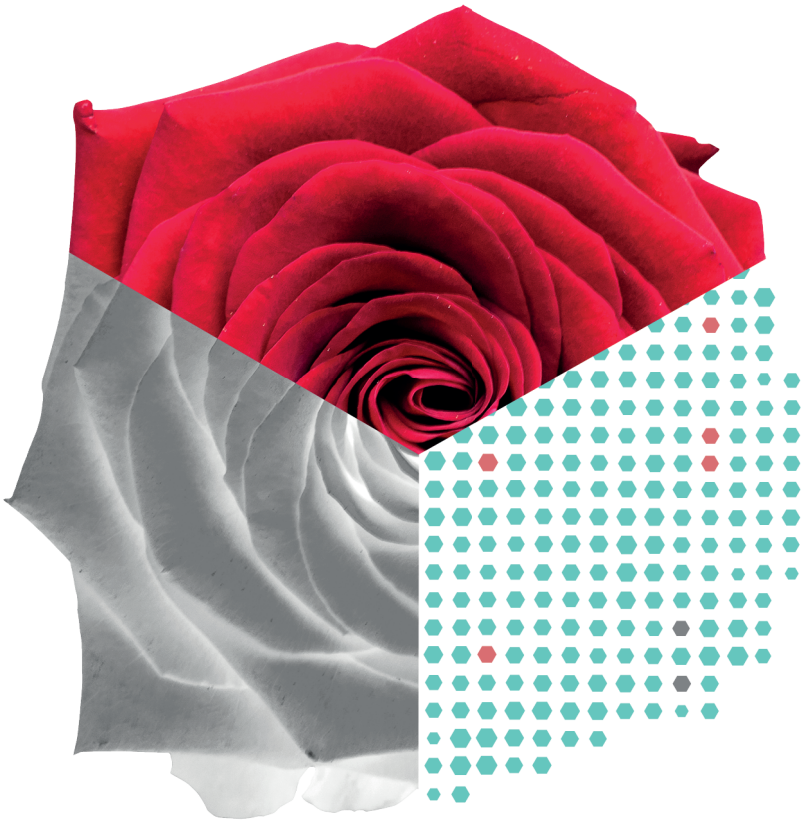
[20] Early Breast Cancer Trialists' Collaborative G, Peto R, Davies C, Godwin J, Gray R, Pan HC, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Lancet. 2012; 379(9814):432-44.

[21] Clarke LD, Plevritis SK, Boer R, Cronin KA, Feuer EJ. A comparative review of CISNET breast models used to analyze U.S. breast cancer incidence and mortality trends. J Natl Cancer Inst Monogr. 2006; (36):96-105.

[22] Boer R, Plevritis S, Clarke L. Diversity of model approaches for breast cancer screening: a review of model assumptions by the Cancer Intervention and Surveillance Network (CISNET) Breast Cancer Groups. Stat Methods Med Res. 2004; 13(6):525-38.

[23] Burnside E, Lee SJ, Bennette C, Near AM, Alagoz O, Huang H, et al. Using Collaborative Simulation Modeling to Develop a Web-based Tool To Support Policy-level Decision-making about Breast Cancer Screening Initiation Age. Medical Decision Making Policy & Practice. In Press. 2017.

[24] Gangnon RE, Sprague BL, Stout NK, Alagoz O, Weedon-Fekjaer H, Holford TR, et al. The contribution of mammography screening to breast cancer incidence trends in the United States: an updated age-period-cohort model. Cancer Epidemiol Biomarkers Prev. 2015; 24(6):905-12.



PART TWO: Quantifying the harms and benefits of age-based breast cancer screening in the United States

Chapter 6

Association of Screening and Treatment With Breast Cancer Mortality by Molecular Subtype in US Women

Sylvia Plevritis, Diego Munoz, Allison Kurian, Oguzhan Alagoz, Aimee M. Near, Sandra J. Lee, Jeroen J. van den Broek, Xuelin Huang, Natasha K. Stout, Clyde B. Schechter, Brain L. Sprague, Harry J. de Koning, Amy Trentham-Dietz, Nicolien T. van Ravesteyn, Ronald Gagnon, Yaojen Chang, Yisheng Li, Mehmet Ali Ergun, Hui Huang, Donald A. Berry, Jeanne S. Mandelblatt.

JAMA. 2018;319(2):154-64.

ABSTRACT

Importance

Given recent advances in screening mammography and adjuvant therapy, quantifying their separate and combined effects on US breast cancer mortality reductions by molecular subtype could guide future decisions to reduce disease burden.

Objective

To evaluate the contributions associated with screening and treatment to breast cancer mortality reductions by molecular subtype based on estrogen-receptor (ER) and human epidermal growth factor receptor 2 (ERBB2, formerly HER2 or HER2/neu).

Design, Setting and Participants

Six Cancer Intervention and Surveillance Network (CISNET) models simulated US breast cancer mortality from 2000 to 2012 using national data on plain-film and digital mammography patterns and performance, dissemination and efficacy of ER/ERBB2-specific treatment, and competing mortality. Multiple US birth cohorts were simulated.

Exposures

Screening mammography and treatment.

Main Outcomes and Measures

The models compared age-adjusted, overall, and ER/ERBB2-specific breast cancer mortality rates between 2000 and 2012 for women aged 30 to 79 years relative to the estimated mortality rate in the absence of screening and treatment (baseline rate); mortality reductions were apportioned to screening and treatment.

Results

In 2000, the estimated reduction in overall breast cancer mortality rate was 37% (model range, 27%-42%) relative to the estimated baseline rate in 2000 of 64 deaths (model range, 56-73) per 100 000 women: 44% (model range, 35%-60%) of this reduction was associated with screening and 56% (model range, 40%-65%) with treatment. In 2012, the estimated reduction in overall breast cancer mortality rate was 49% (model range, 39%-58%) relative to the estimated baseline rate in 2012 of 63 deaths (model range, 54-73) per 100 000 women: 37% (model range, 26%-51%) of this reduction was associated with screening and 63% (model range, 49%-74%) with treatment. Of the 63% associated with treatment, 31% (model range, 22%-37%) was associated with chemotherapy, 27% (model range, 18%-36%) with hormone therapy, and 4% (model range, 1%-6%) with trastuzumab. The estimated relative contributions associated with screening vs treatment

varied by molecular subtype: for ER+/ERBB2−, 36% (model range, 24%-50%) vs 64% (model range, 50%-76%); for ER+/ERBB2+, 31% (model range, 23%-41%) vs 69% (model range, 59%-77%); for ER−/ERBB2+, 40% (model range, 34%-47%) vs 60% (model range, 53%-66%); and for ER−/ERBB2−, 48% (model range, 38%-57%) vs 52% (model range, 44%-62%).

Conclusions and Relevance

In this simulation modeling study that projected trends in breast cancer mortality rates among US women, decreases in overall breast cancer mortality from 2000 to 2012 were associated with advances in screening and in adjuvant therapy, although the associations varied by breast cancer molecular subtype.

INTRODUCTION

Breast cancer mortality rates have been steadily declining over time in the United States (US).[1] Simulation models developed within the Cancer Intervention and Surveillance Network (CISNET) estimated that screening mammography and adjuvant therapy (treatment) contributed approximately equally to the reduction in breast cancer mortality from 1975 to 2000.[2] Since then, mammography has transitioned from plain-film to digital technology optimized for tumor detection.[3,4] At the same time, there have been advances in molecularly-targeted treatments based on expression of estrogen-receptor (ER) and human epidermal growth factor receptor 2 (ERBB2, formerly HER2 or HER2/neu), including aromatase inhibitors for ER+, and trastuzumab for ERBB2+ cancers. In addition, there have been advances in chemotherapy, particularly increasing use of taxanes.[5,6]

It is not known how screening and treatment advances have contributed to recent population-level, molecular subtype-specific breast cancer mortality rates. No single national registry contains sufficient information to assess this progress. Moreover, most clinical trials do not consider both screening and treatment effects, and do not readily translate to population effect. Given these circumstances, simulation modeling can be useful to integrate high-quality data from randomized-controlled trials, large observational studies, and population registries to estimate the relative contributions of advances on population-level mortality.[2]

In this report, six CISNET models compared the separate and combined contribution associated with screening and treatment on US breast cancer mortality rates by molecular subtype from 2000 and 2012.

METHODS

The Institutional Review Board at Georgetown University, the site of the CISNET Breast Cancer Coordinating Center, approved the study as exempt based on the use of de-identified data. The 6 CISNET models were Dana-Farber Cancer Institute (model D)[7], E Erasmus Medical Center (model E)[8], Georgetown University-Albert Einstein College of Medicine (model G-E)[9], MD Anderson Cancer Center (model M),[10] Stanford University (model S),[11,12] and University of Wisconsin-Harvard (model W-H).[13] Compared to earlier analyses[2,14,15] the models portray ER/ERBB2-specific subtypes,[11] include digital screening[3,4] and recent treatment advances,[16] and have updated incidence[17] and competing non-breast cancer mortality.[18] The modeling approach is summarized below; additional details are available in the Supplement and online.[19]

The models incorporated updated estimates of breast cancer incidence[17] and ER/ERBB2-specific survival trends in the *absence* of screening or treatment and then incorpo-

rated information on screening use and molecular subtype-specific treatment patterns to reproduce observed US incidence and mortality trends.[1,20,21] Screen-detection during the preclinical screen-detectable period could result in diagnosis of earlier-stage or smaller tumors than diagnosed via symptomatic detection. This could translate into lower breast cancer mortality. Molecular subtype, age-specific, and stage-specific treatment could reduce the hazards of breast cancer death (models D, GE, M, and S) or result in cure for some cases (models E and W-H).

Model Input Parameters

Each group used a common set of inputs [22] based on their specific model structure, prior research,[15] and assumptions to best reproduce US breast cancer incidence and mortality trends (Supplemental Table 1).[5,6,10-17,22-27] Five models used age-period-cohort (APC) analyses to estimate 1975-2012 breast cancer incidence rates in the absence of screening (baseline incidence rate)[17,25]; model M applied a Bayesian approach to extend 1975-1979 Surveillance Epidemiology and End Results (SEER) rates forward in time with a 4% (SD 0.2%) annual increase. Plain-film and digital mammography sensitivity data from the Breast Cancer Surveillance Consortium (BCSC) for 1994-2012 were used to estimate sensitivity for detection of invasive and DCIS cancers by age group, first vs subsequent screen, and time since last mammogram.

Screening dissemination was derived from national survey data for age at first screen and subsequent screening frequency by birth cohort.[23,24] Plain-film mammography was assumed before 2000. Digital mammography was phased-in starting in 2001 based on data from the BCSC (unpublished data) and the US Food and Drug Administration Mammography Quality Standards Act and Program.[28]

Molecular subtype-specific treatment dissemination was based on SEER patterns-of-care special studies for 1975-1996[26,27] and the National Comprehensive Cancer Network data for 1997 onwards[14,19]. Tamoxifen was used in the 1980s; aromatase inhibitor use began in 1997; taxanes in 1998; and trastuzumab in 2006. Treatment effectiveness was conditioned on stage and ER/ERBB2 status (and age, if applicable), based on clinical trials; all estimates assumed local therapy.[16]

Analyses

Each model simulated mortality rates under four intervention scenarios: 1) no screening or treatment (the baseline mortality rate); 2) screening alone; 3) treatment alone; and 4) combined screening and treatment. Rates were age-adjusted using the 2000 US Standard Population[29] and outcomes were reported for women ages 30-79.

The absolute mortality reductions associated with screening alone, treatment alone, or the combination in a given calendar year were calculated as the difference between the age-adjusted mortality rates predicted with intervention (scenarios 2, 3, or 4) and

the baseline mortality rate in that year (scenario 1). The percent mortality reduction (hereafter referred to as mortality reduction) in a given calendar year was calculated as this difference divided by the baseline mortality rate in that calendar year (scenario 1; Supplemental Table 2).

ER/ERBB2-specific mortality rates were computed by dividing the number of women who died of breast cancer with that subtype by the total breast cancer population at risk. In this manner, rates of all subtypes sum to the overall age-adjusted breast cancer mortality rate.

To estimate the separate contributions associated with screening and treatment to mortality reductions, we considered the modeled effects of screening alone and of treatment alone as a fraction of the combined modeled effect in each calendar year.

The *relative* contribution associated with screening versus treatment to the combination of both was computed as the ratio of the screening alone effect to the sum of the screening alone effect and the treatment alone effect; the relative contribution of treatment was calculated similarly. Alternative approaches for computing these relative contributions were considered, and the main conclusions were unchanged (Supplemental Methods and Supplemental Table 3).

When considering the mortality reductions associated with each treatment intervention (eg, chemotherapy, hormonal therapy, and trastuzumab) to their combination, the relative contribution associated with the various treatments were decomposed by first considering the chemotherapy contribution; then the hormonal therapy contribution for ER+ cases, given chemotherapy contributions; and lastly, the contribution associated with trastuzumab for ERBB2+ cases, given the other therapies.

To estimate relative contributions associated with the most recent advancements, we compared the mortality reduction from 2000 and 2012. We focused on this difference to remove the modeled effect of changes in the baseline rate during this period.

Uncertainty Analysis

All results were reported by model and summarized as the mean and range across models. The range provided a measure of uncertainty because each model has different assumptions and structures to represent unobservable factors such as baseline incidence rate and breast cancer natural history. Results consistent across models were considered robust.

RESULTS

Rates of mammography increased over time (Figure 1A), and plain-film was rapidly replaced by digital mammography starting in 2001 (Figure 1B). Treatment use varied by

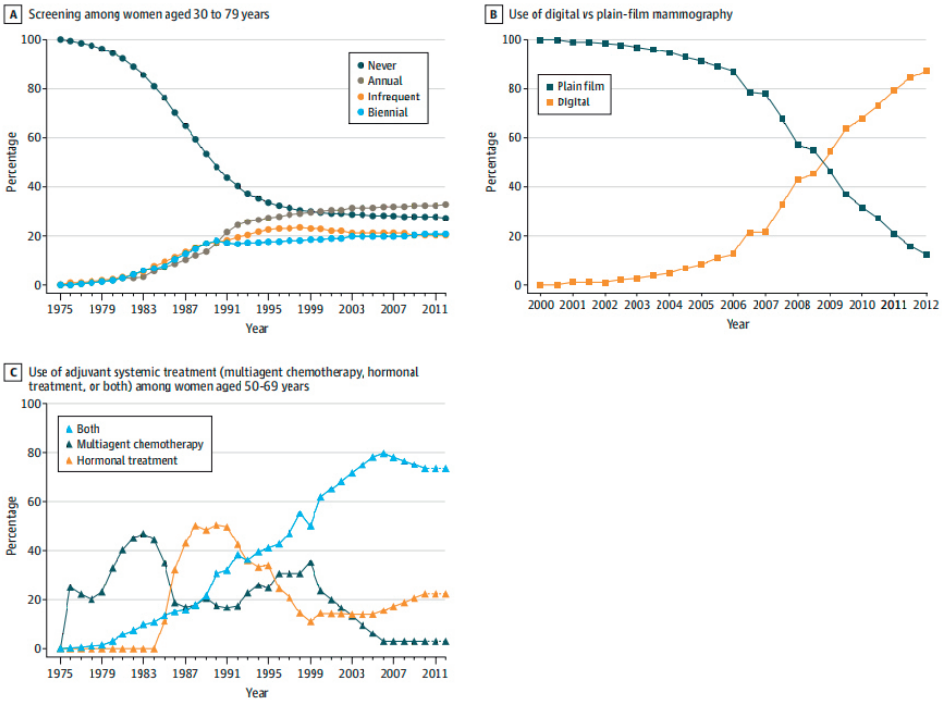


Figure 1 Dissemination of Screening Mammography, Type of Mammography, and Adjuvant Therapy 1975-2012

Panel A shows use of screening among US women ages 30-79 by calendar year based on data from multiple rounds of the National Health Interview Survey over time and Breast Cancer Surveillance Consortium (BCSC) data from 1994-2012. These observed data were used a targets in modeling dissemination of screening and intervals between screens. Note that the rate of never screened includes women ages 30-39.

Panel B illustrates the transition to use of digital vs. plain-film mammography over time using MQSA data on digital mammography facilities from the FDA and the BCSC, which includes over 2.3 million women, aged 30-79, with over 9.5 million mammograms, 95,000 breast cancer cases and 180,000 breast biopsies.

Panel C depicts use of adjuvant systemic treatment dissemination from 1975-2012 for an exemplar stage and set of molecular markers (node positive AJCC 6 stage 2b, ER+/ERBB2-) among women 50 to 69 years of age at diagnosis based on data from SEER special patterns of care studies and the National Comprehenisve Cancer Network. These data were used for all other combinations of ages, stages, and molecular subtypes.

Models used 2010 treatment dissemination data for subsequent years (indicated by dashed segments).

In general, in the 1980's and early 1990's multi-agent chemotherapy included primarily cytoxan, methotrexate, and 5-flourouricil (CMF) regimens; starting in the mid-1990's anthracycline-based regimens were included and increased in use, and in 1997 taxanes could be added to those regimens. Hormonal therapy began with tamoxifen in the 1980's and starting in 1997 also included aromatase inhibitors. Hormonal therapy could be used alone or in combination with multi-agent chemotherapy. Over time, there was an increasing use of both multi-agent chemotherapy and hormonal therapy. For women diagnosed with ERBB2+ tumors (not shown in this example), trastuzumab was disseminated independently of other treatments and, based on its immediate rapid uptake, all ERBB2+ patients were modeled as receiving trastuzumab beginning in year 2006.

molecular subtype, age, and stage, with high rates of dissemination of recent advances (Figure 1C). Incorporating these observed screening and treatment patterns, the models reproduced observed age-adjusted incidence (Supplemental Figure 1) and breast cancer mortality trends from 1975 to 2012 (Figure 2A). Predicted mortality trends for a representative model (Model G-E) illustrate that the mortality reduction associated with treatment alone increased faster than that associated with screening alone over time (Figure 2B).

Overall Breast Cancer Mortality in 2012

With the observed changes in screening technology and treatment regimens, we estimated a 49% (model range: 39%-58%) decrease in overall breast cancer mortality in 2012 relative to the estimated baseline rate in 2012 of 63 deaths (model range, 54-73) per 100,000 women (Table 1, Column 4; Supplemental Table 2). The estimated screening contribution to this mortality reduction was 37% (model range, 26%-51%), while treatment was 63% (model range, 49%-74%). The larger contribution associated with treatment vs. screening in 2012 was predicted in five of six models (Table 1, Columns 7-8).

The estimated 63% (model range, 49%-74%) relative contribution associated with treatment in 2012 consisted of 31% (model range, 23%-37%) from chemotherapy, 27% (model range, 18%-36%) from hormone therapy, and 4% (model range, 1%-6%) from trastuzumab (Table 2).

Molecular Subtype-Specific Breast Cancer Mortality in 2012

The ER+/ERBB2- subtype was estimated to be associated with 64% (model range, 61%-70%) of the overall mortality reduction in 2012 because it was the most common subtype (Supplemental Table 7).

Within-subtype analyses demonstrated significant variations in breast cancer mortality reduction in 2012 (vs estimated subtype-specific baseline rates; Table 1, Column 4). The estimated mortality reduction was largest for the ER+/ERBB2+ subtype at 58% (model range, 46%-71%), followed by the ER+/ERBB2- subtype at 51% (model range, 42%-55%), and the ER-/ERBB2+ subtype at 44% (model range, 33%-55%). The lowest mortality reduction was estimated for the ER-/ERBB2- subtype at 37% (model range, 27%-46%).

The estimated relative contributions associated with screening vs treatment also varied by molecular subtype, ranging from 31% (model range, 23%-41%) versus 69% (model range, 59%-77%) for the ER+/ERBB2+ subtype to 48% (model range, 38%-57%) versus 52% (model range, 43%-62%) for the ER-/ERBB2- subtypes, respectively (Table 1, Columns 7-8). The estimated relative contributions associated with specific treatments varied by subtype (Table 2). For example, for the ER+/ERBB2+ subtype, of the 69% (model range, 59%-77%) relative contribution associated with treatment, 26% (model range, 15%-32%) was associated with chemotherapy, 29% (model range, 23%-36%) with hormone therapy, and 14% (model range, 9%-18%) with trastuzumab (Table 2). For the

ER-/ERBB2- subtype, the 52% (model range, 43%-62%) relative contribution associated with treatment was associated with chemotherapy alone.

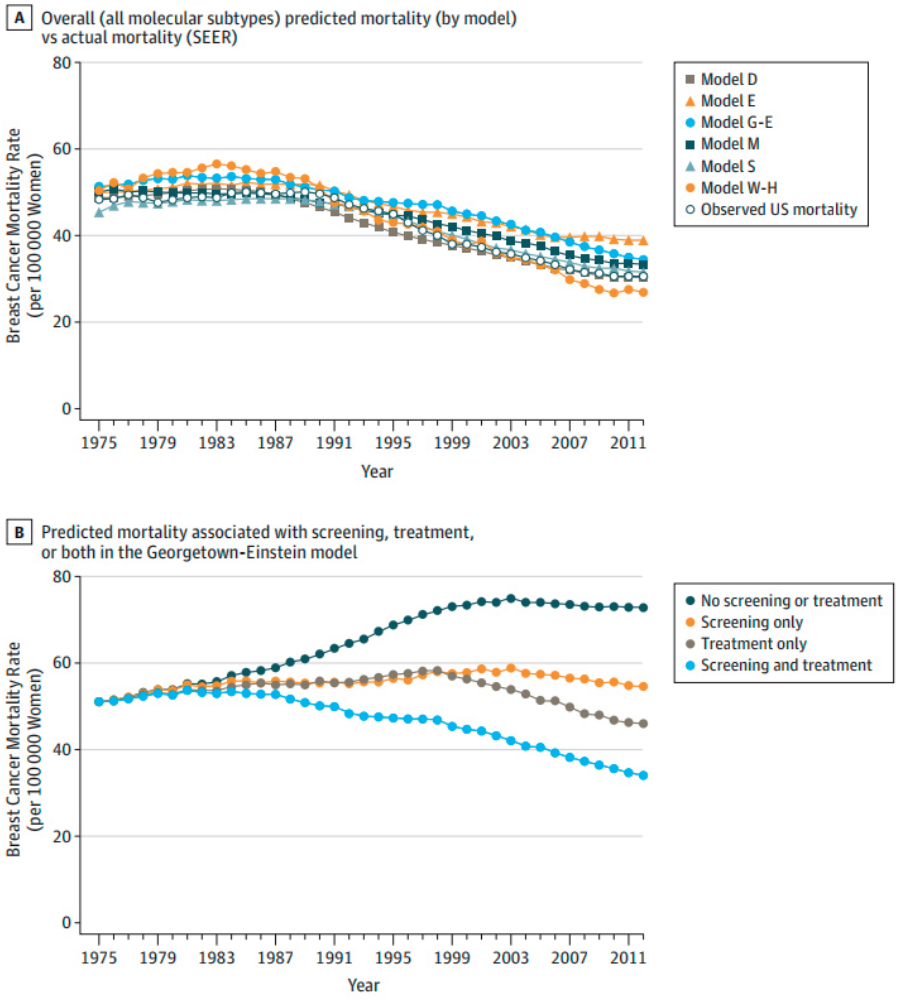


Figure 2 Age-Adjusted Breast Cancer Mortality From 1975-2012 by Model

Panel A compares the predictions the six models for rates of overall breast cancer mortality (all molecular subtypes) to actual US breast cancer mortality among the US population ages 30-79 from 1975 to 2012.

Panel B illustrates predicted breast cancer mortality rates in the US population ages 30-79 from 1975 to 2012 in the absence of screening and adjuvant treatment, presence of screening alone, presence of adjuvant treatment alone, and combination of screening and adjuvant treatment for a representative model (Model Georgetown-Einstein).

Contribution of Screening and Treatment Advances Between 2000 and 2012

The estimated overall breast cancer mortality reduction in 2000 was 37% (model range, 27%-42%) relative to the estimated baseline rate in 2000 of 64 deaths (model range, 56-73) per 100,000 women (Table 3, Column 2; Supplemental Table 2). The estimated overall breast cancer mortality reduction in 2000 was 49% (model range, 39%-58%) relative to the estimated baseline rate in 2012 of 63 deaths (model range, 54-73) per 100,000 women (Table 3, Column 3; Supplemental Table 2). Hence, the estimated difference in the overall breast cancer mortality reduction in 2012 vs. 2000 was 12% (model range, 10%-16%) (Table 3, Column 4; Supplemental Table 5). The estimated relative contribution associated with screening advances to this difference was 17% (2%-31%) (Table 2, Column 5); treatment advances were 83% (69%-98%) (Table 3, sum of Columns 6-8). Of the 83% (69%-98%) treatment-related advances, 38% (model range, 21%-54%) was associated with advances in chemotherapy, largely taxanes; 29% (model range, 9%-44%) was associated with advances in hormone therapy, largely the addition of aromatase inhibitors, and 15% (model range, 4%-25%) with the introduction of trastuzumab (Table 3, Columns 6-8).

Within each molecular subtype, the estimated difference in the breast cancer mortality reductions between 2012 and 2000 was largest for the ER+/ERBB2+ subtype at 19% (model range, 17%-25%) and the smallest for the ER-/ERBB2- subtype at 8% (model range, 5%-11%) (Table 3, Column 4). The estimated relative contribution of screening and treatment to these differences also varied by subtype: the relative contribution of trastuzumab was 41% (model range, 27%-58%) in the ER+/ERBB2+ subtype and 57% (model range, 35%-78%) in the ER-/ERBB2+ subtype (Table 3, Column 8).

To complement the above analysis, we decomposed the overall mortality reduction in 2012 in terms of the contributions associated with advances before 2000 and after 2000 (Supplemental Table 6). Of the 37% (model range, 27%-42%) mortality reduction associated with screening in 2012, 33% (model range, 29%-48%) was associated with screening advances before 2000 and 4% (model range, 1%-8%) after 2000, largely digital mammography. The introduction of trastuzumab was associated with 15% of overall mortality reduction between 2000 and 2012. Of the 31% (model range, 23%-37%) associated with chemotherapy, 22% (model range, 15%-30%) was associated with chemotherapy advances before 2000 and 9% (model range, 7%-14%) after 2000, largely taxanes. Of the 27% mortality reduction (model range, 18%-36%) associated with hormone therapy, 20% (model range, 15%-27%) was associated with advances in hormone therapy before 2000 and 7% (model range, 2%-12%) after 2000, largely from aromatase inhibitors. Supplemental Table 6 provides subtype-specific results.

Table 1 Overall And Subtype-Specific Breast Cancer Mortality Reductions in 2012 Associated with Screening, Treatment, Or Both by Model*

	(A) From Screening Alone	(B) From Treatment Alone	(C) From Combined Screening and Treatment	(D) Percentage of Mortality Reduction Associated with Screening Alone	(E) Percentage of Mortality Reduction Associated with Treatment Alone	(F) Relative Contribution Associated with Screening to Combined Screening and Treatment, %	(G) Relative Contribution Associated with Treatment to Combined Screening and Treatment, %
Operation**	A	B	C	A/C	B/C	A/(A+B)	B/(A+B)
Model				**Overall**			
Dana-Farber	29	28	49	59	57	51	49
Erasmus	18	30	43	41	70	37	63
Georgetown-Einstein	25	37	53	47	69	40	60
MD Anderson	17	29	39	44	73	38	62
Stanford	18	37	50	36	74	33	67
Wisconsin-Harvard	17	49	58	30	84	26	74
Mean	21	35	49	43	71	37	63
				ER+, ERBB2- Subtype			
Dana-Farber	30	30	52	59	58	50	50
Erasmus	18	34	46	39	73	35	65
Georgetown-Einstein	26	39	54	48	71	40	60
MD Anderson	17	31	42	42	75	36	64
Stanford	19	41	53	35	77	31	69
Wisconsin-Harvard	16	51	59	27	86	24	76
Mean	21	38	51	42	73	36	64
				ER+, ERBB2+ Subtype			
Dana-Farber	27	38	57	46	67	41	59
Erasmus	20	42	52	39	82	32	68
Georgetown-Einstein	24	43	58	41	74	36	64
MD Anderson	18	38	46	38	82	32	68
Stanford	17	58	66	26	88	23	77
Wisconsin-Harvard	19	62	71	26	87	23	77
Mean	21	47	58	36	80	31	69

Table 1 Overall And Subtype-Specific Breast Cancer Mortality Reductions in 2012 Associated with Screening, Treatment, Or Both by Model* (continued)

	(A) From Screening Alone	(B) From Treatment Alone	(C) From Combined Screening and Treatment	(D) Percentage of Mortality Reduction Associated with Screening Alone	(E) Percentage of Mortality Reduction Associated with Treatment Alone	(F) Relative Contribution Associated with Screening to Combined Screening and Treatment, %	(G) Relative Contribution Associated with Treatment to Combined Screening and Treatment, %
ER-, ERBB2+ Subtype							
Dana-Farber	25	28	49	52	58	47	53
Erasmus	17	28	41	40	68	37	63
Georgetown-Einstein	25	32	52	48	62	43	57
MD Anderson	15	23	33	45	70	39	61
Stanford	17	25	40	42	63	40	60
Wisconsin-Harvard	23	43	55	41	79	34	66
Mean	20	30	45	45	67	40	60
ER-, ERBB2- Subtype							
Dana-Farber	26	20	40	66	50	57	43
Erasmus	17	22	35	47	64	43	57
Georgetown-Einstein	24	29	46	53	63	45	55
MD Anderson	18	14	27	65	52	56	44
Stanford	18	17	33	53	50	52	48
Wisconsin-Harvard	18	30	42	43	70	38	62
Mean	20	22	37	55	58	48	52

* The column labels are defined as follows: (A) mortality reduction associated with screening alone, relative to the estimated baseline mortality in 2012; (B) mortality reduction associated with treatment alone relative to the estimated baseline mortality in 2012; (C) mortality reduction associated with combined screening and treatment alone, relative to the estimated baseline mortality in 2012 ("combined mortality reduction"); (D) percentage of combined mortality reduced captured by screening alone; (E) percentage of combined mortality reduction captured by treatment alone; (F) relative contribution of screening to combined mortality reduction; (G) relative contribution of treatment to combined mortality reduction. Note: Columns F and G sum to 100%.

** Operation refers to the calculation of the result in the table. For example, column D (Percentage of Mortality Reduction Captured by Screening Alone) is calculated as the result in column A for screening alone divided by the result in column C for the combined mortality reduction with both screening and treatment.

Table 2 Relative Contributions of Treatments to Mortality Reduction in 2012

	Relative Contribution Associated with Chemotherapy, %	Relative Contribution Associated with Hormone Therapy, %	Relative Contribution Associated with Trastuzumab, %
Model		**Overall**	
Dana-Farber	23	24	2
Erasmus	37	25	1
Georgetown-Einstein	37	18	4
MD Anderson	22	34	6
Stanford	34	28	5
Wisconsin-Harvard	33	36	5
Mean	31	27	4
		ER+, ERBB2- Subtype	
Dana-Farber	25	25	0
Erasmus	30	35	0
Georgetown-Einstein	34	24	0
MD Anderson	21	42	0
Stanford	33	36	0
Wisconsin-Harvard	29	47	0
Mean	29	35	0
		ER+, ERBB2+ Subtype	
Dana-Farber	24	23	12
Erasmus	28	30	10
Georgetown-Einstein	32	23	9
MD Anderson	15	36	18
Stanford	30	30	17
Wisconsin-Harvard	25	34	18
Mean	26	29	14
		ER-, ERBB2+ Subtype	
Dana-Farber	36	0	16
Erasmus	45	0	18
Georgetown-Einstein	43	0	11
MD Anderson	24	0	29
Stanford	35	0	25
Wisconsin-Harvard	42	0	23
Mean	37	0	21
		ER-, ERBB2- Subtype	
Dana-Farber	43	0	0
Erasmus	57	0	0
Georgetown-Einstein	53	0	0
MD Anderson	42	0	0
Stanford	48	0	0
Wisconsin-Harvard	62	0	0
Mean	51	0	0

Table 3 Relative Contributions Associated with Advances In Screening and Treatment to The Difference In The Mortality Reduction Between 2000 and 2012

	(A) Mortality Reduction in 2000 (Relative to the estimated baseline mortality rate in 2000), %	(B) Mortality Reduction in 2012 (Relative to the estimated baseline mortality rate in 2012), %	(C) Difference in the Mortality Reduction Between 2000 and 2012, %	Relative Contributions to the Difference in the Mortality Reduction Between 2000 and 2012, % *			
				(D) Relative Contribution Associated with Screening Advances Between 2000 and 2012, %	(E) Relative Contribution Associated with Chemo-therapy Advances Between 2000 and 2012, %	(F) Relative Contribution Associated with Hormone Therapy Advances Between 2000 and 2012, %	(G) Relative Contribution Associated with Trastuzumab, %
Operation*			B-A	D+E+F+G = 100%			
Model				**Overall**			
Dana-Farber	39	49	10	13	34	44	10
Erasmus	32	43	10	31	32	33	4
Georgetown-Einstein	39	53	14	21	54	9	15
MD Anderson	27	39	13	23	21	37	18
Stanford	40	50	10	14	41	20	25
Wisconsin-Harvard	42	58	16	2	48	31	18
Mean	**37**	49	12	17	38	29	15
			ER+, ERBB2- Subtype				
Dana-Farber	43	52	9	14	39	47	0
Erasmus	34	46	13	21	14	64	0
Georgetown-Einstein	41	54	13	29	62	9	0
MD Anderson	29	42	13	24	25	50	0
Stanford	45	53	8	19	46	35	0
Wisconsin-Harvard	45	59	14	3	49	48	0
Mean	39	51	12	19	39	42	0
			ER+, ERBB2+ Subtype				
Dana-Farber	41	57	17	10	19	29	42
Erasmus	33	52	19	24	8	41	27
Georgetown-Einstein	41	58	17	14	46	16	24
MD Anderson	28	46	18	17	6	32	45
Stanford	47	66	19	4	23	14	58
Wisconsin-Harvard	46	71	25	0	29	20	51
Mean	**39**	**58**	**19**	12	22	25	41
			ER-, ERBB2+ Subtype				
Dana-Farber	33	49	16	11	37	0	52
Erasmus	26	41	15	13	37	0	50
Georgetown-Einstein	33	52	19	21	44	0	35
MD Anderson	20	33	13	20	3	0	78
Stanford	26	40	14	0	30	0	70

Table 3 Relative Contributions Associated with Advances In Screening and Treatment to The Difference In The Mortality Reduction Between 2000 and 2012 (continued)

	(A) Mortality Reduction in 2000 (Relative to the estimated baseline mortality rate in 2000), %	(B) Mortality Reduction in 2012 (Relative to the estimated baseline mortality rate in 2012), %	(C) Difference in the Mortality Reduction Between 2000 and 2012, %	Relative Contributions to the Difference in the Mortality Reduction Between 2000 and 2012, % *			
				(D) Relative Contribution Associated with Screening Advances Between 2000 and 2012, %	(E) Relative Contribution Associated with Chemo-therapy Advances Between 2000 and 2012, %	(F) Relative Contribution Associated with Hormone Therapy Advances Between 2000 and 2012, %	(G) Relative Contribution Associated with Trastuzumab, %
Wisconsin-Harvard	33	55	22	0	42	0	58
Mean	**29**	**45**	15	11	32	0	57
ER-, ERBB2- Subtype							
Dana-Farber	34	40	6	13	87	0	0
Erasmus	26	35	10	34	66	0	0
Georgetown-Einstein	35	46	11	14	86	0	0
MD Anderson	22	27	5	41	59	0	0
Stanford	27	33	7	23	77	0	0
Wisconsin-Harvard	32	42	10	9	91	0	0
Mean	**29**	**37**	8	22	78	0	0

* Operation refers to the calculation of the results in the table for each column. The column labels (A through G) are included with each column title.

Details on the computations are included in the Supplemental methods. Briefly, the estimated mean overall mortality reduction associated with combined screening and treatment in 2012 relative to the estimated baseline mortality rate in 2012 (Table 3, Column B) was 49% and in 2000 (Table 3, Column A) it was 37%. Thus, there was an additional 12% mortality reduction in 2012 compared to 2000 (Table 3, Column C).

In 2000, the estimated relative contribution of screening to the mortality reduction associated with combined screening and treatment was 44% (Supplemental Table 4, Row D), hence the mortality reduction associated with screening is 44% of 37% = 16%.

In 2012, the estimated relative contribution of screening to the mortality reduction associated with combined screening and treatment was 37% (Table 1 and Supplemental Table 4, Row M), hence the mortality reduction associated with screening is 37% of 49% = 18%.

The difference in the mortality reduction associated with screening between 2012 and 2000 is 18%-16% = 2%. Hence, the relative contribution of screening advances to the difference in the mortality reduction associated with combined screening and treatment was 2% divided by 12% (Table 3, Column C), giving 17% (Column D). The remainder (83%) is associated with treatment advances. This 83% is distributed by treatment type in columns E-G. Columns D to G total 100%.

DISCUSSION

This model-based analysis provides clinically relevant insights about the separate and combined population contributions associated with screening and treatment advances on reducing breast cancer mortality by molecular subtype. Six independent models found that both screening and treatment were associated with overall and subtype-specific breast cancer mortality declines over time. Between 2000 and 2012, advances in treatment were associated with a larger contribution than screening to overall US breast cancer mortality decreases and for all molecular subtypes except ER-/ERBB2-, the subtype that also had the lowest modeled mortality reduction.

These results build upon past CISNET analyses and other studies that have examined the period before 2000[2,30-32] or considered the role of ER-status.[15,33] The current analysis considered the study period from 2000 to 2012. In this period, digital mammography increased screening sensitivity compared with plain-film mammography, especially for women younger than 50 years and women with dense breasts,[34] and has increased somewhat the number of breast cancer deaths averted with screening.[35] The current results support findings that advances in mammography continue to contribute to reducing breast cancer mortality. It will be important to update the analysis when there is sufficient evidence about the benefits of tomosynthesis or other emerging screening approaches.[36,37]

Even with the recent screening advances, findings from this model-based analysis demonstrate a shift in the relative contributions associated with screening and treatment to breast cancer mortality, with greater contributions associated with treatment in 2012. Recent observational analyses have also found stage-specific survival improvements related to current treatment.[33] The results from this model analysis confirm the benefits at the population level from the discovery and rapid dissemination over this past decade of several new classes of molecularly-targeted therapies, improvements in delivery of standard regimens, and refinements in therapy based on molecular subtype based by ER and ERBB2 status.

A unique contribution of this population-level analysis is how the relative contributions associated with screening and treatment varied by molecular subtype. In 2012, when gains from treatment alone were estimated, treatment alone could have provided roughly 70% of the predicted mortality reduction achieved with both screening and treatment for the all the subtypes expressing the ER and/or ERBB2 receptors. However, screening is likely to remain important even if future treatments could cure all breast cancers, because screening can detect disease at earlier stages where there is less surgical and treatment-related morbidity compared to that with therapy for more advanced stages.

Among the advances in recent adjuvant treatments, advances in chemotherapy with the addition of taxanes were associated with roughly 37% of the difference in overall

breast cancer mortality reduction from 2000 to 2012. Advances in hormone therapy with the addition of aromatase inhibitors had comparable contribution associated with mortality reduction. The contribution associated with trastuzumab was smaller on overall breast cancer mortality (13%), because ERBB2+ cases only account for approximately 20% of all newly diagnosed breast cancers.[38] However, trastuzumab was associated with more than 40% of the difference in mortality reduction between 2000 and 2012 among the ERBB2+ subtypes.

All of the models concluded that the ER-/ERBB2- subtype had the lowest overall modeled mortality reduction over time, although the relative contributions associated with screening and treatment varied somewhat by model, with three of the six models estimating a modestly higher contribution associated with treatment compared to screening in 2012. Prior analysis of SEER data have similar results, with greater mortality declines for those with ER+ vs. ER-tumors.[15,39] Given that treatment advancements are lagging for ER-/ERBB2- cancers, more intensive screening approaches, or screening with different modalities, might be considered for groups at highest risk for this subtype, including African American women. Continued investments to discover molecularly-targeted treatments for the ER-/ERBB2- subgroup remain important to continue to lower breast cancer death rates.

Overall, the models projected that screening and treatment each were associated with continued reductions in breast cancer mortality, but in 2012 treatment was associated with a larger relative proportion than screening of the mortality reductions overall and for all subtypes, except the ER-/ERBB2-. Because ER+ cancers are most prevalent and this group is expected to increase with time,[40] additional advances for this subtype could have the largest impact on reducing the overall population burden of breast cancer. Looking ahead, model-based approaches may continue to be important to evaluate continued population-level progress in reducing the burden of breast cancer through a combination of continued discovery and dissemination of effective molecularly-targeted therapies, invention of novel screening technologies to optimize early detection of aggressive cancer subtypes, and greater ability to identify risk of developing specific molecular subtypes to permit tailored prevention and early detection.

This study has several strengths. First, by synthesizing national and clinical trial data, the results fill an important knowledge gap, especially because current surveillance data systems do not contain information on both screening and treatment. Second, the main findings were robust across six independent models, despite differences in model structures and assumptions. Third, the validity of this comparative modeling approach is supported by the consistency of conclusions across models, and the ability of each model to closely replicate the patterns of observed trends in incidence and mortality.

This research also has several limitations. First, the accuracy of model results depends on the availability of good quality data for input parameters and reasonable assumptions

about unobservable events. For instance, because there are limited long-term clinical trial or registry data on survival by ERBB2 status, the models extrapolated long-term survival. Second, modeled treatment effects were based on efficacy in trials included in the Oxford Overview,[16] so could have slightly over-estimated actual population treatment effects, and the relative contribution of treatment to mortality reductions. Third, each model also made different assumptions about the baseline incidence and natural history of breast cancer, leading to variability in the magnitude of results. Fourth, the models considered only five years of hormonal therapy since recommendations to consider 10-years among women at high-risk of late recurrence were just recently introduced and have not yet been uniformly applied. Future modeling could incorporate the population-level dissemination and effectiveness of longer-term hormonal therapy. Fifth, progesterone-receptor status was not explicitly modeled since it is missing from many data sources. Sixth, subtype results for various racial/ethnic subgroups were not modeled. Understanding interactions between race, ethnicity, and subtype-specific outcomes represents an important future direction.[41] Seventh, the effect of screening and subtype-specific treatment on morbidity and all-cause mortality was not evaluated. Eighth, modeling was based on estimates until 2012, and it is uncertain whether or how well these estimates reflect current breast cancer screening, treatment, or outcomes after 2012.

CONCLUSIONS

In this simulation modeling study that projected trends in breast cancer mortality rates among US women, decreases in overall breast cancer mortality from 2000 to 2012 were associated with advances in screening and in adjuvant therapy, although the associations varied by breast cancer molecular subtypes.

ACKNOWLEDGEMENTS

We thank the Breast Cancer Surveillance Consortium (BCSC) investigators, participating women, mammography facilities, and radiologists for the de-identified data they provided for this study. A list of the BCSC investigators and procedures for requesting BCSC data for research purposes are provided at: http://breastscreening.cancer.gov/. The collection of cancer and vital status data from the BCSC was supported in part by several state public health departments and cancer registries throughout the US. For a full description of these sources, please see: http://breastscreening.cancer.gov/work/acknowledgement.html.

We are also grateful to Ruth Etzioni, PhD, Member of the Fred Hutchinson Cancer Research Center, for reviewing an early draft of the manuscript; Dr. Etzioni did not receive compensation for her review.

Drs. Plevritis and Mandelblatt had full access to all of the data in the study and take full responsibility for the integrity of the data and the accuracy of the data analysis.

The funders of the study had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript or the decision to submit for publication.

REFERENCES

1. National Cancer Institute, Surveillance, Epidemiology, and End-Results Program. Breast Cancer Statistics. https://seer.cancer.gov/statfacts/html/breast.html. Accessed March 31 2017.
2. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *New England Journal of Medicine*. 2005;353(17):1784-1792.
3. Kerlikowske K, Hubbard RA, Miglioretti DL, et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. *Annals of Internal Medicine*. 2011;155(8):493-502.
4. Kerlikowske K, Zhu W, Tosteson AN, et al. Identifying women with dense breasts at high risk for interval cancer: a cohort study. *Annals of Internal Medicine*. 2015;162(10):673-681.
5. Giordano SH, Temin S, Kirshner JJ, et al. Systemic therapy for patients with advanced human epidermal growth factor receptor 2-positive breast cancer: American Society of Clinical Oncology clinical practice guideline. *Journal of Clinical Oncology*. 2014;32(19):2078-2099.
6. National Comprehensive Cancer Network Breast Cancer Guidelines. NCCN guidelines and clinical resources https://www.nccn.org/professionals/physician_gls/f_guidelines.asp - breast Accessed March 31 2017.
7. Lee SJ, Li X, Huang H, Zelen M. Models for breast cancer screening strategies updated for ductal carcinoma in situ and subgroups. *Medical Decision Making*. 2018;38(1s).
8. van den Broek JJ, van Ravesteyn NT, Heijnsdijk EA, de Koning H. Estimating the effects of risk based screening and adjuvant treatment using the miscanfadia continuous tumor growth model for breast cancer. *Medical Decision Making*. 2018;38(1s).
9. Schechter CB, Near AM, Jayasekera J, Chang Y, Mandelblatt JS. structure, function, and applications of the Georgetown-Einstein (GE) breast cancer simulation model. *Medical Decision Making*. 2018;38(1s).
10. Huang X, Li Y, Song J, Berry D. The MD Anderson CISNET model for estimating benefits of adjuvant therapy and screening mammography for breast cancer: an update. *Medical Decision Making*. 2018;38(1s).
11. Munoz D, Plevritis SK. Estimating breast cancer progression features and survival by molecular subtype in the absence of screening and treatment. *Medical Decision Making*. 2018;38(1s).
12. Munoz D, Xu C, Plevritis S. A molecular subtype-specific stochastic simulation model of US breast cancer incidence and mortality trends from 1975 to 2010. *Medical Decision Making*. 2018;38(1s).
13. Alagoz O, Ergun MA, Cevik M, et al. The University Of Wisconsin breast cancer epidemiology simulation model: An update. *Medical Decision Making*. 2018;38(1s).
14. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Annals of Internal Medicine*. 2009;151(10):738-747.
15. Munoz D, Near AM, van Ravesteyn NT, et al. Effects of screening and systemic adjuvant therapy on ER-specific US breast cancer mortality. *Journal of the National Cancer Institute*. 2014;106(11).
16. 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 (London, England)*. 2012;379(9814):432-444.

17. Gangnon RE, Sprague BL, Stout NK, et al. The contribution of mammography screening to breast cancer incidence trends in the United States: an updated age-period-cohort model. *Cancer Epidemiology, Biomarkers & Prevention*. 2015;24(6):905-912.
18. Gangnon RE, Stout NK, Alagoz O, Hampton JM, Sprague BL, Trentham-Dietz A. Contribution of breast cancer to overall mortality for U.S. women. *Medical Decision Making*. 2018;38(1s).
19. Mandelblatt JS, Cronin K, de Koning H, Miglioretti DL, Schechter CS, Stout N. Collaborative modeling of U.S. breast cancer screening strategies. AHRQ Publication No. 14-05201-EF-4. Rockville, MD: U.S. Preventive Services Task Force; 2015. www.uspreventiveservicestaskforce.org/Page/Document/modeling-report-collaborative-modeling-of-us-breast-cancer-1/breast-cancer-screening1. Accessed May 2015.
20. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Mortality - All COD, Aggregated With State, Total U.S. (1969-2013 <Katrina/Rita Population Adjustment>, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2016. Underlying mortality data provided by NCHS (www.cdc.gov/nchs). Accessed March 31 2017.
21. National Center for Health Statistics. Compressed mortality file. https://www.cdc.gov/nchs/data_access/cmf.htm. Accessed March 31, 2017
22. Mandelblatt JS, Near AM, Miglioretti DL, Munoz D, Sprague BK, Trentham-Dietz A, Gangnon R, Kurian AW, Weedon-Fekjaer H, Cronin KA, Plevritis SK. Common model inputs used in CISNET collaborative breast cancer modeling. *Medical Decision Making*. 2018;38(1s).
23. Cronin KA, Mariotto AB, Clarke LD, Feuer EJ. Additional common inputs for analyzing impact of adjuvant therapy and mammography on U.S. mortality. *Journal of the National Cancer Institute Monographs*. 2006(36):26-29.
24. Cronin KA, Yu B, Krapcho M, et al. Modeling the dissemination of mammography in the United States. *Cancer causes & control : CCC*. 2005;16(6):701-712.
25. Holford TR, Cronin KA, Mariotto AB, Feuer EJ. Changing patterns in breast cancer incidence trends. *Journal of the National Cancer Institute Monographs*. 2006(36):19-25.
26. Mariotto A, Feuer EJ, Harlan LC, Wun LM, Johnson KA, Abrams J. Trends in use of adjuvant multi-agent chemotherapy and tamoxifen for breast cancer in the United States: 1975-1999. *Journal of the National Cancer Institute*. 2002;94(21):1626-1634.
27. Mariotto AB, Feuer EJ, Harlan LC, Abrams J. Dissemination of adjuvant multiagent chemotherapy and tamoxifen for breast cancer in the United States using estrogen receptor information: 1975-1999. *Journal of the National Cancer Institute Monographs*. 2006(36):7-15.
28. US Food and Drug Administration. Mammography Quality Standards Act and Program http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/default.htm. June 2013. Silver Spring, Maryland US Dep Health and Human Services. Accessed January 2015.
29. U.S. Bureau of the Census. Population projections of the United States by age, sex, race and hispanic origin: 1995 to 2050. Current Population Reports, P25-1130, U.S. Government Printing Office, Washington, DC, 1996. http//

www.census.gov/prod/1/pop/p25-1130/p251130.pdfExternal Web Site Policy Accessed May 2015.

30. Elkin EB, Hudis C, Begg CB, Schrag D. The effect of changes in tumor size on breast carcinoma survival in the U.S.: 1975-1999. *Cancer.* 2005;104(6):1149-1157.
31. Hermon C, Beral V. Breast cancer mortality rates are levelling off or beginning to decline in many western countries: analysis of time trends, age-cohort and age-period models of breast cancer mortality in 20 countries. *British Journal of Cancer.* 1996;73(7):955-960.
32. Peto R, Boreham J, Clarke M, Davies C, Beral V. UK and USA breast cancer deaths down 25% in year 2000 at ages 20-69 years. *Lancet.* 2000;355(9217):1822.
33. Park JH, Anderson WF, Gail MH. Improvements in US breast cancer survival and proportion explained by tumor size and estrogen-receptor status. *Journal of Clinical Oncology.* 2015;33(26):2870-2876.
34. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *New England Journal of Medicine* 2005;353(17):1773-1783.
35. Stout NK, Lee SJ, Schechter CB, et al. Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography. *Journal of the National Cancer Institute.* 2014;106(6):dju092.
36. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *Journal of the American Medical Association.* 2014;311(24):2499-2507.
37. Lee CI, Lehman CD. Digital breast tomosynthesis and the challenges of implementing an emerging breast cancer screening technology into clinical practice. *Journal of the American College of Radiology.* 2013;10(12):913-917.
38. Howlader N, Altekruse SF, Li CI, et al. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *Journal of the National Cancer Institute.* 2014;106(5).
39. Jatoi I, Chen BE, Anderson WF, Rosenberg PS. Breast cancer mortality trends in the United States according to estrogen receptor status and age at diagnosis. *Journal of Clinical Oncology.* 2007;25(13):1683-1690.
40. Rosenberg PS, Barker KA, Anderson WF. Estrogen receptor status and the future burden of invasive and In situ breast cancers in the United States. *Journal of the National Cancer Institute.* 2015;107(9). djv159, ttps://doi.org/10.1093/jnci/djv159
41. Iqbal J, Ginsburg O, Rochon PA, Sun P, Narod SA. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. *Journal of the American Medical Association.* 2015;313(2):165-173.

Supplementary Online Content

Plevritis SK, Munoz D, Kurian A, et al. Association of screening and treatment with breast cancer mortality by molecular subtype in US women, 2000-2012. *JAMA*. doi:10.1001/jama.2017.19130

eMethods.
eTable 1. Model Parameters
eTable 2. Computation of the Mortality Reduction, Relative to the Baseline Rate
eTable 3. Comparison of 3 Alternative Methods to Compute the Relative Contributions Associated With Screening and Treatment on Overall Breast Cancer Mortality Reduction
eTable 4. Comparison of Breast Cancer Mortality Reduction, Overall and By ER/ERBB2-Subtype in 2000 vs 2012
eTable 5. Relative Contributions Associated With Screening and Treatment Advances on the Difference in the Breast Cancer Mortality Reduction Between 2000 and 2012
eTable 6. Relative Contributions Associated With Screening, Chemotherapy, Hormone Therapy and Trastuzumab to Breast Cancer Mortality Reduction in 2012, Broken Down by Advances Before and After 2000
eTable 7. Breakdown of Overall Breast Cancer Mortality Reduction in 2012 by Molecular Subtype
eFigure 1. Comparison of Model Projections to Observed US Breast Cancer Incidence, for Women Ages 30-79
eFigure 2. Comparison of Model Projections for ER-/ ERBB2-Specific Breast Cancer Mortality Trends Between 1975-2012, for Women Ages 30-79, by Molecular Subtype
eFigure 3. Individual Model Projections for Overall US Breast Cancer Incidence and Mortality (vs SEER) and ER/ERBB2-Subtype-Specific Mortality From 1975-2012, for Women Ages 30-79
eReferences.

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods

Computing the Relative Contributions Associated with Screening and Treatment

In the main text, the relative contribution associated with screening versus treatment to the combination was computed as the ratio of the screening alone mortality reduction divided by the sum of the screening alone mortality reduction and treatment alone mortality reduction; similarly for the relative contribution associated with treatment. Herein, we refer to this approach as "Method A." Two alternative approaches for computing the relative contributions associated with screening and treatment were also considered. In "Method B," we evaluated the relative contributions associated with screening and treatment by first quantifying the contributions associated with screening alone and assigning the remainder of the combined effect to treatment. In "Method C", we evaluated the relative contributions associated with screening and treatment by first quantifying the contributions associated with treatment alone and assigning the remainder of the combined effect to screening. A comparison of all three approaches to compute the relative contributions associated with screening and treatment on overall breast cancer mortality is provided in Supplemental eTable 3. All three approaches provide the same ranking of relative contributions, but results differ because the combination associated with screening and treatment is less than the sum of the contributions associated with screening alone and treatment alone. If the combination was equal to the sum of screening alone and treatment alone, all three methods would give the same result. Because Method A provided a result that was "in-between" Methods B and C, we choose it for the primary analysis.

Computing the Relative Contributions Associated with Screening and Treatment to the Difference in the Reduction Between 2000 and 2012

In Table 3 of the main text, the relative contribution associated with screening and treatment advances to the different in the mortality reduction between 2000 and 2012 are provided. The results in Table 3 are based on the difference in breast cancer mortality reduction in 2012 and breast cancer mortality reduction 2000. Note that the mortality reduction in 2012 is computed relative to the estimated baseline breast cancer mortality in 2012, where the estimate baseline mortality rate in a given calendar year is defined as the estimate mortality rate in that calendar year had there never been screening or adjuvant therapy. Similarly, the mortality reduction in 2000 is computed relative to the estimated baseline breast cancer mortality in 2000. By computing the difference between 2000 and 2012, the baseline effect is removed and the difference estimates the effect of screening and treatment only (not the baseline effect) over this time period. If we did not remove the effect of baseline then the difference in the mortality rate between 2012 and 2000 could be associated with changes in the baseline as well as changes in screening and treatment. Removing the estimated baseline trend provides more robust results for the relative contributions associated with screening and treatment.

To understand how the relative contributions associated with screening and treatment to the difference in the mortality reduction between 2000 and 2012 is computed, we describe the calculations based on overall mortality using the mean results in Table 3. The overall mortality reduction associated with combined screening and treatment was estimated as 37% in 2000 and 49% in 2012, yielding a difference of 12% between 2000 and 2012. In 2000, the relative contribution associated with screening to

the overall mortality reduction was 44% (based on Method A in Supplemental eTable 3), so the mortality reduction associated with screening (vs. baseline) was 44% of 37% = 16% in 2000. In 2012, the relative contribution associated with screening to the overall mortality reduction was 37% (based on Method A in Supplemental Table 3), so the mortality reduction associated with screening (vs. baseline) was 37% of 49% = 18%. The difference in the mortality reduction associated with screening between 2012 and 2000 was 18%- 16% = 2%. This was associated with screening advances (in this case the conversion to digital mammography because the dissemination of screening had not significantly changed). Hence the relative contribution of screening advances to the difference in the mortality reduction associated with combined screening and treatment was estimated as 2% divided by 12%, giving 17%. This leaves 83% associated with treatment advances. **Supplemental eTable 5** provides the results of these calculations for each model, and the mean across the models.

eTable 1. Model Parameters

Parameters	Data	Data Source*
Common Model Parameters		
Incidence in the absence of screening	An age-period-cohort model is used as a starting point for most models (except Model M)	Ref. [1,2]
Mammography dissemination	Screening dissemination is based on the age at first screening and frequency by birth cohort derived from BCSC and NHIS data through 2012	Ref.[3,4]
Proportion of plain film vs. digital mammograms by year	Estimated percent of mammograms in the US that are digital by year from FDA MQSA and BCSC data	Ref.[5,6] BCSC (unpublished data)
Mammography performance	By age, type of screen (initial vs. subsequent), screen interval, and plain film vs. digital	BCSC (unpublished data)
Distribution of ER/ERBB2-status by age and stage	The probability of ER/ERBB2 conditional on age and stage at diagnosis	BCSC (unpublished data)
Survival in the absence of screening and treatment, Overall and by ER/ERBB2	26-year breast cancer survival before adjuvant treatment by joint ER/ERBB2 status, age group, and AJCC/SEER stage or tumor size	Ref.[18]
ER/ERBB2 specific treatment dissemination by year	Based on observed dissemination in the population over time from SEER and the NCCN Outcomes Database (1997-2012)	Ref. [5,7,8] NCCN Outcomes Database (unpublished data)
ER/ERBB2-specific treatment efficacy	Meta-analyses of clinical trial results	Ref.[9]
Non-cancer competing causes of death	Age- and cohort-specific all-cause mortality rates by year	Ref. [10]
Model-specific Parameters		
Tumor sojourn time (or mean tumor doubling time)	Sojourn time by joint ER/ERBB2 status and age group	Ref.[18]
Proportion of DCIS that progresses to invasive cancer	Varies by model	Ref.[5,11-16]
Mean stage dwell time** or tumor growth rates or both	Varies by models based on model structure; can vary by age and/or ER/ERBB2 status	Ref.[11-17]
Screening effects	Stage-shift or change in tumor size between screened and unscreened populations	Ref.[11-16]

* All reference citations refer to those in the main text.
** The mean stage well time is defined as the average time a tumor spends in each stage before progressing to the next.

eTable 2. Computation of the Percent Mortality Reduction, Relative to the Baseline Rate

	Mortality Rate, per 100,000 Women				Mortality Reduction, Relative to Baseline Rate, %		
	No Screening, No Treatment ("Baseline")	Screening Alone	Treatment Alone	Combined Screening and Treatment	Screening Alone	Treatment Alone	Combined Screening and Treatment
Column ID	A	B	C	D	E	F	G
Operation	A	B	C	D	(A-B)/A	(A-C)/A	(A-D)/A
Model	Calendar Year 2000						
Dana Farber	61	44	50	37	27	18	39
Erasmus	65	56	51	44	14	22	32
Georgetown-Einstein	73	58	56	45	21	23	39
MD Anderson	56	48	46	41	13	17	27
Stanford	65	54	47	39	17	28	40
Wisconsin	65	54	45	38	17	30	42
Mean	64	52	49	40	**18**	**23**	**37**
	Calendar Year 2012						
Dana Farber	59	42	43	30	29	28	49
Erasmus	67	56	47	39	18	30	43
Georgetown-Einstein	73	55	46	34	25	37	53
MD Anderson	54	45	39	33	17	29	39
Stanford	63	51	39	31	18	37	50
Wisconsin	63	52	32	27	17	49	58
Mean	63	50	41	32	**21**	**35**	**49**

eTable 3. Comparison of three alternative methods to compute the relative contributions associated with screening and treatment on overall breast cancer mortality reduction in 2012*

				Method A (Main text)		Method B		Method C	
	Mortality reduction in 2012 relative to the estimated baseline mortality rate in 2012, %			Relative contribution associated with screening, %	Relative contribution associated with treatment, %	Relative contribution associated with screening, %	Relative contribution associated with treatment, %	Relative contribution associated with screening, %	Relative contribution associated with treatment, %
	Screening alone	Treatment alone	Combined screening and treatment						
Column ID	A	B	C	D	E	F	G	H	I
Operation	A	B	C	A/(A+B)	B/(A+B)	A/C	1-A/C	1-B/C	B/C
Model	**Overall Breast Cancer Mortality**								
Dana-Farber	29	28	49	51	49	59	41	43	57
Erasmus	18	30	43	37	63	41	59	30	70
Georgetown -Einstein	25	37	53	40	60	47	53	31	69
MD Anderson	17	29	39	38	62	44	56	27	73
Stanford	18	37	50	33	67	36	64	26	74
Wisconsin-Harvard	17	49	58	26	74	30	70	16	84
Mean	**21**	**35**	**49**	**37**	**63**	**43**	**57**	**29**	**71**

* See Supplemental Methods subsection "*Computing the Relative Contributions Associated with Screening and Treatment*" for description of these calculations.

eTable 4. Comparison of breast cancer mortality reduction, overall and by ER/ERBB2-subtype, across models, in 2000 vs 2012

	Mortality reduction in 2000 relative to the estimated baseline mortality rate in 2000, %			Mortality reduction in 2012 relative to the estimated baseline mortality rate in 2012, %			Difference in the mortality reduction between 2012 and 2000, %			Relative contribution associated with screening in 2000, %	Relative contribution associated with treatment in 2000, %	Relative contribution associated with screening in 2012, %	Relative contribution associated with treatment in 2012, %
	Screening alone	Treatment alone	Combined screening and treatment	Screening alone	Treatment alone	Combined screening and treatment	Screening alone	Treatment alone	Combined screening and treatment				
Column ID	A	B	C	D	E	F	G	H	I	J	K	L	M
Operation	A	B	C	D	E	F	D-A	E-B	F-C	A/(A+B)	B/(A+B)	D/(D+E)	E/(D+E)
Model	**Overall**												
Dana-Farber	27	18	39	29	28	49	2	11	10	60	40	51	49
Erasmus	14	22	32	18	30	43	4	8	10	39	61	37	63
Georgetown-Einstein	21	23	39	25	37	53	4	14	14	48	52	40	60
MD Anderson	13	17	27	17	29	39	4	12	13	44	56	38	62
Stanford	17	28	40	18	37	50	1	9	10	38	62	33	67
Wisconsin-Harvard	17	30	42	17	49	58	1	18	16	35	65	26	74
Mean	**18**	**23**	**37**	**21**	**35**	**49**	**3**	**12**	**12**	**44**	**56**	**37**	**63**

eTable 4 (Continued). Comparison of breast cancer mortality reduction, overall and by ER/ERBB2-subtype, across models, in 2000 vs 2012

	Mortality reduction in 2000 relative to the estimated baseline mortality rate in 2000, %			Mortality reduction in 2012 relative to the estimated baseline mortality rate in 2012, %			Difference in the mortality reduction between 2012 and 2000, %			Relative contribution associated with screening in 2000, %	Relative contribution associated with treatment in 2000, %	Relative contribution associated with screening in 2012, %	Relative contribution associated with treatment in 2012, %
	Screening alone	Treatment alone	Combined screening and treatment	Screening alone	Treatment alone	Combined screening and treatment	Screening alone	Screening alone	Treatment alone				
Column ID	A	B	C	D	E	F	G	H	I	J	K	L	M
Operation	A	B	C	D	E	F	D-A	E-B	F-C	A/(A+B)	B/(A+B)	D/(D+E)	E/(D+E)
Model	**ER+, ERBB2- Subtype**												
Dana-Farber	28	21	43	30	30	52	2	9	9	57	43	50	50
Erasmus	15	22	34	18	34	46	4	12	13	40	60	35	65
Georgetown-Einstein	21	25	41	26	39	54	5	13	13	45	55	40	60
MD Anderson	13	19	29	17	31	42	4	12	13	41	59	36	64
Stanford	17	34	45	19	41	53	1	7	8	34	66	31	69
Wisconsin-Harvard	15	35	45	16	51	59	1	16	14	30	70	24	76
Mean	**18**	**26**	**39**	**21**	**38**	**51**	**3**	**11**	**12**	**41**	**59**	**36**	**64**

eTable 4 (Continued). Comparison of breast cancer mortality reduction, overall and by ER/ERBB2-subtype, across models, in 2000 vs 2012

	Mortality reduction in 2000 relative to the estimated baseline mortality rate in 2000, %			Mortality reduction in 2012 relative to the estimated baseline mortality rate in 2012, %			Difference in the mortality reduction between 2012 and 2000, %			Relative contribution associated with screening in 2000, %	Relative contribution associated with treatment in 2000, %	Relative contribution associated with screening in 2012, %	Relative contribution associated with treatment in 2012, %
	Screening alone	Treatment alone	Combined screening and treatment	Screening alone	Treatment alone	Combined screening and treatment	Screening alone	Screening alone	Treatment alone				
Column ID	A	B	C	D	E	F	G	H	I	J	K	L	M
Operation	A	B	C	D	E	F	D-A	E-B	F-C	A/(A+B)	B/(A+B)	D/(D+E)	E/(D+E)
Model	**ER+, ERBB2+ Subtype**												
Dana-Farber	25	21	41	27	38	57	2	17	17	54	46	41	59
Erasmus	14	24	33	20	42	52	6	18	19	36	64	32	68
Georgetown-Einstein	22	27	41	24	43	58	2	16	17	45	55	36	64
MD Anderson	13	18	28	18	38	46	5	20	18	41	59	32	68
Stanford	16	37	47	17	58	66	1	21	19	31	69	23	77
Wisconsin-Harvard	18	31	46	19	62	71	0	31	25	37	63	23	77
Mean	**18**	**26**	**39**	**21**	**47**	**58**	**3**	**21**	**19**	**41**	**59**	**31**	**69**

eTable 4 (Continued)**.** Comparison of breast cancer mortality reduction, overall and by ER/ERBB2-subtype, across models, in 2000 versus 2012

	Mortality reduction in 2000 relative to the estimated baseline mortality rate in 2000, %			Mortality reduction in 2012 relative to the estimated baseline mortality rate in 2012, %			Difference in the mortality reduction between 2012 and 2000, %			Relative contribution associated with screening in 2000, %	Relative contribution associated with treatment in 2000, %	Relative contribution associated with screening in 2012, %	Relative contribution associated with treatment in 2012, %
	Screening alone	Treatment alone	Combined screening and treatment	Screening alone	Treatment alone	Combined screening and treatment	Screening alone	Screening alone	Treatment alone				
Column ID	A	B	C	D	E	F	G	H	I	J	K	L	M
Operation	A	B	C	D	E	F	D-A	E-B	F-C	A/(A+B)	B/(A+B)	D/(D+E)	E/(D+E)
Model	**ER-, ERBB2+ Subtype**												
Dana-Farber	24	14	33	25	28	49	1	15	16	64	36	47	53
Erasmus	14	14	26	17	28	41	3	15	15	51	49	37	63
Georgetown-Einstein	21	16	33	25	32	52	3	17	19	58	42	43	57
MD Anderson	13	11	20	15	23	33	2	12	13	53	47	39	61
Stanford	17	10	26	17	25	40	0	15	14	63	37	40	60
Wisconsin-Harvard	22	16	33	23	43	55	1	27	22	58	42	34	66
Mean	**19**	**13**	**29**	**20**	**30**	**45**	**2**	**17**	**16**	**58**	**42**	**40**	**60**

eTable 4 (Continued)**.** Comparison of breast cancer mortality reduction, overall and by ER/ERBB2-subtype, across models, in 2000 versus 2012

	Mortality reduction in 2000 relative to the estimated baseline mortality rate in 2000, %			Mortality reduction in 2012 relative to the estimated baseline mortality rate in 2012, %			Difference in the mortality reduction between 2012 and 2000, %			Relative contribution associated with screening in 2000, %	Relative contribution associated with treatment in 2000, %	Relative contribution associated with screening in 2012, %	Relative contribution associated with treatment in 2012, %
	Screening alone	Treatment alone	Combined screening and treatment	Screening alone	Treatment alone	Combined screening and treatment	Screening alone	Screening alone	Treatment alone				
Column ID	A	B	C	D	E	F	G	H	I	J	K	L	M
Operation	A	B	C	D	E	F	D-A	E-B	F-C	A/(A+B)	B/(A+B)	D/(D+E)	E/(D+E)
Model	**ER-, ERBB2- Subtype**												
Dana-Farber	25	13	34	26	20	40	1	6	6	65	35	57	43
Erasmus	13	15	26	17	22	35	4	7	10	46	54	43	57
Georgetown-Einstein	22	17	35	24	29	46	2	12	11	56	44	45	55
MD Anderson	14	11	22	18	14	27	4	3	5	57	43	56	44
Stanford	17	12	27	18	17	33	1	5	7	59	41	52	48
Wisconsin-Harvard	16	18	32	18	30	42	2	12	8	48	52	38	62
Mean	**18**	**14**	**29**	**20**	**22**	**37**	**2**	**8**	**8**	**55**	**45**	**48**	**52**

eTable 5. Relative contributions associated with screening and treatment advances on the difference in the breast cancer mortality reduction between 2000 and 2012*

				Model**						
Year	Metric	Row ID	Operation	D	E	G-E	M	S	W-H	Mean
2000	Mortality Reduction in 2000 Relative to Baseline in 2000, Screening Alone, %	A	A	27	14	21	13	17	17	**17**
	Mortality Reduction in 2000 Relative to Baseline in 2000, Treatment Alone, %	B	B	18	22	23	17	28	30	**23**
	Mortality Reduction in 2000 Relative to Baseline in 2000, Combined Screening and Treatment, %	C	C	39	32	39	27	40	42	**37**
	Relative Contribution Associated with Screening, %	D	A/(A+B)	60	39	48	44	38	35	**44**
	Relative Contribution Associated with Treatment, %	E	B/(A+B)	40	61	52	56	62	65	**56**
	Mortality Reduction Associated with Screening given Combination, %	F	D*C	24	13	19	12	15	15	**16**
	Mortality Reduction Associated with treatment given combination, %	G	E*C	16	20	21	15	25	27	**21**
2012	Mortality Reduction Relative to Baseline, Screening Alone, %	H	H	29	18	25	17	18	17	**21**
	Mortality Reduction Relative to Baseline, Treatment Alone, %	I	I	28	30	37	29	37	49	**35**
	Mortality Reduction Baseline, Combined Screening and Treatment, %	J	J	49	43	53	39	50	58	**49**
	Relative Contribution Associated with Screening, %	K	H/(H+I)	51	37	40	38	33	26	**37**
	Relative Contribution Associated Treatment, %	L	I/(H+I)	49	63	60	62	67	74	**63**
	Mortality Reduction Associated with Screening given Combination, %	M	K*J	25	16	22	15	16	15	**18**
	Mortality Reduction associated with Treatment given Combination, %	N	L*J	24	27	32	24	34	43	**31**
2000 vs 2012	Difference in Mortality Reduction Between 2000 and 2012, %	Q	J-C	10	10	14	13	10	16	**12**
	Difference in the Mortality Reduction Associated with Screening Advances Between 2000 and 2012, %	O	M-F	1	3	3	3	1	0	**2**
	Difference in the Mortality Reduction Associated with Treatment Advances Between 2000 and 2012, %	P	N-G	9	7	11	10	9	15	**10**
	Relative Contribution Associated with Screening Advances Between 2000 and 2012, %	R	O/Q	13	31	21	24	14	2	**17**
	Relative Contribution Associated with Treatment Advances Between 2000 and 2012, %	S	P/Q	87	69	79	76	86	98	**83**

* See Supplemental Methods subsection "*Computing the Relative Contributions of Screening and Treatment to the Difference in the Reduction Between Two Calendar Years*" for description of these calculations.

** *Abbreviations*: Model D is Dana Farber; Model E is Erasmus; Model G-E is Georgetown-Einstein; Model M is MD Anderson; Model S is Stanford; Model W-H is Wisconsin-Harvard.

eTable 6. Relative contributions associated with screening, chemotherapy, hormone therapy and trastuzumab to breast cancer mortality reduction in 2012, broken down by advances before and after 2000*

Relative Contributions Associated with Mortality Reduction in 2012, Percent							
	Screening Advances before 2000	Screening Advances after 2000	Chemo-therapy Advances before 2000	Chemo-therapy Advances after 2000	Hormone Therapy Advances before 2000	Hormone Therapy Advances after 2000	Trast-uzumab
Model	**Overall**						
Dana-Farber	48	3	16	7	15	9	2
Erasmus	29	8	30	8	17	8	1
Georgetown-Einstein	35	5	23	14	16	2	4
MD Anderson	30	8	15	7	22	12	6
Stanford	30	3	26	8	24	4	5
Wisconsin-Harvard	26	1	20	13	27	9	5
Mean	33	4	22	9	20	7	4
	ER+, ERBB2- Subtype						
Dana-Farber	48	2	19	6	17	8	0
Erasmus	29	6	26	4	17	18	0
Georgetown-Einstein	35	6	21	13	22	2	0
MD Anderson	29	8	13	8	27	16	0
Stanford	28	3	26	7	30	5	0
Wisconsin-Harvard	23	1	17	12	35	11	0
Mean	32	4	20	8	25	10	0
	ER+, ERBB2+ Subtype						
Dana-Farber	38	3	18	6	15	8	12
Erasmus	23	9	25	3	15	15	10
Georgetown-Einstein	32	4	19	14	19	5	7
MD Anderson	25	7	12	2	23	13	18
Stanford	22	1	24	7	26	4	17
Wisconsin-Harvard	23	0	14	10	27	7	18
Mean	27	4	19	7	21	9	14
	ER-, ERBB2+ Subtype						
Dana-Farber	44	4	25	12	0	0	16
Erasmus	32	5	31	13	0	0	18
Georgetown-Einstein	38	7	30	14	0	0	11
MD Anderson	34	8	25	1	0	0	32
Stanford	40	0	24	11	0	0	25
Wisconsin-Harvard	34	0	25	17	0	0	24
Mean	37	4	27	11	0	0	21

eTable 6 (Continued)**.** Relative contributions associated with screening, chemotherapy, hormone therapy and trastuzumab to breast cancer mortality reduction in 2012, broken down by advances before and after 2000*

Relative Contributions Associated with Mortality Reduction in 2012, Percent							
	Screening Advances before 2000	Screening Advances after 2000	Chemo-therapy Advances before 2000	Chemo-therapy Advances after 2000	Hormone Therapy Advances before 2000	Hormone Therapy Advances after 2000	Trast-uzumab
	ER-, ERBB2- Subtype						
Dana-Farber	55	2	30	13	0	0	0
Erasmus	34	9	40	18	0	0	0
Georgetown-Einstein	43	3	35	19	0	0	0
MD Anderson	46	10	28	15	0	0	0
Stanford	47	5	33	15	0	0	0
Wisconsin-Harvard	36	2	39	23	0	0	0
Mean	44	5	34	17	0	0	0

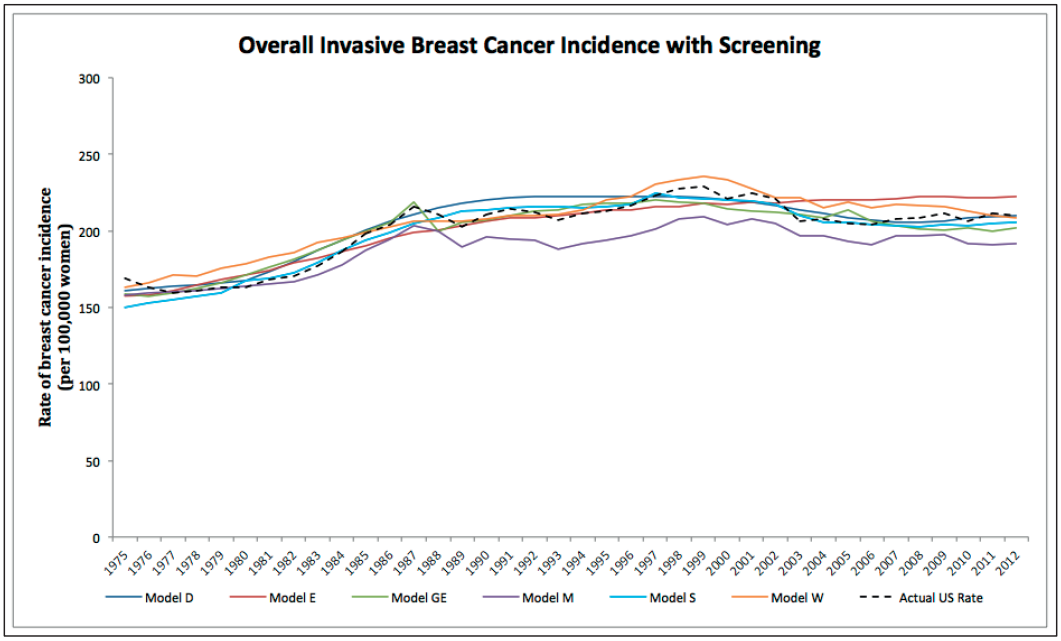
*Row sum is 100%, within rounding error.

eTable 7. Breakdown of overall breast cancer mortality reduction in 2012 by molecular subtype*

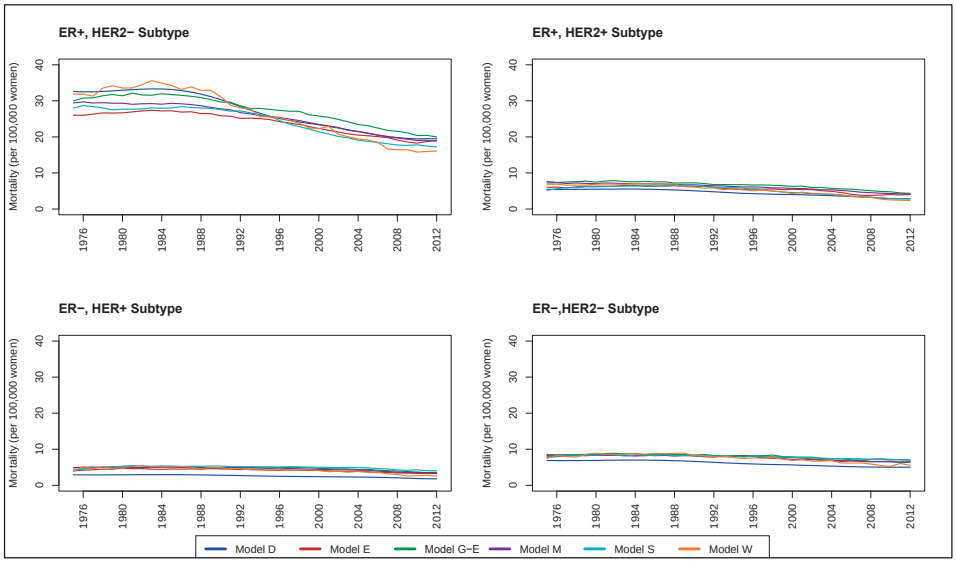
Model	ER+/ERBB2- Subtype	ER+/ERBB2+ Subtype	ER-/ERBB2+ Subtype	ER-/ERBB2- Subtype
Dana-Farber	70	13	6	11
Erasmus	62	17	10	12
Georgetown-Einstein	62	15	9	14
MD Anderson	61	17	9	13
Stanford	65	16	8	11
Wisconsin-Harvard	66	15	8	11
Mean	**64**	**16**	**8**	**12**

*Row sum is 100%.

eFigure 1. Comparison of model projections to actual US breast cancer incidence, for women ages 30-79, invasive cancer only

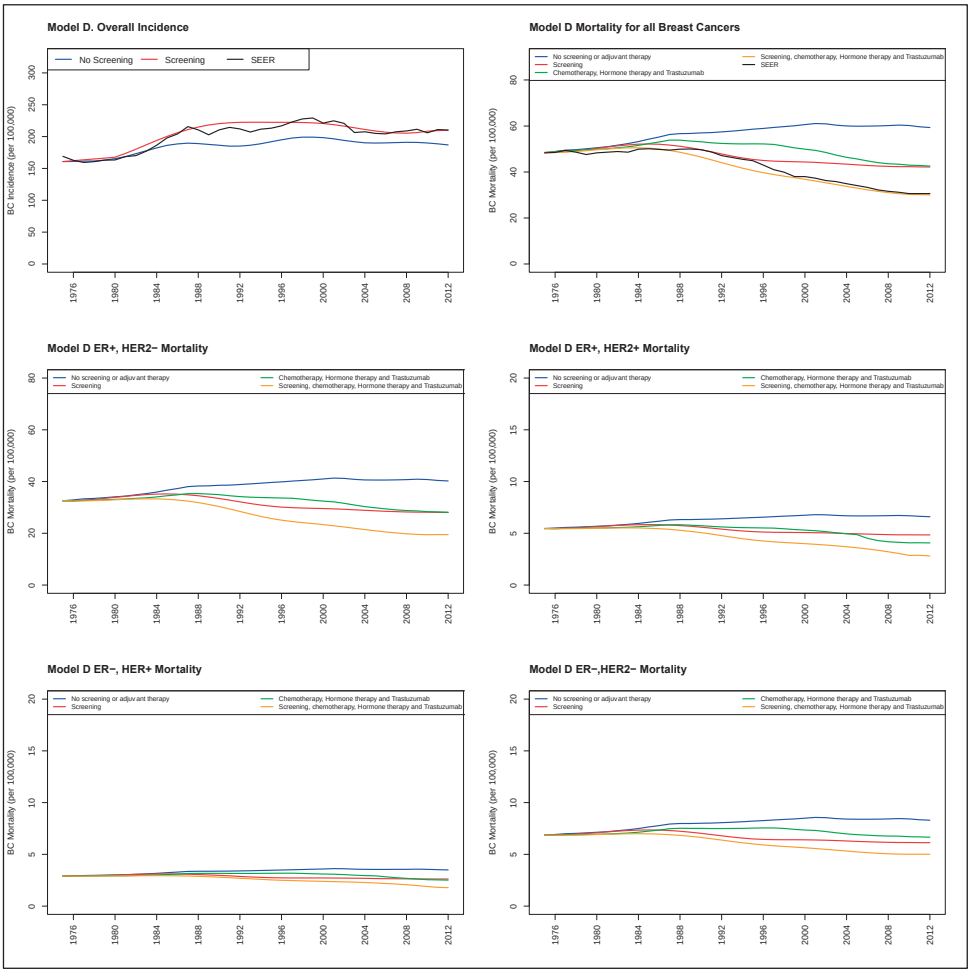


eFigure 2. Comparison of model projections for ER-/ ERBB2-specific breast cancer mortality trends between 1975-2012, for women ages 30-79, by molecular subtype. (Upper left) ER+/ERBB2-, (upper right) ER+/ERBB2+, (lower left) ER-/ERBB2+, (lower right) ER-/ERBB2- subtypes.

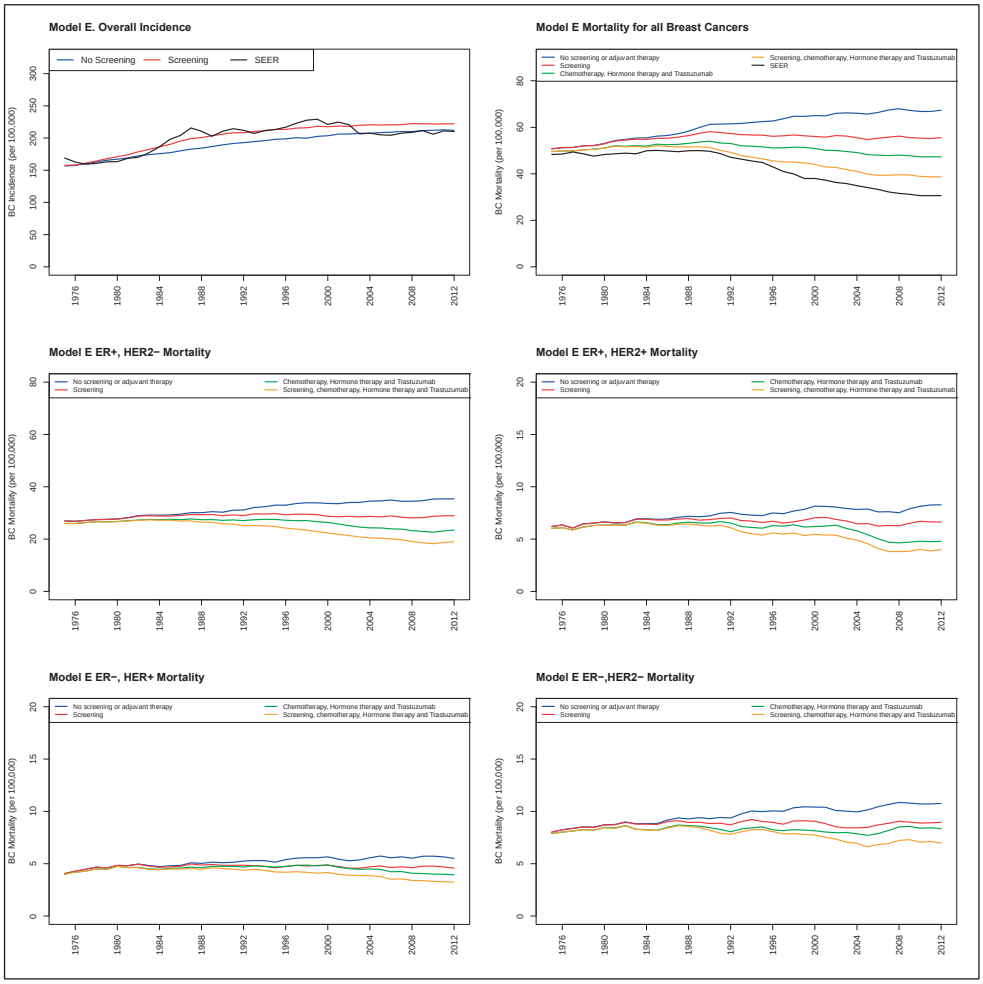


eFigure 3. Individual model projections for overall US breast cancer incidence and mortality (vs. SEER) and ER/ERBB2-subtype-specific mortality from 1975-2012, for women ages 30-79*

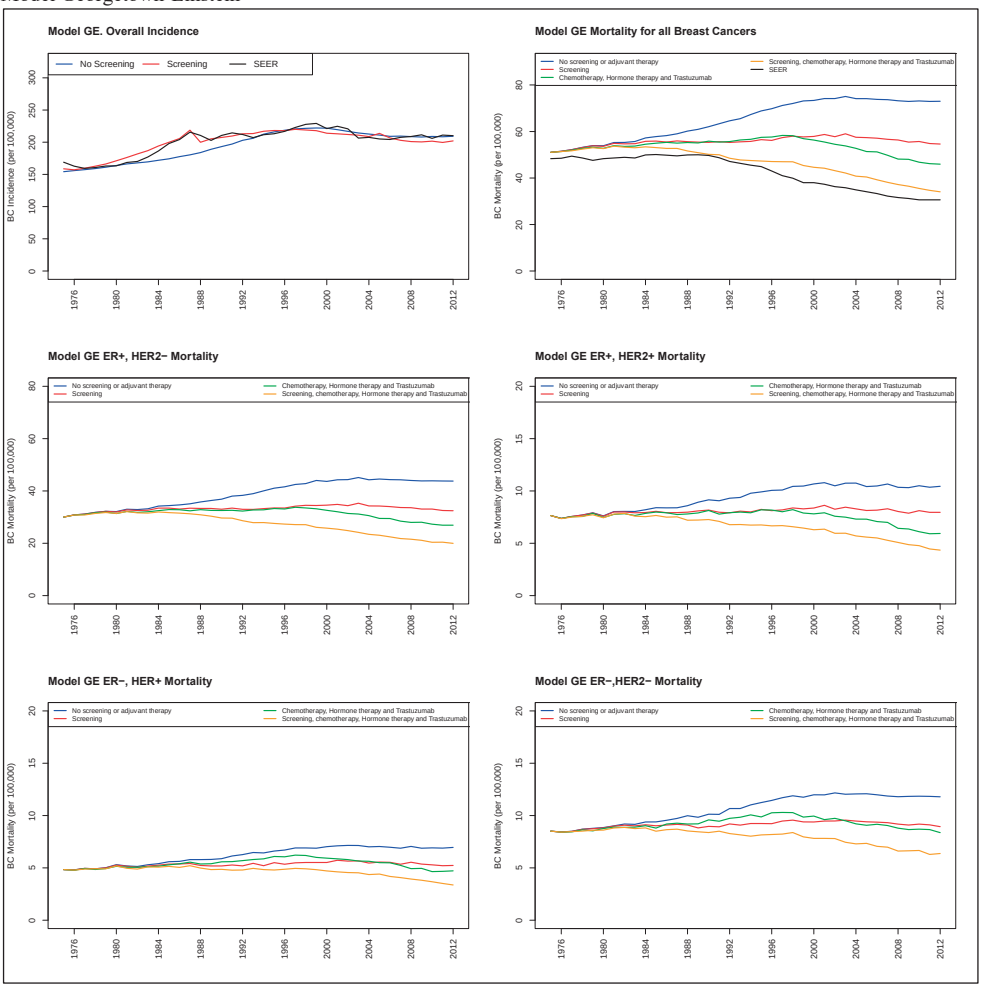
Model Dana-Farber



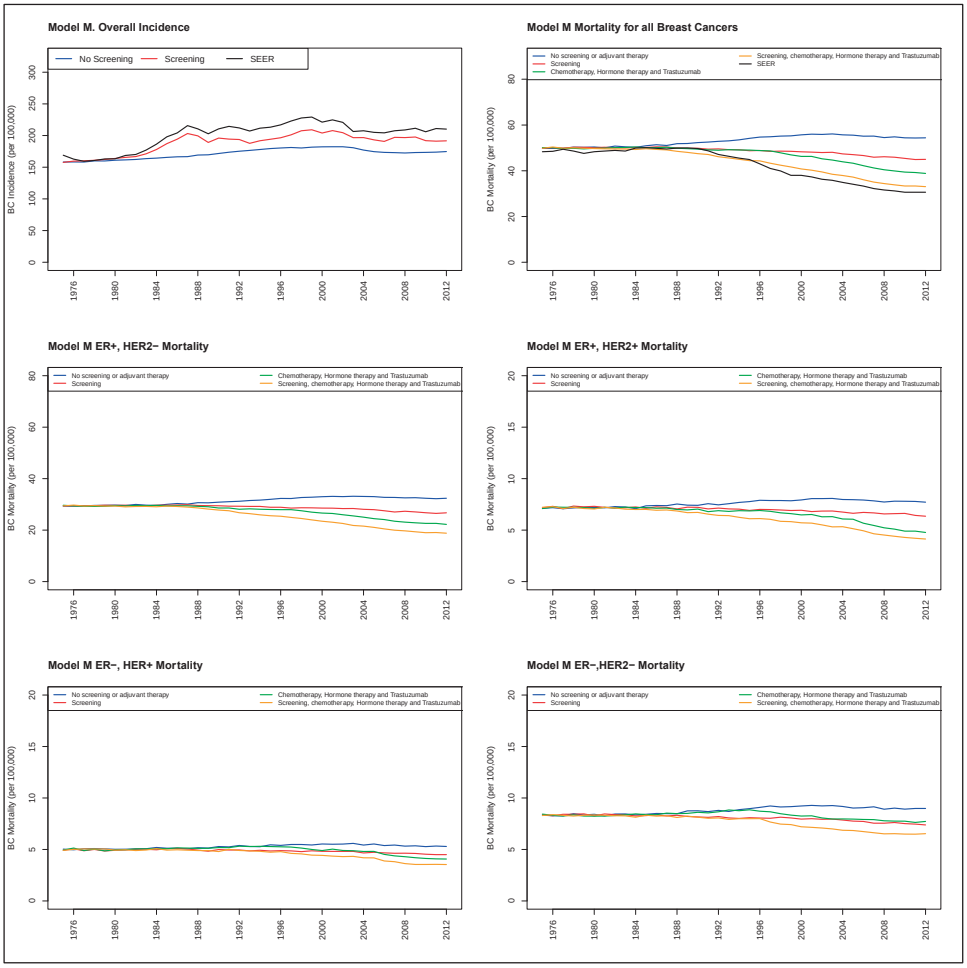
Model Erasmus



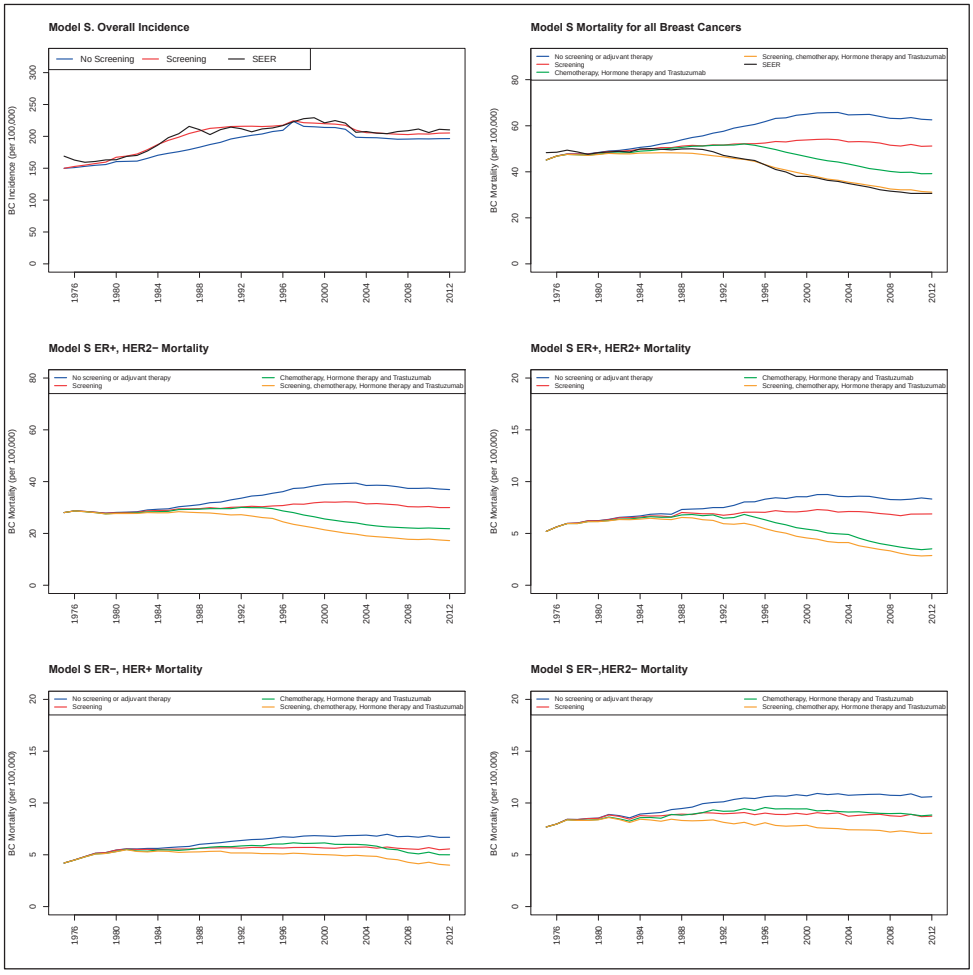
Model Georgetown-Einstein



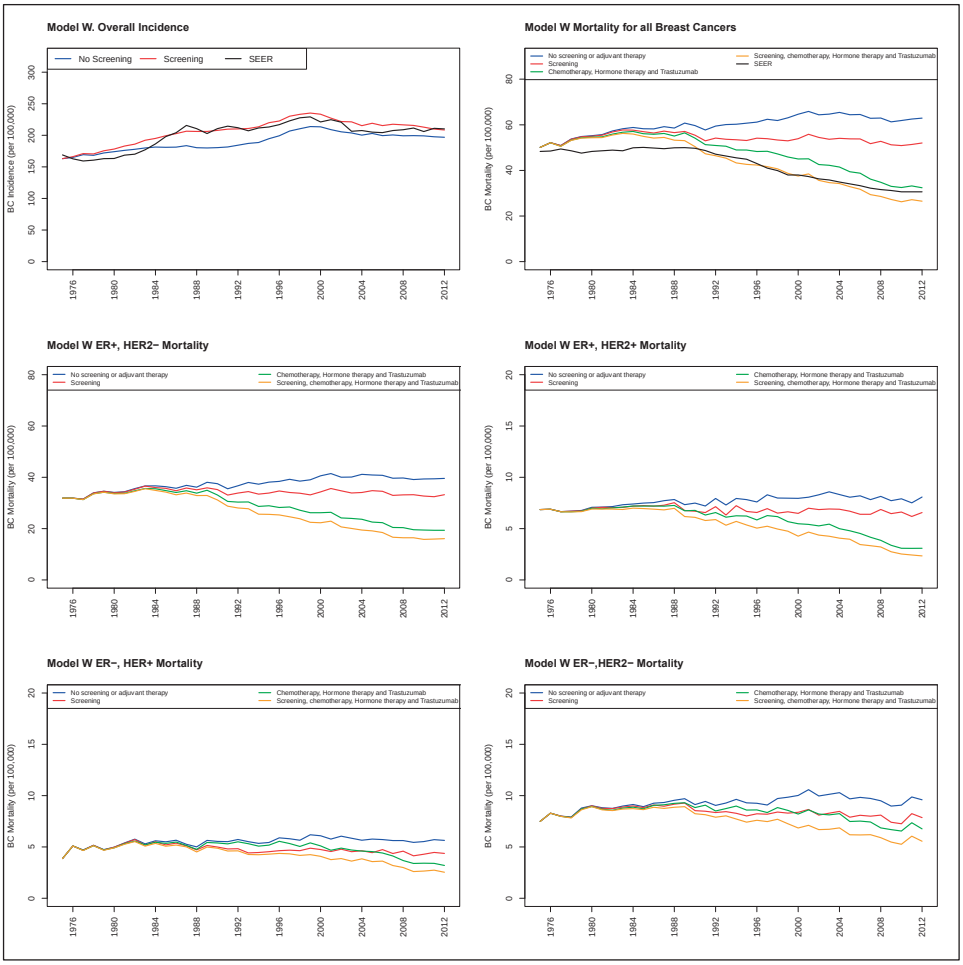
Model MD Anderson



Model Stanford



Model Wisconsin-Harvard



*** Legend for Supplemental Figure 3**: (upper two panels) Individual model projections of breast cancer incidence and mortality rates vs. SEER rates to 2012, with modeled incidence reported in the presence and absence of screening; (lower four panels) Individual model projections by ER/ERBB2 under 4 scenarios: (i) no screening and treatment, (ii) screening alone, (iii) treatment alone, (iv) screening and treatment combined. Subtype-specific comparison to SEER is not possible because ER and ERBB2 status were not jointly reported over this period.

eReferences.

1. Gangnon RE, Sprague BL, Stout NK, et al. The contribution of mammography screening to breast cancer incidence trends in the United States: an updated age-period-cohort model. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 2015;24(6):905-912.
2. Holford TR, Cronin KA, Mariotto AB, Feuer EJ. Changing patterns in breast cancer incidence trends. *Journal of the National Cancer Institute Monographs.* 2006(36):19-25.
3. Cronin KA, Mariotto AB, Clarke LD, Feuer EJ. Additional common inputs for analyzing impact of adjuvant therapy and mammography on U.S. mortality. *Journal of the National Cancer Institute Monographs.* 2006(36):26-29.
4. Cronin KA, Yu B, Krapcho M, et al. Modeling the dissemination of mammography in the United States. *Cancer causes & control : CCC.* 2005;16(6):701-712.
5. Mandelblatt JS, Near AM, Miglioretti DL, Munoz D, Sprague BK, Trentham-Dietz A, Gangnon R, Kurian AW, Weedon-Fekjaer H, Cronin KA, Plevritis SK. Common model inputs used in CISNET collaborative breast cancer modeling. Medical Decision Making. In Press 2017.
6. US Food and Drug Administration. http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/default.htm. June 2013. Silver Spring, Maryland US Dep Health and Human Services. Accessed January 2015.
7. Mariotto A, Feuer EJ, Harlan LC, Wun LM, Johnson KA, Abrams J. Trends in use of adjuvant multi-agent chemotherapy and tamoxifen for breast cancer in the United States: 1975-1999. *Journal of the National Cancer Institute.* 2002;94(21):1626-1634.
8. Mariotto AB, Feuer EJ, Harlan LC, Abrams J. Dissemination of adjuvant multiagent chemotherapy and tamoxifen for breast cancer in the United States using estrogen receptor information: 1975-1999. *Journal of the National Cancer Institute Monographs.* 2006(36):7-15.
9. 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 (London, England).* 2012;379(9814):432-444.
10. Gangnon RE, Stout NK, Alagoz O, Hampton JM, Sprague BL, Trentham-Dietz A. Contribution of Breast Cancer to Overall Mortality for U.S. Women. Medical Decision Making. In Press. 2017.
11. Lee SJ, Li X, Huang H, Zelen M. Models for breast cancer screening strategies updated for ductal carcinoma in situ and subgroups. Medical Decision Making. In Press. 2017.
12. van den Broek JJ, van Ravesteyn NT, Heijnsdijk EA, de Koning H. Estimating the effects of risk based screening and adjuvant treatment using the miscan-fadia continuous tumor growth model for breast cancer. Medical Decision Making. In Press. 2017.
13. Schechter CB, Near AM, Jayasekera J, Chang Y, Mandelblatt JS. structure, function, and applications of the Georgetown-Einstein (GE) breast cancer simulation model. Medical Decision Making. In Press. 2017.
14. Huang X, Li Y, Song J, Berry D. The MD Anderson CISNET model for estimating benefits of adjuvant therapy and screening mammography for breast cancer: an update. Medical Decision Making. In Press. 2017.
15. Munoz D, Xu C, Plevritis SK. A molecular subtype-specific stochastic simulation model of US breast cancer incidence and mortality trends from 1975 to 2000. Medical Decision Making. In Press. 2017.
16. Alagoz O, Ergun MA, Cevik M, et al. The University Of Wisconsin breast cancer epidemiology simulation model: An update. Medical Decision Making, In Press. 2017.
17. Munoz D, Near AM, van Ravesteyn NT, et al. Effects of screening and systemic adjuvant therapy on ER-specific US breast cancer mortality. *Journal of the National Cancer Institute.* 2014;106(11).
18. Munoz D, Plevritis SK. Estimating breast cancer progression features and survival by molecular subtype in the absence of screening and treatment. Medical Decision Making. In Press. 2017.

Chapter 7

Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies.

Mandelblatt JS, Stout NK, Schechter CB, van den Broek JJ, Miglioretti DL, Krapcho M, Trentham-Dietz A, Munoz D, Lee SJ, Berry DA, van Ravesteyn NT, Alagoz O, Kerlikowske K, Tosteson AN, Near AM, Hoeffken A, Chang Y, Heijnsdijk EA, Chisholm G, Huang X, Huang H, Ergun MA, Gangnon R, Sprague BL, Plevritis S, Feuer E, de Koning HJ, Cronin KA.

Annals of Internal Medicine. 2016 Feb 16;164(4):215-25.

ABSTRACT

Background

Controversy persists about optimal mammography screening strategies.

Objective

To evaluate screening outcomes, taking into account advances in mammography and treatment of breast cancer.

Design

Collaboration of six simulation models using national data on incidence, digital mammography performance, treatment effects, and other-cause mortality.

Setting and Patients

The average-risk US female population and sub-groups with varying risk, breast density, or comorbidity.

Setting

Unites States

Patients

Average-risk U.S. female population and subgroups with varying risk, breast density, or comorbidity

Interventions

Eight strategies differing by age at which screening starts (40, 45, 50 years) and screening interval (annual, biennial, and hybrid [annual in the 40s and biennial thereafter]); all strategies assumed 100% adherence and stopped at age 74.

Measurements

Benefits (breast cancer-specific mortality reduction, breast cancer deaths averted, life-years and quality-adjusted life years); number of mammograms used; harms (false-positive results, benign biopsies, and overdiagnosis); and ratios of harms (or use) and benefits (efficiency) per 1000 screens.

Results

Biennial strategies were consistently the most efficient for average-risk women. Biennial screening from ages 50-74 avoided a median of 7 breast cancer deaths vs. no screening; annual screening from ages 40-74 years avoided an additional 3 deaths, but yielded

1988 more false-positives and 7 more overdiagnoses per 1,000 women screened. Annual screening from ages 50-74 was inefficient (similar benefits but more harms than other strategies). For groups with a 2- to 4-fold increased risk, annual screening from age 40 had similar harms and benefits as screening average-risk women biennially from 50-74. For groups with moderate or severe comorbidity, screening could stop at age 66 to 68 years.

Limitations

Other imaging technologies, polygenic risk, and nonadherence were not considered.

Conclusion

Biennial screening for breast cancer is efficient for average-risk populations. Decisions regarding starting ages and intervals will ultimately depend on population characteristics and the decision-makers' weight given to the harms and benefits of screening.

Primary Funding Source

National Institutes of Health

INTRODUCTION

Despite decades of mammography screening for early breast cancer detection, there is no consensus on optimal strategies, target populations, or the magnitude of harms and benefits.(1-11) The 2009 US Preventive Services Task Force recommended biennial film mammography from ages 50-74, and suggested shared decision-making about screening in the 40's.(12) Since that recommendation was formulated, there have been some new data regarding screening benefits,(2,6,8,9,11,13,14) digital mammography has essentially replaced plain film,(15) and increasingly effective breast cancer systemic treatment regimens have become standard.(16) There has also been growing interest in consumer preferences and personalized screening approaches.(17-20). These factors could each affect the outcomes of breast cancer screening programs and/or alter policy decisions about population screening strategies.(17)

Modeling can inform screening policy decisions since it uses the best available evidence to evaluate a wide range of strategies, while holding selected conditions (e.g., treatment effects) constant, facilitating strategy comparisons.(21,22) Modeling also provides a quantitative summary of outcomes in different groups and assesses how preferences affect results. Collaboration of several models provides a range of plausible effects and illustrates the impact of differences in model assumptions on results.(1,7,23)

We used six well-established simulation models to synthesize current data to examine the outcomes of digital mammography screening at various starting ages and intervals among average-risk women. We also examined how breast density, risk, or comorbidity levels affect results, and whether preferences for health states related to screening and its downstream consequences affected conclusions.

METHODS

Strategies

We evaluated eight strategies that varied by starting age (40, 45, 50) and interval (annual, biennial, and hybrid [annual in the 40's and biennial thereafter]); all strategies stop screening at age 74. We included "no screening" as a baseline.

Model Descriptions

The models used to evaluate the screening strategies were developed within the Cancer Intervention and Surveillance Modeling Network (CISNET) (24-30) and the research was institutional review board approved. The models included model D (Dana-Farber Cancer Institute, Boston, Massachusetts), model E (Erasmus Medical Center, Rotterdam, the Netherlands), model GE (Georgetown University Medical Center, Washington, DC and

Albert Einstein College of Medicine, Bronx, New York), model M (MD Anderson Cancer Center, Houston, Texas), model S (Stanford University, Stanford, California), and model W (University of Wisconsin, Madison, Wisconsin and Harvard Medical School, Boston, Massachusetts).

Since earlier analyses,(1) the models have undergone substantial revision to reflect advances in breast cancer control, including portrayal of molecular subtypes based on estrogen receptor (ER) and human epidermal growth factor-2 receptor (HER2) status;(23) current population incidence (31) and competing non-breast cancer mortality; digital screening; and the most current therapies.(32) All models except model S include ductal carcinoma in-situ (DCIS).

The general modeling approach is summarized below; full details including approach, construction, data sources, assumptions, and implementation are available at: https://resources.cisnet.cancer.gov/registry and at (33). Additional information is available on request and the models are available for use via collaboration.

The models begin with estimates of breast cancer incidence (31) and ER/HER2-specific survival trends *without* screening or adjuvant treatment and then overlay data on screening and molecular subtype-specific adjuvant treatment to generate observed US population incidence and breast cancer-specific mortality trends.(1,7,17,23,34) Breast cancers have a distribution of preclinical screen-detectable periods (sojourn time) and clinical detection points. Digital mammography performance characteristics depend on age, first vs. subsequent screen, time since last mammogram, and breast density. ER/HER2 status is assigned at diagnosis based on stage and age. Molecular subtype- and stage-specific treatment reduces the hazard of breast cancer death (models D, GE, M, and S) or results in a cure for some cases (models E and W). Women can die of breast cancer or other causes. Screen detection of cancer during the preclinical screen-detectable period can result in the identification (and treatment) of earlier-stage or smaller tumors than might occur via clinical detection, with a corresponding reduction in breast cancer mortality.

We used a cohort of women born in 1970 with average-risk and average breast density and follow them from age 25 (since breast cancer is rare before this age [0.08% of cases]) until death or age 100.

Model Input Parameters

The models used a common set of age-specific variables for breast cancer incidence, digital mammography performance, treatment effects, and average and comorbidity-level specific-non-breast cancer causes of death.(20,33,35) The parameter values are available at: http://www.uspreventiveservicestaskforce.org/Page/Document/modeling-report-collaborative-modeling-of-us-breast-cancer-1/breast-cancer-screening1.(33) In addition, each group included model-specific inputs (or intermediate outputs) to represent preclinical detectable times, lead-time, and age- and ER/HER2-specific stage

distribution in screen- vs. non-screen-detected women on the basis of their specific model structure.(1,7,23-30) These model-specific parameters were based on assumptions about combinations of values that reproduced US trends in incidence and breast cancer-specific mortality, including proportions of DCIS that were nonprogressive and would not be detected without screening. Models M and W also assumed some small nonprogressive invasive cancers. The models adopted an age-period-cohort modeling approach to project breast cancer incidence rates in the absence of screening;(31,36) Model M used 1975-79 SEER rates. The models assumed 100% adherence to screening and receipt of the most effective treatment to isolate the effect of varying screening strategies.

Four models used age-specific digital mammography sensitivity values observed in the Breast Cancer Surveillance Consortium (BCSC) for detection of invasive and DCIS cancers combined (model S only uses data for invasive cancers). Separate values were used for initial and subsequent mammography by screening interval using standard BCSC definitions: annual includes data from screens occurring within 9-18 months of the prior screen and biennial includes data on screens within 19-30 months.(37,38) Model D used these data as input variables (28) and models GE, S, and W used the data for calibration.(24,25,27) Models E and M fit estimates from the BCSC and other data.(26,29)

Women with ER-positive tumors received five years of hormonal therapy and an anthracycline-based regimen accompanied by a taxane. Women with ER-negative invasive tumors received anthracycline-based regimens with a taxane. Those with HER2-positive tumors also received trastuzumab. Women with ER-positive DCIS received hormonal therapy.(16) Treatment effectiveness was based on clinical trials and was modeled as a reduction in breast cancer-specific mortality risk or increase in the proportion cured compared to ER/HER2-specific survival in the absence of adjuvant treatment.(32)

Benefits

Screening benefits (vs. no screening or incremental to other strategies) included percent breast cancer mortality reduction, breast cancer deaths averted, and life-years (LYs) and quality-adjusted life-years (QALYs) gained because of averted or delayed breast cancer death. Benefits (and harms) were accumulated from ages 40-100 years to capture the lifetime impact of screening.

We considered preferences, or utilities to account for morbidity from screening and treatment. A disutility for age- and gender-specific general population health was first applied to quality-adjust life years.(39) These were further adjusted to account for additional decrements in life years related to undergoing screening (-0.006 for one week), evaluation of a positive screen (-0.105 for five weeks), undergoing initial treatment by stage (for the first 2 years after diagnosis), and experiencing distant disease (for the last year of life for all women who die of breast cancer) (see Supplement Table 1).(33,40,41)

Use and Harms

Use of services focused on the number of mammograms required for the screening strategy. Harms included false-positive mammograms, benign biopsies, and overdiagnosis. False-positive mammogram rates were calculated as mammograms read as abnormal or needing further work-up in women without cancer divided by the total number of screening mammograms. Benign biopsies were defined as biopsies among women with false-positive screening results; we assume 100% compliance with biopsy recommendations.(42) Overdiagnosis was defined as all cases that would not have been clinically detected in the absence of screening because of lack of progressive potential or death from competing non-breast cancer mortality. The impact of overdiagnosis on QALYs was captured by the disutility of being treated for cancer but dying of other causes.

Statistical Analysis

For each model, strategies were ranked by the number of mammograms performed. We report the median use, benefits, and harms and range across models. We also obtained an efficiency frontier by plotting the sequence of points that represent the largest incremental percent breast cancer mortality reduction (or LYs or QALYs) per mammogram performed or harm entailed. Screening strategies that fell on this frontier were considered the most efficient (i.e., have the steepest slope such that no alternative exists that provides more benefit with less use/fewer harms).

Three models (E, GE, and W) also evaluated results based on combinations of breast cancer risk and density. Risk levels included: 1.3 (e.g., nulliparity or age at first live birth >30); (18,43) 2.0 (e.g., family history of one first degree relative); (18) or 4.0 times higher than average-risk (e.g., 2 or more first degree relatives).(18,44) Greater risk levels, such as seen with BRCA 1/2 mutations, were not considered since such groups have specific screening guidelines. We made the simplifying assumption that risk affected incidence, but not other aspects of disease.

Breast density was modeled as entirely fatty ("a"), scattered density ("b"), heterogeneously dense ("c") and extremely dense ("d"). Based on observed age-specific prevalence rates, density was assigned at age 40, and remained the same or decreased by one level at age 50 and again at age 65.(45) Density modified mammography sensitivity and specificity based on age, interval, and first vs. subsequent screening.(33) Density also modified the age-group specific (40-49, 50-64, and 65+) risk of developing breast cancer compared to average population density in the age-group (BCSC unpublished data).(44,46) Density was assumed to not affect molecular subtype or disease natural history. Density results were grouped into low ("a and b") and high density ("c and d") for presentation. The risk- and density-specific results were also compared to those for screening average-risk and density groups biennially from 50-74, since many guideline groups accept the latter.

In other analyses, two models (model E and GE) examined the impact of comorbidity on screening cessation using comorbidity-specific life expectancy. Examples of conditions that placed women in severe and moderate comorbidity groups included congestive heart failure and diabetes, respectively; the specific conditions and their associated life expectancies have been previously reported.(20,35,47) We compared results for continuing to screen biennially past age 74 among women with no or low comorbidity or stopping earlier than 74 for those with moderate or high comorbidity. These analyses included women who survived and were breast cancer-free up until the point where screening was to be extended or stopped.

Four models evaluated whether high disutility values would eliminate screening benefits. Finally, we evaluated the ability of the models to independently predict external trends and results (Supplement Figure 1 and Supplement Table 2).

Role of the Funding Source

We worked with US Preventive Services Task Force and Agency for Healthcare Research and Quality to develop the research questions. NCI investigators (KC, EF) collaborated in their role as scientific project officers. The agencies had no role in the study conduct or decision to submit the manuscript for publication.

RESULTS

Benefits in the Average-risk Population

The models produced consistent rankings of the screening strategies (Table 1). For instance, biennial screening from ages 50 to 74 yielded a median 25.8% reduction in breast cancer mortality compared to no screening (range: 24.1%-31.8). Annual screening led to slightly greater reductions in mortality than biennial strategies. However, biennial strategies maintained a median of 79.8%-81.3% of the breast cancer-specific mortality reduction of annual screening (range 68.3-98.9%) (Supplement Table 3).

Biennial screening also maintained the majority of annual benefits for LYS and QALYs and quality-adjustment did not change the ranking of strategies. Across all strategies, the largest decrement from quality-adjustment to life years was related to declines in general health as women aged; smaller decrements occurred due to the disutility of undergoing diagnostic evaluation of an abnormal screening exam and for having cancer. The disutility associated with screening itself had minimal impact on QALYs. (see 33)

The incremental benefits of initiating screening at age 40 were slightly greater than starting at age 50 in terms of breast cancer deaths averted with both annual and biennial screening (median 1.3 [range: 1.1-1.7] and 1.0 [0.8-1.7] per 1000 women screened, respectively) (Table 3). Initiating screening at age 45 yielded benefits intermediate be-

Table 1 Ranking of Benefits (Percent Breast Cancer Mortality Reduction, LYs, QALYs) by Model and Screening Strategy Per 1000 Women Screened

Strategies	Results per 1000 Women Screened							
	# of screens*	Percent breast cancer mortality reduction (vs. no screening) by model[1]						Median (range across models)
		D	E	G-E	M	S	W	
B 50-74	11,127	25.6%	26.0%	31.8%	26.8%	24.1%	25.4%	25.8% (24.1-31.8)
B 45-74	13,212	26.6%	27.6%	33.9%	28.4%	25.9%	26.7%	27.2% (25.9-33.9)
H 45-74	15,966	27.7%	29.7%	35.9%	29.2%	27.3%	30.1%	29.5% (27.3-35.9)
B 40-74	16,013	28.3%	30.3%	35.9%	31.9%	28.2%	30.5%	30.4% (28.2-35.9)
H 40-74	20,884	29.0%	32.3%	37.9%	31.7%	29.3%	32.8%	32.0% (29.0-37.9)
A 50-74	21,318	32.1%	33.9%	37.6%	27.1%	29.1%	35.3%	33.0% (27.1-37.6)
A 45-74	26,136	34.2%	37.6%	41.6%	29.4%	32.3%	39.1%	35.9% (29.4-41.6)
A 40-74	31,038	35.5%	40.1%	43.6%	32.5%	34.4%	42.6%	37.8% (32.5-43.6)

[1]Without screening, the median probability of dying of breast cancer is 2.50% (range 1.50-3.20%). Thus, if a particular screening strategy leads to a 30% reduction in breast cancer mortality, this means that the probability of breast cancer mortality was reduced from 2.50% to 1.75%. This translates into 7.5 deaths averted per 1000 women screened. The absolute reduction in breast cancer deaths (i.e., deaths averted) vs. no screening for each strategy is included in Table 2.

Strategies	Results per 1000 Women Screened							
	# of screens*	Years of Life Gained (vs. no screening) by model						Median (range across models)
		D	E	G-E	M	S	W	
B 50-74	11,127	153.8	94.0	140.5	146.5	104.2	74.6	122.4 (74.6-153.8)
B 45-74	13,212	168.4	107.7	161.2	171.3	115.2	84.0	138.2 (84.0-171.3)
H 45-74	15,966	175.3	117.9	170.2	171.4	125.1	95.7	147.7 (95.7-175.3)
B 40-74	16,013	183.7	123.7	172.4	194.8	131.6	98.8	152.0 (98.8-194.8)
H 40-74	20,884	191.1	137.6	187.2	211.5	141.0	110.9	164.1 (110.9-211.5)
A 50-74	21,318	180.0	125.9	167.3	156.3	133.3	104.3	144.8 (104.3-180.0)
A 45-74	26,136	201.3	149.3	196.7	177.8	154.2	123.0	166.0 (123.0-201.3)
A 40-74	31,038	217.1	168.8	213.5	218.1	170.1	140.5	191.8 (140.5-218.1)

Strategies	Results per 1000 Women Screened							
	# of screens*	QALYs Gained (vs. no screening) by model						Median (range across models)
		D	E	G-E	M	S	W	
B 50-74	11,127	114.5	67.3	100.1	109.6	71.9	47.1	86.0 (47.1-114.5)
B 45-74	13,212	123.8	75.6	114.4	129.4	78.8	51.9	96.6 (51.9-129.4)
H 45-74	15,966	126.6	80.9	118.3	128.5	84.5	58.3	101.4 (58.3-128.5)
B 40-74	16,013	133.7	85.4	120.1	148.1	89.1	60.4	104.6 (60.4-148.1)
H 40-74	20,884	134.2	91.0	126.1	159.4	92.5	64.8	109.3 (64.8-159.4)
A 50-74	21,318	127.0	84.1	111.4	113.2	87.5	62.4	99.5 (62.4-127.0)
A 45-74	26,136	138.9	97.3	129.5	129.4	99.5	71.7	114.5 (71.7-138.9)
A 40-74	31,038	146.6	107.3	137.2	160.6	107.6	80.0	122.4 (80.0-160.6)

A=Annual B=Biennial H=Hybrid
*Strategies are ranked from the least to the most mammograms, where the number of mammograms is the median across models. Not all possible mammograms in the age interval are obtained since some women die from other causes before screening would occur.
†Model Group Abbreviations: D (Dana Farber Cancer Center), E (Erasmus Medical Center), G-E (Georgetown U. –Einstein COM.), M (M.D. Anderson Cancer Center), S (Stanford U.), W (University of Wisconsin/Harvard)
‡Grey shaded areas in the table show strategies that are inferior or inefficient ("dominated") within a specific model; a strategy is classified as inferior or inefficient if there is another strategy that results in an equal or higher benefit (either percent mortality decline; LYG; or QALYs) with fewer harms (e.g., average screening exams).
§QALYs are adjusted for general health, diagnosis, screening and treatment.
||100% of women receive adjuvant systemic therapy based on recommended stage, ER/HER2-specific adjuvant therapy for pre- and post-menopausal women.

tween beginning at 40 and 50, although there were slightly greater incremental benefits when starting at age 45 (vs. 50) than starting at age 40 (vs. 45) (e.g., 10.6 vs. 8.0 and 15.4 vs. 7.9 QALYs for biennial and annual strategies, respectively) (Table 1).

Harms in the Average-risk Population

All models projected more false-positive results, benign biopsies, and overdiagnosed cases under annual vs. biennial schedules and starting earlier than age 50 (Table 2). For instance, if biennial screening began at age 40 instead of age 50, for every 1000 women screened there would be a median of 1 more death averted, but 576 more false-positive results, 58 benign biopsies, and 2 additional overdiagnosed cases. Compared to screening initiation at age 45, starting screening at age 40 had 1 or fewer added deaths averted depending on interval, but more incremental harms.

Efficiency Frontiers for Average-risk Populations

Efficiency frontier plots were used to graphically depict the balance between the number of mammograms and benefits (life years gained) of screening strategies. Biennial strategies starting at either age 40, 45, and 50 were all efficient (Figure 1, Supplemental Figure 2). Points that were close to, but fell below the frontier were less efficient than those on the frontier line. For example, compared to the point on the efficient frontier for biennial screening at age 45, the hybrid strategy of annual screening at 45 was less efficient than biennial screening starting at 40. This is because the hybrid strategy at 45 would require 405.8 more mammograms to gain an additional life year for every 1000 women screened compared to biennial screening at 45, while biennial screening starting at 40 only requires 189.5 extra mammograms to gain an additional life year.

Finally, annual screening from ages 50 to 74 was consistently inferior to other strategies (i.e., was inefficient, or dominated) since it yielded the same or fewer benefits than the next least intensive strategy depending on the measure of benefits, but required

Table 2 Lifetime Benefits and Harms of Screening Strategies based on Starting Ages and Screening Intervals

Strategy	Median number (range across models) per 1000 women screened (vs. no screening)*					
	Screens	Breast cancer deaths averted	False-positive screens	Benign breast biopsies	Over-diagnosed cases (invasive and DCIS) † ‡	Percent of all cases over-diagnosed † ‡
Biennial						
50-74	11,127	7 (4-9)	953 (830-1325)	146 (120-205)	19 (11-34)	12% (8–22)
45-74	13,212	8 (4-9)	1220 (930-1599)	168 (120-221)	19 (11-34)	12% (8–22)
40-74	16,013	8 (5-10)	1529 (1100-1976)	204 (140-264)	21 (12-38)	13% (9–24)
Hybrid						
45-74	15,966	8 (5-9)	1520 (1160-1968)	190 (140-250)	21 (12-40)	13% (8–25)
40-74	20,884	9 (5-10)	2106 (1480-2623)	245 (170-309)	23 (12-44)	14% (9–27)
Annual						
50-74	21,318	9 (5-10)	1798 (1706-2445)	228 (219-317)	25 (12-68)	15% (8–36)
45-74	26,136	9 (6-11)	2355 (2185-3087)	247 (230-329)	28 (12-74)	17% (9–38)
40-74	31,038	10 (6-11)	2941 (2550-3742)	303 (260-388)	30 (13-77)	18% (9–39)

*In all scenarios, 100% of women receive adjuvant systemic therapy based on recommended stage, ER/HER2-specific adjuvant therapy for pre- and post-menopausal women.

†Over-diagnosed cases are defined as cases that would not have been clinically detected in the absence of screening

(i.e., cases that do not die from breast cancer because of lack of progressive potential or death from competing non-breast cancer mortality). The result includes DCIS and invasive overdiagnosis. Over-diagnosis is calculated by comparing cases detected in the screening scenario to those detected in the non-screened scenario. Model S is excluded since it does not include DCIS. The percent overdiagnosis is calculated as the percent of all cases detected in the screening strategy that are overdiagnosis.

‡The upper range for all over diagnosis estimates is based on model M results. Model M generates very high overdiagnosis based on the assumption that incidence in the absence of screening has essentially remained flat since 1975-79, with virtually all of the increases over time attributable to screening. The other models use some form of an age-period-cohort model for incidence in the absence of screening, where some of the increases in incidence are due to screening and some to changes in risk factors (e.g., use of hormone replacement therapy), generating lower rates of overdiagnosis. Other sources of variation across models are related to assumptions about the proportions of DCIS cases that never progress to invasive cancer or the number of early invasive cancers that might be nonprogressive. Generally, models that assume higher proportions of DCIS and/or invasive cancer to be nonprogressive generate higher estimates of overdiagnosis than models that assume less nonprogressive disease. Unfortunately, the underlying incidence in the absence of screening and the proportion and types of tumors that are nonprogressive are unknown and unobservable. Therefore, the different results across models based on their respective assumptions provide a range of possible overdiagnosis.

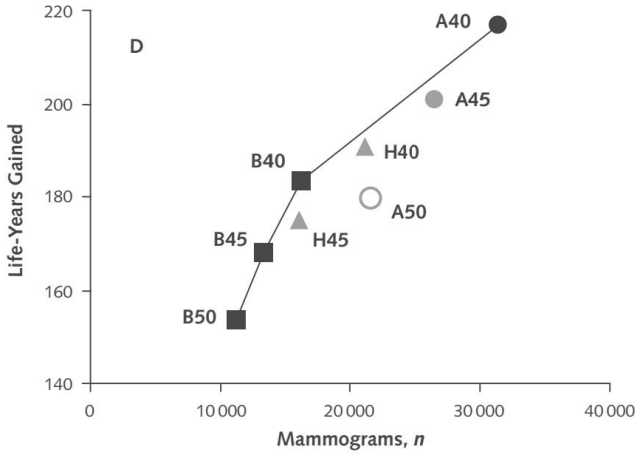


Figure 1 Efficiency frontier for life-years gained versus mammograms performed per 1000 women in model D (Dana-Farber Cancer Institute).
Legend for Figure 1. Efficiency Frontier
Efficiency frontier graphs for all models are shown in Appendix Figure 2 (available at ww.annals.org). This graph plots the average gain in life-years per additional mammogram performed per 1000 women for each screening strategy (vs. no screening) in model D. Biennial strategies are indicated with a square; hybrid strategies (annual in the 40s followed by biennial from 50 to 74 years of age) with a triangle; and annual strategies with a circle. Efficient strategies were plotted (i.e., those in which increases in mammography use resulted in greater life-years gained than the next less intensive strategy). The line represents the "efficiency frontier" by joining efficient strategies in which increases in mammography use resulted in greater life-years gained than the next less intensive efficient strategy. Strategies on this line would be considered efficient because they achieve the greatest gain in benefit (life-years gained) per harm or use of mammograms. Strategies that use more mammograms but still have small benefits (i.e., a shallower slope than the next best strategy) are considered to be less efficient (i.e., weakly dominated). When and if the slope in the efficiency frontier plot levels off, it means that the additional life-years gained per increase in mammography are small relative to the previous strategies and could indicate a point at which additional screening might be considered as having a low return (or additional benefit). There is no definitive inflection point across the models for the strategies or metrics evaluated. Black strategies are efficient; gray strategies close to the efficiency frontier are less efficient; and open gray strategies are inefficient (inferior, or dominated). Reference (33) provides efficiency frontiers for other harm and benefit metrics.

more mammograms or entailed more harms. These above patterns were generally seen with other harm and benefit metrics (see Supplement Figure 2).

Sensitivity Analyses for Average-Risk Populations

Varying the disutilities for usual health, screening, diagnosis, and treatment did not affect strategy rankings for average-risk populations and QALY gains persisted under all screening strategies, although their magnitude decreased.

Table 3 Incremental Changes in Breast Cancer Deaths Averted by Interval, Age of Screening Initiation, and Model

	Annual		Biennial	
	Number of breast cancer deaths averted/1000 women (% breast cancer mortality reduction)			
Model	Start at 40 vs. 50	Start at 45 vs. 50	Start at 40 vs. 50	Start at 45 vs. 50
D	1.1 (3.4%)	0.6 (2.1%)	0.9 (2.7%)	0.3 (1.0%)
E	1.5 (6.2%)	0.9 (3.6%)	1.0 (4.3%)	0.4 (1.6%)
G-E	1.5 (6.0%)	1.0 (4.0%)	1.0 (4.1%)	0.5 (2.2%)
M	1.7 (5.3%)	0.7 (2.3%)	1.7 (5.1%)	0.5 (1.6%)
S	1.1 (5.2%)	0.7 (3.1%)	0.9 (4.1%)	0.4 (1.7%)
W	1.1 (7.3%)	0.6 (3.8%)	0.8 (5.1%)	0.2 (1.3%)
Median	1.3 (5.7%)	0.7 (3.4%)	1.0 (4.2%)	0.4 (1.6%)

*Incremental difference between starting at age 40 or 45 vs. 50. Annual is comparing A40-74 (or 45-74) to A50-74; biennial is comparing B40-74 (or 45-74) to B50-74. Hybrid strategies are compared to B50-74, therefore for those incremental comparisons the hybrid results are the same as the annual results

Harms and Benefits by Risk Level

The balance of harms and benefits differed by risk group, with women who had higher-risk having lower rates of false-positives and higher gains from screening than lower-risk groups. Screening higher-risk women also yielded a lower proportion of overdiagnosed cases per breast cancer death averted than screening average-risk women. However, annual screening from ages 50 to 74 had the same or less benefit and more harms than other strategies at all risk levels.(33)

For women with a 2- to 4-fold increase in risk, annual screening starting at age 40 or 45 had similar or more favorable harm-to-benefit ratios (based on false-positives) as biennial screening of average-risk women from 50-74. For instance, for every 1000 average-risk women screened biennially from 50-74, there would be 226.5 (range: 169.9-267.0) false-positives per death averted. If women with a two-fold increase in risk began annual screening at age 40, their corresponding ratio would be slightly more favorable at 200.7 (range: 177.5-232.2). For women with a 1.3-fold increase in risk, biennial screening starting at age 40 had similar harm-to-benefit ratios as biennial screening of average-risk women from ages 50-74.

Benefits and Harms by Breast Density Group

Breast density (low vs. high) changed absolute benefits, but annual screening from 50-74 remained inefficient across breast density groups. Women in the low-density group had a greater proportion of their cancers detected due to greater digital mammography sensitivity, and therefore a greater breast cancer-specific mortality reduction than the high-density group. However, women in the high-density group had a greater absolute

number of cancers detected because their risk of cancer was higher, leading to more life years saved among women in the high-density than the low-density group (33)).

Benefits and Harms by Comorbidity

For women with no comorbidity, biennial screening could continue to age 78 or 80 and still have similar harm-to-benefit ratios as screening women with average non-breast cancer mortality biennially from 50-74. However, for women with moderate to severe comorbidity, the comparable ratios were equivalent at about age 68 (33).

DISCUSSION

This study used six established models to estimate the potential efficacy of different US breast cancer screening strategies. All six models demonstrated that screening initiation at age 40 has some benefits for average-risk populations, but also higher levels of harms than strategies starting at age 50. The findings also suggest that comorbidity levels could be used to tailor the age of screening cessation. Biennial screening strategies were the most efficient, but annual screening could be considered from ages 40-74 in groups with a two to four-fold higher than average-risk.

Results from all models indicated that digital mammography screening of average-risk women in their 40's modestly lowers breast cancer-specific mortality and extends the length and quality of life, even after considering disutilities related to the screening process. The absolute benefits of starting screening in the 40's varied somewhat based on model structure and assumptions, but were consistent with observations from randomized trials.(6) However, starting at age 40 vs. 45 was associated with increasing incremental harms relative to the increase in benefits. Thus, decisions about initiating screening before age 50 may depend on the weight attached to screening benefits and harms.

Consistent with other analyses of screening upper age limits,(20,48-50) and other recommendations,(12,51) our results suggested that the balance of harms and benefits of screening was affected by competing non-breast cancer mortality, so that age of screening cessation could be tailored by comorbidity levels.

Similar to our 2009 analysis,(1) biennial strategies are most consistently efficient. Screening annually from ages 50-74 had the same or fewer benefits for any given harm for all population groups in virtually all models, and would be considered inefficient. However, annual screening in the 40's followed by biennial screening at age 50, or the most intensive schedule evaluated (annual screening 40-74) were also efficient or close to being efficient. Additionally, annual screening of women with a two- to four-fold increased risk (e.g., due to non-BRCA related family history) from ages 40-74 had com-

parable harm-to-benefit ratios as did biennial screening from age 50 to 74 in average-risk populations.

The results also suggest that benefits of screening vary by breast density, at least when grouped into low/high categories. Women with dense breasts have a higher risk of developing cancer and absolute detection rate, but lower relative detection. (19,52) This is because digital sensitivity, while optimized for density, is still lower in women with dense than non-dense breasts.(53-56) Improving outcomes for women with dense breasts (55) may require new innovations in imaging (57-60) or identification of risk biomarkers.(61,62)

This analysis extends our prior work by explicitly considering overdiagnosis as a screening harm. Depending on screening strategy, the models estimated that 2-12% of invasive and 30% to 50% of DCIS cases might represent overdiagnosis. While the models differed in absolute estimates, they agreed on how overdiagnosis affected the ranking of strategies and the finding that the majority of overdiagnosed cases were DCIS. The model results for overdiagnosis are not directly comparable to other published estimates (8,63) since the models followed women for their entire lives. The models also made assumptions about unobservable input parameters related to natural history. While there is no agreement on methods to estimate overdiagnosis (64) or on its true rate,(65,66) there is agreement that it is an important harm. Active surveillance for DCIS with a low risk of progression is one potential future approach to reduce harms from DCIS overdiagnosis. More information is also needed on consumer knowledge of and willingness to risk overdiagnosis.(67)

Overall, this study has several important strengths including collaboration of six long-established, independent modeling groups, use of well-calibrated models that reproduce temporal epidemiological trends and a screening trial result, inclusion of digital technology, incorporation of increasingly effective treatments, and consideration of quality of life, risk factors, breast density, and comorbidity.(68) The conclusions about the ranking of screening strategies are robust and should provide greater credibility than inferences based on one model alone.

Our study also had limitations. First, to evaluate program efficacy we assumed 100% adherence to screening, prompt evaluation of abnormal results, and full use of optimal treatment. Actual benefits will fall short of our projected results since adherence is not perfect. Second, we only focused on hybrid strategies for women in their 40's. Alternative hybrid strategies may be important to examine in future research. Third, the analysis also did not consider other imaging technologies for average-risk populations or for groups with high breast density, such as ultrasound, (69) computer-aided detection,(70) tomosynthesis, or magnetic resonance imaging (MRI). Data on tomosynthesis performance and needs for radiologist re-training are still emerging.(58) Fourth, we did not model any radiation-induced breast cancers due to more intensive mammography

schedules.(71) Fifth, we assumed that risk factors influenced the incidence of disease, but not its natural history. Sixth, certain risk factors, such as family history, are age-dependent in their effects.(18,72) Since we held relative risk levels constant over age, our benefit estimates could be over- or under-estimated for specific risk factors.(17) Seventh, we did not consider polygenic risk,(73,74) or explicitly model menopausal status; we used age 50 as a proxy for the average age of menopause. Additionally, the analysis did not include screening program costs or utility estimates specific to some of the newest treatments. Finally, compared to our earlier research,(1) the models all estimated similar, but somewhat greater breast cancer-specific mortality reductions (for example, a median 22% vs. 25.8% reduction with biennial screening from 50-74 in 2009 vs. current models, respectively). The primary reasons for this modeled improvement relate to the increased sensitivity of digital vs. film mammography, advances in molecular-targeted therapies, and changes in underlying breast cancer trends.

Overall, the six models conclude that biennial screening strategies are the most efficient. Choices about optimal ages of initiation (and cessation) and screening intervals will ultimately depend on program goals, the weight attached by the decision-maker to screening harms and benefits,(75) and considerations of efficiency.

Acknowledgments

The authors thank CISNET consultants Elizabeth Burnside, Allison Kurian, Donald Weaver, and Diana Buist for review of earlier versions of the models for structure, parameters, and assumptions for clinical face validity; Jennifer Croswell from the Agency for Healthcare Research and Quality for assistance with analysis and review of results; members of the U.S. Preventive Services Task Force and the Oregon Evidence-based Practice Center for comments on earlier versions of this research; and Jessica Garshell for data processing. The authors thank Adrienne Ryans for manuscript preparation.

The authors thank John Wong, Stuart Baker, William Lawrence, and Tom Trikalinos for helpful suggestions on earlier versions of this paper.

POTENTIAL CONFLICTS OF INTEREST: NONE DISCLOSED

REFERENCES

(1) Mandelblatt J, Cronin K, Bailey S, Berry DA, de Koning H, Draisma G et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. Ann Inten Med 2009; 151(10):738-747.

(2) Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2013; 6:CD001877.

(3) Biller-Andorno N, Juni P. Abolishing mammography screening programs? A view from the Swiss Medical Board. N Engl J Med 2014; 370(21):1965-1967.

(4) Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet 2002; 359(9310):909-919.

(5) Tabar L, Vitak B, Chen HH, Duffy SW, Yen MF, Chiang CF et al. The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. Radiol Clin North Am 2000; 38(4):625-651.

(6) Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. Lancet 2006; 368(9552):2053-2060.

(7) Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M et al. Effect of screening and adjuvant therapy on mortality from breast cancer. New Engl J Med 2005; 353(17):1784-1792.

(8) Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. BMJ 2014; 348:g366.

(9) Paci E, Broeders M, Hofvind S, Puliti D, Duffy SW. European breast cancer service screening outcomes: a first balance sheet of the benefits and harms. Cancer Epidemiol Biomarkers Prev 2014; 23(7):1159-1163.

(10) Smith RA. The value of modern mammography screening in the control of breast cancer: understanding the underpinnings of the current debates. Cancer Epidemiol Biomarkers Prev 2014; 23(7):1139-1146.

(11) Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F et al. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med 2015; 372(24):2353-2358.

(12) Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2009; 151(10):716-236.

(13) Broeders M, Moss S, Nystrom L, Njor S, Jonsson H, Paap E et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. J Med Screen 2012; 19 Suppl 1:14-25.

(14) Moss SM, Wale C, Smith R, Evans A, Cuckle H, Duffy SW. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. Lancet Oncol 2015. Sep;16(9):1123-32.

(15) US Food and Drug Administration. Available from: http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/default.htm. June 2013. Silver Spring, Maryland. US Department of Health and Human Services. Accessed January 2015.

(16) National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice

Guidelines in Oncology - Breast Cancer. Available from: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp. 2014. Accessed January 2015.

(17) van Ravesteyn NT, Miglioretti DL, Stout NK, Lee SJ, Schechter CB, Buist DS et al. Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: a comparative modeling study of risk. Ann Intern Med 2012; 156(9):609-617.

(18) Nelson HD, Zakher B, Cantor A, Fu R, Griffin J, O'Meara ES et al. Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. Ann Intern Med 2012; 156(9):635-648.

(19) Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. Ann Intern Med 2011; 155(1):10-20.

(20) Lansdorp-Vogelaar I, Gulati R, Mariotto AB, Schechter CB, de Carvalho TM, Knudsen AB et al. Personalizing age of cancer screening cessation based on comorbid conditions: model estimates of harms and benefits. Ann Intern Med 2014; 161(2):104-112.

(21) de Gelder R, Heijnsdijk EA, Fracheboud J, Draisma G, de Koning HJ. The effects of population-based mammography screening starting between age 40 and 50 in the presence of adjuvant systemic therapy. Int J Cancer 2015; 137(1):165-172.

(22) Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR. Assessing the effectiveness of health interventions for cost-effectiveness analysis. Panel on Cost-Effectiveness in Health and Medicine. J Gen Intern Med 1997; 12(9):551-558.

(23) Munoz D, Near AM, van Ravesteyn NT, Lee SJ, Schechter CB, Alagoz O et al. Effects of screening and systemic adjuvant therapy on ER-specific US breast cancer mortality. J Natl Cancer Inst 2014 Sep 24;106(11).

(24) Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL. The Wisconsin Breast Cancer Epidemiology Simulation Model. J Natl Cancer Inst Monogr 2006;(36):37-47.

(25) Mandelblatt J, Schechter CB, Lawrence W, Yi B, Cullen J. The SPECTRUM population model of the impact of screening and treatment on U.S. breast cancer trends from 1975 to 2000: principles and practice of the model methods. J Natl Cancer Inst Monogr 2006; 36:47-55.

(26) Berry DA, Inoue L, Shen Y, Venier J, Cohen D, Bondy M et al. Modeling the impact of treatment and screening on U.S. breast cancer mortality: a Bayesian approach. J Natl Cancer Inst Monogr 2006;(36):30-36.

(27) Plevritis SK, Sigal BM, Salzman P, Rosenberg J, Glynn P. A stochastic simulation model of U.S. breast cancer mortality trends from 1975 to 2000. J Natl Cancer Inst Monogr 2006;(36):86-95.

(28) Lee S, Zelen M. A stochastic model for predicting the mortality of breast cancer. J Natl Cancer Inst Monogr 2006;(36):79-86.

(29) Tan SY, van Oortmarssen GJ, de Koning HJ, Boer R, Habbema JD. The MISCAN-Fadia continuous tumor growth model for breast cancer. J Natl Cancer Inst Monogr 2006;(36):56-65.

(30) Clarke LD, Plevritis SK, Boer R, Cronin KA, Feuer EJ. A comparative review of CISNET breast models used to analyze U.S. breast cancer incidence and mortality trends. J Natl Cancer Inst Monogr 2006;(36):96-105.

(31) Gangnon RE, Sprague BL, Stout NK, Alagoz O, Weedon-Fekjaer H, Holford TR et al. The Contribution of Mammography Screening to Breast

Cancer Incidence Trends in the United States: An Updated Age-period-cohort Model. Can Epi Biomarkers Prev 2015. Jun;24(6):905-12

(32) Peto R, Davies C, Godwin J, Gray R, Pan HC, Clarke M 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):432-444.

(33) Mandelblatt JS, Cronin K, de Koning H, Miglioretti DL, Schechter CS, Stout N. Modleing Technical Report. Collaborative modeling of U.S. breast cancer screening strategies. AHRQ Publication No. 14-05201-EF-4. April 2015 Available from: http://www.uspreventiveservicestaskforce.org/Page/Document/modeling-report-collaborative-modeling-of-us-breast-cancer-1/breast-cancer-screening1. Accessed May 2015.

(34) Chang Y, Schechter CB, van Ravesteyn NT, Near AM, Heijnsdijk EA, Adams-Campbell L et al. Collaborative modeling of the impact of obesity on race-specific breast cancer incidence and mortality. Breast Cancer Res Treat 2012; 136:823-835.

(35) Cho H, Klabunde CN, Yabroff KR, Wang Z, Meekins A, Lansdorp-Vogelaar I et al. Comorbidity-adjusted life expectancy: a new tool to inform recommendations for optimal screening strategies. Ann Intern Med 2013; 159(10):667-676.

(36) Holford TR, Cronin KA, Mariotto AB, Feuer EJ. Changing patterns in breast cancer incidence trends. J Natl Cancer Inst Monogr 2006;(36):19-25.

(37) Kerlikowske K, Zhu W, Hubbard RA, Geller B, Dittus K, Braithwaite D et al. Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy. JAMA Intern Med 2013; 173(9):807-816.

(38) Dittus K, Geller B, Weaver DL, Kerlikowske K, Zhu W, Hubbard R et al. Impact of mammography screening interval on breast cancer diagnosis by menopausal status and BMI. J Gen Intern Med 2013; 28(11):1454-1462.

(39) Hanmer J, Lawrence WF, Anderson JP, Kaplan RM, Fryback DG. Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores. Med Decis Making 2006; 26(4):391-400.

(40) Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG. Retrospective cost-effectiveness analysis of screening mammography. J Natl Cancer Inst 2006; 98(11):774-782.

(41) de Haes JC, de Koning HJ, Van Oortmarssen GJ, van Agt HM, De Bruyn AE, van der Maas PJ. The impact of a breast cancer screening programme on quality-adjusted life-years. Int J Cancer 1991; 49:538-544.

(42) Rosenberg RD, Yankaskas BC, Abraham LA, Sickles EA, Lehman CD, Geller BM et al. Performance benchmarks for screening mammography. Radiology 2006; 241(1):55-66.

(43) Kerlikowske K, Walker R, Miglioretti DL, Desai A, Ballard-Barbash R, Buist DS. Obesity, mammography use and accuracy, and advanced breast cancer risk. J Natl Cancer Inst 2008; 100(23):1724-1733.

(44) Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer Inst 2006; 98(17):1204-1214.

(45) Sprague BL, Gangnon RE, Burt V, Trentham-Dietz A, Hampton JM, Wellman RD et al. Prevalence of mammographically dense breasts in the United States. J Natl Cancer Inst 2014; 106(10).

(46) Tice JA, O'Meara ES, Weaver DL, Vachon C, Ballard-Barbash R, Kerlikowske K. Benign breast disease, mammographic breast density, and the risk of breast cancer. J Natl Cancer Inst 2013; 105(14):1043-1049.

(47) Mariotto AB, Wang Z, Klabunde CN, Cho H, Das B, Feuer EJ. Life tables adjusted for comorbidity more accurately estimate noncancer survival for recently diagnosed cancer patients. J Clin Epidemiol 2013; 66(12):1376-1385.

(48) Walter LC, Covinsky KE. Cancer screening in elderly patients: a framework for individualized decision making. JAMA 2001; 285(21):2750-2756.

(49) Schonberg MA, Davis RB, McCarthy EP, Marcantonio ER. External validation of an index to predict up to 9-year mortality of community-dwelling adults aged 65 and older. J Am Geriatr Soc 2011; 59(8):1444-1451.

(50) van Ravesteyn N, Stout NK, Schechter CB, Heijnsdijk EAM, Alagoz O, Trentham-Dietz A et al. Benefits and harms of mammography screening after age 74 years: model estimates of overdiagnosis. J Nat Cancer Inst 2015; 107(7).

(51) Smith RA, Cokkinides V, Brawley OW. Cancer screening in the United States, 2012: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin 2012; 62(2):129-142.

(52) Vilaprinyo E, Forne C, Carles M, Sala M, Pla R, Castells X et al. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. PLoS One 2014; 9(2):e86858.

(53) Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005; 353(17):1773-1783.

(54) Kerlikowske K, Hubbard RA, Miglioretti DL, Geller BM, Yankaskas BC, Lehman CD et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. Ann Intern Med 2011; 155(8):493-502.

(55) Kerlikowske K, Zhu W, Tosteson AN, Sprague BL, Tice JA, Lehman CD et al. Identifying women with dense breasts at high risk for interval cancer: a cohort study. Ann Intern Med 2015; 162(10):673-681.

(56) Braithwaite D, Zhu W, Hubbard RA, O'Meara ES, Miglioretti DL, Geller B et al. Screening outcomes in older US women undergoing multiple mammograms in community practice: does interval, age, or comorbidity score affect tumor characteristics or false positive rates? J Natl Cancer Inst 2013; 105(5):334-341.

(57) Lee CI, Lehman CD. Digital breast tomosynthesis and the challenges of implementing an emerging breast cancer screening technology into clinical practice. J Am Coll Radiol 2013; 10(12):913-917.

(58) Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA 2014; 311(24):2499-2507.

(59) Park JY, Yi SY, Park HJ, Kim MS, Kwon HJ, Park NH et al. Breast-specific gamma imaging: correlations with mammographic and clinicopathologic characteristics of breast cancer. AJR Am J Roentgenol 2014; 203(1):223-228.

(60) Rechtman LR, Lenihan MJ, Lieberman JH, Teal CB, Torrente J, Rapelyea JA et al. Breast-specific gamma imaging for the detection of breast cancer in dense versus nondense breasts. AJR Am J Roentgenol 2014; 202(2):293-298.

(61) Matamala N, Vargas MT, Gonzalez-Campora R, Minambres R, Arias JI, Menendez P et al. Tumor microRNA expression

profiling identifies circulating microRNAs for early breast cancer detection. Clin Chem. 2015 Aug;61(8):1098-106.

(62) Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst 2015; 107(5).

(63) Welch HG, Passow HJ. Quantifying the benefits and harms of screening mammography. JAMA Intern Med 2014; 174(3):448-454.

(64) Etzioni R, Gulati R, Mallinger L, Mandelblatt J. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. Ann Intern Med 2013; 158(11):831-838.

(65) Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. BMJ 2015; 350:g7773.

(66) Etzioni R, Xia J, Hubbard R, Weiss NS, Gulati R. A reality check for overdiagnosis estimates associated with breast cancer screening. J Natl Cancer Inst 2014; 106(12).

(67) Moynihan R, Nickel B, Hersch J, Beller E, Doust J, Compton S et al. Public opinions about overdiagnosis: a national community survey. PLoS One 2015; 10(5):e0125165.

(68) Elmore JG, Harris RP. The harms and benefits of modern screening mammography. BMJ 2014; 348:g3824.

(69) Sprague BL, Stout NK, Schechter C, van Ravesteyn NT, Cevik M, Alagoz O et al. Benefits, Harms, and Cost-Effectiveness of Supplemental Ultrasonography Screening for Women With Dense Breasts. Ann Intern Med 2015 Feb 3;162(3):157-66

(70) Fenton JJ, Xing G, Elmore JG, Bang H, Chen SL, Lindfors KK et al. Short-term outcomes of screening mammography using computer-aided detection: a population-based study of medicare enrollees. Ann Intern Med 2013; 158(8):580-587.

(71) Miglioretti DL, Lange J, van Ravesteyn N, van den Broek JJ, Lee CI, Melnikow J et al. Radiation-Induced Breast Cancer and Breast Cancer Death From Mammography Screening [Abstract]. Rockville, MD: Agency for Healthcare Research and Quality. 2015.

(72) Trentham-Dietz A, Sprague BL, Hampton JM, Miglioretti DL, Nelson HD, Titus LJ et al. Modification of breast cancer risk according to age and menopausal status: a combined analysis of five population-based case-control studies. Breast Cancer Res Treat 2014; 145(1):165-175.

(73) Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet 2013; 45(4):392.

(74) Stevens KN, Vachon CM, Couch FJ. Genetic susceptibility to triple-negative breast cancer. Cancer Res 2013; 73(7):2025-2030.

(75) Hoffmann TC, Del MC. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA Intern Med 2015; 175(2):274-286.

Appendix Model validation

Each model has a different structure and assumptions and some varying input variables, so no single method can be used to validate results against an external standard. Therefore, we used several approaches. First, considering actual screening and treatment patterns instead of the efficacy strategies simulated in the base case, we compared model projections of incidence, breast cancer–specific mortality, and stage distribution with those reported by the Surveillance, Epidemiology, and End Results program for 1975 to 2010. In our previous work, results of each model accurately projected trends for incidence and breast cancer–specific mortality by ER status for 1975 to 2000 (23). Next, we approximated the Age screening trial (6), assuming perfect adherence to invitations for annual screening with 13-year follow-up of women aged 40 to 49 years (6). Finally, we examined the consistency of results across models. Using inputs for actual dissemination of screening and treatment in the United States, the models captured the major trends in incidence and the general shape of breast cancer–specific mortality decreases over time (Appendix Figure 1). They also closely matched current stage distribution (not shown) and the Age trial results (Appendix Table 2) (6, 33). Thus, the models replicated patterns of observed US incidence and breast cancer–specific mortality over time. The models also estimated similar breast cancer–specific mortality reduction as that observed among women aged 40 to 49 years who actually attended screening in the Age trial, although the model results are slightly more optimistic than the trial because the models assume 100% screening and use of the most effective systemic regimens (6). Overall, use of 6 models to project a range of plausible screening outcomes provides implicit cross-validation, with the range of results from the models as a measure of uncertainty.

Appendix Table 1. Utility Input Parameter Values

Utilities for Cancer-Related States*

State	Utility	Disutility (Worst Case 150%, 200%)	Duration	Unit
Cancer treatment for local or DCIS	0.9	0.1 (0.15, 0.20)	2	Year
Cancer treatment for regional	0.75	0.25 (0.375, 0.50)	2	Year
Cancer treatment for distant	0.6	0.4 (0.6, 0.8)	Until death	-
Screening attendance (routine screening)	0.994	0.006 (0.009, 0.012)	1	Week
Diagnostic phase (evaluation of positive screen)	0.895	0.105 (0.158, 0.210)	5	Weeks

Age-Specific Utilities for General Health in U.S. Women†

Age	Healthy Base Value (Range)
20 y	0.913 (0.905-0.920)
25 y	0.913 (0.905-0.920)
30 y	0.893 (0.886-0.900)
35 y	0.893 (0.886-0.900)
40 y	0.863 (0.855-0.871)
45 y	0.863 (0.855-0.871)
50 y	0.837 (0.829-0.846)
55 y	0.837 (0.829-0.846)
60 y	0.811 (0.800-0.822)
65 y	0.811 (0.800-0.822)
70 y	0.771 (0.758-0.784)
75 y	0.771 (0.758-0.784)
80 y	0.724 (0.701-0.747)
85 y	0.724 (0.701-0.747)

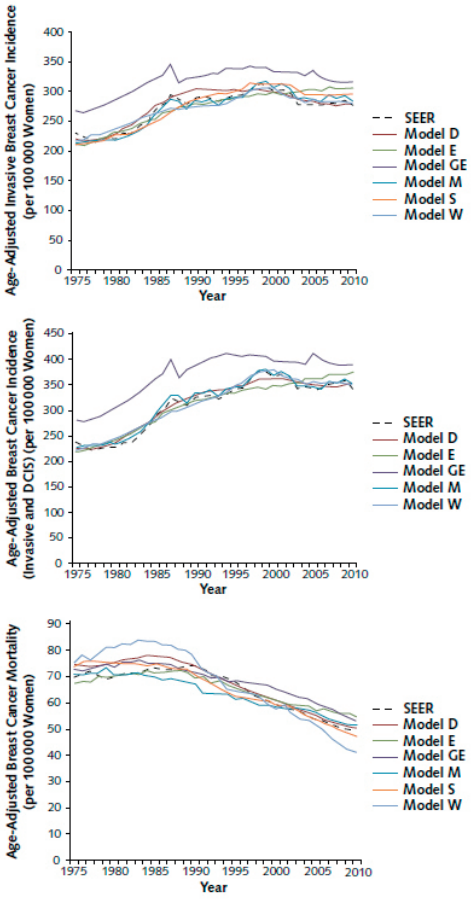
DCIS = ductal carcinoma in situ.
* From references 40 and 41.
† Values from the EuroQoL-5D quality-of-life questionnaire (39).

Appendix Table 2. Approximation of the Age Trial With 13-y Follow-up, by Model*

Model	Relative Risk for Breast Cancer Death With 100% Screening†
D	0.75
E	0.73
GE	0.65
M	0.72
S	0.69
W	0.71
Median (range)	0.72 (0.65-0.75)

D = Dana-Farber Cancer Institute; E = Erasmus Medical Center; GE = Georgetown University Medical Center and Albert Einstein College of Medicine; M = MD Anderson Cancer Center; S = Stanford University; W = University of Wisconsin and Harvard Medical School. * Projection of relative risk of breast cancer death with annual screening from age 40 to 49 y; biennial at age 50 and 52 y versus a control group with biennial screening at age 50 and 52 y. Because the models are estimating mortality reduction with actual screening, model estimates are most comparable to the Age trial results (6) among women who actually attended screening. Model results show more benefit than observed in the trial because the models assume that 100% of women complied with the trial-specified screening schedule. In reality, not all women who were invited attended screening, and among those who attended, many did not attend all scheduled screening rounds. In addition, the models assumed 100% receipt of the most effective treatments. † Age trial invitation results (intention to treat): relative risk, 0.83 (95% CI, 0.66–1.04). Age trial results for women who actually were screened: relative risk, 0.76 (CI, 0.51–1.01).

Appendix Figure 1. **Modeled versus observed incidence of breast cancer and breast cancer–specific mortality in women aged 40 to 100 years.**



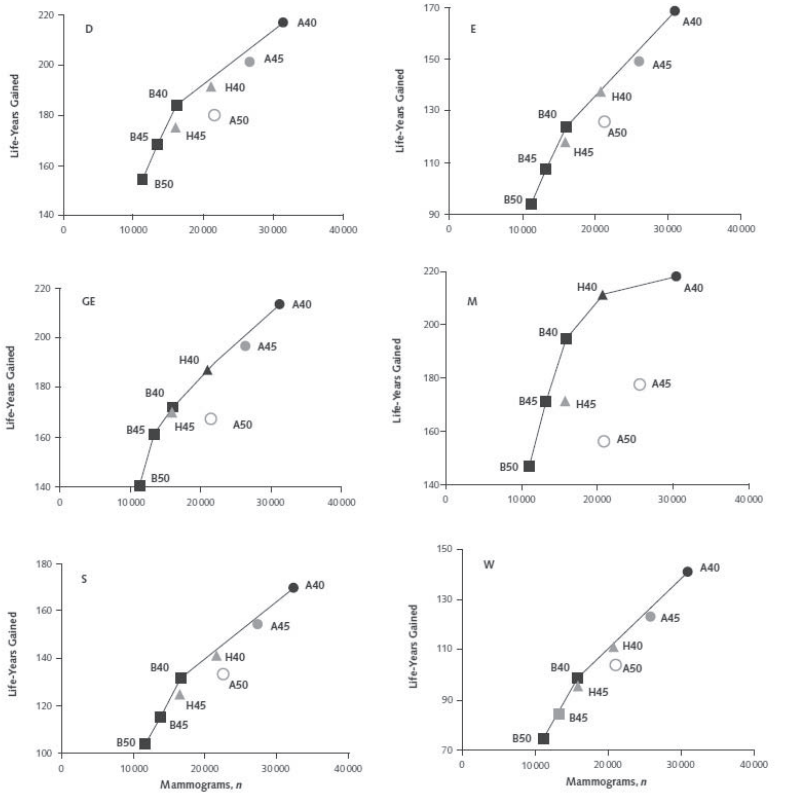
The models closely estimate observed U.S. trends in incidence of invasive disease (top), incidence of invasive disease and DCIS (middle)*, and breast cancer–specific mortality (bottom). Using inputs for actual dissemination of screening and treatment in the United States, the models all captured the major trends in incidence over time. Early increases with the advent of mammography in the mid-1980s are seen, followed by a downturn in the 2000s and then a leveling off. The models also captured the general shape of decreases in breast cancer–specific mortality over time. All models show an increase in incidence with the introduction of mammography screening. Model GE has a steep peak in incidence in 2005 owing to the specific method for capturing the transition from plain film to digital mammography, because digital mammography has higher sensitivity and detection of ductal carcinoma in situ than plain film mammography; other models include a more gradual transition surrounding this period. D = Dana- Farber Cancer Institute; DCIS = ductal carcinoma in situ; E = Erasmus Medical Center; GE = Georgetown University Medical Center and Albert Einstein College of Medicine; M = MD Anderson Cancer Center; QALY = quality-adjusted life-year; S = Stanford University; SEER = Surveillance, Epidemiology, and End Results; W = University of Wisconsin and Harvard Medical School. * Model S does not include DCIS.

Appendix Table 3. Annual Mortality Reduction Maintained by Biennial Screening, by Strategy and Model

Age at Screening	Mortality Reduction, %						
	Model D	Model E	Model GE	Model M*	Model S	Model W	Median
50-74†	79.8	76.7	84.6	98.9	82.8	72.0	81.3
45-74‡	77.8	73.4	81.5	96.6	80.2	68.3	79.0
40-74§	79.7	75.6	82.3	98.2	82.0	71.6	80.8

D = Dana-Farber Cancer Institute; E = Erasmus Medical Center; GE = Georgetown University Medical Center and Albert Einstein College of Medicine; M = MD Anderson Cancer Center; S = Stanford University; W = University of Wisconsin and Harvard Medical School. * Model M does not include a natural history component. On the basis of a combination of assumptions about underlying incidence trends in the absence of screening, it essentially yields a long lead time for invasive cancer; thus, all cancers found with annual screening can also be detected with biennial screening. † Percentage of reduction with annual screening in women aged 50-74 y that is maintained by biennial screening in women aged 50-74 y is calculated as the percent mortality reduction with biennial screening in women aged 50-74 y divided by the percent mortality reduction with annual screening in women aged 50-74 y. § Percentage of reduction with annual screening in women aged 45-74 y that is maintained by biennial screening in women aged 45-74 y is calculated as the percent mortality reduction with biennial screening in women aged 45-74 y divided by the percent mortality reduction with annual screening in women aged 45-74 y. § Percentage of reduction with annual screening in women aged 40-74 y that is maintained by biennial screening in women aged 40-74 y is calculated as the percent mortality reduction with biennial screening in women aged 40-74 y divided by the percent mortality reduction with annual screening in women aged 40-74 y.

Appendix Figure 2. Efficiency frontier for life-years gained versus mammograms performed for each screening strategy, by model.



The average gain in life-years per additional mammogram performed per 1000 women for each screening strategy (vs. no screening). Biennial strategies are indicated with a square; hybrid strategies (annual in the 40s followed by biennial from 50 to 74 years of age) with a triangle; and annual strategies with a circle. Efficient strategies were plotted (those in which increases in mammography use resulted in greater life-years gained than the next least-intensive strategy). The line represents the "efficiency frontier" by joining efficient strategies in which increases in mammography use resulted in greater life-years gained than the next less intensive efficient strategy. Strategies on this line would be considered efficient because they achieve the greatest gain in benefit (life years gained) per harm or use of mammograms. Strategies that use more mammograms but still have small benefits (i.e., a shallower slope than the next best strategy) are considered to be less efficient (i.e., weakly dominated). When and if the slope in the efficiency frontier plot levels off, it means that the additional life-years gained per increase in mammography are small relative to the previous strategies and could indicate a point at which additional screening might be considered as having a low return (or additional benefit). There is no definitive inflection point across the models for the strategies or metrics evaluated. Black strategies are efficient; gray strategies close to the efficiency frontier are less efficient; and open gray strategies are inefficient (inferior, or dominated). Reference 33 provides efficiency frontiers for other harm and benefit metrics. D = Dana-Farber Cancer Institute; E = Erasmus Medical Center; GE = Georgetown University Medical Center and Albert Einstein College of Medicine; M = MD Anderson Cancer Center; QALY = quality-adjusted life-year; S = Stanford University; W = University of Wisconsin and Harvard Medical School.

Chapter 8

Radiation-Induced Breast Cancer Incidence and Mortality From Digital Mammography Screening: A Modeling Study.

Miglioretti DL, Lange J, van den Broek JJ, Lee CI, van Ravesteyn NT, Ritley D, Kerlikowske K, Fenton JJ, Melnikow J, de Koning HJ, Hubbard RA.

Annals of Internal Medicine. 2016 Feb 16;164(4):205-14. doi: 10.7326/M15-1241.

ABSTRACT

Background

Estimates of risk for radiation-induced breast cancer from mammography screening have not considered variation in dose exposure or diagnostic work-up after abnormal screening results.

Objective

To estimate distributions of radiation-induced breast cancer incidence and mortality from digital mammography screening, considering exposure from screening and diagnostic mammography and dose variation across women.

Design

Two simulation-modeling approaches.

Setting

U.S. population.

Patients

Women aged 40-74 years.

Interventions

Annual or biennial digital mammography screening from age 40, 45, or 50 years until age 74 years.

Measurements

Lifetime breast cancer deaths averted (benefits) and radiation-induced breast cancer incidence and mortality (harms) per 100,000 women screened.

Results

Annual screening of 100,000 women aged 40 to 74 years was projected to induce 125 breast cancers (95% confidence interval [CI]=88–178) leading to 16 deaths (95% CI=11–23) relative to 968 breast cancer deaths averted by early detection from screening. Women exposed at the 95th percentile were projected to develop 246 radiation-induced breast cancers leading to 32 deaths per 100,000 women. Women with large breasts requiring extra views for complete breast examination (8% of population) were projected to have higher radiation-induced breast cancer incidence and mortality (266 cancers, 35 deaths per 100,000 women), compared to women with small or average breasts (113

cancers, 15 deaths per 100,000 women). Biennial screening starting at age 50 reduced risk of radiation-induced cancers 5-fold.

Limitations

Life-years lost from radiation-induced breast cancer could not be estimated.

Conclusions

Radiation-induced breast cancer incidence and mortality from digital mammography screening are affected by dose variability from screening, resultant diagnostic work-up, initiation age, and screening frequency. Women with large breasts may have a greater risk for radiation-induced breast cancer.

Funding source

Agency for Healthcare Research and Quality, U.S. Preventive Services Task Force, National Cancer Institute.

Primary research and data collection for the American College of Radiology Imaging Network (ACRIN) Digital Mammographic Imaging Screening Trial (DMIST) were supported by the National Cancer Institute (U01 CA80098, U01 CA80098-S1, U01 CA79778, and U01 79778-S1).

INTRODUCTION

Exposure to ionizing radiation from repeated mammography examinations may increase breast cancer risk (1, 2). Radiation-induced breast cancer incidence and mortality associated with recommended screening strategies are suggested to be low relative to breast cancer deaths prevented (3-5). However, prior projected population risks were based on exposure from screening only and assumed only four standard views per screen at the mean radiation dose. Evaluations of screening programs should consider full episodes of care including diagnostic work-up prompted by an abnormal screening result (6). False-positive recalls, breast biopsies, and short-interval follow-up examinations are relatively common in the United States and add radiation exposure from diagnostic mammography (7). Some subgroups of women, such as obese women and women with dense breasts, are more likely to have additional evaluations (7-9), which may increase their risk for radiation-induced cancer.

When risk for radiation-induced breast cancer is being evaluated, it may also be important to consider variation in radiation dose from a single examination. Examinations vary in the number of views performed and dose per view; therefore, some women receive more than the mean dose. The American College of Radiology Imaging Network DMIST (Digital Mammographic Imaging Screening Trial) found an average radiation dose to the breast of 1.86 mGy to the breast from a single digital mammography screening view (10), but dose per view varied widely from 0.15 to 13.4 mGy (**Supplemental Content**) and 21% of digital screening examinations used more than four views (10). Radiation dose is strongly correlated with compressed breast thickness; thus, women with large breasts women tend to receive higher doses per view and may require more than four views for complete examination (10, 11). Women with breast augmentation receive implant-displacement views in addition to standard screening views, which doubles their radiation dose (12). Woman may have repeated views because of movement artifacts or improper breast positioning.

We estimated the distribution of cumulative radiation dose and associated breast cancer risk from full screening episodes to identify subgroups of women who may have a greater risk for radiation-induced cancer because they have factors contributing to greater doses per examination or frequent false-positive screening results that lead to additional radiation exposure from subsequent diagnostic work-up. Using population-based data from the Breast Cancer Surveillance Consortium (BCSC) (13), we estimated the probability of a false-positive screening result followed by additional imaging evaluation, short-interval follow-up, or biopsy. We used data from the BCSC, DMIST, and other sources in 2 simulation models to estimate radiation exposure and radiation-induced breast cancer incidence and mortality associated with 8 potential screening strategies

with different starting ages (40, 45, or 50 years) and screening intervals (annual, biennial, or a hybrid strategy).

METHODS

Screening Strategies

We used 2 complementary stochastic modeling approaches to evaluate the following 8 strategies for screening with digital mammography:

1. Annual screening from ages 40-74, 45-74, and 50-74 years.
2. Biennial screening from ages 40-74, 45-74, and 50-74 years.
3. Hybrid strategy of annual screening from ages 40-49 or 45-49 and biennial screening from ages 50-74 years.

We included the hybrid strategies because more frequent screening has been advocated for younger and premenopausal women due to their greater prevalence of dense breasts and more aggressive tumors, resulting in a greater risk for interval cancer, than older women (14-17). Outcomes were breast cancer deaths averted (benefits) and radiation-induced breast cancer incidence and mortality (harms) associated with a lifetime of mammography screening relative to no screening.

Simulation Modeling Approaches

Figure 1 summarizes our approach. We used 2 complementary stochastic modeling approaches to simulate mammography events associated with radiation exposure and outcomes for a population adherent with each of the 8 screening strategies. The first approach used the Microsimulation of Screening Analysis–Fatal Diameter (MISCAN-Fadia) model (18), which is a detailed natural history model of breast cancer. This approach provided estimates of breast cancer incidence and mortality with and without screening to contextualize estimates of radiation-induced breast cancer cases. Although MISCAN-Fadia models the average effects of screening on a population level, it does not model correlation among repeated mammography results in individual women or the specific types of work-up after an abnormal screening result; thus, it cannot be used to estimate the distribution of cumulative radiation exposure from both screening mammography and subsequent diagnostic work-up among women. Therefore, we developed a new simulation model that provides woman-level exposure histories that were not available from the MISCAN-Fadia model. This new model captures exposure heterogeneity by simulating mammography results and subsequent work-up in each woman and allowing for variability in radiation exposure and breast size.

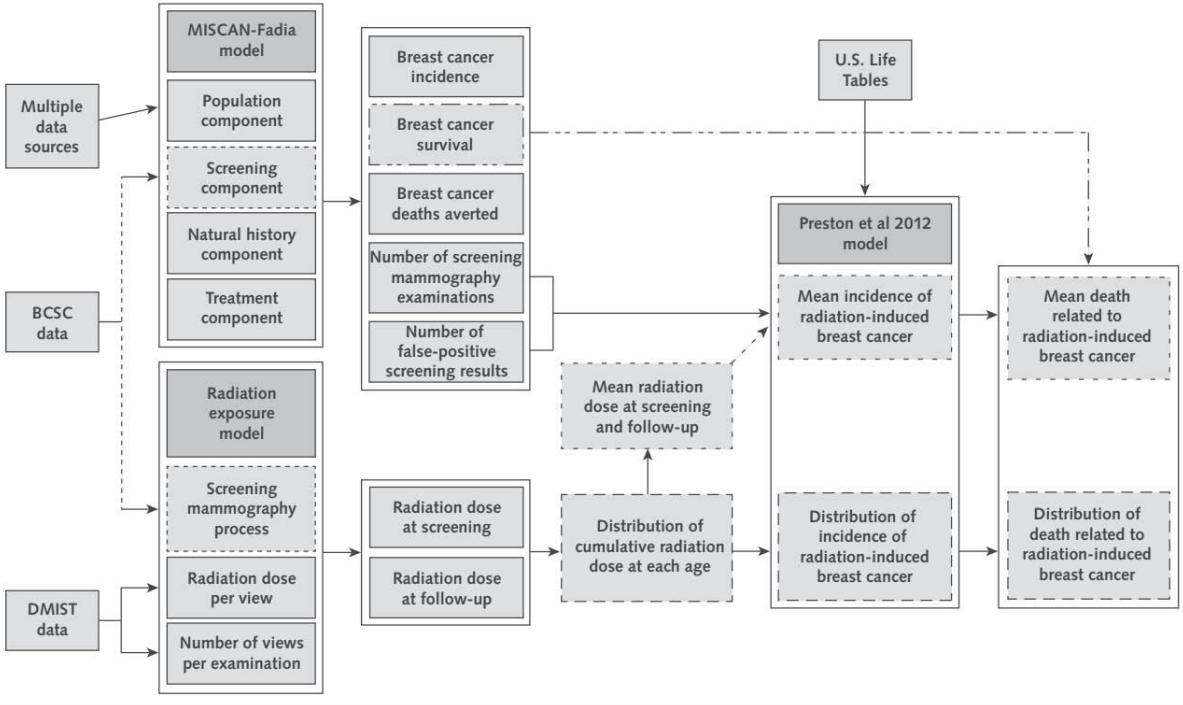


Figure 1 Schematic of 2 modeling approaches used to simulate mammography events and outcomes associated with 8 screening strategies.
Estimates of the number of screening examinations and false-positive results from the MISCAN-Fadia model were combined with the mean radiation dose from the radiation exposure model to estimate mean incidence of radiation-induced breast cancer. Estimates of the probability distribution of cumulative radiation dose at each age among women from the radiation exposure model were used to estimate the probability distribution of radiation-induced breast cancer incidence. Radiation-induced breast cancer incidence was combined with breast cancer survival estimates from the MISCAN-Fadia model to estimate radiation-induced breast cancer mortality. BCSC = Breast Cancer Surveillance Consortium; DMIST = Digital Mammographic Imaging Screening Trial; MISCAN-Fadia = Microsimulation of Screening Analysis–Fatal Diameter.

MISCAN-Fadia Simulation Model

The MISCAN-Fadia microsimulation model simulates individual life histories of women with and without breast cancer in the presence and absence of screening from birth to death from breast cancer or other causes. The model has been described in detail elsewhere (18) and information about the model can be found online (http://cisnet.cancer.gov/); inputs and assumptions are described in our report for the draft USPSTF recommendations (19). In brief, on the basis of BCSC data on sensitivity of digital mammography screening, cancer detection rates, and cancer stage at detection, we estimated thresholds at which tumors become screen-detectable. Screening sensitivity and specificity depended on age, breast density, and screening interval. Breast cancer risk depended on age and breast density. The effect of screening on breast cancer natural history was assessed by modeling continuous tumor growth, in which tumors detected before they reached their fatal diameter were cured and those detected past their fatal diameter led to breast cancer death. We assumed that all women received the mean dose per screening examination and, if recalled, the mean dose associated with diagnostic work-up after a false-positive screening result, both of which were estimated from the radiation exposure model. We also projected breast cancer incidence and mortality with and without screening.

Radiation Exposure Simulation Model

Full details including approach, data sources, and assumptions are available in the **Supplemental Content**. In brief, for each of the 8 screening strategies, we simulated woman-level factors and screening-related events for 100 000 women.

Woman-level factors: Each woman was assigned a compressed breast thickness from the DMIST distribution (**Supplemental Table 2**). Women with a compressed breast thickness of 7.5 cm or greater (8% of DMIST population) were assumed to have large breasts that required extra views for complete examination. On the basis of distributions seen in the BCSC, each woman was assigned a baseline Breast Imaging Reporting and Data System (12) density at the start of screening, which could potentially decrease by 1 category at ages 50 and 65 years (20) (**Supplemental Table 4**).

Evaluation of a positive screening exam: For each screening strategy, we simulated events after a positive screening result that did not lead to a diagnosis of breast cancer (Figure 2) to focus on risk for first breast cancer induced by radiation. We modeled the probability of each event by using data from digital mammography done at BCSC facilities from 2003 to 2011 on women aged 40 to 74 years without a history of breast cancer or cancer diagnosed within 1 year after the examination. At each screening, a woman's probability of recall for additional imaging was based on age, breast density, screening interval, prior screening results, and a woman-specific random effect. If recalled, the probability of referral to biopsy, short-interval follow-up, or return to routine screening was based on age, breast density, and screening interval.

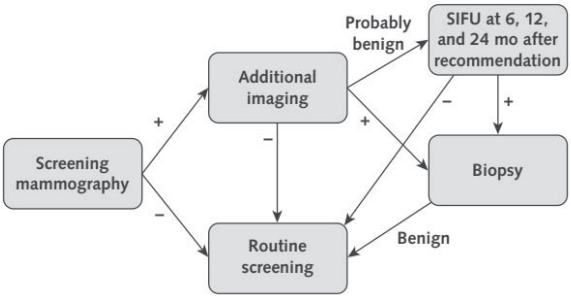


Figure 2 Screening mammography process.
Short-interval follow-up (SIFU) examinations included unilateral diagnostic views on the recalled breast at 6 mo after the initial SIFU recommendation. The examinations included unilateral diagnostic views on the recalled breast plus bilateral routine screening views at 12 and 24 mo after the initial SIFU recommendation for women who received annual screening and 24 mo after the initial SIFU recommendation for those who received biennial screening. The routine screening views could result in recall for additional imaging to work up a new finding, followed by a recommendation for another SIFU examination or tissue biopsy.

Radiation dose: For each screening and diagnostic event, we sampled the number of screening mammography views from the DMIST distribution (**Supplemental Table 1**) and number of views for diagnostic work-up on the basis of expert opinion, conditional on compressed breast thickness (**Supplemental Table 3)**. assumed different distributions of views for women with and without large breasts. We randomly sampled the radiation dose per view on the basis of the DMIST distribution conditional on the woman's compressed breast thickness (**Supplemental figure 1**). For each age, we calculated total breast-level dose by multiplying half the number of views of both breasts by the dose per view. We report the mean and the 5th, 25th, 75th, and 95th percentiles (to quantify exposure leading to increased risk for radiation induced breast cancer) for the number of mammography views and associated dose from each screening examination and all follow-up mammograms within 1 year of a screening examination **Supplemental Table 9**.

Radiation-induced breast cancer incidence and mortality

We estimated radiation-induced breast cancer incidence using the excess absolute risk model from pooled analysis of four cohorts by Preston and colleagues (1), the preferred model for estimating radiation-induced breast cancer incidence (2, 21). Details are provided in the Supplemental Content. Women in these cohorts received cumulative radiation doses of 20 mGy or greater. This level of cumulative radiation exposure is reached after 2 to 4 years of mammography screening and diagnostic work-up (**Supplemental Table 9**). This model assumes that excess risk of radiation-induced breast cancer increases linearly with increasing radiation dose within the exposure ranges from mammography. In addition, risk decreases with increasing age at exposure, especially after

age 50 (a surrogate for menopause) and increases with age, the highest incidence of radiation-induced breast cancer late in life. We modeled the latency period for developing radiation-induced breast cancer using a logistic function that phases in increased breast cancer risk between 4 and 11 years after exposure (21). We estimated radiation-induced breast cancer mortality by multiplying radiation-induced breast cancer incidence by the age-specific case-fatality rates derived from MISCAN-Fadia and assuming 100% adherence to screening and available treatment. We assumed that breast cancers induced by radiation is screen-detected at the same rate as non-induced cancer. We approximated Confidence Intervals (CI) by re-estimating risk using the upper and lower 95% CIs for the risk coefficient, β, because this uncertainty dominates the uncertainty in estimated risk (2, 21).

The MISCAN-Fadia model was programmed in Delphi (Borland). All other analyses were done in R, version 3.1.0 (R Foundation for Statistical Computing) and SAS version 9.4 (SAS Institute).

Role of the Funding Source

This study was funded by the Agency for Healthcare Research and Quality under a contract to support the work of the U.S. Preventive Services Task Force and by the National Cancer Institute. Investigators worked with Task Force members and Agency staff to develop the scope, analytic framework, and key questions. The funding source had no role in model input selection, data synthesis, or data analysis. Agency staff provided project oversight and reviewed the report to ensure that the analysis met methodological standards. The authors are solely responsible for the content and the decision to submit the manuscript for publication.

RESULTS

Radiation exposure

Most radiation exposure from screening and subsequent diagnostic work-up was due to the screening examination (**Supplemental Table 9**). Diagnostic work-up accounted for only 10% of the mean annual radiation dose but 24% of the dose for women with exposure at the 95th percentile. On average, women with large breasts were exposed to 2.3 times more radiation than those with small or average-sized breasts.

Radiation-induced breast cancer incidence and breast cancer death

Risk estimates corresponding to mean exposures were similar for the 2 modeling approaches (Table 1); therefore, we focus on results from the radiation exposure model. We projected that annual screening and diagnostic work-up of 100 000 women aged

Table 1 Comparison of lifetime attributable risks of radiation-induced breast cancer and breast cancer death (per 100,000 women) from two modeling approaches.

Screening Strategy	MISCAN-Fadia Model	Radiation-Exposure Model		
	Mean (95% CI)	Mean (95% CI)	5th percentile (95% CI)	95th percentile (95% CI)
	Lifetime Attributable Risk of Radiation-Induced Breast Cancer (Per 100,000 Women)			
Biennial screening				
Ages 50-74 y	28 (20, 40)	27 (19, 38)	13 (9, 19)	55 (39, 78)
Ages 45-74 y	44 (31, 62)	45 (31, 64)	21 (15, 30)	92 (65, 130)
Ages 40-74 y	67 (47, 96)	68 (48, 97)	33 (23, 47)	138 (97, 196)
Hybrid strategy				
A45-49 y, B50-74 y	57 (40, 81)	59 (41, 84)	29 (20, 41)	118 (82, 168)
A40-49 y, B50-74 y	101 (71, 143)	89 (62, 126)	44 (31, 62)	177 (125, 251)
Annual screening				
Ages 50-74 y	54 (39, 75)	49 (34, 69)	25 (17, 35)	97 (68, 139)
Ages 45-74 y	85 (59, 121)	81 (57, 115)	41 (29, 58)	159 (111, 226)
Ages 40-74 y	129 (90, 183)	125 (88, 178)	64 (44, 90)	246 (171, 349)
Lifetime Attributable Risk of Radiation-Induced Breast Cancer Death (Per 100,000 Women)				
Biennial screening				
Ages 50-74 y	5 (3, 7)	4 (3, 6)	2 (2, 3)	9 (6, 13)
Ages 45-74 y	8 (5, 11)	8 (5, 11)	4 (3, 5)	16 (11, 22)
Ages 40-74 y	12 (8, 17)	12 (8, 17)	6 (4, 8)	24 (17, 34)
Hybrid strategy				
A45-49 y, B50-74 y	10 (7, 14)	10 (7, 14)	5 (3, 7)	20 (14, 29)
A40-49 y, B50-74 y	18 (13, 25)	15 (11, 22)	8 (5, 11)	31 (22, 44)
Annual screening				
Ages 50-74 y	7 (5, 10)	7 (5, 9)	3 (2, 5)	13 (9, 19)
Ages 45-74 y	11 (8, 16)	11 (8, 15)	5 (4, 8)	21 (15, 30)
Ages 40-74 y	16 (12, 23)	16 (11, 23)	8 (6, 12)	32 (22, 45)

40 to 74 years (35 screening examinations per woman) would induce an average of 125 breast cancer cases (95% CI, 88 to 178), resulting in 16 deaths (CI, 11 to 23) (Table 1). Risk projections varied widely, with 100 000 women exposed at the 5th percentile projected to develop 64 radiation-induced cancer cases (CI, 44 to 90), resulting in 8 deaths (CI, 6 to 12), and 100 000 women exposed at the 95th percentile projected to develop 246 radiation-induced cases of cancer (CI, 171 to 349), resulting in 32 deaths (CI, 22 to 45). Women with large breasts requiring extra views for complete examination had more than twice as many cases of radiationinduced breast cancer (mean, 266 cases [CI, 186 to 380]) and breast cancer deaths (mean, 35 deaths [CI, 24 to 50]) than women with small or average-sized breasts (113 breast cancer cases [CI, 79 to 161] and 15 breast cancer deaths [CI, 10 to 21]) (Table 2). Starting screening at age 50 years and following a biennial

Table 2 Mean, 5th percentile, and 95th percentile (95% confidence intervals) of lifetime attributable risks (per 100,000 women) of radiation-induced breast cancer and breast cancer death, by breast size, for different screening strategies.

Screening Strategy	Small or average breasts			Large breasts		
	Mean (95% CI)	5th percentile (95% CI)	95th percentile (95% CI)	Mean (95% CI)	5th percentile (95% CI)	95th percentile (95% CI)
	Lifetime Attributable Risk of Radiation-Induced Breast Cancer (Per 100,000 Women)					
Biennial screening						
Ages 50-74 y	24 (17, 35)	13 (9, 18)	43 (30, 61)	57 (40, 82)	28 (19, 40)	108 (77, 154)
Ages 45-74 y	40 (28, 57)	21 (15, 30)	72 (50, 102)	95 (67, 135)	46 (32, 65)	181 (128, 259)
Ages 40-74 y	61 (43, 87)	33 (23, 46)	107 (76, 152)	144 (100, 205)	71 (49, 101)	266 (188, 384)
Hybrid strategy						
A45-49 y, B50-74 y	53 (37, 75)	29 (20, 41)	91 (64, 130)	125 (87, 178)	60 (43, 88)	233 (162, 335)
A40-49 y, B50-74 y	80 (56, 114)	43 (31, 62)	137 (96, 195)	189 (132, 269)	95 (65, 134)	351 (244, 495)
Annual screening						
Ages 50-74 y	44 (31, 62)	25 (17, 35)	74 (52, 105)	104 (73, 149)	53 (37, 76)	187 (131, 267)
Ages 45-74 y	73 (51, 103)	40 (28, 57)	122 (85, 174)	173 (121, 245)	88 (62, 126)	315 (221, 445)
Ages 40-74 y	113 (79, 161)	63 (44, 89)	189 (133, 268)	266 (186, 380)	136 (95, 193)	487 (339, 700)
Lifetime Attributable Risk of Radiation-Induced Breast Cancer Death (Per 100,000 Women)						
Biennial screening						
Ages 50-74 y	4 (3, 6)	2 (1, 3)	7 (5, 10)	10 (7, 14)	5 (3, 7)	18 (13, 26)
Ages 45-74 y	7 (5, 10)	4 (3, 5)	12 (9, 17)	16 (11, 23)	8 (5, 11)	31 (22, 44)
Ages 40-74 y	11 (7, 15)	6 (4, 8)	19 (13, 26)	25 (17, 35)	12 (8, 17)	46 (33, 67)
Hybrid strategy						
A45-49 y, B50-74 y	9 (6, 13)	5 (3, 7)	16 (11, 22)	21 (15, 31)	10 (7, 15)	40 (28, 57)
A40-49 y, B50-74 y	14 (10, 20)	8 (5, 11)	24 (17, 34)	33 (23, 47)	16 (11, 23)	61 (42, 86)
Annual screening						
Ages 50-74 y	6 (4, 9)	3 (2, 5)	10 (7, 14)	14 (10, 20)	7 (5, 10)	25 (18, 36)
Ages 45-74 y	10 (7, 14)	5 (4, 8)	16 (11, 23)	23 (16, 33)	12 (8, 17)	42 (29, 59)
Ages 40-74 y	15 (10, 21)	8 (6, 12)	25 (17, 35)	35 (24, 50)	18 (12, 25)	63 (44, 91)

strategy (13 screening examinations) greatly reduced risk for radiation-induced breast cancer and breast cancer death (Table 1). Compared with annual screening from age 40 to 74 years, biennial screening from age 50 to 74 years was projected to cause approximately one fifth of the radiation-induced breast cancer cases (mean, 125 cases [CI, 88 to 178] vs. 27 cases [CI, 19 to 38] per 100 000 women, respectively, and 266 cases [CI, 186 to 380] vs. 57 cases [CI, 40 to 82] per 100 000 women with large breasts) (Table 2).

Breast cancer deaths averted per radiation-induced cancer:

From the MISCAN-Fadia model, we projected that 16 947 breast cancer cases would be diagnosed from age 40 years through death per 100 000 women screened annually

Table 2 Mean, 5th percentile, and 95th percentile (95% confidence intervals) of lifetime attributable risks (per 100,000 women) of radiation-induced breast cancer and breast cancer death, by breast size, for different screening strategies.

Screening Strategy	Small or average breasts			Large breasts		
	Mean (95% CI)	5th percentile (95% CI)	95th percentile (95% CI)	Mean (95% CI)	5th percentile (95% CI)	95th percentile (95% CI)
	Lifetime Attributable Risk of Radiation-Induced Breast Cancer (Per 100,000 Women)					
Biennial screening						
Ages 50-74 y	24 (17, 35)	13 (9, 18)	43 (30, 61)	57 (40, 82)	28 (19, 40)	108 (77, 154)
Ages 45-74 y	40 (28, 57)	21 (15, 30)	72 (50, 102)	95 (67, 135)	46 (32, 65)	181 (128, 259)
Ages 40-74 y	61 (43, 87)	33 (23, 46)	107 (76, 152)	144 (100, 205)	71 (49, 101)	266 (188, 384)
Hybrid strategy						
A45-49 y, B50-74 y	53 (37, 75)	29 (20, 41)	91 (64, 130)	125 (87, 178)	60 (43, 88)	233 (162, 335)
A40-49 y, B50-74 y	80 (56, 114)	43 (31, 62)	137 (96, 195)	189 (132, 269)	95 (65, 134)	351 (244, 495)
Annual screening						
Ages 50-74 y	44 (31, 62)	25 (17, 35)	74 (52, 105)	104 (73, 149)	53 (37, 76)	187 (131, 267)
Ages 45-74 y	73 (51, 103)	40 (28, 57)	122 (85, 174)	173 (121, 245)	88 (62, 126)	315 (221, 445)
Ages 40-74 y	113 (79, 161)	63 (44, 89)	189 (133, 268)	266 (186, 380)	136 (95, 193)	487 (339, 700)
Lifetime Attributable Risk of Radiation-Induced Breast Cancer Death (Per 100,000 Women)						
Biennial screening						
Ages 50-74 y	4 (3, 6)	2 (1, 3)	7 (5, 10)	10 (7, 14)	5 (3, 7)	18 (13, 26)
Ages 45-74 y	7 (5, 10)	4 (3, 5)	12 (9, 17)	16 (11, 23)	8 (5, 11)	31 (22, 44)
Ages 40-74 y	11 (7, 15)	6 (4, 8)	19 (13, 26)	25 (17, 35)	12 (8, 17)	46 (33, 67)
Hybrid strategy						
A45-49 y, B50-74 y	9 (6, 13)	5 (3, 7)	16 (11, 22)	21 (15, 31)	10 (7, 15)	40 (28, 57)
A40-49 y, B50-74 y	14 (10, 20)	8 (5, 11)	24 (17, 34)	33 (23, 47)	16 (11, 23)	61 (42, 86)
Annual screening						
Ages 50-74 y	6 (4, 9)	3 (2, 5)	10 (7, 14)	14 (10, 20)	7 (5, 10)	25 (18, 36)
Ages 45-74 y	10 (7, 14)	5 (4, 8)	16 (11, 23)	23 (16, 33)	12 (8, 17)	42 (29, 59)
Ages 40-74 y	15 (10, 21)	8 (6, 12)	25 (17, 35)	35 (24, 50)	18 (12, 25)	63 (44, 91)

CI, confidence interval; y, years; A, annual screening at ages 40-50 or 45-50 and B, biennial screening at 50-74 years.

from age 40 to 74 years (data not shown). The number of breast cancer deaths averted ranged from 627 per 100 000 women screened biennially from age 50 to 74 years to 968 per 100 000 women screened annually from age 40 to 74 years (Table 3). For biennial screening from age 50 to 74 years, we projected a mean of 23 breast cancer deaths averted for each radiation-induced case of breast cancer (CI, 16 to 33) (5th percentile, 48; 95th percentile, 11) and 140 breast cancer deaths averted for each radiation induced breast cancer death (CI, 98 to 199) (5th percentile, 289; 95th percentile, 68). For annual screening from age 40 to 74 years, these ratios were lower, at 8 breast cancer deaths

Table 3 Number of breast cancer deaths averted by screening 100,000 women and ratio of number of breast cancer deaths averted per number (mean, 5th percentile, and 95th percentile) of radiation-induced breast cancers and of radiation-induced breast cancer deaths.

Strategy	Number of breast cancer deaths averted by screening	Overall			Small or average breasts	Large breasts
		Mean (95% CI)	5th Percentile (95% CI)	95th Percentile (95% CI)	Mean (95% CI)	Mean (95% CI)
	Ratio of Breast Cancer Deaths Averted per Radiation-Induced Breast Cancer					
Biennial screening						
Ages 50-74 y	627	23 (16, 33)	48 (34, 69)	11 (8, 16)	26 (18, 37)	11 (8, 16)
Ages 45-74 y	666	15 (10, 21)	31 (22, 45)	7 (5, 10)	17 (12, 24)	7 (5, 10)
Ages 40-74 y	732	11 (8, 15)	22 (16, 32)	5 (4, 8)	12 (8, 17)	5 (4, 7)
Hybrid strategy						
A45-49 y, B50-74 y	717	12 (9, 17)	25 (17, 35)	6 (4, 9)	14 (10, 19)	6 (4, 8)
A40-49 y, B50-74 y	780	9 (6, 13)	18 (12, 25)	4 (3, 6)	10 (7, 14)	4 (3, 6)
Annual screening						
Ages 50-74 y	819	17 (12, 24)	33 (23, 47)	8 (6, 12)	19 (13, 27)	8 (6, 11)
Ages 45-74 y	907	11 (8, 16)	22 (16, 32)	6 (4, 8)	12 (9, 18)	5 (4, 8)
Ages 40-74 y	968	8 (5, 11)	15 (11, 22)	4 (3, 6)	9 (6, 12)	4 (3, 5)
	Ratio of Breast Cancer Deaths Averted per Radiation-Induced Breast Cancer Death					
Biennial screening						
Ages 50-74 y	627	140 (98, 199)	289 (203, 415)	68 (48, 97)	155 (109, 221)	66 (46, 93)
Ages 45-74 y	666	87 (61, 125)	184 (130, 263)	43 (30, 60)	97 (68, 139)	41 (29, 59)
Ages 40-74 y	732	62 (44, 89)	128 (90, 183)	31 (22, 44)	69 (48, 98)	29 (21, 42)
Hybrid strategy						
A45-49 y, B50-74 y	717	71 (50, 102)	145 (102, 207)	35 (25, 51)	79 (56, 113)	33 (23, 48)
A40-49 y, B50-74 y	780	51 (36, 72)	102 (72, 146)	25 (18, 36)	56 (40, 80)	24 (17, 34)
Annual screening						
Ages 50-74 y	819	123 (86, 176)	242 (171, 346)	62 (43, 89)	136 (96, 195)	58 (40, 83)
Ages 45-74 y	907	84 (60, 121)	167 (118, 239)	43 (30, 61)	94 (66, 134)	39 (28, 57)
Ages 40-74 y	968	59 (42, 85)	117 (82, 167)	30 (21, 43)	66 (46, 94)	28 (20, 40)

CI, confidence interval; y, years; A, annual screening at ages 40-50 or 45-50 and B, biennial screening at 50-74 years.

averted per radiation-induced case of breast cancer (CI, 5 to 11) (5th percentile, 15; 95th percentile, 4) and 59 breast cancer deaths averted per radiation-induced breast cancer death among all women (CI, 42 to 85) (5th percentile, 117; 95th percentile, 30). For annual screening from age 40 to 74 years of women with large breasts, ratios were even lower, at 4 breast cancer deaths averted per radiation-induced case of breast cancer (CI, 3 to 5) and 28 per radiation induced breast cancer death (CI, 20 to 40).

DISCUSSION

We improved previous estimates of the potential harms from radiation exposure of screening strategies for breast cancer by using methods that more fully represent the experience of women who have routine digital screening mammography. Our models included radiation exposure from diagnostic evaluations prompted by abnormal screening results and incorporated variation in dose at each screening and diagnostic examination. In addition to the mean, we reported the 5th and 95th percentiles of the population distribution to highlight that some women have risk that is substantially lower or higher than average because of variation in radiation exposure. Most of the increased risk was due to screening examinations with more than 4 views and higher-than-average doses per view. We used DMIST data to model the number of views per screening examination and to incorporate the increased radiation dose per view for thicker compressed breasts. However, even for a given compressed breast thickness, some women received greater doses than others, which was probably due to greater breast density that required more radiation for penetration. Because women with large breasts may require more views per examination and tend to receive a greater dose per view, breast size was an important factor in determining radiation exposure and associated risk. Another reason for greater radiation exposure is false-positive results; additional imaging performed to work up false-positive results accounted for one fourth of the radiation dose received by women at the 95th percentile compared with only one tenth of the radiation dose received by women at the mean.

Relative to a projected 16 947 breast cancer cases diagnosed per 100 000 women aged 40 years or older with annual screening, we estimate that the number of breast cancer cases induced by screening is probably very small, even for women with the greatest radiation exposures. However, relative to the number of breast cancer deaths averted with screening, radiation induced breast cancer incidence is not trivial. Most concerning are numbers projected for annual screening and screening before age 50 years of women with large breasts requiring extra views for complete examination, who have more than twice the risk for radiation induced breast cancer as women with small or average-sized breasts. Although we did not model this explicitly, women with breast augmentation should also have twice the risk for radiation-induced breast cancer because they receive implant-displacement views in addition to standard screening views, resulting in a minimum of 8 views per examination compared with the standard 4 views (12).

The benefit–harm ratio in terms of breast cancer deaths averted per radiation-induced case of breast cancer could be improved by initiating screening at age 50 years instead of 40 years, thereby reducing risk for radiation-induced breast cancer by 60%, or by using biennial screening, which would cut the risk in half compared with annual screening.

Doing both (screening biennially from age 50 to 74 years) would reduce the risk almost 5-fold compared with annual screening from age 40 to 74 years. Several steps should be taken to further improve the benefit–harm ratio. Current efforts to reduce the radiation dose per view should continue. Radiology staff should strive to minimize the number of additional views performed and to reduce false-positive rates, which are much higher in the United States than many other countries, suggesting room for improvement (22-25). Radiation doses from diagnostic mammography could be avoided for certain screen-detected masses amenable to ultrasonography work-up alone. In addition, facilities should ensure that large breasts are imaged using larger detector sizes to minimize the need for extra views for complete examination.

Hendrick (3) also estimated incidence and mortality of radiation-induced breast cancer using DMIST data but used the mean dose for 4 views without accounting for additional radiation exposure from additional screening views received by 21% of women or from diagnostic follow-up imaging. He projected that annual screening of 100 000 women from age 40 to 80 years with an examination-level dose of 3.7 mGy would induce 72 breast cancer cases leading to 20 deaths. For women screened annually from age 40 to 74 years, we estimated fewer breast cancer deaths (16 deaths per 100 000 women), despite more radiation-induced breast cancer cases (125 cases per 100 000 women), because we optimistically assumed 100% adherence to the screening regimen and use of available treatments. In particular, we assumed that 10% to 19% of women diagnosed with breast cancer between ages 40 and 74 years would die of the disease (depending on the screening scenario) compared with recent estimates of more than 23% (26). Thus, we may have underestimated the number of radiation-induced breast cancer deaths. Yaffe and Mainprize (4) projected that screening 100 000 women annually from age 40 to 55 years and biennially thereafter to age 74 years with a dose of 3.7 mGy would induce 86 breast cancer cases and 11 deaths. In comparison, we projected that screening 100 000 women annually from age 40 to 49 years and biennially thereafter to age 74 years would induce 89 breast cancer cases and 15 deaths. Our estimates are probably greater because we accounted for some screening examinations having more than 4 views and for radiation exposure from diagnostic work-up.

Doses from current digital mammography systems may be lower than doses from older DMIST units. Nevertheless, DMIST doses may still be conservative because, similar to most prior studies, dose estimates assumed breast compositions of 50% glandular tissue, which probably underestimates dose by 8% to 18% (27, 28). Although Mammography Quality Standards Act inspections suggest that doses for a digital mammography view decreased 2.5% between 2007 and 2009 (29), these doses were measured with phantoms simulating breasts with a compressed breast thickness at the 30th percentile in DMIST. Radiation dose is highly correlated with compressed breast thickness, which may increase over time with increasing population body mass index (BMI) (30).

The use of digital breast tomosynthesis for screening is increasing in the United States (31). Doses from breast tomosynthesis vary by the strategy; however, the 3-dimensional acquisition results in a radiation dose similar to or slightly greater than standard digital mammography (28, 32, 33). Most U.S. practices offering screening tomosynthesis combine it with digital mammography, which at least doubles doses and the risk for radiation-induced breast cancer. Software approved by the U.S. Food and Drug Administration to generate synthetic 2-dimensional views from tomosynthesis acquisitions will probably eliminate the need for standard digital mammography views and their associated radiation exposure (34); however, the rate at which this software will diffuse into clinical practice is unknown. Estimating radiation-induced cancer risks associated with tomosynthesis screening is further complicated by the expectation that this method will decrease recall rates and potentially eliminate the need for diagnostic mammography to work up some imaging findings (35-41).

Our study had several limitations. We had inadequate information on the percentage of women requiring more than 4 views for complete breast examination. In DMIST, 21% of women required more than 4 screening views (10), although most received only 1 or 2 extra views, probably because of patient movement or poor positioning. On the basis of the observed distribution of compressed breast thickness and number of views, we assumed that 8% of women received extra views because they had large breasts. Of note, the early generation mammography systems used in DMIST had smaller image detectors (10). Most modern units have larger detectors; therefore, the percentage of women requiring extra views because of large breast size is probably less than 8%.

We could not calculate life-years lost due to radiation-induced breast cancer, which may occur later in life than deaths prevented from screening. Because of lack of data, we did not model the association between breast size and the probability of a false-positive result; thus, we may have underestimated exposure from additional work-up in women with large breasts because obese women may be 20% more likely than normal-weight women to have false-positive results (9). We also assumed that the number of breast cancer deaths averted with screening did not vary by breast size; however, screening may prevent more deaths among postmenopausal obese women (who tend to have large breasts) because they have a greater risk for advanced disease (42). In addition, we did not model the association between breast density and radiation dose per view because of lack of representative data. Probabilities for events after screening mammography were based on point estimates from models that used the best available data and did not account for uncertainty due to model misspecification or inherent variability in parameter estimates. We could not estimate 95% CIs for deaths averted with screening because of the computational complexity of the MISCAN-Fadia model and because many input parameters of the model (such as tumor growth rate) are unobservable and

therefore have unknown distributions. We also made several simplifying assumptions (supplementary material).

In conclusion, population projections of radiation induced breast cancer incidence and mortality from mammography screening are affected by variability in doses from screening and resultant diagnostic examinations, age at screening initiation, and screening frequency. Our study suggests that women with large breasts or breast augmentation receive greater radiation doses and may have a greater risk for radiation induced breast cancer and breast cancer death. Radiology practices should strive to ensure that large breasts are imaged with large detectors with the fewest number of views possible.

Acknowledgments

The authors thank Benjamin Herman, PhD, at the American College of Radiology Imaging Network coordinating center for providing the DMIST data and helping with data interpretation; John Boone, PhD, Professor and Vice Chair of Radiology and Professor of Biomedical Engineering at the University of California, Davis, Medical Center, for helpful input and suggestions on our modeling strategy; Chris Tachibana, PhD, from the Group Health Research Institute for scientific editing; and an anonymous reviewer from the American College of Radiology Imaging Network for his/her comments on an earlier draft.

Disclaimer: The authors are responsible for study design, analysis and interpretation of the data, writing of the manuscript, and the decision to submit the manuscript for publication. The findings and conclusions in this article are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the funding sources.

Supplemental Table 1 Comparison of lifetime attributable risks of radiation-induced breast cancer and breast cancer death (per 100,000 women) from two modeling approaches

Strategy	MISCAN-Fadia	Radiation Exposure Model
	Mean (95% CI)	Mean (95% CI)
	Lifetime Attributable Risk of Breast Cancer (Per 100,000 Women)	
Biennial screening		
Ages 50-74 y	28 (20, 40)	27 (19, 38)
Ages 45-74 y	44 (31, 62)	45 (31, 64)
Ages 40-74 y	67 (47, 96)	68 (48, 97)
Hybrid strategy		
A45-49 y, B50-74 y	57 (40, 81)	59 (41, 84)
A40-49 y, B50-74 y	101 (71, 143)	89 (62, 126)
Annual screening		
Ages 50-74 y	54 (39, 75)	49 (34, 69)
Ages 45-74 y	85 (59, 121)	81 (57, 115)
Ages 40-74 y	129 (90, 183)	125 (88, 178)
	Lifetime Attributable Risk of Breast Cancer Death (Per 100,000 Women)	
Biennial screening		
Ages 50-74 y	5 (3, 7)	4 (3, 6)
Ages 45-74 y	8 (5, 11)	8 (5, 11)
Ages 40-74 y	12 (8, 17)	12 (8, 17)
Hybrid strategy		
A45-49 y, B50-74 y	10 (7, 14)	10 (7, 14)
A40-49 y, B50-74 y	18 (13, 25)	15 (11, 22)
Annual screening		
Ages 50-74 y	7 (5, 10)	7 (5, 9)
Ages 45-74 y	11 (8, 16)	11 (8, 15)
Ages 40-74 y	16 (12, 23)	16 (11, 23)

CI, confidence interval; y, years; A, annual screening at ages 40-50 or 45-50 and B, biennial screening at 50-74 years.

REFERENCES

1. Preston DL, Mattsson A, Holmberg E, Shore R, Hildreth NG, Boice JD, Jr. Radiation effects on breast cancer risk: a pooled analysis of eight cohorts. Radiat Res. 2002;158(2):220-35. Epub 2002/07/11. PubMed PMID: 12105993.
2. Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation and National Research Council. Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2. Washington, D.C.: The National Academies Press; 2006.
3. Hendrick RE. Radiation doses and cancer risks from breast imaging studies. Radiology. 2010;257(1):246-53. Epub 2010/08/26. doi: 10.1148/radiol.10100570. PubMed PMID: 20736332.
4. Yaffe MJ, Mainprize JG. Risk of Radiation-induced Breast Cancer from Mammographic Screening. Radiology. 2010;258(1):98-105. Epub 2010/11/18. doi: radiol.10100655 [pii] 10.1148/radiol.10100655. PubMed PMID: 21081671.
5. Feig SA, Hendrick RE. Radiation risk from screening mammography of women aged 40-49 years. J Natl Cancer Inst Monogr. 1997(22):119-24. Epub 1997/01/01. PubMed PMID: 9709287.
6. Harris RP, Sheridan SL, Lewis CL, Barclay C, Vu MB, Kistler CE, et al. The harms of screening: a proposed taxonomy and application to lung cancer screening. JAMA internal medicine. 2014;174(2):281-5. doi: 10.1001/jamainternmed.2013.12745. PubMed PMID: 24322781.
7. Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Ann Intern Med. 2011;155(8):481-92. Epub 2011/10/19. doi: 10.7326/0003-4819-155-8-201110180-00004. PubMed PMID: 22007042; PubMed Central PMCID: PMC3209800.
8. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med. 2003;138(3):168-75. PubMed PMID: 12558355.
9. Elmore JG, Carney PA, Abraham LA, Barlow WE, Egger JR, Fosse JS, et al. The association between obesity and screening mammography accuracy. Arch Intern Med. 2004;164(10):1140-7. PubMed PMID: 15159273.
10. Hendrick RE, Pisano ED, Averbukh A, Moran C, Berns EA, Yaffe MJ, et al. Comparison of acquisition parameters and breast dose in digital mammography and screen-film mammography in the American College of Radiology Imaging Network digital mammographic imaging screening trial. AJR Am J Roentgenol. 2010;194(2):362-9. Epub 2010/01/23. doi: 10.2214/ajr.08.2114. PubMed PMID: 20093597.
11. Wells CL, Slanetz PJ, Rosen MP. Mismatch in breast and detector size during screening and diagnostic mammography results in increased patient radiation dose. Acad Radiol. 2014;21(1):99-103. doi: 10.1016/j.acra.2013.10.005. PubMed PMID: 24331271.
12. American College of Radiology. American College of Radiology Breast Imaging Reporting and Data System Atlas (BI-RADS® Atlas). Reston, VA: American College of Radiology; 2013.
13. Ballard-Barbash R, Taplin SH, Yankaskas BC, Ernster VL, Rosenberg RD, Carney PA, et al. Breast Cancer Surveillance

Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol. 1997;169(4):1001-8. PubMed PMID: 9308451.

14. Miglioretti D, Zhu W, Kerlikowske K, Sprague BL, Onega T, Buist DSM, et al. Risk of less-favorable breast tumor characteristics with biennial versus annual mammography. JAMA Oncology. In Press.
15. Buist DS, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40-49 years. J Natl Cancer Inst. 2004;96(19):1432-40. PubMed PMID: 15467032.
16. Tabar L, Faberberg G, Day NE, Holmberg L. What is the optimum interval between mammographic screening examinations? An analysis based on the latest results of the Swedish two-county breast cancer screening trial. Br J Cancer. 1987;55(5):547-51. Epub 1987/05/01. PubMed PMID: 3606947.
17. Tabar L, Fagerberg G, Chen HH, Duffy SW, Gad A. Tumour development, histology and grade of breast cancers: prognosis and progression. Int J Cancer. 1996;66(4):413-9. Epub 1996/05/16. doi: 10.1002/(sici)1097-0215(19960516)66:4<413::aid-ijc1>3.0.co;2-z. PubMed PMID: 8635853.
18. Tan SY, van Oortmarssen GJ, de Koning HJ, Boer R, Habbema JD. The MISCAN-Fadia continuous tumor growth model for breast cancer. J Natl Cancer Inst Monogr. 2006(36):56-65. Epub 2006/10/13. doi: 10.1093/jncimonographs/lgj009. PubMed PMID: 17032895.
19. Mandelblatt J, Cronin K, de Koning HJ, Miglioretti DL, Schechter C, Stout NC. Collaborative Modeling of U.S. Breast Cancer Screening Strategies. Rockville, MD: Agency for Healthcare Research and Quality: AHRQ Publication No. 14-05201-EF-4.; 2015 [updated 4/2015 Accessed 9/2/2015]. Available from: http://www.uspreventiveservicestaskforce.org/Page/Document/modeling-report-collaborative-modeling-of-us-breast-cancer-1/breast-cancer-screening1.
20. Sprague BL, Gangnon RE, Burt V, Trentham-Dietz A, Hampton JM, Wellman RD, et al. Prevalence of mammographically dense breasts in the United States. J Natl Cancer Inst. 2014;106(10). Epub 2014/09/14. doi: 10.1093/jnci/dju255. PubMed PMID: 25217577.
21. Berrington de Gonzalez A, Iulian Apostoaei A, Veiga LH, Rajaraman P, Thomas BA, Owen Hoffman F, et al. RadRAT: a radiation risk assessment tool for lifetime cancer risk projection. J Radiol Prot. 2012;32(3):205-22. doi: 10.1088/0952-4746/32/3/205. PubMed PMID: 22810503.
22. Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, et al. Comparison of screening mammography in the United States and the United kingdom. JAMA. 2003;290(16):2129-37.
23. Jacobsen KK, Abraham L, Buist DS, Hubbard RA, O'Meara ES, Sprague BL, et al. Comparison of cumulative false-positive risk of screening mammography in the United States and Denmark. Cancer epidemiology. 2015;39(4):656-63. Epub 2015/05/28. doi: 10.1016/j.canep.2015.05.004. PubMed PMID: 26013768.
24. Elmore JG, Nakano CY, Koepsell TD, Desnick LM, D'Orsi CJ, Ransohoff DF. International variation in screening mammography interpretations in community-based programs. J Natl Cancer Inst. 2003;95(18):1384-93. Epub 2003/09/18. PubMed PMID: 13130114; PubMed Central PMCID: PMCPmc3146363.
25. Hofvind S, Vacek PM, Skelly J, Weaver DL, Geller BM. Comparing screening

mammography for early breast cancer detection in Vermont and Norway. J Natl Cancer Inst. 2008;100(15):1082-91. Epub 2008/07/31. doi: 10.1093/jnci/djn224. PubMed PMID: 18664650; PubMed Central PMCID: PMCPmc2720695.

26. American Cancer Society. Cancer Facts & Figures 2015 Atlanta: American Cancer Society; 2015 [updated 2015 Accessed 9/2/2015]. Available from: http://www.cancer.org/Research/CancerFacts-Figures/index.
27. Yaffe MJ, Boone JM, Packard N, Alonzo-Proulx O, Huang SY, Peressotti CL, et al. The myth of the 50-50 breast. Med Phys. 2009;36(12):5437-43. Epub 2010/01/26. PubMed PMID: 20095256.
28. Olgar T, Kahn T, Gosch D. Average glandular dose in digital mammography and breast tomosynthesis. ROFO Fortschr Geb Rontgenstr Nuklearmed. 2012;184(10):911-8. PubMed PMID: 22711250.
29. FDA. Trends in Mammography Dose and Image Quality 1974-2009 2015 [updated 10/30/2014 Accessed 05/04/2015]. Available from: http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/FacilityScorecard/ucm326264.htm.
30. Robinson M, Kotre CJ. Trends in compressed breast thickness and radiation dose in breast screening mammography. Br J Radiol. 2008;81(963):214-8. Epub 2008/02/14. doi: 10.1259/bjr/90916004. PubMed PMID: 18270295.
31. Hardesty LA, Kreidler SM, Glueck DH. Digital breast tomosynthesis utilization in the United States: a survey of physician members of the Society of Breast Imaging. J Am Coll Radiol. 2014;11(6):594-9. Epub 2014/04/10. doi: 10.1016/j.jacr.2013.11.025. PubMed PMID: 24713501.
32. Svahn TM, Houssami N, Sechopoulos I, Mattsson S. Review of radiation dose estimates in digital breast tomosynthesis relative to those in two-view full-field digital mammography. Breast. 2015;24(2):93-9. Epub 2015/01/03. doi: 10.1016/j.breast.2014.12.002. PubMed PMID: 25554018.
33. Feng SS, Sechopoulos I. Clinical digital breast tomosynthesis system: dosimetric characterization. Radiology. 2012;263(1):35-42. Epub 2012/02/15. doi: 10.1148/radiol.11111789. PubMed PMID: 22332070.
34. Lee CI, Lehman CD. Digital breast tomosynthesis and the challenges of implementing an emerging breast cancer screening technology into clinical practice. J Am Coll Radiol. 2013;10(12):913-7. doi: 10.1016/j.jacr.2013.09.010. PubMed PMID: 24295940.
35. McCarthy AM, Kontos D, Synnestvedt M, Tan KS, Heitjan DF, Schnall M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. J Natl Cancer Inst. 2014;106(11). doi: 10.1093/jnci/dju316. PubMed PMID: 25313245.
36. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R, Jr. Implementation of breast tomosynthesis in a routine screening practice: an observational study. AJR Am J Roentgenol. 2013;200(6):1401-8. Epub 2013/05/25. doi: 10.2214/ajr.12.9672. PubMed PMID: 23701081.
37. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA. 2014;311(24):2499-507. doi: 10.1001/jama.2014.6095. PubMed PMID: 25058084.
38. Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, et al. Comparison of Digital Mammography Alone and Digital Mammography Plus Tomosyn-

thesis in a Population-based Screening Program. Radiology. 2013;267(1):47-56. doi: 10.1148/radiol.12121373. PubMed PMID: 23297332.

39. Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol. 2013;14(7):583-9. Epub 2013/04/30. doi: 10.1016/s1470-2045(13)70134-7. PubMed PMID: 23623721.
40. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. Radiology. 2013;269(3):694-700. Epub 2013/08/01. doi: 10.1148/radiol.13130307. PubMed PMID: 23901124.
41. Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. AJR Am J Roentgenol. 2014;203(3):687-93. Epub 2014/06/12. doi: 10.2214/ajr.14.12642. PubMed PMID: 24918774.
42. Kerlikowske K, Walker R, Miglioretti DL, Desai A, Ballard-Barbash R, Buist DS. Obesity, mammography use and accuracy, and advanced breast cancer risk. J Natl Cancer Inst. 2008;100(23):1724-33. doi: 10.1093/jnci/djn388. PubMed PMID: 19033562.

Supplement. Supplemental Material

Radiation Exposure Model

For each screening strategy, we simulated screening-related events for 100,000 women from starting age through 74. For each woman, we:

1. Randomly sampled breast density, compressed breast thickness, and a woman-specific random effect for false-positive mammogram probabilities. Determined breast size from compressed breast thickness.

2. Randomly sampled screening results and resulting diagnostic events, conditional on age, breast density, current screening interval, prior screening results, and the woman-specific random effect.

3. Randomly sampled number of views per screening examination and, if recalled, diagnostic events, conditional on breast size, and randomly sampled breast dose per view conditional on compressed breast thickness.

4. Summed the number of mammographic views across events and calculated total dose based on sampled dose per view for each year of age.

5. Estimated radiation-induced breast cancer incidence and mortality through age 100 or death based on total dose at each age.

Data sources

Data were from the Breast Cancer Surveillance Consortium (BCSC) and the American College of Radiology Imaging Network (ACRIN) digital mammographic imaging screening trial (DMIST). The BCSC (13) (http://breastscreening.cancer.gov) has prospectively collected data including patient characteristics and radiology information from community-based facilities since 1994. Characteristics of women are comparable to the US population (43). Breast cancer diagnoses and tumor characteristics are obtained by linking to pathology databases; regional Surveillance, Epidemiology, and End Results (SEER) programs; and state tumor registries.

ACRIN DMIST was powered to compare the screening accuracy of digital and screen-film mammography (44, 45). For this paired trial, 49,528 women provided informed consent to receive both modalities between October 2001 and November 2003. For quality assurance, compressed breast thickness, breast dose, and number of additional views performed were recorded on a subset of examinations at 33 sites. The ACRIN coordinating center provided the distribution for number of views for 5,021 digital examinations and the joint distribution between dose and compressed breast thickness for 19,205 digital mammography views from 4,876 digital examinations.

Breast Size, Compressed Breast Thickness, And Number Of Views Per Examination

We estimated the percentage of women with large breasts based on the number of views and compressed breast thickness observed in DMIST (**Appendix Tables 1 and 3**). Based on expert opinion, we assumed all women with 5 views and a portion of women with 6 views received these extra views due to issues with positioning or movement. In contrast, we assumed a portion of women with 6 views and all women with 7 or more views received extra views because they had large breasts. To estimate the percentage of women with large breasts, we chose a threshold of compressed breast thickness 7.5 cm or larger, consistent with the percentage of women having 6 or more views. This resulted in 8.1% of women having large breasts and 35% of examinations with 6 views being performed in women with large breasts.

DMIST has information only on number of views for screening examinations. For diagnostic examinations and procedure types, we obtained the typical number of views from expert opinion of a radiologist who specializes in breast imaging and scaled the distribution for screening examinations from DMIST based on the typical number of views for that diagnostic exam or procedure type relative to the typical number of screening views. For numbers of rescaled views that were not integers (e.g., 5 views /4 views = 1.3 views), we reassigned women into adjacent groups so the resultant mean number of views was unchanged (e.g., for 1.3 views, we assumed 70% received 1 view and 30% received 2 views). For example, the typical number of screening views is 4 (2 per breast). From DMIST, we estimated that 86% of women without large breasts received 4 views, 9% received 5 views, and 5% received 6 views. Typically, two magnification views are used for an additional evaluation of a positive screening mammogram. Thus, to calculate the distribution of diagnostic views, we halved the number of views from the screening distribution. This resulted in 86% of women receiving 2 views, 9% receiving 2.5 views, and 5% receiving 3 views. We reassigned the 9% of women with 2.5 views to half receiving 2 and half receiving 3. This gave a final distribution for number of magnification views of 91% receiving 2 views and 9% receiving 3 views. Distributions are in **Appendix Table 4**.

Breast density

We assigned a baseline Breast Imaging-Reporting and Data System (BI-RADS) (12) density at the start of screening according to distributions observed in the BCSC (20) (**Appendix Table 2**). At age 50 and 65 years, we allowed breast density to potentially decrease by one category based on transition probabilities that maintain the marginal distributions of density by age (**Appendix Table 2**). We did not account for the inverse relationship between breast density and breast size due to lack of information on the association between event probabilities (i.e., short interval follow-up (SIFU) examinations) and breast size.

Radiation Dose

Radiation dose depends on compressed breast thickness, which depends on breast size. For each woman, we sampled a dose per view based on the distribution observed in DMIST given her compressed breast thickness (**Appendix Figure**). Magnification views have higher radiation dose than standard mammography views (46, 47); however, only part of the breast is typically irradiated (48). Therefore, we assumed the same dose for all views. This assumption is supported by data from Boone, Nosratieh, and Seibert in the 2013 Society for Breast Imaging newsletter (http://www.sbi-online.org/NEWS.aspx). For women with large breasts who receive extra views, most glandular tissue is irradiated on all views; therefore, summing the doses per view for an exam-specific dose for each breast was reasonable (10, 11). We assumed the total bilateral dose per view was half the dose per single breast, as in Law and Faulkner (48). Thus, to calculate the total bilateral dose at each year of age, we summed the total number of views on both breasts from screening and associated diagnostic work-up within the following year, and divided in half.

Events following A Positive Screening exam

Figure 2 in manuscript summarizes possible events following a screening mammogram (12). At each screening mammogram, a woman's probability of recall for additional imaging was based on age, breast density, screening interval, prior screening mammogram results, and a woman-specific random effect. If recalled, the probability of referral to biopsy, short interval follow-up (SIFU), or return to routine screening was based on age, breast density, and screening interval. Following BI-RADS guidelines (12), women recommended for SIFU received diagnostic views at 6, 12, and 24 months after screening mammogram, regardless of screening interval. At each SIFU exam, the probability of a biopsy recommendation was based on age and breast density. Women with a SIFU recommendation also continued to receive bilateral screening views according to their screening schedule with recall and subsequent follow-up recommendations assigned using the probabilities for all BCSC screening exams. A woman assigned to SIFU following recall from screening views restarted the SIFU sequence; otherwise, she continued according to her assigned SIFU schedule. Biopsy type was randomly assigned based on BCSC distributions to be fine needle aspiration, core biopsy, excisional biopsy, or core and excisional biopsy, based on age and breast density. Fine needle aspirations resulted in no additional mammography. For core and excisional biopsies, we randomly assigned ultrasound or stereotactic guidance based on proportions observed in BCSC. Women returned to routine screening following a benign biopsy based on suggestions that 6-month follow-up imaging after biopsy has no benefit (49).

We modeled the probability of events following a screening mammogram using BCSC data. We included digital mammograms of women

aged 40-74 without a history of breast cancer or cancer diagnosed within 1 year after the exam and without breast augmentation. Most analyses included mammograms conducted from 2003 to 2011 with at least one year of complete cancer capture available following the screening exam. Mammograms were classified as screening or SIFU based on the indication given by the radiologist or technologist. For screening mammograms, we excluded unilateral exams and exams performed less than 9 months after a prior mammogram or breast ultrasound exam to avoid misclassifying diagnostic exams as screens. Screening mammograms were classified as annual exams if the previous mammogram was 9–18 months prior and as biennial if 19–30 months prior. We excluded screening mammograms conducted more than 30 months after a prior mammogram because we were interested in estimating events in annual and biennial screeners.

To estimate the recall rates for additional imaging, we included 613,797 digital screening mammograms with sufficient information on prior false-positive results. We defined a recall based on a positive initial BI-RADS assessment (12, 50). We estimated the probability of being recalled for additional views on a screening mammogram using logistic regression including age at exam; BI-RADS breast density; mammogram number (first, second, or third or more); and screening interval for subsequent screens (annual vs. biennial). For second exams, we also included the prior screening result and for third or subsequent exams, we included the prior two screening results. We included a woman-specific random effect to allow for additional correlation of recall across a woman's entire screening regimen and report results for a random effect of 0, corresponding to median rates. Results are in **Supplement Table 1**. To evaluate the model fit, we compared results to prior estimates of cumulative false-positive rates after 10 rounds of screening using a different method (51) and got similar results.

To estimate the probability of events following an abnormal mammogram, we included 725,433 digital mammograms with information on the specific type of recommendation at the end of all imaging work-up. We estimated the probability of recommended follow-up (either return to routine screening, SIFU, or biopsy) after recalled screening mammogram using multinomial logistic regression including age at exam, BI-RADS breast density, and screening interval as predictors. Results are in **Supplement Table 2**.

We estimated the probability of a biopsy recommendation following a SIFU exam using 21,124 SIFU exams. We classified exams as having a biopsy recommendation based on the final BI-RADS assessment and recommendations at the end of all imaging work-up. We fit a logistic regression model including BI-RADS density and age at exam as predictors. Results are in **Supplement Table 3**.

To estimate the distribution for type of biopsy following a positive screening mammogram or SIFU exam, we included 2,284 women with biopsies within 100 days following a positive screening or SIFU conducted in the most recent years available (2010 and 2011), because use of core biopsy instead of excisional biopsy has increased over time. We grouped biopsy events as core biopsy (no excisional, may also include fine

needle aspirations), excisional biopsy (no core biopsy, may also include fine needle aspirations), core and excisional (may also include fine needle aspirations), and fine needle aspiration only. We modeled biopsy type using multinomial logistic regression, including BI-RADS density and age at mammogram as predictors. To estimate the type of biopsy guidance distributions, we selected all biopsies within 100 days of a positive final assessment of a screening or SIFU mammogram. Given inconsistencies in excisional biopsy guidance records, we limited our data to core biopsies only and calculated the proportion of ultrasound and stereotactic biopsies in this sample. Results are in **Supplement Table 4**.

Simplifying Assumptions

Radiation dose depends on the mammography machine used (10), but we could not include this factor in our modeling due to lack of data. Estimates of the U.S. distribution of manufacturers are protected market share information. However, the majority of digital machines used by BCSC facilities are Hologic, which had the highest dose per view but the fewest exams with more than 4 views in DMIST (10). The majority of machines used in DMIST were not Hologic. Thus, if the BCSC is reflective of the U.S., we would likely have underestimated dose for women with small or average breasts but may have overestimated dose for women with large breasts because they would be less likely to need extra views for complete breast examination. We may have slightly underestimated dose due to diagnostic imaging for several reasons. Our estimates of the number of views used for diagnostic evaluations may be conservative, because we assumed that every abnormal screening examination identified only one finding needing diagnostic views and we did not include repeat whole breast views. Also, the chance of repeat images is likely higher for diagnostic spot magnification views for subtle calcifications or masses, and for large-breasted women. In these instances, the technologist may require several images to position small or subtle findings within the field of view. Moreover, magnification spot views require greater exposure time for optimal image resolution, making patient movement more likely.

Radiation-Induced Breast Cancer Incidence And Mortality

The incidence of radiation-induced breast cancer was modeled using the excess absolute risk model from pooled analysis of four cohorts by Preston et al.(1), the preferred model for estimating radiation-induced breast cancers (2, 21). The model formula from page 234 of Preston et al. (1) is

$$\beta D exp(a/50)^{\eta}$$

where β is the risk coefficient per 10,000 person years-Gy, estimated as 10 with 95% confidence interval (CI) 7.0–14.2. D is dose in Gy, e is age at exposure; a is attained age; and η is 3.5 for $a \leq 50$ and 1.0 otherwise. Similar to Berrington et al. (21), we modeled the latency period for

developing radiation-induced breast cancer using a logistic function with shape parameter 0.75, which phases in increased breast cancer risk between 4 and 11 years after exposure. We did not apply a dose and dose-rate effectiveness factor (2) because the Preston 2002 model (1) included data from two cohorts with radiation exposures from high-dose-rate X-rays similar to those used for mammography screening. Also, Preston et al. (1) found no evidence that fractionated exposures result in lower breast cancer risk than acute exposures. We adjusted for competing causes of death using US general population life tables for women (52). Radiation-induced breast cancer mortality was estimated by multiplying radiation-induced breast cancer incidence by the non-radiation induced breast cancer age-specific case-fatality rates derived from MISCAN-Fadia assuming 100% adherence to screening and current treatment. We assumed that breast cancers induced by radiation are screen detected at the same rate as non-induced cancers. Uncertainty ranges were estimated by re-estimating radiation-induced breast cancer risk using the upper and lower 95% CIs for the risk coefficient, β, given this uncertainty dominates the uncertainty in estimated risk (2, 21).

Supplemental Results

From the radiation exposure model simulation results, women who obtained screening annually from ages 40-74 years received an average of 5.0 mammography views (5th percentile=4 views, 95th percentile=9 views) and a dose of 4.8 mGy (5th percentile=2.3 views, 95th percentile=10.7 mGy) from each screening exam and all diagnostic work-up prompted by that screen within a 1-year period (**Appendix Table 5**). The mean dose from screening views was 4.3 mGy (5th percentile=2.2 views, 95th percentile=8.3 mGy), and the mean dose from all diagnostic work-up among women with a false-positive screen was 4.5 mGy (5th percentile=1.7 views, 95th percentile=10.7 mGy). Women with large breasts undergoing annual screening received a mean of 8.4 views (5th percentile=6.0 views, 95th percentile=14.0) and mean dose of 10.0 mGy (5th percentile=4.6 views, 95th percentile=20.8 mGy) from each screening exam plus all diagnostic work-up prompted by that screen, compared to 4.7 views (5th percentile=4 views, 95th percentile=8 views) and 4.3 mGy (5th percentile=2.2 mGy, 95th percentile=8.4 mGy) for women without large breasts.

Supplement Table 1. **Probability of a false-positive recall (median and interquartile range) by age, BI-RADS breast density, screening round, and prior screening results among women aged 40-74 years with digital mammography from 2003-2011 and no cancer diagnosis within 1-year follow-up period, estimated from the Breast Cancer Surveillance Consortium.**

		Screening Schedule = Annual				Screening Schedule = Biennial			
Screening round and prior screening results	**Age, years**	**Almost entirely fat**	**Scattered fibro. densities**	**Hetero. dense**	**Extremely dense**	**Almost entirely fat**	**Scattered fibro. densities**	**Hetero. dense**	**Extremely dense**
		Probability of False-Positive Screening Mammogram, Median (Interquartile Range)							
Round 1	**40-44**	13 (8,20)%	19 (12,29)%	23 (15,35)%	19 (12,28)%	13 (8,20)%	19 (12,29)%	23 (15,35)%	19 (12,28)%
	45-49	18 (11,28)%	27 (18,39)%	32 (21,45)%	26 (17,38)%	18 (11,28)%	27 (18,39)%	32 (21,45)%	26 (17,38)%
	50-54	15 (9,23)%	22 (14,33)%	27 (17,39)%	21 (14,32)%	15 (9,23)%	22 (14,33)%	27 (17,39)%	21 (14,32)%
Round 2									
No prior FP	**40-49**	7 (4,11)%	11 (6,17)%	13 (8,21)%	10 (6,16)%	7 (4,12)%	12 (7,18)%	14 (9,22)%	11 (7,18)%
	50-59	5 (3,9)%	9 (5,14)%	11 (7,17)%	8 (5,14)%	6 (3,10)%	9 (6,15)%	12 (7,19)%	9 (5,15)%
	60-74	5 (3,8)%	8 (5,13)%	10 (6,16)%	7 (4,12)%	5 (3,9)%	8 (5,14)%	10 (6,17)%	8 (5,13)%
Prior FP	**40-49**	6 (4,10)%	10 (6,16)%	12 (7,19)%	9 (6,15)%	7 (4,11)%	11 (6,17)%	13 (8,21)%	10 (6,16)%
	50-59	5 (3,8)%	8 (5,13)%	10 (6,16)%	8 (5,12)%	5 (3,9)%	9 (5,14)%	11 (6,17)%	8 (5,13)%
	60-74	4 (3,7)%	7 (4,11)%	9 (5,14)%	7 (4,11)%	5 (3,8)%	8 (4,12)%	9 (6,15)%	7 (4,12)%
Round 3+									
Past two results TNs	**40-49**	5 (3,8)%	8 (5,13)%	10 (6,16)%	7 (4,12)%	5 (3,9)%	8 (5,14)%	10 (6,17)%	8 (5,13)%
	50-59	4 (2,6)%	6 (4,10)%	8 (5,13)%	6 (3,10)%	4 (2,7)%	7 (4,11)%	8 (5,14)%	6 (4,11)%
	60-74	3 (2,6)%	5 (3,9)%	7 (4,11)%	5 (3,9)%	4 (2,6)%	6 (3,10)%	7 (4,12)%	6 (3,9)%
Past two results TN then FP	**40-49**	6 (4,10)%	10 (6,16)%	12 (7,20)%	9 (6,15)%	7 (4,11)%	11 (6,17)%	13 (8,21)%	10 (6,16)%
	50-59	5 (3,8)%	8 (5,13)%	10 (6,16)%	8 (5,13)%	5 (3,9)%	9 (5,14)%	11 (7,17)%	8 (5,14)%
	60-74	4 (3,7)%	7 (4,12)%	9 (5,14)%	7 (4,11)%	5 (3,8)%	8 (5,13)%	10 (6,16)%	7 (4,12)%
Past two results FP then TN	**40-49**	6 (4,10)%	10 (6,16)%	12 (7,20)%	9 (6,15)%	7 (4,11)%	11 (6,17)%	13 (8,21)%	10 (6,16)%
	50-59	5 (3,8)%	8 (5,13)%	10 (6,16)%	8 (5,13)%	5 (3,9)%	9 (5,14)%	11 (7,17)%	8 (5,13)%
	60-74	4 (3,7)%	7 (4,12)%	9 (5,14)%	7 (4,11)%	5 (3,8)%	8 (5,12)%	10 (6,15)%	7 (4,12)%
Past two results FPs	**40-49**	18 (11,27)%	26 (17,38)%	31 (21,44)%	25 (16,37)%	19 (12,29)%	28 (18,40)%	33 (22,46)%	27 (18,39)%
	50-59	14 (9,23)%	22 (14,33)%	27 (17,39)%	21 (13,32)%	16 (10,24)%	24 (15,35)%	28 (19,41)%	23 (14,34)%
	60-74	13 (8,20)%	20 (12,30)%	24 (15,35)%	19 (12,29)%	14 (8,22)%	21 (13,32)%	26 (17,37)%	20 (13,31)%

BI-RADS = Breast Imaging Reporting and Data Systems; FP, false positive; Fibro = fibroglandular**;** Hetero = heterogeneously

Estimates are based on a mixed effects logistic regression model, and the interquartile range reflects heterogeneity among women based on quartiles of the woman-specific random effect distribution.

Supplement Table 2. **Probability of subsequent events given a false-positive recall, by age and BI-RADS breast density, among women aged 40-74 years with digital mammography from 2003-2011 and no cancer diagnosis within 1-year follow-up period, estimated from the Breast Cancer Surveillance Consortium.**

		Screening Schedule = Annual				Screening Schedule = Biennial			
	Age, years	Almost entirely fat	Scattered fibro. densities	Hetero. dense	Extremely dense	Almost entirely fat	Scattered fibro. densities	Hetero. dense	Extremely dense
Recommendation after false-positive screening mammogram									
Return to normal interval follow-up									
Round 1	**40-44**	52%	56%	55%	50%	52%	56%	55%	50%
	45-49	43%	47%	46%	41%	43%	47%	46%	41%
	50-54	46%	50%	49%	43%	46%	50%	49%	43%
Round 2+	**40-49**	69%	72%	72%	67%	64%	68%	67%	62%
	50-59	66%	70%	69%	64%	61%	65%	64%	58%
	60-74	66%	70%	69%	64%	61%	65%	64%	59%
Short interval follow-up									
Round 1	**40-44**	32%	30%	29%	29%	32%	30%	29%	29%
	45-49	38%	36%	34%	34%	38%	36%	34%	34%
	50-54	29%	28%	26%	26%	29%	28%	26%	26%
Round 2+	**40-49**	20%	19%	18%	19%	23%	21%	20%	21%
	50-59	21%	20%	19%	20%	24%	22%	21%	22%
	60-74	22%	20%	19%	20%	24%	22%	21%	22%
Biopsy									
Round 1	**40-49**	16%	14%	16%	21%	16%	14%	16%	21%
	50-59	19%	17%	20%	25%	19%	17%	20%	25%
	60-74	25%	22%	25%	31%	25%	22%	25%	31%
Round 2+	**40-49**	11%	9%	10%	14%	13%	11%	13%	17%
	50-59	12%	11%	12%	16%	15%	13%	15%	20%
	60-74	12%	10%	12%	16%	15%	13%	15%	19%

BI-RADS = Breast Imaging Reporting and Data Systems; FP = false positive; Fibro = fibroglandular; Hetero = heterogeneously
Due to rounding, some percentages may not add to 100%.

***Supplement Table 3*. Probability of a biopsy at a short-interval follow-up (SIFU) exam among women aged 40-74 years with SIFU exam from 2003-2011 and no cancer diagnosis within 1-year follow-up period, estimated from the Breast Cancer Surveillance Consortium.**

BI-RADS breast density	Age, years	Probability of biopsy at SIFU exam
Almost entirely fat	40-49	2.8%
	50-59	2.9%
	60-74	2.9%
Scattered fibro. densities	40-49	3.3%
	50-59	3.5%
	60-74	3.4%
Heterogeneously dense	40-49	5.0%
	50-59	5.2%
	60-74	5.2%
Extremely dense	40-49	6.2%
	50-59	6.6%
	60-74	6.5%

BI-RADS = Breast Imaging Reporting and Data Systems; SIFU = short interval follow-up; Fibro = fibroglandular

Supplement Table 4. **Distribution of type of biopsy, by BI-RADS breast density and age, among women aged 40-74 years with a biopsy recommendation from a digital mammography examination from 2003-2011 and no cancer diagnosis within 1-year follow-up period, estimated from the Breast Cancer Surveillance Consortium.**

		Type of biopsy (row %)			
BI-RADS Density	**Age, years**	**Core**	**Excisional**	**Core + excisional**	**Fine needle aspiration only**
Almost entirely fat	40-49	71%	3%	3%	23%
	50-59	70%	3%	3%	24%
	60-74	71%	3%	5%	21%
Scattered fibro. densities	40-49	73%	9%	5%	13%
	50-59	74%	7%	5%	14%
	60-74	73%	8%	7%	13%
Heterogeneously dense	40-49	74%	10%	6%	10%
	50-59	75%	8%	6%	11%
	60-74	73%	8%	9%	9%
Extremely dense	40-49	72%	13%	6%	9%
	50-59	73%	11%	6%	9%
	60-74	71%	12%	9%	8%

BI-RADS = Breast Imaging Reporting and Data Systems; Fibro = fibroglandular
Due to rounding, some percentages may not add to 100%.

References

43. **Sickles EA, Miglioretti DL, Ballard-Barbash R, Geller BM, Leung JW, Rosenberg RD, et al.** Performance benchmarks for diagnostic mammography. Radiology. 2005;235:775-90. [PMID: 15914475]

44. **Pisano ED, Gatsonis CA, Yaffe MJ, Hendrick RE, Tosteson AN, Fryback DG, et al.** American College of Radiology Imaging Network digital mammographic imaging screening trial: objectives and methodology. Radiology. 2005;236:404-12. [PMID: 15961755]

45. **Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al; Digital Mammographic Imaging Screening Trial (DMIST) Investigators Group.** Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005;353:1773-83. [PMID: 16169887]

46. **Law J.** Breast dose from magnification films in mammography. Br J Radiol. 2005;78:816-20. [PMID: 16110103]

47. **Koutalonis M, Delis H, Pascoal A, Spyrou G, Costaridou L, Panayiotakis G.** Can electronic zoom replace magnification in mammography? A comparative Monte Carlo study. Br J Radiol. 2010;83:569-77. [PMID: 20603409] doi:10.1259/bjr/21753020

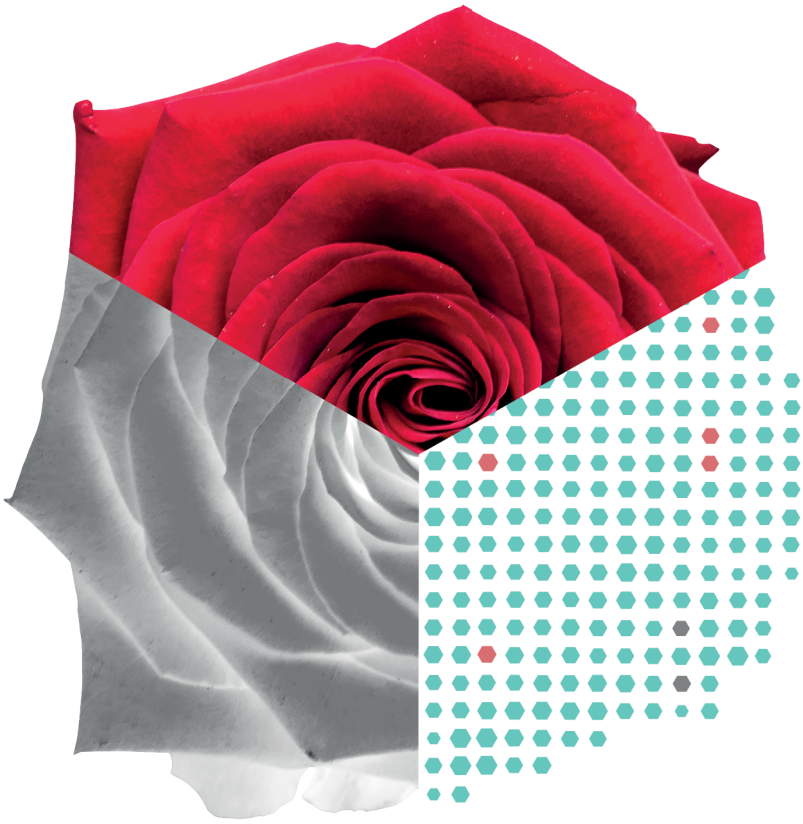
48. **Law J, Faulkner K.** Radiation benefit and risk at the assessment stage of the UK Breast Screening Programme. Br J Radiol. 2006;79:479-82. [PMID: 16714749]

49. **Johnson JM, Johnson AK, O'Meara ES, Miglioretti DL, Geller BM, Hotaling EN, et al.** Breast cancer detection with short-interval follow-up compared with return to annual screening in patients with benign stereotactic or U.S.-guided breast biopsy results. Radiology. 2015;275:54-60. [PMID: 25423143] doi:10.1148/radiol.14140036

50. **Breast Cancer Surveillance Consortium.** BCSC Glossary of Terms: BCSC; 2009. Accessed at http://breastscreening.cancer.gov/data/bcsc_data_definitions.pdf on 9 March 2015.

51. **Hubbard RA, Miglioretti DL, Smith RA.** Modelling the cumulative risk of a false-positive screening test. Stat Methods Med Res. 2010;19:429-49. [PMID: 20356857] doi:10.1177/0962280209359842

52. **Arias E, Curtin LR, Wei R, Anderson RN.** U.S. decennial life tables for 1999–2001, United States life tables. Natl Vital Stat Rep. 2008;57:1-36. [PMID: 18972722]



PART THREE: Projecting the harms and benefits of risk-based breast cancer screening in the United States

Chapter 9

Tailoring Breast Cancer Screening Intervals by Breast Density and Risk for Women Aged 50 Years or Older: Collaborative Modeling of Screening Outcomes.

Trentham-Dietz A, Kerlikowske K, Stout NK, Miglioretti DL, Schechter CB, Ergun MA, van den Broek JJ, Alagoz O, Sprague BL, van Ravesteyn NT, Near AM, Gangnon RE, Hampton JM, Chandler Y, de Koning HJ, Mandelblatt JS, Tosteson AN; Breast Cancer Surveillance Consortium and the Cancer Intervention and Surveillance Modeling Network.

Annals of Internal Medicine. 2016 Nov 15;165(10):700-712. doi: 10.7326/M16-0476.

ABSTRACT

Background

Biennial screening is generally recommended for average-risk women aged 50 to 74 years, but tailored screening may provide greater benefits.

Objective

To estimate outcomes for various screening intervals after age 50 based on breast density and risk for breast cancer.

Design

Collaborative simulation modeling using national incidence, breast density, and screening performance data.

Setting

United States

Patients

Women aged 50 years or older with various combinations of breast density and relative risk (RR) of 1.0, 1.3, 2.0, or 4.0.

Interventions

Annual, biennial, or triennial digital mammography screening from ages 50 to 74 years (vs. no screening) and ages 65 to 74 years (vs. biennial digital mammography from ages 50 to 64 years)

Measurements

Lifetime breast cancer deaths, life expectancy and quality-adjusted life-years (QALYs), false-positive mammograms, benign biopsy results, overdiagnosis, cost-effectiveness, and ratio of false-positive results to breast cancer deaths averted

Results

Screening benefits and overdiagnosis increase with breast density and RR. False-positive mammograms and benign results on biopsy decrease with increasing risk. Among women with fatty breasts or scattered fibroglandular density and an RR of 1.0 or 1.3, breast cancer deaths averted were similar for triennial versus biennial screening for both age groups (50 to 74 years, median of 3.4 to 5.1 vs. 4.1 to 6.5 deaths averted; 65 to 74 years, median of 1.5 to 2.1 vs. 1.8 to 2.6 deaths averted). Breast cancer deaths averted increased with annual versus biennial screening for women aged 50 to 74 years at all

levels of breast density and an RR of 4.0, and those aged 65 to 74 years with heterogeneously or extremely dense breasts and an RR of 4.0. However, harms were almost 2-fold higher. Triennial screening for the average-risk subgroup and annual screening for the highest-risk subgroup cost less than $100 000 per QALY gained

Limitations

Models did not consider women younger than 50 years, those with an RR less than 1, or other imaging methods.

Conclusions

Average-risk women with low breast density undergoing triennial screening and higher-risk women with high breast density receiving annual screening will maintain a similar or better balance of benefits and harms than average-risk women receiving biennial screening.

Primary Funding Source

National Cancer Institute

INTRODUCTION

Debate surrounding breast cancer screening for women in their 40s continues; however, there is a greater consensus about U.S. guidelines for average-risk women 50 or older (1, 2), with groups now recommending biennial mammography from ages 50 or 55 to 74 years (3, 4). Biennial screening is supported by clinical trials (5, 6), observational studies (5, 7), and modeling results (8). Present recommendations also acknowledge that implementing screening in clinical practice should involve shared decision making to consider preferences, risk levels, and breast density (3, 4). However, data to guide clinicians and women in making personalized decisions about screening intervals based on such factors are limited.

Observational data (7, 9) and modeling studies (10, 11) suggest that annual screening may be more effective than biennial screening for women at high risk for breast cancer due to dense breasts and other risk factors, further, triennial screening may retain most of the benefit of biennial screening but may be less harmful and more cost-effective for low-risk women with low density. However, past empirical research on alternative screening intervals did not include mortality outcomes (12). Moreover, most prior modeling studies have relied on single models (10, 11), data on film-screen mammography and older treatment regimens (10, 11, 13), and did not consider changes in breast density as women age (10), or triennial intervals (8).

To fill this gap, the Cancer Intervention and Surveillance Modeling Network (14) collaborating with the Breast Cancer Surveillance Consortium (BCSC) (a longstanding network of 6 U.S. breast imaging registries with links to tumor and pathology registries (15)), used 3 well-established models to evaluate various screening intervals for digital mammography among subgroups of women based on age, risk, and breast density. Outcomes were projected for women aged 50 (or 65) years who were deciding whether to initiate (or continue) biennial screening until age 74 years or to have annual or triennial screening. Study results are intended to inform discussions about implementing tailored breast cancer screening intervals to maximize screening benefits while minimizing harms.

METHODS

Overview of Breast Cancer Screening Strategies

The study included the following 3 microsimulation models: Model E (Erasmus Medical Center, Rotterdam, Netherlands), Model GE (Georgetown University Medical Center, Washington, DC; and Albert Einstein College of Medicine, Bronx, New York), and Model W (University of Wisconsin–Madison, Madison, Wisconsin; and Harvard Medical School,

Boston, Massachusetts). These models were either exempt from human subjects review or approved by review boards at each institution.

The models used a lifetime horizon to evaluate screening strategies for 2 populations, women aged 50 years who were starting screening for the first time and those aged 65 years who had received biennial screening from ages 50 to 64 years. We selected these populations because there is a consensus on screening women in their 50s and because at age 65 years, increases in competing mortality risks and decreases in breast density might alter the balance of benefits and harms.

Strategies for each age group varied by screening interval (annual, biennial, and triennial) and were compared with no screening. These intervals were applied to population subgroups based on combinations of the following 4 breast density levels, as defined by the American College of Radiology's Breast Imaging Reporting and Data System: almost entirely fat ("a"), scattered fibroglandular density ("b"), heterogeneously dense ("c"), or extremely dense ("d") (16)], and 4 exemplar relative risk (RR) levels, which incorporated risk factors other than breast density. These levels represent common risk factors considered alone or in combination: 1.0 (average), 1.3 (for example, postmenopausal obesity) (17-27), 2.0 (for example, history of benign breast biopsy results), and 4.0 (history of lobular carcinoma in situ) (25-29) (**Appendix Table 1**). Populations with risk suggestive of mutations in breast cancer susceptibility genes 1 and 2 were not included in these analyses.

Model Overview

The models shared common inputs but used different structures and underlying assumptions (**Appendix Table 2**) (8, 14). They started with estimates of age-specific breast cancer incidence (31) and survival trends specific to breast cancer stage, estrogen receptor (ER) status, and human epidermal growth factor receptor 2 (HER2) status (30) all without screening or adjuvant treatment. Incidence in the absence of screening was calibrated from an age–period–cohort model that accounted for changes in underlying risk (for example, secular patterns in postmenopausal hormone use) (31). Tumors had a range of preclinical periods during which they could be detected by screening (that is, sojourn times). Data on screening and ER/HER2-specific adjuvant treatment were added to generate breast cancer–specific incidence and mortality (14). Models have been validated using data from the U.K. Age trial During the preclinical detectable period, screening could result in the identification and treatment of earlier-stage or smaller tumors and lead to a reduction in breast cancer mortality reduction (**Appendix Figure 1**). All models assumed that a portion of ductal carcinoma in situ lesions was non-progressive and nonlethal; model W also considered that some types of small invasive cancer would not progress.

Model Input Parameters

The models used a common set of age-specific variables for population demographics (32), breast cancer natural history and risk (30, 31, 33-36), digital mammography (37, 38), breast density, treatment (39-41), mortality (30), costs (42, 43), and quality of life (**Table 1 and Appendix Table 2)** (14, 44-46). Each model also included parameters to represent preclinical detectable times, lead time, and age- and ER/HER2-specific stage distribution in screen- versus non–screendetected cancer based on each model's specific structure. These model-specific parameters were based on assumptions about combinations of values that reproduced U.S. trends in breast cancer incidence and breast cancer–specific mortality from 1975 to 2010 in the SEER (Surveillance, Epidemiology, and End Results) program (47). To isolate the effect of various screening strategies, all models assumed 100% adherence to screening and receipt of the most effective treatment. The population included women born in 1970 and followed until death. This birth cohort was chosen because these women experience modern conditions (for example, digital mammography performance, treatment effectiveness, and competing mortality) and for consistency with recent collaborative modeling reports (8). In each simulation, subgroups of women were followed from age 25 years until death or age 100 years. Subgroups were defined on the basis of combinations of 4 RR levels (1.0, 1.3, 2.0, and 4.0) and 4 breast density levels, with the combination of breast density levels and other factors treated multiplicatively. The risk level modified the underlying breast cancer incidence in the absence of screening. We assumed that risk level was constant over age and did not affect other model parameters. Women were assigned to either the same breast density category or the next lower category at ages 50 and 65 years based on observed age-specific prevalence in the BCSC(27, 48). Density also affected mammography performance (**Table 1 and Appendix Table 3**).

Digital mammography sensitivity and specificity were based on age, initial or subsequent screening, screening interval, and breast density using BCSC data (Table 1 and Appendix Table 3). Models GE and W used these data for calibration, and model E fit estimates from the BCSC and other sources Specificity data were used to estimate rates of false-positive mammograms. The BCSC rates of biopsy recommendations were applied to these estimates to calculate the number of benign biopsy results. Treatment effectiveness was based on clinical trials and modeled as a reduction in mortality risk (model GE) or an increase in the proportion cured (models E and W) compared with age-, stage-, and ER/HER2- specific survival in the absence of therapy (39). Women died of either breast cancer or other causes.

Screening Outcomes

Primary outcomes were lifetime benefits and harms; secondary outcomes were use of services and costs. Benefits included breast cancer deaths averted and life-years and

Table 1 Model Input Parameters

Parameter	Description	Data Source
Population Demographics		
Birth cohorts	1970 birth cohort	(32)
Natural History of Breast Cancer		
Incidence in the absence of screening	An age-period-cohort model is used as a starting point for calibration to observed SEER Program rates.	(31)
Stage distribution	Stage distribution among clinically-detected and digital screen-detected women by age group (<50, 50–64, ≥65 years), screening round (first, subsequent), and screening interval (annual, biennial, triennial).	BCSC data from 1994–2013 (digital from 2003–2013)
ER/HER2 joint distribution	Probability of ER/HER2 conditional on age and stage at diagnosis.	BCSC
Sojourn time	Sojourn time by joint ER/HER2 status and age.	(30)
Mean stage dwell time/tumor growth rates	Varies by models; can vary by age and/or ER/HER2 status.	(33-35)
Breast Cancer Screening		
Mammography use	Assume all women are screened by digital mammography.	(37, 38)
Sensitivity/ detection rates of digital screening	Sensitivity of initial and subsequent digital mammography by age group, screening interval (annual, biennial, triennial), and breast density. See Appendix Table 3.	BCSC
Specificity	False-positive mammograms are calculated as the difference between the overall number of positive mammograms in a screening scenario minus the number of positive mammograms among breast cancer cases.	BCSC
Prevalence of breast density	Prevalence of breast density (BI-RADS a, b, c, d) by age group. Density is assigned at age 40 years and can decrease by one level or remain the same at age 50 years and again at age 65 years.	BCSC
Risk levels for density	Risk of breast cancer based on BI-RADS relative to average density by age group.	BCSC
Risk levels for factors other than density	RR=1 is used at the referent for average population. RR=1.3, 2.0, and 4.0 are used as levels associated with common risk factors.	(36)
Breast Cancer Treatment		
Treatment use	Assume receipt of and adherence to the most effective available treatment specific to age, stage and ER/HER2 status.	1997–2010 (40, 41)
Treatment effects	Meta-analyses of clinical trial results.	(39)
Survival		
Breast cancer survival	26-year breast cancer survival before adjuvant treatment by joint ER/HER2 status, age group, and AJCC/SEER stage or tumor size	(30)
Non-breast cancer mortality	Age- and cohort-specific all-cause mortality rates by year.	Vanness D, Personal communication, 2015
Costs		
Screening mammogram	$138.28	Medicare reimbursement
Work-up after false-positive mammogram	Imaging costs: $141.42 (all ages). Biopsy costs by age: $1,354.05 for ages 50-64; $1,361.39 for ages 65-74; and $1,442.19 for ages 75-100. Biopsies applied to 10.6% of women screened within each age group.	(42)

Table 1 Model Input Parameters (continued)

Parameter	Description	Data Source
Work-up after true positive mammogram	By age: $2,154.58 for ages 50-64; $2,166.52 for ages 65-74; and $1,826.80 for ages 75-100.	(42)
Breast cancer treatment	By stage during initial treatment: $13,695.67 for DCIS and local stage; $25,893.77 for regional stage; and $39,990.86 for distant stage. During the last year of life among women with cancers that were not cured/progressed, depending on stage at diagnosis: $37,070.10 for DCIS and local stage; $43,878.64 for regional stage; and $61,544.91 for distant stage.	(43)
Utilities		
Healthy women	Age-specific quality of life utilities among women without breast cancer.	(45)
Screening mammogram	0.994 for 1 week	(44)
Diagnostics after positive mammogram	0.895 for 5 weeks	(44)
Cancer treatment	By stage: 0.9 for 2 years for DCIS and local stage; 0.75 for 2 years for regional stage; and 0.6 until death for distant stage.	(46)

Abbreviations: AJCC, American Joint Committee on Cancer; BCSC, Breast Cancer Surveillance Consortium; BI-RADS, Breast Imaging Reporting and Data System; DCIS, ductal carcinoma in situ; ER, estrogen receptor; HER2, human epidermal growth factor 2; RR, relative risk; SEER, Surveillance, Epidemiology, and End Results.
Note: Not all models use all parameters; some models use parameters as direct inputs and others use them as a target for calibration or other estimation (See Appendix Table 2).

quality-adjusted life-years (QALYs) gained. The QALYs were based on utilities for the general U.S. population estimated both with and without adjustments for having a screening examination (0.006 for 1 week per examination = –1 hour per examination) and having a positive screening result and undergoing diagnostic evaluation (0.0105 for 5 weeks = –8.8 hours). We also adjusted for breast cancer treatment (Table 1).

Harms included false-positive mammograms, benign biopsies, and overdiagnosis. The rate of false-positive mammograms was the number read as abnormal in women without cancer divided by the total. Benign biopsies were defined as a biopsy recommendation among women with false-positive screening results (49). Overdiagnosis was defined as screen-detected cancer that would not have been diagnosed in a woman's lifetime in the absence of mammography (14, 50).

Costs (reported in 2014 U.S. dollars) were estimated based on the number of mammograms; evaluation of positive mammograms, including additional imaging or biopsy among women with cancer and those with false-positive mammograms; and stage-specific cancer treatments based on Medicare reimbursement schedules and published studies (Table 1).

Table 2 Lifetime benefits of screening annually, biennially or triennially per 1000 women screened by relative risk, breast density, and age group across 3 models.

		Breast cancer deaths averted vs. no screening, median (range across models)			Life years gained vs. no screening, median (range across models) *		
Density	**RR**	**Triennial**	**Biennial**	**Annual**	**Triennial**	**Biennial**	**Annual**
Ages 50-74†							
Almost entirely fatty	1	3.4 (1.8-3.6)	4.1 (2.4-4.3)	4.7 (3.2-5.6)	50 (35-64)	64 (47-73)	84 (62-85)
	1.3	4.4 (2.4-4.6)	5.3 (3.1-5.5)	6.0 (4.1-7.1)	64 (46-82)	82 (60-94)	108 (80-109)
	2	6.4 (3.6-7.0)	8.0 (4.8-8.0)	9.1 (6.2-10.3)	94 (69-124)	120 (92-142)	159 (122-163)
	4	11.0 (7.2-13.1)	13.8 (9.2-15.0)	17.2 (12.0-17.7)	164 (136-235)	209 (177-269)	277 (233-309)
Scattered fibroglandular	1	4.0 (2.9-5.9)	5.2 (3.8-6.8)	6.9 (5.1-7.9)	59 (56-107)	77 (74-123)	106 (101-143)
	1.3	5.1 (3.7-7.5)	6.5 (4.9-8.7)	8.7 (6.6-10.1)	75 (72-137)	97 (95-158)	134 (129-184)
	2	7.2 (5.6-11.2)	9.2 (7.4-12.9)	12.3 (9.9-15.0)	109 (107-204)	144 (139-236)	194 (191-275)
	4	11.5 (10.8-20.2)	14.7 (13.9-23.3)	19.4 (18.4-27.0)	207 (175-372)	269 (227-430)	360 (308-502)
Heterogeneously dense	1	4.8 (3.3-8.4)	6.3 (4.4-9.8)	8.4 (6.1-11.7)	72 (64-149)	94 (86-175)	130 (122-210)
	1.3	6.0 (4.2-10.7)	7.7 (5.6-12.4)	10.4 (7.8-14.8)	90 (82-190)	117 (110-223)	161 (155-267)
	2	8.3 (6.3-15.5)	10.6 (8.3-18.1)	14.3 (11.6-21.6)	124 (122-278)	162 (162-326)	230 (224-392)
	4	12.4 (11.4-26.5)	15.8 (15.1-31.0)	21.0 (20.8-37.1)	221 (192-485)	294 (248-568)	411 (338-685)
Extremely dense	1	5.1 (3.1-9.9)	6.5 (4.2-11.7)	8.9 (6.0-14.4)	75 (61-174)	98 (82-206)	138 (121-255)
	1.3	6.2 (4.0-12.5)	8.0 (5.4-14.7)	10.9 (7.7-18.1)	93 (79-219)	122 (106-261)	170 (155-323)
	2	8.4 (5.9-17.9)	10.8 (7.9-21.1)	14.7 (11.4-26.0)	127 (115-317)	166 (155-376)	231 (226-468)
	4	12.0 (10.4-29.3)	15.4 (14.0-34.7)	20.5 (20.2-42.9)	204 (187-534)	277 (242-634)	402 (332-789)
Ages 65-74‡							
Almost entirely fatty	1	1.5 (0.8-1.6)	1.8 (1.0-2.0)	2.3 (1.4-2.4)	16 (11-21)	19 (15-26)	26 (21-31)
	1.3	1.9 (1.0-2.0)	2.3 (1.4-2.6)	3.0 (1.9-3.1)	20 (14-27)	24 (19-34)	34 (27-40)
	2	2.7 (1.5-3.0)	3.2 (2.1-3.9)	4.3 (2.8-4.4)	28 (20-40)	33 (29-50)	47 (41-59)
	4	4.2 (2.6-5.4)	5.1 (3.8-7.0)	6.8 (5.0-8.0)	44 (37-71)	54 (52-92)	73 (73-107)
Scattered fibroglandular	1	1.7 (1.1-2.3)	2.1 (1.6-2.9)	3.0 (2.2-3.4)	18 (17-30)	23 (22-39)	33 (32-45)
	1.3	2.1 (1.5-2.9)	2.6 (2.1-3.7)	3.6 (2.9-4.3)	22 (21-38)	30 (27-49)	42 (39-58)
	2	2.9 (2.1-4.2)	3.5 (3.0-5.4)	4.9 (4.1-6.3)	30 (29-55)	43 (36-71)	60 (53-84)
	4	4.0 (3.6-7.2)	5.3 (4.9-9.4)	7.2 (6.8-10.9)	50 (41-96)	74 (50-124)	102 (73-146)
Heterogeneously dense	1	2.0 (1.2-3.6)	2.5 (1.8-4.7)	3.6 (2.5-5.7)	21 (17-47)	26 (25-62)	38 (37-75)
	1.3	2.5 (1.5-4.5)	3.0 (2.2-5.9)	4.3 (3.2-7.1)	26 (21-59)	32 (31-77)	47 (47-95)
	2	3.2 (2.2-6.4)	3.9 (3.2-8.4)	5.5 (4.6-10.1)	33 (31-84)	46 (40-111)	66 (60-135)
	4	4.0 (3.6-10.1)	5.4 (4.8-13.3)	7.6 (6.7-16.1)	50 (40-134)	76 (49-176)	109 (72-216)
Extremely dense	1	2.0 (1.1-4.3)	2.5 (1.7-5.9)	3.6 (2.4-7.3)	21 (16-57)	26 (24-77)	39 (36-97)
	1.3	2.4 (1.4-5.4)	3.0 (2.1-7.3)	4.3 (3.1-9.1)	25 (20-72)	31 (30-96)	46 (45-122)
	2	3.0 (2.0-7.5)	3.7 (3.0-10.1)	5.3 (4.4-12.6)	31 (29-99)	43 (38-134)	64 (57-170)
	4	3.5 (3.3-11.2)	4.9 (4.3-15.1)	7.3 (6.0-18.9)	46 (36-149)	70 (43-202)	105 (64-257)

Abbreviations: RR, relative risk.

* Life years gained are undiscounted.

† Screening is initiated at age 50.

‡ Women who are currently 65 and have been screened biennially from 50-64.

Statistical Analysis

For each age group modeled (≥50 and ≥65 years), there were 16 possible population subgroups based on combinations of breast cancer risk and density. Benefits and harms for each strategy were compared with no screening for every 1000 women screened. No screening was assumed to occur before age 50 years in all analyses. Screening strategies for women aged 65 to 74 years assumed that they received biennial mammography during ages 50 to 64 years. We report the median benefits and harms and the range across models as a measure of uncertainty. In secondary analyses, the ratio of false-positive mammograms to breast cancer deaths averted was calculated as a metric of the tradeoffs of harms to benefits. We also estimated the incremental costs per QALY for each strategy and population risk–density subgroup. For this estimate, the change in cost was divided by the change in benefit (for example, QALYs) when each more costly screening strategy was compared with the strategy with the next lowest cost within the subgroup. Costs and QALYs were discounted at 3% per year, and QALYs included screening and work-up adjustments. Screening strategies were considered cost-effective with a common threshold of $100 000 per QALY gained (51).

Role of the Funding Source

The National Cancer Institute funded this research but had no role in the design or conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

RESULTS

The results of all 3 models illustrate that across intervals and age groups, screening (vs. no screening) (Appendix Table 4, available at www.annals.org) had a greater absolute benefit in terms of breast cancer deaths averted, life-years gained, and QALYs gained among 2 groups of women: those with dense breasts and those at higher RR within each breast density group (Tables 2 and 3). Adjustments for screening harms did not affect the ordering of screening strategies by QALY.

Women Starting Screening at Age 50

For all screening intervals, as risk and breast density increased, the benefits (breast cancer deaths averted, life-years gained, and QALYs gained) of screening increased and the harms (false-positive mammograms and benign biopsy results but not overdiagnosis) decreased with greater risk (Tables 2 to 4).Among average-risk women with fatty breasts (RR, 1.0 or 1.3), biennial screening, compared with no screening, in women aged 50 to 74

Table 3 Lifetime QALY benefits of screening annually, biennially or triennially per 1000 women screened by relative risk, breast density, and age group with and without screening and work-up adjustments.

		QALYs gained with screening and work-up adjustments vs. no screening, median (range across models) *			QALYs gained without screening and work-up adjustments vs. no screening, median (range across models)*		
Density	**RR**	**Triennial**	**Biennial**	**Annual**	**Triennial**	**Biennial**	**Annual**
Ages 50-74†							
Almost entirely fatty	1	32 (21-44)	41 (29-49)	51 (36-51)	37 (26-50)	48 (35-58)	63 (47-66)
	1.3	43 (29-59)	54 (39-66)	69 (50-71)	47 (34-65)	61 (45-75)	80 (61-86)
	2	65 (46-93)	82 (63-105)	106 (81-116)	70 (52-99)	89 (69-114)	118 (93-131)
	4	118 (98-183)	150 (128-209)	194 (168-234)	123 (103-190)	157 (135-217)	206 (180-249)
Scattered fibroglandular	1	36 (35-76)	47 (47-86)	60 (60-92)	43 (43-85)	57 (56-99)	78 (78-115)
	1.3	48 (47-101)	63 (62-114)	82 (81-126)	55 (55-110)	73 (72-127)	99 (98-148)
	2	75 (71-155)	100 (92-178)	132 (123-200)	83 (78-164)	110 (102-190)	149 (140-222)
	4	153 (122-292)	199 (158-336)	264 (212-386)	160 (129-301)	209 (168-349)	280 (228-407)
Heterogeneously dense	1	44 (40-110)	57 (55-126)	75 (73-143)	52 (49-120)	69 (66-141)	95 (94-169)
	1.3	57 (54-143)	74 (74-165)	100 (98-190)	65 (63-153)	85 (85-180)	120 (119-216)
	2	86 (83-215)	114 (108-249)	159 (145-292)	94 (91-225)	125 (119-263)	178 (165-317)
	4	164 (133-384)	220 (173-448)	305 (233-533)	172 (141-394)	230 (184-461)	322 (251-556)
Extremely dense	1	47 (40-131)	62 (54-154)	84 (77-185)	54 (47-140)	71 (64-166)	100 (94-206)
	1.3	60 (54-169)	79 (73-199)	108 (104-240)	67 (61-177)	88 (82-211)	124 (120-261)
	2	85 (83-248)	112 (112-293)	161 (153-358)	92 (90-257)	121 (121-305)	176 (169-379)
	4	154 (129-425)	210 (169-503)	302 (231-622)	161 (136-433)	218 (178-514)	317 (246-641)
Ages 65-74‡							
Almost entirely fatty	1	9 (6-15)	11 (8-18)	15 (11-20)	11 (8-16)	13 (10-21)	19 (15-24)
	1.3	12 (8-19)	15 (11-24)	20 (16-27)	14 (10-21)	17 (14-27)	24 (20-31)
	2	18 (13-30)	22 (19-38)	29 (26-42)	20 (15-31)	24 (21-40)	34 (30-47)
	4	30 (25-55)	37 (36-71)	50 (49-81)	32 (27-57)	39 (38-73)	54 (53-85)
Scattered fibroglandular	1	10 (10-21)	13 (12-27)	18 (17-29)	13 (13-23)	17 (16-30)	24 (23-36)
	1.3	13 (13-28)	18 (16-35)	25 (22-39)	16 (15-30)	22 (19-39)	31 (28-46)
	2	19 (19-42)	28 (23-53)	38 (32-60)	21 (21-44)	31 (26-57)	44 (38-67)
	4	35 (28-75)	52 (33-96)	71 (47-111)	37 (30-77)	55 (37-99)	77 (53-117)
Heterogeneously dense	1	12 (10-35)	15 (14-45)	20 (20-52)	15 (13-38)	19 (18-49)	27 (27-60)
	1.3	15 (13-45)	20 (18-58)	27 (26-68)	18 (16-47)	24 (22-62)	35 (33-76)
	2	21 (20-65)	30 (25-85)	43 (36-100)	24 (23-67)	34 (29-89)	50 (43-108)
	4	35 (27-105)	54 (33-138)	76 (46-167)	38 (30-107)	57 (36-142)	82 (52-174)
Extremely dense	1	12 (10-44)	15 (15-58)	21 (21-72)	15 (12-46)	18 (18-62)	27 (27-78)
	1.3	15 (13-55)	20 (18-74)	28 (27-91)	18 (15-57)	23 (21-77)	34 (32-98)
	2	20 (20-78)	29 (24-104)	43 (35-130)	22 (22-80)	32 (27-107)	48 (41-136)
	4	33 (24-118)	51 (29-159)	76 (43-201)	35 (26-120)	53 (32-162)	80 (47-206)

Abbreviations: QALY, quality-adjusted life year; RR, relative risk.

* QALYs gained are undiscounted.

† Screening is initiated at age 50.

‡ Women who are currently 65 and have been screened biennially from 50-64.

Table 4 Lifetime harms of screening annually, biennially or triennially per 1000 women screened by relative risk, breast density, and age group

Density and RR	False-positives vs. no screening, median (range across models)		
	Triennial	Biennial	Annual
Ages 50-74†			
Almost entirely fatty			
1	489 (424-616)	618 (613-858)	1101 (1094-1548)
1.3	484 (420-611)	612 (606-851)	1089 (1081-1536)
2	471 (412-600)	598 (590-836)	1062 (1051-1507)
4	438 (390-571)	564 (547-794)	996 (972-1429)
Scattered areas of fibroglandular density			
1	781 (693-935)	1009 (991-1326)	1806 (1776-2440)
1.3	767 (683-922)	994 (972-1309)	1776 (1740-2406)
2	734 (662-894)	963 (929-1267)	1714 (1659-2329)
4	649 (613-818)	888 (818-1158)	1568 (1452-2123)
Heterogeneously dense			
1	917 (822-1064)	1197 (1171-1524)	2123 (2080-2829)
1.3	894 (807-1043)	1174 (1141-1493)	2078 (2023-2771)
2	842 (775-995)	1125 (1073-1424)	1984 (1896-2642)
4	715 (703-875)	1016 (906-1248)	1778 (1585-2308)
Extremely dense			
1	732 (652-849)	939 (925-1200)	1668 (1647-2225)
1.3	712 (638-827)	917 (898-1169)	1626 (1597-2167)
2	666 (608-780)	872 (839-1102)	1540 (1487-2039)
4	555 (543-663)	776 (697-933)	1359 (1223-1719)
Ages 65-74‡			
Almost entirely fatty			
1	145 (137-169)	209 (206-227)	413 (395-459)
1.3	142 (135-166)	206 (202-224)	405 (388-453)
2	135 (130-160)	198 (193-217)	387 (373-438)
4	119 (118-145)	178 (169-197)	340 (335-399)
Scattered areas of fibroglandular density			
1	230 (225-278)	343 (333-375)	667 (648-757)
1.3	223 (220-271)	335 (322-366)	645 (632-741)
2	209 (206-257)	317 (298-348)	597 (597-704)
4	180 (166-225)	276 (239-299)	520 (480-607)
Heterogeneously dense			
1	273 (260-329)	407 (397-432)	794 (760-875)
1.3	262 (250-319)	394 (381-417)	762 (735-845)
2	238 (230-298)	367 (346-384)	693 (684-779)
4	182 (181-254)	302 (264-311)	580 (528-617)

Benign biopsies vs. no screening, median (range across models)			Over-diagnosis vs. no screening, median (range across models)*		
Triennial	**Biennial**	**Annual**	**Triennial**	**Biennial**	**Annual**
79 (68-106)	91 (91-136)	127 (127-191)	11 (9-17)	12 (11-20)	17 (12-24)
78 (67-106)	91 (90-135)	126 (125-190)	12 (11-21)	15 (11-26)	21 (12-31)
76 (66-104)	89 (88-133)	123 (122-187)	17 (11-31)	22 (11-37)	30 (12-44)
71 (63-99)	84 (81-126)	116 (113-177)	27 (11-53)	35 (11-63)	49 (12-75)
126 (111-158)	150 (147-206)	209 (206-296)	13 (11-22)	17 (11-27)	23 (12-35)
123 (110-156)	148 (144-203)	206 (202-292)	16 (11-28)	20 (11-34)	29 (12-44)
118 (107-152)	143 (138-197)	199 (193-283)	21 (10-39)	28 (11-48)	39 (12-62)
105 (99-140)	132 (122-181)	183 (169-259)	31 (11-60)	40 (12-74)	56 (13-95)
163 (146-195)	178 (174-235)	266 (261-365)	16 (10-20)	20 (11-26)	28 (12-38)
159 (144-191)	174 (169-230)	261 (254-358)	19 (10-25)	24 (11-32)	34 (12-46)
150 (138-183)	167 (160-220)	249 (238-342)	25 (10-34)	32 (11-44)	45 (13-63)
128 (126-162)	152 (136-194)	224 (200-301)	32 (11-49)	41 (12-63)	57 (14-89)
130 (116-156)	139 (137-185)	209 (206-288)	16 (10-17)	21 (11-22)	31 (12-32)
127 (113-152)	136 (133-181)	204 (200-281)	19 (10-21)	26 (11-27)	37 (12-39)
119 (108-144)	129 (125-171)	193 (186-265)	26 (10-26)	34 (11-35)	47 (13-53)
99 (97-123)	116 (104-146)	171 (154-225)	32 (10-37)	41 (12-49)	56 (15-74)
22 (20-25)	29 (29-32)	45 (43-51)	5 (4-8)	6 (5-11)	9 (5-13)
21 (20-25)	29 (28-31)	45 (43-50)	7 (4-10)	8 (5-14)	11 (5-17)
20 (20-24)	28 (27-30)	43 (41-48)	9 (4-15)	11 (5-20)	15 (6-25)
18 (18-22)	25 (24-28)	37 (37-44)	14 (5-25)	16 (6-34)	22 (7-41)
34 (34-42)	48 (47-52)	73 (71-83)	7 (4-10)	8 (5-15)	12 (5-20)
33 (33-41)	47 (45-51)	71 (69-81)	8 (4-13)	10 (5-19)	14 (6-24)
31 (31-39)	44 (42-49)	66 (66-77)	11 (5-18)	13 (6-26)	18 (6-34)
27 (25-34)	39 (33-42)	57 (53-67)	14 (5-27)	17 (7-38)	23 (8-50)
46 (44-56)	57 (56-61)	95 (91-105)	8 (4-10)	10 (5-14)	14 (6-20)
45 (43-54)	55 (53-58)	91 (88-101)	10 (5-12)	12 (6-17)	17 (7-25)
41 (39-51)	51 (48-54)	83 (82-93)	12 (5-16)	15 (7-23)	21 (8-33)
31 (31-43)	42 (37-44)	70 (63-74)	13 (6-22)	16 (8-32)	22 (10-46)

Table 4 Lifetime harms of screening annually, biennially or triennially per 1000 women screened by relative risk, breast density, and age group (continued)

Density and RR	False-positives vs. no screening, median (range across models)		
	Triennial	Biennial	Annual
Extremely dense			
1	202 (187-239)	295 (291-312)	583 (553-631)
1.3	193 (179-231)	284 (279-298)	559 (532-604)
2	175 (161-214)	263 (253-268)	507 (491-544)
4	133 (118-180)	197 (191-221)	404 (383-412)

Abbreviations: RR, relative risk.

* Over-diagnosed cases are defined as cases that would not have been clinically detected in the absence of screening. The value includes DCIS and invasive over-diagnosis. Over-diagnosis is calculated by comparing cases detected in the screening scenario to those detected in the unscreened scenario.

† Per 1000 women compared to no screening at any age.

‡ Per 1000 women compared to biennial mammograms 50-64 with no subsequent screening.

years averted a median of 4.1 and 5.3 breast cancer deaths per 1000 women screened, respectively. In average-risk women with scattered fibroglandular density (RR, 1.0 or 1.3), biennial screening compared with no screening averted a median of 5.2 and 6.5 breast cancer deaths, respectively (Table 2). Screening outcomes were similar for triennial screening compared with no screening in average-risk women with low-breast density; for every 1000 women screened, the median number of breast cancer deaths averted ranged from 3.4 to 5.1. Screening triennially compared with biennially for average-risk women with low breast density resulted in a median ranging from 21% to 23% fewer false-positive mammograms, 13% to 17% fewer benign biopsies, and 8% to 20% fewer overdiagnosed cases (Table 4). Among women with fatty breasts (RR, 2.0), triennial screening, compared with biennial screening, averted a median of 1.6 breast cancer deaths per 1000 screened. In women with scattered fibroglandular density (RR, 2.0), triennial screening, compared with biennial screening, averted 2 breast cancer deaths per 1000 women screened. Thus, 1000 women with fatty breasts (RR, 2.0) and 1000 women with scattered fibroglandular density (RR, 2.0) would have 9 rounds of triennial screening resulting in 6.4 and 7.2 breast cancer deaths averted, 471 and 734 false-positive mammograms, and 76 and 118 biopsy results, respectively; for 13 rounds of biennial screening, we noted 8.0 and 9.2 breast cancer deaths averted, 598 and 963 false-positive mammograms, and 89 and 143 biopsy results, respectively.

The benefits of more frequent screening increased as density increased and RR increased to 2 or greater. For example, biennial screening, compared with no screening, among women aged 50 to 74 years in subgroups with an RR of 2 and heterogeneously dense breasts resulted in a median of 10.6 breast cancer deaths averted and 1125

Benign biopsies vs. no screening, median (range across models)			Over-diagnosis vs. no screening, median (range across models)*		
Triennial	Biennial	Annual	Triennial	Biennial	Annual
34 (32-41)	41 (41-44)	70 (66-76)	7 (4-9)	10 (6-11)	15 (7-17)
33 (30-39)	40 (39-42)	67 (64-72)	9 (5-10)	12 (6-13)	18 (7-20)
30 (27-36)	37 (35-37)	61 (59-65)	12 (5-12)	15 (7-17)	21 (9-27)
23 (20-31)	28 (27-31)	49 (46-49)	12 (6-16)	14 (9-24)	20 (11-38)

false-positive mammograms per 1000 women screened. If these women received annual rather than biennial screening, a median of 3.7 more deaths could have been averted; however, false-positive mammograms would increase almost 2-fold (1984 vs. 1125 false-positive mammograms per 1000 women screened). Breast cancer deaths averted per

1000 women screened were highest with annual screening for women ages 50 to 74 years with all levels of breast density and an RR of 4.0; averted deaths ranged from 17.2 in women with fatty breasts to 20.5 in women with extremely dense breasts.

The Figure (top) is an exemplar model showing the ratio of harms and benefits for subgroups of women with different levels of risk and density screened from ages 50 to 74 years. Compared with the ratios projected for biennial screening of average-risk women from ages 50 to 74 years regardless of breast density, annual screening has a similar or better ratio when the RR is 2 or greater across all density groups. Triennial screening has similar or better ratios of harms and benefits than biennial screening for average-risk women regardless of breast density in nearly all of the RR and density subgroups because false-positive mammograms are reduced with triennial screening, and the magnitude of breast cancer deaths averted is similar or slightly lower than with biennial screening.

Women at Age 65

The different intervals among women aged 65 to 74 years had similar patterns of benefits and harms across subgroups as observed for screening during ages 50 to 74 years but with lower absolute magnitudes (Tables 2 to 4 and Figure, bottom). If women changed from biennial to triennial screening at age 65 years, fewer than a median of 1 less death per 1000 women screened was averted for all RRs and density subgroups. The exception was women with an RR of 4 and heterogeneously or extremely dense breasts; a median of 1.4 fewer breast cancer deaths were averted in this group (Table 2). For example, continuing biennial screening among average-risk women (RR, 1.0 or 1.3) and women with fatty breasts or scattered fibroglandular density averted a median of 1.8 to 2.3 deaths for women with fatty breasts and 2.1 to 2.6 deaths for women with scattered fibroglandular density for every 1000 women screened (Table 2); switching to triennial

Table 5 Incremental costs per quality-adjusted life year gained* by breast density, risk level, screening interval, and age for 3 models.

Density	RR	Screening Frequency	Age 50-74			Age 65-74		
			Model E	Model W	Model GE	Model E	Model W	Model GE
Fatty	1.0	Triennial	**68,777**	117,753	**43,098**	100,058	131,294	**27,639**
		Biennial	122,007	123,132	232,710	109,587	212,665	104,235
		Annual	389,195	586,116	Dom	435,881	516,979	>1,000,000
	1.3	Triennial	**50,231**	**83,220**	**27,022**	**70,716**	**92,938**	**15,785**
		Biennial	**83,577**	**86,426**	133,826	**75,433**	135,221	**67,152**
		Annual	231,495	309,654	>1,000,000	258,193	286,643	799,501
	2.0	Triennial	**30,910**	W Dom	**10,364**	**42,229**	**58,276**	**3,004**
		Biennial	**50,526**	**50,084†**	**65,297**	**46,300**	**70,911**	**32,912**
		Annual	122,540	148,375	392,745	141,183	146,961	263,493
	4.0	Triennial	**14,969**	**22,663**	**†**	**19,130**	W Dom	**†**
		Biennial	**22,802**	**23,295**	**19,932**	**21,242**	**30,054**	**5,331**
		Annual	**54,906**	**56,451**	**95,362**	**69,089**	**62,251**	**76,840**
Scat-tered	1.0	Triennial	**69,714**	**72,156**	**18,509**	W Dom	**55,051**	**14,112**
		Biennial	111,605	**75,673**	104,454	101,612	Dom	**61,723**
		Annual	317,991	288,199	Dom	382,578	612,349	>1,000,000
	1.3	Triennial	**50,010**	**51,493**	**9,683**	W Dom	**60,785**	**5,449**
		Biennial	**75,416**	**53,967**	**63,057**	**72,488**	**73,479**	**39,636**
		Annual	186,322	171,038	488,376	224,322	201,088	450,818
	2.0	Triennial	**31,053**	**29,757**	**641**	W Dom	**38,299**	**†**
		Biennial	**43,721**	**31,198**	**28,182**	**42,160**	**42,347**	**15,956**
		Annual	**96,584**	**85,607**	144,723	120,188	104,553	159,293
	4.0	Triennial	**15,414**	**12,179**	**†**	W Dom	**17,507**	Dom
		Biennial	**19,733**	**13,116**	**5,116**	**20,076**	**17,977**	**‡**
		Annual	**44,019**	**32,452**	**39,105**	**61,818**	**39,362**	**42,660**
Het. dense	1.0	Triennial	**57,924**	W Dom	**8,016**	W Dom	**75,197**	**611**
		Biennial	**85,241**	**60,333†**	**50,421**	**85,145**	**96,863**	**23,104**
		Annual	222,789	185,805	268,798	279,586	290,534	179,689
	1.3	Triennial	**42,324**	**41,815**	**2,179**	**60,235**	**54,355**	**†**
		Biennial	**61,309**	**42,551**	**31,442**	**61,760**	**60,225**	**11,809**
		Annual	137,983	116,700	134,915	169,196	174,243	97,850
	2.0	Triennial	**26,726**	**23,375**	**†**	W Dom	**31,637**	**†**
		Biennial	**35,235**	**26,574**	**12,543**	**36,446**	**36,762**	**478**
		Annual	**75,747**	**57,557**	**56,331**	**99,035**	**85,503**	**44,784**
	4.0	Triennial	**13,432**	**8,534**	Dom	W Dom	**13,298**	Dom
		Biennial	**16,745**	**9,256**	**‡**	**18,673**	**13,814**	**‡**
		Annual	**36,845**	**22,339**	**14,716**	**57,264**	**32,355**	**8,752**
Dense	1.0	Triennial	**50,563**	**52,953**	**3,017**	W Dom	**63,918**	**†**
		Biennial	**68,216**	**55,420**	**27,942**	**75,917**	**77,061**	**8,555**

Table 5 Incremental costs per quality-adjusted life year gained* by breast density, risk level, screening interval, and age for 3 models. (continued)

Density	RR	Screening Frequency	Age 50-74			Age 65-74		
			Model E	Model W	Model GE	Model E	Model W	Model GE
		Annual	148,014	129,536	89,425	187,329	203,860	60,177
	1.3	Triennial	**37,937**	**36,486**	†	W Dom	**45,929**	†
		Biennial	**49,172**	**40,051**	**16,293**	**55,033**	**52,754**	**2,547**
		Annual	101,399	**87,230**	**56,264**	130,774	130,339	**36,740**
	2.0	Triennial	**24,715**	**20,626**	†	W Dom	**26,367**	Dom
		Biennial	**30,291**	**23,683**	**4,631**	**35,097**	**32,766**	‡
		Annual	**60,577**	**47,687**	**25,753**	**82,794**	**71,187**	**15,070**
	4.0	Triennial	**13,169**	**7,130**	Dom	W Dom	**10,180**	Dom
		Biennial	**14,856**	**7,823**	‡	**19,207**	**11,669**	Dom
		Annual	**31,433**	**18,224**	**4,407**	**52,645**	**26,834**	§

Note: Incremental ratios **bold** if values are <$100,000, a common threshold for least costly and most effective strategies (dominant). Unless otherwise indicated, triennial strategies are compared to no screening. Breast density categories shown as: fatty, almost entirely fat; scattered, scattered fibroglandular density; het. dense, heterogeneously dense; and dense, extremely dense.
Abbreviations: RR, relative risk; Dom, more expensive and less effective (strongly dominated); W Dom, more expensive and more effective but less efficient (weakly dominated).
*Costs and quality-adjusted life years discounted at 3% per year. Quality-adjusted life years include disutility from participation in screening mammography.
†Strategy with no screening is strongly dominated. Triennial is the least costly strategy for comparison.
‡Strategy with biennial screening is the least costly.
§Strategy with annual screening is the least costly

screening averted a median of 1.5 to 1.9 deaths for women with fatty breasts and 1.7 to 2.1 deaths for women with scattered fibroglandular density. Switching from biennial to annual screening increased the median number of breast cancer deaths averted to 2 or more for women with heterogeneously or extremely dense breasts and an RR of 4.

As was the case for screening in women aged 50 to 74 years, the ratio of harms (measured as false-positive mammograms) and benefits (breast cancer deaths averted) for annual screening in women aged 65 to 74 years was similar to or better (lower) than that seen in biennial screening of average-risk women with an RR of 2 or greater in all density subgroups; exceptions were rare (Figure, bottom). Triennial screening also had a lower or more favorable ratio than biennial screening because it reduces false-positive mammograms, and the magnitude of breast cancer deaths averted is the same or slightly lower. Continuing biennial screening has a similar balance as triennial screening for most subgroups as seen for average-risk groups, regardless of breast density

Cost-Effectiveness

When we used a common threshold of $100 000 per QALY, triennial strategies were the only costeffective strategies for subgroups with average risk and low breast density (fatty breasts or scattered fibroglandular density) in both age groups (Table 5). Biennial strategies were cost-effective for most density subgroups at average or intermediate risk (RR, 1.3 or 2.0). Annual strategies were only consistently cost-effective across models for subgroups with an RR of 4, regardless of density, or an RR of 2 or greater and heterogeneously or extremely dense breasts.

DISCUSSION

This collaborative modeling study shows that risk and density level can be useful for guiding tailored screening recommendations. For average-risk women in low-density subgroups, which comprise a large proportion of the population, triennial screening provides a reasonable balance of benefits and harms and is costeffective. Annual screening has a favorable balance of benefits and harms and would be considered costeffective for subgroups of women aged 50 years with risk levels that are 2 to 4 times the average and that have heterogeneously or extremely dense breasts. Benefits of screening women with heterogeneously dense breasts (at any interval) were greater than screening those with extremely dense breasts at each risk level, reflecting increased risk but fewer missed cases of cancer than screening women with extremely dense breasts. The same patterns are seen for women aged 65 years such that subgroups at average risk with low breast density can consider triennial screening. In contrast, the few women who remain at higher risk might benefit from annual screening. Of note, biennial screening maintains an acceptable balance of outcomes and is also cost-effective for women with an RR of 1.3 or 2 as long as they are not in the highest-density groups. Screening benefits and harms exist on a continuum across age, risk, and density, with the optimal screening interval depending on women's values and preferences for benefits and harms.

Current U.S. screening guidelines focus on the average-risk population and generally recommend biennial screening for women in their 50s or older (3, 4). These new modeling results support this recommendation for women who do not have either higher-than average risk and high breast density or average to low risk and low breast density. Annual screening has been suggested for high-risk women (4). The current results provide further guidance on the specific combinations of RRs and breast density after age 50 years that identify the subgroups in which annual screening should be considered; these subgroups are estimated to constitute fewer than 1% of the population at both ages 50 and 65 years (BCSC; Miglioretti DL. Personal communication. 2016).

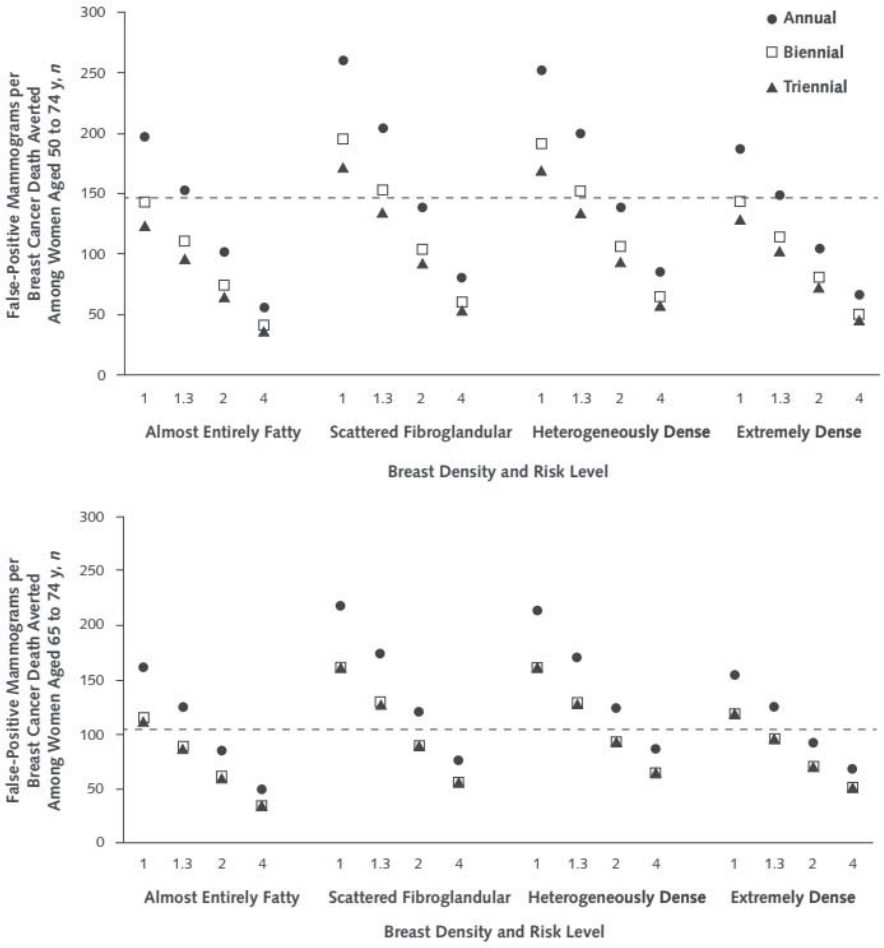


Figure 1 False-positives mammograms per breast cancer death averted for women (A) aged 50-74 and (B) aged 65-74 according to screening frequency and risk level (relative risk group, breast density) using an exemplar model (Model E). Values for all screening frequencies compared to the scenario with no mammography screening. Values for ages 65-74 assume all women received biennial screening during ages 50-64. Dashed lines show this value for women with average density and average risk receiving biennial screening (147.7 for ages 50-74 and 105.8 for ages 65-74). Having fewer false-positives per death averted than this level, i.e., a value below the dashed line, would be more favorable.

Although triennial screening is routinely used in several countries (52, 53), this interval has not been considered in the United States. Our modeling suggests that triennial screening has a similar balance of benefits and harms compared with biennial screening in some groups. Decisions about using triennial versus biennial screening for average-risk women in the lowdensity subgroups result in fewer false-positive mammograms,

biopsies, and overdiagnosis with minimal effect on breast cancer deaths averted. Others have noted that triennial screening can be cost-effective for average-risk women or those with an RR of 2 or less aged 60 to 79 years with fatty breasts or scattered fibroglandular density(10, 11). We found that 12% of women aged 50 years and 20% of those aged 65 years have low breast density (fatty breasts and scattered fibroglandular density) and an RR of 1.0 or 1.3 (BCSC; Miglioretti DL. Personal communication. 2016).

Breast cancer screening guidelines include an upper limit based on age or life expectancy (3, 4, 54). Although we did not evaluate comorbidity, our study results suggest that screening intervals for older women should be based on competing causes of mortality, breast cancer risk, and changes in breast density associated with aging. The ability to tailor screening based on density may become increasingly feasible with the trend toward mandated standard reporting of breast density to women after a mammogram. Because our results show that the RR of breast cancer in combination with breast density has a strong influence on the net benefit of mammography at all screening intervals, evaluation of different risk assessment tools will be important in this context.

Although the models provide new data and have consistent conclusions, several caveats should be considered. First, the 3 models used common inputs but varied in how these data were implemented based on model structure. These variations led to differences in the absolute values for outcome metrics. For example, based on assumptions about temporal trends in underlying incidence, models with the lowest projected incidence estimate fewer breast cancer deaths averted than those with higher incidence. This analysis includes 3 of 6 Cancer Intervention and Surveillance Modeling Network breast models and is an extension of work conducted by all 6 groups(8). Second, because the analytic goal was to determine screening efficacy, the models assumed 100% adherence to screening and use of the most effective modern treatments. Actual benefits will fall short of those projected under these assumptions. Third, we did not explicitly consider lower-than-average risk (that is, RR <1). It will be important to extend our analyses to lower-risk groups because most U.S. women have an RR less than 1 across all density subgroups (70% of women aged 50 years and 66% aged 65 years) (BCSC; Miglioretti DL. Personal communication. 2016). By extension, our current findings suggest that triennial screening would be a reasonable option for lower-than-average risk women with fatty breasts or scattered fibroglandular density. Fourth, we did not model the effect of screening from ages 40 to 49 years, other combinations of ages and intervals, or carriers of breast cancer susceptibility genes 1 and 2. Whether the lack of strategies incorporating screening women in their 40s would affect the balance of benefits and harms against longer (or shorter) screening intervals after age 50 years is unclear. Fifth, although 2 age groups and change in density between age groups were considered, our results do not provide guidance for women whose risk changes over time; modeling change in risk with aging is an important area for future research. Sixth, we used RR rather than absolute risk

level because our simulation models were better suited for this approach. Absolute risk calculators are commonly available (27, 55-57), and the suitability of these calculators to assign risk to personalize screening intervals should continue to be evaluated. Finally, we did not evaluate alternative or supplemental imaging.

Overall, this comparative modeling study illustrates consistent patterns in benefits and harms that could be useful for guiding shared decision making and tailoring screening intervals. The results show that for all screening intervals, benefits and harms change with risk and breast density. Further, the threshold to decide on the screening interval will depend on individual preference(1). Assessing breast density and breast cancer risk can identify subgroups of average-risk women with low breast density who can consider triennial screening and higher-risk women with high breast density who may benefit from annual screening.

Reproducible Research Statement: *Study protocol*: Not available. *Statistical code*: Detailed information about the models is available online at http://cisnet.cancer.gov/breast/profiles.html and in reference (14). *Data set*: Input and output data from the models are available at reference (14) and by contacting Dr. Trentham-Dietz at trentham@wisc.edu.

REFERENCES

1. Kerlikowske K. Progress Toward Consensus on Breast Cancer Screening Guidelines and Reducing Screening Harms. JAMA Intern Med. 2015;175(12):1970-1.
2. Siu AL, Bibbins-Domingo K, Grossman DC, LeFevre ML, Force USPST. Convergence and Divergence Around Breast Cancer Screening. Ann Intern Med. 2016;164(4):301-2.
3. Siu AL, Force USPST. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2016;164(4):279-96.
4. Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. JAMA. 2015;314(15):1599-614.
5. Myers ER, Moorman P, Gierisch JM, Havrilesky LJ, Grimm LJ, Ghate S, et al. Benefits and Harms of Breast Cancer Screening: A Systematic Review. JAMA. 2015;314(15):1615-34.
6. Nelson HD, Pappas M, Cantor A, Griffin J, Daeges M, Humphrey L. Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med. 2016;164(4):256-67.
7. Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, et al. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med. 2015;372(24):2353-8.
8. Mandelblatt JS, Stout NK, Schechter CB, van den Broek JJ, Miglioretti DL, Krapcho M, et al. Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies. Ann Intern Med. 2016;164(4):215-25.
9. Kerlikowske K, Zhu W, Hubbard RA, Geller B, Dittus K, Braithwaite D, et al. Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy. JAMA Intern Med. 2013;173(9):807-16.
10. Vilaprinyo E, Forne C, Carles M, Sala M, Pla R, Castells X, et al. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. PloS one. 2014;9(2):e86858.
11. Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. Annals of Internal Medicine. 2011;155(1):10-20.
12. Kerlikowske K, Zhu W, Tosteson AN, Sprague BL, Tice JA, Lehman CD, et al. Identifying women with dense breasts at high risk for interval cancer: a cohort study. Ann Intern Med. 2015;162(10):673-81.
13. van Ravesteyn NT, Miglioretti DL, Stout NK, Lee SJ, Schechter CB, Buist DS, et al. Tipping the Balance of Benefits and Harms to Favor Screening Mammography Starting at Age 40 Years: A Comparative Modeling Study of Risk. Annals of Internal Medicine. 2012;156(9):609-17.
14. Mandelblatt J, Cronin KA, De Koning H, Miglioretti DL, Schechter C, Stout N, et al. Collaborative Modeling of U.S. Breast Cancer Screening Strategies. Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services, Rockville, MD; AHRQ Publication No. 14-05201-EF-4, December 2015. Available from: http://www.uspreventiveservicestaskforce.org/Home/GetFile/1/16255/collabmodelingbc/pdf. Accessed May 2016.
15. Breast Cancer Surveillance Consortium [Internet]. Rockville, MD: National Cancer Institute [2015 Jul 06; cited 2016 May

17]. Available from: http://breastscreening.cancer.gov.
16. American College of Radiology. The American College of Radiology Breast Imaging Reporting and Data System (BI-RADS). 4th ed. Reston, VA: American College of Radiology; 2003.
17. Munsell MF, Sprague BL, Berry DA, Chisholm G, Trentham-Dietz A. Body mass index and breast cancer risk according to postmenopausal estrogen-progestin use and hormone receptor status. Epidemiol Rev. 2014;36(1):114-36.
18. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol. 2012;13(11):1141-51.
19. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW, Jr., et al. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer. 2002;87(11):1234-45.
20. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Alcohol Consumption and Ethyl Carbamate. Lyon, France: World Health Organization; 2010.
21. Chlebowski RT, Anderson GL, Aragaki AK, Prentice R. Breast Cancer and Menopausal Hormone Therapy by Race/Ethnicity and Body Mass Index. J Natl Cancer Inst. 2016;108(2).
22. Chlebowski RT, Anderson GL, Gass M, Lane DS, Aragaki AK, Kuller LH, et al. Estrogen plus progestin and breast cancer incidence and mortality in postmenopausal women. JAMA. 2010;304(15):1684-92.
23. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results. From the Women's Health Initiative randomized controlled trial. JAMA. 2002;288(3):321-33.
24. Trentham-Dietz A, Sprague BL, Hampton JM, Miglioretti DL, Nelson HD, Titus LJ, et al. Modification of breast cancer risk according to age and menopausal status: a combined analysis of five population-based case-control studies. Breast Cancer Res Treat. 2014;145(1):165-75.
25. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet. 2001;358(9291):1389-99.
26. Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA. Family history and the risk of breast cancer: a systematic review and meta-analysis. Int J Cancer. 1997;71(5):800-9.
27. Tice JA, Miglioretti DL, Li CS, Vachon CM, Gard CC, Kerlikowske K. Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer. J Clin Oncol. 2015;33(28):3137-43.
28. Dyrstad SW, Yan Y, Fowler AM, Colditz GA. Breast cancer risk associated with benign breast disease: systematic review and meta-analysis. Breast Cancer Res Treat. 2015;149(3):569-75.
29. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 2015;107(5).
30. Munoz D, Near AM, van Ravesteyn NT, Lee SJ, Schechter CB, Alagoz O, et al. Effects of screening and systemic adjuvant therapy on ER-specific US breast

cancer mortality. J Natl Cancer Inst. 2014;106(11).

31. Gangnon RE, Sprague BL, Stout NK, Alagoz O, Weedon-Fekjaer H, Holford TR, et al. The contribution of mammography screening to breast cancer incidence trends in the United States: an updated age-period-cohort model. Cancer Epidemiol Biomarkers Prev. 2015;24(6):905-12.
32. Carter SB, Gartner SS, Haines MR, Olmstead AL, Sutch R, Wright G. Historical Statistics of the United States, Volume One: Population. New York: Cambridge University Press; 2006.
33. Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL. The Wisconsin Breast Cancer Epidemiology Simulation Model. J Natl Cancer Inst Monogr. 2006(36):37-47.
34. Mandelblatt J, Schechter CB, Lawrence W, Yi B, Cullen J. The SPECTRUM population model of the impact of screening and treatment on U.S. breast cancer trends from 1975 to 2000: principles and practice of the model methods. J Natl Cancer Inst Monogr. 2006(36):47-55.
35. Tan SY, van Oortmarssen GJ, de Koning HJ, Boer R, Habbema JD. The MISCAN-Fadia continuous tumor growth model for breast cancer. J Natl Cancer Inst Monogr. 2006(36):56-65.
36. Nelson HD, Zakher B, Cantor A, Fu R, Griffin J, O'Meara ES, et al. Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. Ann Intern Med. 2012;156(9):635-48.
37. Cronin KA, Mariotto AB, Clarke LD, Feuer EJ. Additional common inputs for analyzing impact of adjuvant therapy and mammography on U.S. mortality. J Natl Cancer Inst Monogr. 2006(36):26-9.
38. Cronin KA, Yu B, Krapcho M, Miglioretti DL, Fay MP, Izmirlian G, et al. Modeling the dissemination of mammography in the United States. Cancer Causes Control. 2005;16(6):701-12.
39. Early Breast Cancer Trialists' Collaborative Group, Peto R, Davies C, Godwin J, Gray R, Pan HC, 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.
40. Mariotto AB, Feuer EJ, Harlan LC, Abrams J. Dissemination of adjuvant multiagent chemotherapy and tamoxifen for breast cancer in the United States using estrogen receptor information: 1975-1999. J Natl Cancer Inst Monogr. 2006(36):7-15.
41. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Breast Cancer. 2015.
42. Stout NK, Lee SJ, Schechter CB, Kerlikowske K, Alagoz O, Berry D, et al. Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography. J Natl Cancer Inst. 2014;106(6):dju092.
43. Yabroff KR, Lamont EB, Mariotto A, Warren JL, Topor M, Meekins A, et al. Cost of care for elderly cancer patients in the United States. J Natl Cancer Inst. 2008;100(9):630-41.
44. de Haes JC, de Koning HJ, van Oortmarssen GJ, van Agt HM, de Bruyn AE, van Der Maas PJ. The impact of a breast cancer screening programme on quality-adjusted life-years. Int J Cancer. 1991;49(4):538-44.
45. Hanmer J, Lawrence WF, Anderson JP, Kaplan RM, Fryback DG. Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores. Med Decis Making. 2006;26(4):391-400.
46. Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG. Retrospective cost-effectiveness

analysis of screening mammography. J Natl Cancer Inst. 2006;98(11):774-82.
47. Surveillance Epidemiology and End Results (SEER) Program. (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 9 Regs Research Data, Nov 2014 Sub (1973-2012) <Katrina/Rita Population Adjustment> - Linked to County Attributes - Total U.S., 1969-2013 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2015, based on the November 2014 submission. .
48. Sprague BL, Gangnon RE, Burt V, Trentham-Dietz A, Hampton JM, Wellman RD, et al. Prevalence of mammographically dense breasts in the United States. J Natl Cancer Inst. 2014;106(10).
49. Rosenberg RD, Yankaskas BC, Abraham LA, Sickles EA, Lehman CD, Geller BM, et al. Performance benchmarks for screening mammography. Radiology. 2006;241(1):55-66.
50. van Ravesteyn NT, Stout NK, Schechter CB, Heijnsdijk EA, Alagoz O, Trentham-Dietz A, et al. Benefits and harms of mammography screening after age 74 years: model estimates of overdiagnosis. J Natl Cancer Inst. 2015;107(7).
51. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness--the curious resilience of the $50,000-per-QALY threshold. N Engl J Med. 2014;371(9):796-7.
52. Giordano L, von Karsa L, Tomatis M, Majek O, de Wolf C, Lancucki L, et al. Mammographic screening programmes in Europe: organization, coverage and participation. J Med Screen. 2012;19 Suppl 1:72-82.
53. Von Karsa L, Antilla A, Ronco G, Ponti A, Malila N, Arbyn M, et al. Cancer Screening in the European Union: Report on the implementation of the Council Recommendation on cancer screening. First Report. European Communities; 2008.
54. Lansdorp-Vogelaar I, Gulati R, Mariotto AB, Schechter CB, de Carvalho TM, Knudsen AB, et al. Personalizing age of cancer screening cessation based on comorbid conditions: model estimates of harms and benefits. Ann Intern Med. 2014;161(2):104-12.
55. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989;81(24):1879-86.
56. Rosner B, Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. J Natl Cancer Inst. 1996;88(6):359-64.
57. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med. 2004;23(7):1111-30.

Appendix Table 1. Examples of Risk Factors for Women Aged 50-74 y Corresponding to the Relative Risk Levels in the Simulation Models

Published Risk Estimates	Risk Group	Comparison Group*	Reference
Relative risk in simulation models: 1.3			
1.12	Age at menopause ≥55 y	Age at menopause 50-54 y	18
1.19	Age at menarche <11 y	Age at menarche 13 y	18
1.21-1.46	≥25 g alcohol per day	No alcoholic beverage consumption	19, 20
1.25-1.28	Use of estrogen plus progestin postmenopausal hormones	Never use of postmenopausal hormones	21-23
1.30-1.52	Nulliparity or age at first full-term pregnancy ≥25 y	Age at first full-term pregnancy <22 y	24
1.42	Body mass index ≥30 kg/m^2 among never-users of postmenopausal hormones	Body mass index <25 kg/m^2 among never-users of postmenopausal hormones	17
1.47-1.99	1 first-degree family member diagnosed with breast cancer	No family members diagnosed with breast cancer	25-27
Relative risk in simulation models: 2.0			
1.44-2.07	History of benign breast disease, not otherwise specified	No history of benign breast disease	27, 28
1.66-2.02	History of proliferative disease without atypia	No history of benign breast disease	27, 28
2.3-3.9	≥2 first-degree family members with breast cancer	No family members diagnosed with breast cancer	25-27
Relative risk in simulation models: 4.0			
3.29-5.84	History of lobular carcinoma in situ	No history of benign breast disease	27, 28
3.36	Highest 1% of polygenic risk score	40th-60th percentile of polygenic risk score	29
3.93	History of atypical hyperplasia, not otherwise specified	No history of benign breast disease	28

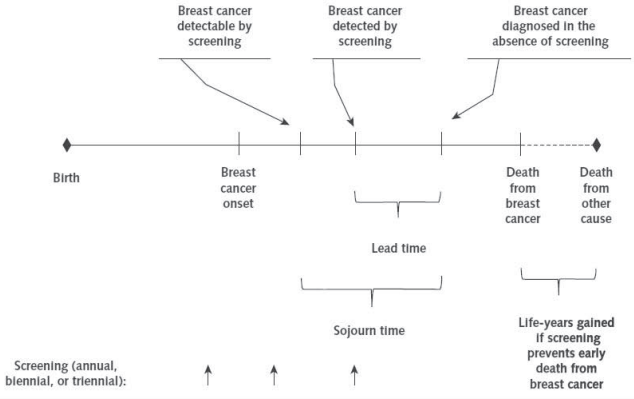
* Risk estimates were based on the comparison group consisting of the largest proportion of women, i.e., "average risk." Women with reduced risk were not modeled, including women who engaged in regular moderate–vigorous physical activity with age at menarche >13 y or age at menopause <50 y with almost entirely fat breast density (Breast Imaging Reporting and Data System category = "a") or who breastfed for ≥1 y. Relative risks associated with breast density categories are shown in Appendix Table 3.

Appendix Table 2. Summary of Model Features*

Feature	Model		
	E	GE	W
Natural history of cancer	Continuous tumor growth	Stage transition	Continuous tumor growth
Tumors obligated to progress	DCIS nonobligate; invasive obligate	DCIS nonobligate; invasive obligate	DCIS and some small invasive are nonobligate; larger invasive obligate
SEER breast cancer data used for model calibration (1975-2010)	Incidence	Incidence	Incidence and mortality
Implementation of screening benefit	Smaller tumor size	Younger age and earlier stage	Younger age and smaller tumor size
Implementation of treatment benefit	Cure fraction based on fatal diameter	Hazard reduction	Cure fraction
Factors affecting treatment benefit	ER and HER2; age; year of and size at diagnosis	ER and HER2; age; year of and stage at diagnosis	ER and HER2; age; year of and stage at diagnosis
Model software program†	Delphi	C++	C++

DCIS = ductal carcinoma in situ; E = Erasmus Medical Center; ER = estrogen receptor; GE = Georgetown University Medical Center and Albert Einstein College of Medicine; HER2 = human epidermal growth factor receptor 2; SEER = Surveillance, Epidemiology, and End Results; W = University of Wisconsin–Madison and Harvard Medical School. * Adapted from reference 14. Additional information is available at https://resources.cisnet.cancer.gov/registry/site-summary/breast. † Combined output from all 3 models was analyzed using SAS software, version 9.4 (SAS Institute).

Appendix Figure. Schema representing breast cancer natural history and screening as simulated in the models.



Sojourn time is the duration of the preclinical, screen-detectable phase of the tumor. Lead time is the interval from screen detection to the time of clinical diagnosis, which is when the tumor would have surfaced without screening. See Appendix Table 2 for the description of the implementation of screening benefit in the 3 simulation models.

Appendix Table 3. Age-Specific Model Input Parameters, by Breast Density and Screening Round for Digital Mammography Performance*

Age, by Breast Density	Density Prevalence	Density Relative Risk†	Screening Frequency	Sensitivity for Invasive Cancer	Sensitivity for DCIS	Specificity	Biopsy After False-Positive Mammogram, ‡
Almost entirely fatty							
50-64 y	0.097	0.50	First	0.948	0.955	0.903	23
			Annual	0.868	0.921	0.948	11
			Biennial	0.921	0.943	0.944	14
			Triennial	0.922	0.921	0.935	15
≥65 y	0.135	0.61	First	0.963	0.955	0.916	23
			Annual	0.903	0.922	0.955	11
			Biennial	0.943	0.944	0.952	14
			Triennial	0.944	0.921	0.944	15
Scattered fibroglandular density							
50-64 y	0.464	0.84	First	0.930	0.949	0.843	23
			Annual	0.826	0.912	0.912	11
			Biennial	0.895	0.937	0.906	14
			Triennial	0.895	0.912	0.892	15
≥65 y	0.533	0.94	First	0.950	0.950	0.863	23
			Annual	0.871	0.913	0.924	11
			Biennial	0.924	0.937	0.919	14
			Triennial	0.924	0.912	0.907	15
Heterogeneously dense							
50-64 y	0.376	1.25	First	0.876	0.965	0.812	23
			Annual	0.716	0.938	0.894	12
			Biennial	0.818	0.956	0.886	14
			Triennial	0.819	0.938	0.870	17
≥65 y	0.300	1.28	First	0.909	0.965	0.836	23
			Annual	0.782	0.938	0.908	12
			Biennial	0.865	0.956	0.901	14
			Triennial	0.865	0.938	0.887	17
Extremely dense							
50-64 y	0.063	1.53	First	0.822	0.944	0.857	23
			Annual	0.623	0.904	0.921	12
			Biennial	0.747	0.930	0.915	14
			Triennial	0.748	0.903	0.903	17
≥65 y	0.032	1.45	First	0.868	0.944	0.876	23
			Annual	0.702	0.904	0.932	12
			Biennial	0.808	0.931	0.927	14
			Triennial	0.809	0.904	0.916	17

DCIS = ductal carcinoma in situ. * Annual mammography was defined as 9- to 18-mo intervals; biennial mammography was defined as 19- to 30-mo intervals; triennial mammography was defined as

31- to 42-mo intervals. Data were obtained from the Breast Cancer Surveillance Consortium. † Age-specific relative risk for breast cancer associated with breast density; reference group is women with average density. ‡ Corrected for missing data.

Appendix Table 4. Median (Range) Lifetime Breast Cancer Outcomes per 1000 Women in the Absence of Screening, by Relative Risk, Breast Density, and Age Group From 3 Simulation Models*

Relative Risk, by Density	Ages 50-74 y†			Ages 65-74 y‡		
	Breast Cancer Deaths	Life-Years§	QALYs§	Breast Cancer Deaths	Life-Years§	QALYs§
Almost entirely fatty						
1.0	11.7 (7.2-14.0)	33.5 (32.1-33.7)	26.5 (25.4-26.7)	8.9 (4.5-9.5)	20.7 (19.3-21.0)	15.8 (14.7-16.0)
1.3	15.0 (9.2-17.8)	33.5 (32.0-33.7)	26.5 (25.4-26.6)	11.3 (5.8-11.9)	20.7 (19.2-21.0)	15.8 (14.7-16.0)
2.0	22.6 (13.7-25.4)	33.4 (31.9-33.6)	26.4 (25.3-26.5)	16.6 (8.5-16.8)	20.7 (19.2-20.9)	15.8 (14.6-15.9)
4.0	41.9 (25.3-42.0)	33.2 (31.6-33.2)	26.2 (25.0-26.2)	25.4 (14.7-30.0)	20.6 (19.1-20.7)	15.7 (14.5-15.8)
Scattered fibroglandular density						
1.0	19.2 (11.7-20.4)	33.4 (32.0-33.6)	26.4 (25.4-26.6)	13.1 (7.3-14.0)	20.7 (19.2-20.9)	15.8 (14.7-15.9)
1.3	24.5 (14.9-25.4)	33.4 (31.9-33.5)	26.4 (25.3-26.5)	16.6 (9.2-17.1)	20.7 (19.2-20.9)	15.8 (14.6-15.9)
2.0	35.2 (21.9-36.1)	33.2 (31.7-33.3)	26.3 (25.1-26.3)	22.8 (12.9-24.2)	20.6 (19.1-20.8)	15.7 (14.6-15.8)
4.0	53.6 (38.6-64.1)	32.8 (31.4-32.9)	25.9 (24.8-26.0)	31.1 (20.7-41.0)	20.5 (19.0-20.6)	15.6 (14.5-15.7)
Heterogeneously dense						
1.0	26.1 (15.7-29.2)	33.3 (31.9-33.5)	26.4 (25.3-26.4)	17.9 (9.0-21.8)	20.7 (19.2-20.8)	15.8 (14.6-15.9)
1.3	31.8 (19.8-36.9)	33.3 (31.8-33.3)	26.3 (25.2-26.3)	21.3 (11.2-27.2)	20.7 (19.1-20.8)	15.7 (14.6-15.8)
2.0	42.8 (28.6-53.1)	33.0 (31.6-33.1)	26.1 (25.0-26.1)	27.2 (15.6-38.2)	20.6 (19.1-20.6)	15.7 (14.5-15.7)
4.0	61.1 (48.5-88.9)	32.4 (31.2-32.7)	25.5 (24.7-25.8)	33.6 (24.0-59.4)	20.4 (19.0-20.5)	15.5 (14.4-15.6)
Extremely dense						
1.0	29.6 (16.9-37.2)	33.3 (31.8-33.3)	26.3 (25.2-26.3)	19.8 (9.5-28.6)	20.7 (19.1-20.8)	15.8 (14.6-15.8)
1.3	35.8 (21.4-46.6)	33.2 (31.7-33.2)	26.2 (25.1-26.3)	23.4 (11.9-35.3)	20.7 (19.1-20.7)	15.7 (14.6-15.7)
2.0	47.3 (30.7-66.0)	32.8 (31.5-33.0)	25.9 (24.9-26.1)	29.2 (16.5-48.4)	20.5 (19.0-20.6)	15.6 (14.5-15.7)
4.0	64.9 (51.5-105.5)	32.1 (31.1-32.6)	25.3 (24.6-25.7)	34.3 (25.1-70.6)	20.2 (18.9-20.5)	15.3 (14.4-15.6)

QALY = quality-adjusted life-year.

* Values are median numbers (range across models).

† Screening was initiated at age 50 y.

‡ Women who were currently age 65 y and have been screened previously biennially from ages 50-64 y.

§ Undiscounted.

Chapter 10

Personalizing Breast Cancer Screening Based on Family History & Polygenic Risk.

Jeroen J. van den Broek, Clyde B. Schechter, Nicolien T. van Ravesteyn, A. Cecile J.W. Janssens, Michael C. Wolfson, Amy Trentham-Dietz, Jacques Simard, Douglas F. Easton, Jeanne S. Mandelblatt, Peter Kraft, Harry J. de Koning

Submitted

ABSTRACT

Background

Current breast cancer screening guidelines are age-based. However, at any given age there is variability in breast cancer risk. This study assessed screening approaches using first-degree family history (FH) and polygenic risk scores (PRS) to identify women for risk-based screening.

Method

Two established breast cancer models assessed the impact of risk-based screening on breast cancer deaths, life years gained, false-positive mammograms, and overdiagnoses for the 1985 U.S. female birth cohort. Digital mammography screening strategies varying in initiation age (30, 35, 40, 45, 50) and interval (annual, hybrid, biennial, triennial) were evaluated for women differing in risk due to their family history and European-ancestry PRS. The benefits and harms of risk-based screening were compared to current age-based guidelines for biennial screening from 50-74.

Results

Under the most effective screening strategies, the estimated increase in life years gained and reduction in breast cancer mortality due to risk-based screening were 6% and 3% (FH), 19% and 11% (PRS), and 24% and 14% (PRS+FH). The predicted rate of false-positives and overdiagnoses of 917 and 14.5 per 1.000 women screened over their lifetimes for age-based screening increased to 997 and 14.9 (FH), 1154 and 15.9 (PRS), 1157 and 16.3 (PRS+FH).

Conclusion

European-ancestry women at increased risk due to family history or polygenic risk could consider risk-based screening strategies starting before age 50 depending on their attitude towards the harms and benefits of breast cancer screening.

INTRODUCTION

Regular mammography screening has been shown to reduce breast cancer mortality. (1) However, it is uncertain whether current age-based screening recommendations (2, 3) are optimal, as there is variability in breast cancer risk at any given age. The risk of developing breast cancer approximately doubles for women with a first-degree family member with breast cancer.(4) Approximately 20% of the familial risk is attributable to high- or moderate penetrance mutations in genes including BRCA1, BRCA2, PALB2, ATM, and CHEK2.(5, 6) The majority of the remaining 80% is due to a combination of more common variations in the DNA sequence, e.g., single nucleotide polymorphisms (SNPs). Currently, about 170 common breast cancer risk SNPs have been identified.(7) While these individual variants are associated with small to modest risks, their combined effects considered as a polygenic risk score (PRS) can be substantial and could achieve a level useful for population screening.(8, 9)

The U.S. Preventive Services Task Force recommends that women discuss their individual risk and screening options with their healthcare providers, yet there are limited data to inform such discussions. Two ongoing trials, My-PEBS and the WISDOM trial are presently testing age-based vs. risk-based screening approaches that include genetic markers and family history information, but results are not expected until 2024-2025. (10) A recent study modeled the use of polygenic risk scores to determine the cost-effectiveness of screening women triennially above a certain risk threshold in the UK.(11) However, there are no studies that have estimated the impact of screening strategies tailored to risk from family history of breast cancer, polygenic risk, or both. To fill this gap, two established Cancer Intervention and Surveillance Modeling Network (CISNET) models which were used to inform current breast cancer screening guidelines(12, 13), estimated the lifetime effects of screening based on family history status and polygenic risk.(2) The projections in this study are intended to inform screening policy and provide background for clinical discussions about risk-based breast cancer screening.

METHODS

Model overview

Breast cancer simulation models developed by the Erasmus University Medical Center (14) and the Georgetown University-Albert Einstein College of Medicine (15) evaluated the lifetime effects of different screening strategies among the 1985 U.S. female birth cohort. Model descriptions and detailed information on model inputs and validation have been described.(12, 16-18) (Appendix 1)

Briefly, both models used incidence from four breast cancer molecular subtypes based on estrogen receptor (ER) and human epidermal growth factor receptor (HER)2 status.(17) Screen-detection of breast cancer was modeled using digital mammography sensitivity and stage distributions reported by the Breast Cancer Surveillance Consortium.(19) Treatment impact was derived from systematically reviewed treatment effectiveness (20) and reduced the probability of death from breast cancer. At any time, women diagnosed with breast cancer could die of the disease or competing other cause mortality. To evaluate the potential efficacy of different screening strategies, the models assumed that all women received genetic testing and were screened according to the selected strategy, and, if diagnosed with cancer, received sub-type specific adjuvant therapy.

Screening strategies

Nineteen digital mammography screening strategies that varied by age at initiation of screening (30, 35, 40, 45, 50) and screening interval (annual, biennial, triennial, hybrid) were evaluated. Hybrid strategies screen annually before age 50 and biennially starting at age 50. All strategies stopped screening at age 74. The primary comparator was biennial screening from ages 50 to 74 as this strategy is supported in the screening guidelines in many developed countries. (2, 21, 22)

Risk stratification

Family history

We defined five family history groups: women who learned in age ranges 30-39, 40-49, 50-64, 65-100 that they had a first-degree relative with breast cancer; and women with no family history of breast cancer in their lifetimes. Using the age-specific distribution of family history in the National Health Interview Survey and associated risk levels observed in the Collaborative Breast Cancer Study (CBCS)(23), breast cancer risk was adjusted accordingly in the models.(Table 1)

Polygenic risk

Stratification based on polygenic risk used a polygenic risk score based on 77 SNPs, as defined by Mavaddat et al.(24) As a sensitivity analysis, we considered a PRS that included a larger number of common genetic variants (167 SNPs).(7) The polygenic risk scores are based on a multiplicative relative risk model for the joint effects of the SNPs, and are hence defined as the sum of risk alleles weighted by their effect size as estimated in the combined European ancestry Genome Wide Association Studies (GWAS) data.(7, 24) We established seven PRS groups spanning risk levels from 0 to 10 times the U.S. population average. (Table 1) Risk group prevalence was calculated by simulating the distribution of risk as a function of the PRS to match that of Mavaddat et al. Using the cut-off risk levels of the seven groups, we calculated the number of women in each group, for details see Appendix 2.

Table 1 Prevalence and relative risk (RR) according to polygenic risk score and family history of breast cancer.

Family history (FH) age groups **	**Risk relative to population average**	**% of all women**
FH positive between 30 and 39	2.19	4.7%
FH positive between 40 and 49	1.73	4.2%
FH positive between 50 and 64	1.39	5.9%
FH positive at age 65 or older	1.34	2.3%
No positive FH in life	0.79	82.9%
Polygenic risk groups *	**Risk relative to population average**	**% of all women**
Polygenic risk group 1	$0.0 < RR \leq 0.5$	9.5%
Polygenic risk group 2	$0.5 < RR \leq 1.0$	49.4%
Polygenic risk group 3	$1.0 < RR \leq 1.5$	27.7%
Polygenic risk group 4	$1.5 < RR \leq 2.0$	9.4%
Polygenic risk group 5	$2.0 < RR \leq 3.0$	3.5%
Polygenic risk group 6	$3.0 < RR \leq 5.0$	0.4%
Polygenic risk group 7	$5.0 < RR \leq 10.0$	0.0 % *

* Based on the 77-SNP polygenic risk score (23), very few women would have 5 to 10-fold increased breast cancer risk.
**A positive first-degree family history was modeled as an increase in risk at the first age-year of each age-group.

Analysis

In total we examined 47 potential risk-groups; five family history, seven polygenic risk, and 35 combinations of both. We first used model projections on the harms (overdiagnoses and false-positives), and benefits (life-years gained and breast cancer deaths averted) of biennial screening *average-risk* women aged 50-74 as the benchmark for the outcomes of current screening guidelines. Overdiagnosis was defined as screen-detected cases (invasive + in situ) that would not have been diagnosed in the absence of screening. Next, we estimated the harms and benefits of the 19 screening strategies (described above) in each risk group. Among these comparisons, we selected the set of strategies that maximized the overall number of life-years gained, while maintaining a similar, or better, ratio of screens to life-years gained as seen with the baseline approach of biennial screening all women from 50-74. This methodology insured that risk-based screening would only increase the number of screens if the associated life-years gained increased at least proportionally. The overall population impact was quantified by accumulating the harms and benefits of the individual risk groups.

Sensitivity Analyses

To test the impact of improved polygenic risk scores on the harms and benefits of risk-based screening, we conducted sensitivity analyses of a PRS derived from 167 independent SNPs from the largest breast cancer GWAS to date.(7) In addition, since part of

the additional benefits of risk-based screening may accrue from an increased number of screens, we analyzed what the impact of polygenic risk-based screening could be if the total number of screens was fixed to the number performed with guideline screening (13 screens per woman with biennial 50-74).

RESULTS

Age-based guidelines

If all women among the 1985 US birth cohort undergo age-based biennial screening from ages 50 to 74, the models project an average of 11,157 screens and 124 life-years gained (range across models: (103 – 146) and 7 (6.4 – 7.6) breast cancer deaths averted for 1.000 women screened over their lifetimes vs. no-screening. These results provide a benchmark of 90 screens per life year gained of current screening guidelines. This threshold was used to select risk-based strategies with equal or better trade-off between screens and life-years gained.

Risk-based screening: family history

Women with a known family history of breast cancer before they reach age 50 (8.9% of all women) were screened biennially starting at either age 30 or at 40, depending on the age at which they first learned about breast cancer in a first-degree relative.(Table 2) This was estimated to lead to 44% more life-years gained and 24% reduction in breast cancer deaths relative to current screening guidelines. However, overdiagnoses increased by 26% and the number of false-positives doubled. The overall impact of a family history-based screening approach was modest due to the low prevalence of breast cancer family history in the population: 0.2 fewer cancer deaths and 7 additional life-years per 1,000 women screened over a lifetime.

Risk-based screening: polygenic risk

Next, we considered screening strategies targeted to polygenic risk. The optimal strategy that was selected for each polygenic risk group is given in Table 3 and Figure 1. Overall, polygenic risk-based screening was estimated to increase the number of screens by 17%, life-years gained by 19% and reduced breast cancer deaths by 11% compared to screening all women biennially from ages 50 to 74. The harms of screening such as overdiagnoses and false positives increased by 10% and 26%. In absolute numbers, using polygenic risk to personalize screening strategies was estimated to lead to 0.7 fewer cancer deaths and 24 additional life-years per 1,000 women screened.(Table 3)

Table 2 Benefits and harms of mammography screening based on breast cancer family history. Outcomes presented as average of two models per 1,000 women screened.

Row	Risk group based on family history (FH) of breast cancer		Screening strategy	Number of screens	Life years gained *	Breast cancer deaths averted *	Over diagnoses	False positives	Screens / life year gained	Life years gained / overdiagnoses
1	*Average risk population*		Biennial 50-74	11157	124	7.0	14.5	917	90	8.6
2	Positive FH ages 30-39	Guideline	Biennial 50-74	10815	168	9.3	16.5	892	64	10.2
3		Risk-based strategy	Biennial 30-74	20528	254	11.9	21.7	2079	81	11.7
4			*% change (3) vs. (2)*	*90%*	*51%*	*28%*	*31%*	*133%*	*25%*	*16%*
5	Positive FH ages 40-49	Guideline	Biennial 50-74	10904	168	9.3	16.7	901	65	10.1
6		Risk-based strategy	Biennial 40-74	15713	228	11.3	20.3	1468	69	11.3
7			*% change (6) vs. (5)*	*44%*	*36%*	*20%*	*21%*	*63%*	*6%*	*12%*
8	Positive FH < age 50	Guideline	Biennial 50-74	10857	168	9.3	16.6	896	65	10.1
9	Risk-based strategies (rows 3,6)		Biennial 30/40-74	18256	242	11.6	21.0	1791	75	11.5
10			*change (9) vs. (8)*	*68%*	*44%*	*24%*	*26%*	*100%*	*17%*	*14%*
11	Positive FH 50-64	Guideline/risk-based	Biennial 50-74	11054	162	9.0	16.8	908	68	9.6
12	Positive FH 65+	Guideline/risk-based	Biennial 50-74	11185	129	7.5	16.8	915	87	7.7
13	No FH during life	Guideline/risk-based	Biennial 50-74	11236	105	5.8	15.7	919	107	6.7
14	**FH groups aggregated (rows 3,6,11,12,13)**			11813	131	7.2	14.9	997	90	8.8
15			*% change (14) vs. (1)*	*6%*	*6%*	*3%*	*3%*	*9%*	*0%*	*3%*

*The life-years gained and breast cancer deaths averted are relative to the life-years and breast cancer deaths of women at the same level of risk who are not screened.

Risk-based screening: polygenic risk and family history

Finally, we considered risk stratification by polygenic risk and family history simultaneously, defining 35 risk groups. (Table 4, optimal strategies given in Appendix 3) At the population level, risk-based screening using polygenic risk and family history was estimated to lead to 31 additional life years gained, 1 fewer breast cancer death and 1.8 additional overdiagnoses per 1,000 women screened.

Sensitivity Analyses

We also considered the effect of using an enhanced polygenic risk score of 167 SNPs instead of 77 SNPs. The percentage of women undergoing a different screening strategy based on the 167-SNP PRS distribution was small: 7.6%.(Appendix 2) The estimated number of mammograms decreased by approximately 1%, the number of overdiagnoses,

false positives, breast cancer deaths and life years gained remained virtually unchanged. (Table 4) Without increasing the number of mammograms of guideline (biennial 50-74) screening, a PRS screening approach based on 77 SNPs still gained 9% additional life years and 3% more breast cancer deaths averted.

Table 3 Benefits and harms of mammography screening based on polygenic risk. Outcomes presented as average of two models per 1,000 women screened.

Row	Risk group based on polygenic risk score (PRS)		Screening strategy	Number of screens	Life years gained *	Breast cancer deaths averted *	Over diagnoses	False positives	Screens / life year gained	Life years gained / overdiagnoses
1	*Average risk population*		Biennial 50-74	11157	124	7.0	14.5	917	90	8.6
2	PRS7 (5.0 < RR < 10.0)	Guideline	Biennial 50-74	8886	512	27.5	28.0	726	17	18.3
3	Risk-based strategy		Annual 30-74	35214	959	42.8	53.3	3648	37	18.0
4			*% change (3) vs. (2)*	*296%*	*87%*	*56%*	*90%*	*403%*	*112%*	*-2%*
5	PRS6 (3.0 < RR < 5.0)	Guideline	Biennial 50-74	9897	352	19.3	24.8	811	28	14.2
6	Risk-based strategy		Annual 35-74	32835	616	28.8	42.5	3253	53	14.5
7			*% change (6) vs. (5)*	*232%*	*75%*	*50%*	*72%*	*301%*	*89%*	*2%*
8	PRS5 (2.0 < RR < 3.0)	Guideline	Biennial 50-74	10469	252	13.9	21.0	859	42	12.0
9	Risk-based strategy		an40-50,bi50-74	19574	358	17.2	26.2	1954	55	13.7
10			*% change (9) vs. (8)*	*87%*	*42%*	*23%*	*25%*	*127%*	*31%*	*14%*
11	PRS4 (1.5 < RR < 2.0)	Guideline	Biennial 50-74	10844	183	10.2	17.7	891	59	10.3
12	Risk-based strategy		Biennial 40-74	15646	242	12.1	21.2	1462	65	11.4
13			*% change (12) vs. (11)*	*44%*	*32%*	*18%*	*20%*	*64%*	*9%*	*10%*
14	PRS3 (1.0 < RR < 1.5)	Guideline	Biennial 50-74	11091	137	7.6	15.2	912	81	9.0
15	Risk-based strategy		Biennial 40-74	15923	180	9.0	18.1	1487	89	9.9
16			*% change (15) vs. (14)*	*44%*	*32%*	*18%*	*19%*	*63%*	*9%*	*10%*
17	PRS2 (0.5 < RR < 1.0)	Guideline	Biennial 50-74	11332	90	5.1	12.4	932	126	7.3
18	Risk-based strategy		Biennial 50-74	11332	90	5.1	12.4	932	126	7.3
19			*% change (18) vs. (17)*	*0.0%*	*0.0%*	*0.0%*	*0.0%*	*0.0%*	*0.0%*	*0%*
20	PRS1 (0.0 < RR < 0.5)	Guideline	Biennial 50-74	11588	40	2.3	9.0	953	290	4.4
21	Risk-based strategy		Triennial 50-74	8020	34	1.9	8.3	705	238	4.1
22			*% change (21) vs (20)*	*-31%*	*-16%*	*-15%*	*-8%*	*-26%*	*-18%*	*-8%*
23	**PRS groups aggregated (rows 3,6,9,12,15,18,21)**			13011	148	7.7	15.9	1154	88	9.3
24			*% change (23) vs. (1)*	*17%*	*19%*	*11%*	*10%*	*26%*	*-2%*	*9%*

*The life-years gained and breast cancer deaths averted are relative to the life-years and breast cancer deaths of women at the same level of risk who are never screened.

Table 4 Benefits and harms comparison of mammography screening based on breast cancer family history, polygenic risk score, and family history combined with polygenic risk – for both the primary analysis and the sensitivity analyses. Outcomes presented as average of two models per 1,000 women screened.

Row	Risk group	Screening strategy	Number of screens	Life years gained *	Breast cancer deaths averted *	Over diagnoses	False positives	Screens / life year gained	Life years gained / overdiagnoses
1	*Average risk population*	Biennial 50-74	11157	124	7.0	14.5	917	90	8.6
2	*FH groups aggregated*	Risk-based strategies	11813	131	7.2	14.9	997	90	8.8
3		*% change (2) vs. (1)*	*6%*	*6%*	*3%*	*3%*	*9%*	*0%*	*3%*
4	PRS groups *aggregated*	Risk-based strategies	13011	148	7.7	15.9	1154	88	9.3
5		*% change (4) vs. (1)*	*17%*	*19%*	*11%*	*10%*	*26%*	*-2%*	*9%*
6	*PRS + FH groups aggregated*	Risk-based strategies	13032	155	8.0	16.3	1157	84	9.5
7		*% change (6) vs. (1)*	*17%*	*24%*	*14%*	*14%*	*26%*	*-6%*	*10%*
8	***Sensitivity Analyses***								
9	PRS groups *aggr. (167 SNPs)*	Risk-based strategies	12948	149	7.7	15.9	1150	87	9.4
10		*% change (9) vs. (1)*	*16%*	*20%*	*11%*	*10%*	*25%*	*-3%*	*9%*
11	*PRS + FH aggr. (167 SNPs)*	Risk-based strategies	12722	154	7.9	16.2	1128	83	9.5
12		*% change (11) vs. (1)*	*14%*	*24%*	*14%*	*12%*	*23%*	*-8%*	*11%*
13	*Redistributing screens* **	Risk-based strategies	11172	135	7.2	14.2	968	85.7	10.3
14		*% change (13) vs. (1)*	*0%*	*9%*	*3%*	*-2%*	*6%*	*-7%*	*13%*

**The life-years gained* and breast cancer deaths averted are relative to the life-years and breast cancer deaths of women at the same level of risk who are never screened.

** The redistributing screens scenario represents a scenario where the number of screens of current screening guidelines (in row 1) is not *increased, rather redistributed across the population based on the polygenic risk scores. The selection of screening strategies in this scenario are given in Appendix 2.*

DISCUSSION

This is the first modeling study to quantify the harms and benefits of breast cancer screening based on polygenic risk and family history. Using two tumor type-specific natural history models including sensitivity and specificity of digital mammography, we show that risk-based screening has greater projected benefits when based on polygenic risk scores than family history. The screening approach combining polygenic risk scores and family history resulted in the maximal improvement in outcomes compared to current

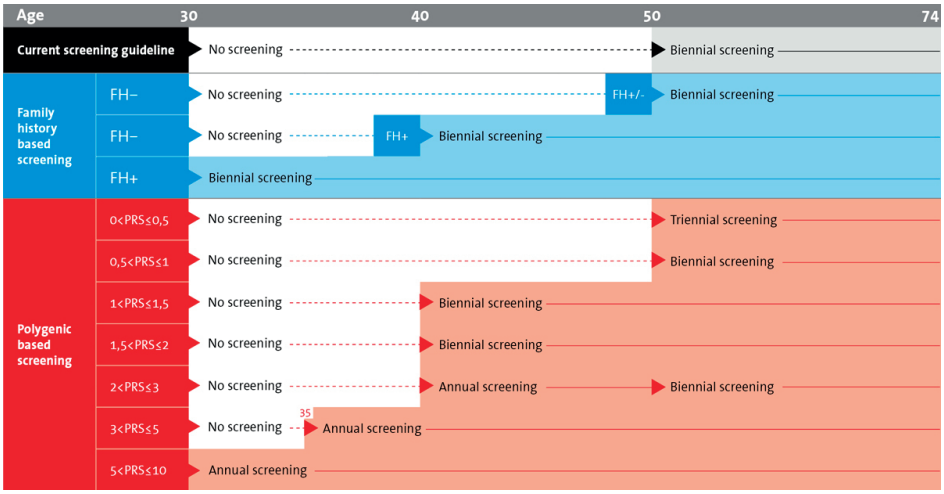


Figure 1 Selected screening strategies in the family history (blue), and polygenic risk (red) screening approaches. Starting age and interval of screening, as well as family history status may change by age (top row).
The screening strategies selected for the 35 risk-groups in the polygenic risk combined with breast cancer family history screening approach can be found in Appendix 3.

age-based screening guidelines. The risk group-specific outcomes suggest that high risk women could initiate screening at an earlier age, and that women with below-average risk could consider screening at longer intervals than current age-based guidelines. Inclusion of additional, more recently identified SNP into the models only modestly improved the benefits and harms. With the inclusion of more SNPs in the future, there is still potential for a PRS to further improve the discriminatory performance. Notably, polygenic risk used in combination with breast density, and reproductive, lifestyle, and hormonal factors is likely to improve risk prediction and the harm-benefit ratio for stratified screening.(25)

Current age-based guidelines recommend that women should discuss screening with healthcare providers to select the best approach for their individual risk.(2, 21) Our analysis extends this advice by providing specific screening strategies that could be considered in practice based on genetic risk factors. Our data suggest that among higher than average-risk women (i.e., twice the average population risk), initiating screening at an earlier age (<50) is likely to provide greater benefits than harm.

Our results are consistent with our previous work on risk-based screening based on more common classical risk factors (26, 27) and prior research in other countries. In Spain, Vilaprinyo and colleagues performed an analysis using four risk-groups based on breast density, family history, and personal history of breast biopsy to guide screening for women aged 40-85.(28) Recently, Pashayan used a life-table model to assess risk-based screening for women 50-85 in the United Kingdom based on polygenic risk profile.(11)

Like our results, both studies concluded that risk-based screening strategies were more efficient and had lower harm-benefit ratios than age-based screening.

While our results, and the results of others lend support to risk-based screening, our approach was unique in evaluating whether the associated increases in benefits were merely attributed to the increase in the number of screening examinations. In the sensitivity analyses, we demonstrated that, redistributing the guideline-concordant number of screens across all women, increased life-years gained and breast cancer deaths averted more than overdiagnoses and false-positives would increase. However, this also showed that a large part of the projected benefit-increase was explained by the greater number of mammograms as screening increases cancer detection.

Implementing breast cancer screening based on polygenic risk and family history status would require a one-time saliva sample to establish a polygenic risk profile. The result, together with a questionnaire about family history could assist women in making choices about more personalized screening options. Knowledge of genetic susceptibility to breast cancer could guide early detection strategies. However, ethical aspects of genetic testing such as patient autonomy, accessibility, possible (unknown) differential effects across ancestries, should be considered before the implementation or recommendation of polygenic risk-based screening. Overall, it will be essential to develop and evaluate polygenic risk models for non-European ancestry women.

This study has several important strengths including consistent results across two well-established simulation models, use of U.S. national data, and evaluation of polygenic risk and family history information to personalize breast cancer screening. There are also several caveats that should be considered in evaluating the results. First, we did not explicitly model the effects of rare but higher risk variants in genes such as BRCA1, BRCA2, PALB2, CHEK2 or ATM that could be used to tailor screening strategies. Mutations in genes BRCA1 and BRCA2 confer exceptionally high risk, and carriers of mutations in these genes are typically advised to undergo annual screening with both MRI and mammography, starting at an early age.(29) MRI has higher sensitivity than mammography but is associated with a higher false-positive rate. We anticipate that if MRI were to be used as screening modality in the setting of higher than average polygenic risk, projected benefits would be larger but false-positives and possibly overdiagnoses would increase as well. Second, while we account for differing tumor natural history by ER/HER2 tumor status, the models assumed that polygenic risk did not affect tumor type and tumor progression since there are insufficient data to inform modeling of variation in natural history. However, an increasing number of SNP associations are known to differ by tumor subtype, particularly ER-status (30, 31), and there is some data showing that the PRS has differential effects by mode of detection (32). It is possible in due course that a separate PRS consisting of SNPs associated with faster or slower growing tumors may inform screening intensity. Third, we did not explicitly consider second degree family members

with breast cancer, nor the use of breast density or other risk factors which may have potential value for risk-based screening.(26) Fourth, to test the efficacy of risk-based screening, we assumed 100% uptake of genetic testing, screening, and treatment. Fifth, the effectiveness of screening in combination with treatment in women under age 40 has been assessed in case-control studies, but not in a randomized controlled trial. Finally, the PRS used in this study was developed using data primarily from Caucasian women of European ancestry. Screening strategies should be re-assessed for minority groups as genetic databases evolve.

Overall, this research showed that more breast cancer deaths could be prevented and lives extended for select, but identifiable, groups of women at high risk due to their family history of breast cancer and polygenic risk. These results are intended to inform continued debates about optimal breast cancer screening strategies and could begin to guide patient-provider discussions in routine clinical practice.

REFERENCES

1. Independent U. K. Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet. 2012;380(9855):1778-86.
2. Siu AL, On behalf of the U. S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2016;164(4):279-96.
3. Viale PH. The American Cancer Society Guidelines on Screening for Breast Cancer: What's New? J Adv Pract Oncol. 2015;6(6):508-10.
4. Collaborative Group on Hormonal Factors in Breast C. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet. 2001;358(9291):1389-99.
5. Thompson D, Easton D. The genetic epidemiology of breast cancer genes. Journal of mammary gland biology and neoplasia. 2004;9(3):221-36.
6. Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, et al. Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med. 2015;372(23):2243-57.
7. Michailidou K, Lindstrom S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017;551(7678):92-4.
8. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med. 2008;358(26):2796-803.
9. Burton H, Chowdhury S, Dent T, Hall A, Pashayan N, Pharoah P. Public health implications from COGS and potential for risk stratification and screening. Nat Genet. 2013;45(4):349-51.
10. Esserman LJ, Study W, Athena I. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. NPJ Breast Cancer. 2017;3:34.
11. Pashayan N, Morris S, Gilbert FJ, Pharoah PDP. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. JAMA Oncol. 2018.
12. Mandelblatt JS, Stout NK, Schechter CB, van den Broek JJ, Miglioretti DL, Krapcho M, et al. Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies. Ann Intern Med. 2016;164(4):215-25.
13. Mandelblatt JS, Cronin KA, Bailey S, Berry DA, Koning H J de, Draisma G, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms Ann Int Med. 2009;151:738-47.
14. van den Broek JJ, van Ravesteyn NT, Heijnsdijk EA, de Koning HJ. Estimating the effects of risk-based screening and adjuvant treatment using the MISCAN-Fadia continuous tumor growth model for breast cancer. Medical Decision Making. 2018.
15. Schechter CB, Near AM, Jayasekera J, Chang Y, Mandelblatt JS. Structure, Function, and Applications of the Georgetown-Einstein (GE) Breast Cancer Simulation Model Medical Decision Making. 2018.
16. van den Broek JJ, van Ravesteyn NT, Mandelblatt JS, Huang H, Ergun MA, Burnside ES, et al. Comparing CISNET Breast Cancer Incidence and Mortality Predictions to Observed Clinical Trial Results of Mammography Screening from Ages 40 to 49. Med Decis Making. 2018;38(1_suppl):140S-50S.

17. Plevritis SK, Munoz D, Kurian AW, Stout NK, Alagoz O, Near AM, et al. Association of Screening and Treatment With Breast Cancer Mortality by Molecular Subtype in US Women, 2000-2012. JAMA. 2018;319(2):154-64.
18. Mandelblatt JS, Near AM, Miglioretti DL, Munoz D, Sprague BL, Trentham-Dietz A, et al. Common Model Inputs used in CISNET Collaborative Breast Cancer Modeling Medical Decision Making. 2018.
19. Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DS, Kerlikowske K, et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. Radiology. 2017;283(1):49-58.
20. Early Breast Cancer Trialists' Collaborative G, Peto R, Davies C, Godwin J, Gray R, Pan HC, 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.
21. Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. JAMA. 2015;314(15):1599-614.
22. Ebell MH, Thai TN, Royalty KJ. Cancer screening recommendations: an international comparison of high income countries. Public Health Rev. 2018;39:7.
23. Shiyanbola OO, Arao RF, Miglioretti DL, Sprague BL, Hampton JM, Stout NK, et al. Emerging Trends in Family History of Breast Cancer and Associated Risk. Cancer Epidemiol Biomarkers Prev. 2017;26(12):1753-60.
24. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 2015;107(5).
25. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med. 2019.
26. Trentham-Dietz A, Kerlikowske K, Stout NK, Miglioretti DL, Schechter CB, Ergun MA, et al. Tailoring Breast Cancer Screening Intervals by Breast Density and Risk for Women Aged 50 Years or Older: Collaborative Modeling of Screening Outcomes. Ann Intern Med. 2016;165(10):700-12.
27. van Ravesteyn NT, Miglioretti DL, Stout NK, Lee SJ, Schechter CB, Buist DS, et al. Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: a comparative modeling study of risk. Ann Intern Med. 2012;156(9):609-17.
28. Vilaprinyo E, Forne C, Carles M, Sala M, Pla R, Castells X, et al. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. PLoS One. 2014;9(2):e86858.
29. Paluch-Shimon S, Cardoso F, Sessa C, Balmana J, Cardoso MJ, Gilbert F, et al. Prevention and screening in BRCA mutation carriers and other breast/ovarian hereditary cancer syndromes: ESMO Clinical Practice Guidelines for cancer prevention and screening. Ann Oncol. 2016;27(suppl 5):v103-v10.
30. Shimelis H, LaDuca H, Hu C, Hart SN, Na J, Thomas A, et al. Triple-Negative Breast Cancer Risk Genes Identified by Multigene Hereditary Cancer Panel Testing. J Natl Cancer Inst. 2018.
31. Milne RL, Kuchenbaecker KB, Michailidou K, Beesley J, Kar S, Lindstrom S, et al. Identification of ten variants associated with risk of estrogen-receptor-

negative breast cancer. Nat Genet. 2017;49(12):1767-78.

32. Li J, Ugalde-Morales E, Wen WX, Decker B, Eriksson M, Torstensson A, et al. Differential Burden of Rare and Common Variants on Tumor Characteristics, Survival, and Mode of Detection in Breast Cancer. Cancer Res. 2018;78(21):6329-38.

APPENDIX

1. **Model descriptions**
2. **Prevalence and relative risk calculations for the polygenic risk score groups**
3. **Strategies in the combined polygenic risk score and family history screening approach**
4. **Individual model predictions**

APPENDIX 1

Model Descriptions

The Institutional Review Board at Georgetown University approved the study as exempt based on the use of de-identified data. Detailed descriptions of model inputs and model validation have been described in detail elsewhere.(11, 17-19)

Model Erasmus University Medical Center (Model E)

Model E, also known as MISCAN-Fadia which is an acronym for Microsimulation Screening Analysis – Fatal Diameter is a breast cancer simulation model that uses continuous tumor growth to simulate the natural history of breast cancer. The model simulates individual life histories from birth to death, with and without breast cancer, in the presence and in the absence of screening and treatment. Life histories are simulated according to discrete events such as birth, tumor inception, metastasis, and death from breast cancer or death from other causes. Model E consists of four main components: demography, natural history of breast cancer, screening, and treatment. Screening impact on the natural history of breast cancer is assessed by simulating continuous tumor growth and the "fatal diameter" concept. This concept implies that tumors diagnosed at a size that is between the screen detection threshold and the fatal diameter are cured, while tumors diagnosed at a diameter larger than the fatal tumor diameter metastasize and lead to breast cancer death. MISCAN-Fadia includes different natural histories for molecular subtypes based on a tumor's ER status and HER-2 status.

Model Georgetown University-Albert Einstein College of Medicine (Model GE)

Model GE is a continuous-time, event-driven microsimulation of single-life histories of women utilizing a parallel universes approach. The parallel universes approach starts with the generation of a basic life history for each simulated woman in the absence of any screening or adjuvant treatment. The effects of each screening and adjuvant treatment strategy under study are then simulated starting using the exact same basic life

history. In this manner, the outputs for the different screening and adjuvant treatment strategies are matched pairs. The approach for simulating breast cancer natural history is phenomenological, relying on dates, stage, and age of clinical and screen detection for a tumor molecular subtype without explicitly modeling tumor growth. The model accommodates differences in natural history associated with estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) biomarkers, as well as conventional breast cancer risk factors. Breast cancer incidence depends on age, time period, and birth cohort, and is modified based on risk. The incidence includes a subset of ductal carcinoma in situ (DCIS) tumors that never surface clinically and eventually regress.

APPENDIX 2

Prevalence and relative risk calculation for polygenic risk score groups (77-SNPs PRS)

We modeled the distribution of risk relative to the average woman without a family history (RR*) as a function of polygenic risk and family history:

$$RR^* = Lognormal\left(FH\chi\left(\mu_i + \frac{\sigma_i}{2}\right) - \left(\frac{\sigma_i}{2}\right)^2 , \sigma_i\right)$$

Where FHx is an indicator for first degree family history of breast cancer (yes=1, no=0), γ is the log relative risk of family history (adjusted for polygenic effects), and σ_i is the log relative risk associated with a one standard deviation change in the polygenic risk score in age group i.

We used the following parameter values for μ_i and σ_i from Table 3 in Mavaddat (2015) JNCI.

	Sigma (σ)		
Age	**All cancers**	**ER+**	**ER-**
< 40	0.46 (0.38 – 0.53)	0.56 (0.47 – 0.65)	0.48 (0.36 – 0.59)
40-49	0.46 (0.42 – 0.50)	0.53 (0.48 – 0.57)	0.36 (0.29 – 0.43)
50-59	0.48 (0.45 – 0.51)	0.54 (0.50 – 0.57)	0.37 (0.32 – 0.43)
≥60	0.41 (0.38 – 0.43)	0.44 (0.41 – 0.47)	0.36 (0.31 – 0.42)
All ages	*0.44 (0.42 – 0.46)*	*0.49 (0.47 - 0.51)*	*0.37 (0.34 – 0.40)*

To model the distribution of risk relative to the population average, we consider RR = RR*/mean(RR*), where the mean of RR* is taken over the joint distribution of family history and polygenic risk in the population. The distributions of risk relative to the population average for various subgroups are displayed in the following figures:

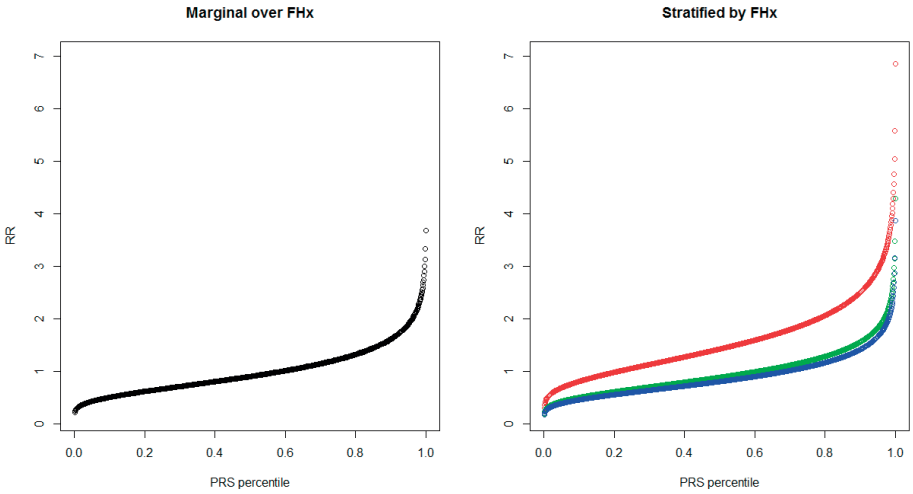


Figure 1: Distribution of breast cancer relative risk as a function of PRS (left) and family history + PRS (right). Red represents FH+ < age 40 women, green: FH+ between ages 40 and 65, blue: no FH in life.

In the screening approach based on a polygenic risk score **(77-SNPs)** combined with family history of breast cancer, the prevalence distribution of women among the risk-groups is as follows:

	Polygenic risk score groups (low to high polygenic risk)						
FH group	**0.0-0.5**	**0.5-1.0**	**1.0-1.5**	**1.5-2.0**	**2.0-3.0**	**3.0-5.0**	**5.0-10.0**
30 < FH+ < 40	0.9%	20.2%	33.5%	23.4%	17.6%	4.2%	0.3%
40 < FH+ < 49	0.9%	20.3%	33.5%	23.3%	17.5%	4.3%	0.3%
50 < FH+ < 64	9.6%	51.0%	27.7%	8.4%	3.0%	0.3%	0.0%
65 < FH+ < 100	9.7%	50.9%	27.7%	8.4%	3.0%	0.3%	0.0%
No FH in life	14.0%	54.8%	23.3%	5.9%	1.8%	0.2%	0.0%

Sensitivity analyses

In the sensitivity analyses screening based on a polygenic risk score consisting of 167 SNPs instead of 77 SNPs, we used a sigma (σ) of 0.48 instead of 0.44. The prevalence distribution of women among the polygenic risk groups based on the 77-SNP and 167-SNP polygenic risk score is as follows. The total percentage of women who end up in a different risk group based on the 167-SNP PRS compared to the 77-SNP PRS is 7.6%.

Polygenic risk group	77-SNPs	167-SNPs	Screening strategy
Polygenic risk group 1 (0.0 < Relative Risk < 0.5)	9.5%	12.3%	Triennial 50-74
Polygenic risk group 2 (0.5 < Relative Risk < 1.0)	49.4%	47.3%	Biennial 50-74
Polygenic risk group 3 (1.0 < Relative Risk < 1.5)	27.7%	26.0%	Biennial 40-74
Polygenic risk group 4 (1.5 < Relative Risk < 2.0)	9.4%	9.4%	Biennial 40-74
Polygenic risk group 5 (2.0 < Relative Risk < 3.0)	3.5%	4.3%	Hybrid 40-74
Polygenic risk group 6 (3.0 < Relative Risk < 5.0)	0.4%	0.6%	Annual 35-74
Polygenic risk group 7 (5.0 < Relative Risk < 10.0)	0.0 % *	0.0 % *	Annual 30-74

* Based on the 77-SNP and 167-SNP polygenic risk score [21], practically zero women would have 5 to 10-fold increased breast cancer risk.

In the screening approach based on a polygenic risk score **(167-SNPs)** combined with family history of breast cancer, the prevalence distribution of women among the risk-groups is as follows:

	Polygenic risk score groups (low to high polygenic risk)						
FH group	**0.0-0.5**	**0.5-1.0**	**1.0-1.5**	**1.5-2.0**	**2.0-3.0**	**3.0-5.0**	**5.0-10.0**
30 < FH+ < 40	1.6%	22.1%	31.3%	21.6%	17.6%	5.4%	0.5%
40 < FH+ < 49	1.6%	22.1%	31.4%	21.6%	17.5%	5.4%	0.5%
50 < FH+ < 64	12.0%	48.5%	26.2%	9.0%	3.8%	0.5%	0.1%
65 < FH+ < 100	12.0%	48.6%	26.1%	9.0%	3.8%	0.6%	0.0%
No FH in life	17.5%	51.9%	21.8%	6.3%	2.3%	0.3%	0.0%

In the screening approach that redistributed the number of screens of biennial 50-74 across the different polygenic risk groups, the following screening strategies were selected:

Polygenic risk group	Screening strategy
Polygenic risk group 1 (0.0 < Relative Risk < 0.5)	No screening
Polygenic risk group 2 (0.5 < Relative Risk < 1.0)	Biennial 50-74
Polygenic risk group 3 (1.0 < Relative Risk < 1.5)	Biennial 45-74
Polygenic risk group 4 (1.5 < Relative Risk < 2.0)	Biennial 45-74
Polygenic risk group 5 (2.0 < Relative Risk < 3.0)	Hybrid 40-74
Polygenic risk group 6 (3.0 < Relative Risk < 5.0)	Hybrid 40-74
Polygenic risk group 7 (5.0 < Relative Risk < 10.0)	Annual 30-74

APPENDIX 3

Selected screening strategies in the approach combining polygenic risk scores (PRS) with family history (FH) of breast cancer.

Risk group		Screening strategy
PRS group 1 (0.0 < RR < 0.5)	FH positive between 30 and 39	Triennial 30-74
	FH positive between 40 and 49	Triennial 40-74
	FH positive between 50 and 64	Triennial 50-74
	FH positive at age 65 or older	Triennial 50-74
	No positive FH in life	Triennial 50-74
PRS group 2 (0.5 < RR < 1.0)	FH positive between 30 and 39	Biennial 35-74
	FH positive between 40 and 49	Biennial 40-74
	FH positive between 50 and 64	Biennial 50-74
	FH positive at age 65 or older	Biennial 50-74
	No positive FH in life	Biennial 50-74
PRS group 3 (1.0 < RR < 1.5)	FH positive between 30 and 39	Biennial 30-74
	FH positive between 40 and 49	Biennial 40-74
	FH positive between 50 and 64	Biennial 45-74
	FH positive at age 65 or older	Biennial 45-74
	No positive FH in life	Biennial 45-74
PRS group 4 (1.5 < RR < 2.0)	FH positive between 30 and 39	Annual 30-50 + Biennial 50-74
	FH positive between 40 and 49	Annual 40-50 + Biennial 50-74
	FH positive between 50 and 64	Biennial 40-74
	FH positive at age 65 or older	Biennial 40-74
	No positive FH in life	Biennial 40-74
PRS group 5 (2.0 < RR < 3.0)	FH positive between 30 and 39	Annual 30-74
	FH positive between 40 and 49	Annual 40-74
	FH positive between 50 and 64	Annual 40-50 + Biennial 50-74
	FH positive at age 65 or older	Annual 40-50 + Biennial 50-74
	No positive FH in life	Annual 40-50 + Biennial 50-74
PRS group 6 (3.0 < RR < 5.0)	FH positive between 30 and 39	Annual 30-74
	FH positive between 40 and 49	Annual 35-74
	FH positive between 50 and 64	Annual 35-74
	FH positive at age 65 or older	Annual 35-74
	No positive FH in life	Annual 35-74
PRS group 7 (5.0 < RR < 10.0)	FH positive between 30 and 39	Annual 30-74
	FH positive between 40 and 49	Annual 30-74
	FH positive between 50 and 64	Annual 30-74
	FH positive at age 65 or older	Annual 30-74
	No positive FH in life	Annual 30-74

APPENDIX 4

Benefits and harms projections of mammography screening based on breast cancer family history. Outcomes for model E (Erasmus) per 1,000 women screened.

Risk group based on family history (FH) of breast cancer	Screening strategy	Number of screens	Life years gained	Breast cancer deaths averted	Over diagnoses	False positives
Positive FH ages 30-39	Biennial 50-74	10754	141.5	8.6	21.2	889
Risk-based strategy	Biennial 30-74	20466	212.1	10.9	28.5	2083
Positive FH ages 40-49	Biennial 50-74	10859	141.3	8.6	21.4	897
Risk-based strategy	Biennial 40-74	15670	182.7	10.0	26.4	1472
Positive FH ages 50-64	Biennial 50-74 (risk-based)	11025	126.0	7.9	21.6	911
Risk-based strategy	Biennial 40-74 (risk-based)	11143	104.3	6.6	21.6	911
Positive FH ages 65+	Biennial 50-74 (risk-based)	11197	88.3	5.4	19.3	920

Benefits and harms projections of mammography screening based on breast cancer family history. Outcomes for model GE (Georgetown-Einstein) per 1,000 women screened.

Risk group based on family history (FH) of breast cancer	Screening strategy	Number of screens	Life years gained	Breast cancer deaths averted	Over diagnoses	False positives
Positive FH ages 30-39	Biennial 50-74	10875	194.7	10.0	11.9	895
Risk-based strategy	Biennial 30-74	20590	296.6	13.0	14.9	2076
Positive FH ages 40-49	Biennial 50-74	10949	195.2	10.1	12.1	906
Risk-based strategy	Biennial 40-74	15755	274.3	12.5	14.1	1464
Positive FH ages 50-64	Biennial 50-74 (risk-based)	11084	197.2	10.2	12.1	906
Risk-based strategy	Biennial 40-74 (risk-based)	11227	154.2	8.4	12.1	919
Positive FH ages 65+	Biennial 50-74 (risk-based)	11275	121.6	6.2	12.1	919

Benefits and harms projections of mammography screening based on polygenic risk scores. Outcomes for model E (Erasmus) per 1,000 women screened.

Risk group based on polygenic risk score (PRS)	Screening strategy	Number of screens	Life years gained	Breast cancer deaths averted	Over diagnoses	False positives
PRS7 (5.0 < RR < 10.0)	Biennial 50-74	9009	361	21.3	44.4	750
Risk-based strategy	Annual 30-74	35105	763	36.1	91.9	3668
PRS6 (3.0 < RR < 5.0)	Biennial 50-74	9889	267	16.2	37.9	820
Risk-based strategy	Annual 35-74	32616	517	26.0	70.4	3255
PRS5 (2.0 < RR < 3.0)	Biennial 50-74	10429	200	12.2	30.2	863
Risk-based strategy	An40-50,bi50-74	19165	281	14.6	38.6	1930
PRS4 (1.5 < RR < 2.0)	Biennial 50-74	10799	148	9.1	23.5	892
Risk-based strategy	Biennial 40-74	15604	191	10.6	28.3	1467
PRS3 (1.0 < RR < 1.5)	Biennial 50-74	11048	113	7.0	18.4	912
Risk-based strategy	Biennial 40-74	15889	145	8.1	21.9	1490
PRS2 (0.5 < RR < 1.0)	Biennial 50-74	11298	75	4.7	12.6	932
Risk-based strategy	Biennial 50-74	11298	75	4.7	12.6	932
PRS1 (0.0 < RR < 0.5)	Biennial 50-74	11567	34	2.1	5.8	953
Risk-based strategy	Triennial 50-74	8005	27	1.7	4.5	758

Benefits and harms projections of mammography screening based on polygenic risk scores. Outcomes for model GE (Georgetown-Einstein) per 1,000 women screened.

Risk group based on polygenic risk score (PRS)	Screening strategy	Number of screens	Life years gained	Breast cancer deaths averted	Over diagnoses	False positives
PRS7 (5.0 < RR < 10.0)	Biennial 50-74	8763	664	33.6	11.6	702
Risk-based strategy	Annual 30-74	35323	1154	49.4	14.6	3628
PRS6 (3.0 < RR < 5.0)	Biennial 50-74	9905	436	22.4	11.7	802
Risk-based strategy	Annual 35-74	33054	715	31.7	14.7	3252
PRS5 (2.0 < RR < 3.0)	Biennial 50-74	10509	303	15.7	11.8	855
Risk-based strategy	An40-50,bi50-74	19982	436	19.8	13.8	1979
PRS4 (1.5 < RR < 2.0)	Biennial 50-74	10890	218	11.3	11.9	889
Risk-based strategy	Biennial 40-74	15688	292	13.6	14.1	1458
PRS3 (1.0 < RR < 1.5)	Biennial 50-74	11133	160	8.3	12.0	911
Risk-based strategy	Biennial 40-74	15957	215	10.0	14.4	1483
PRS2 (0.5 < RR < 1.0)	Biennial 50-74	11367	105	5.5	12.1	932
Risk-based strategy	Biennial 50-74	11367	105	5.5	12.1	932
PRS1 (0.0 < RR < 0.5)	Biennial 50-74	11609	46	2.4	12.3	953
Risk-based strategy	Triennial 50-74	8035	41	2.1	12.1	651

Benefits and harms projections of mammography screening based on polygenic risk scores and breast cancer family history combined. Outcomes for model E (Erasmus) per 1,000 women screened.

Risk group based on polygenic risk score (PRS) and breast cancer family history (FH)		Screening strategy	Number of screens	Life years gained	Breast cancer deaths averted	Over diagnoses	False positives
PRS group 7 (5.0 < RR < 10.0)							
FH+ ages 30-39		Biennial 50-74	8093	313	18.2	37	675
	Risk-based strategy	Annual 30-74	32362	903	38.5	96	3419
FH+ ages 40-49		biennial 50-74	8413	335	19.4	41	701
	Risk-based strategy	Annual 30-74	33415	826	37.1	99	3522
FH+ ages 50-64		biennial 50-74	8876	327	19.4	47	740
	Risk-based strategy	Annual 30-74	34847	720	34.2	97	3650
FH+ ages 65+		biennial 50-74	9169	288	17.5	46	763
	Risk-based strategy	Annual 30-74	35564	659	31.5	93	3708
PRS group 6 (3.0 < RR < 5.0)							
FH+ ages 30-39		biennial 50-74	9117	258	15.4	37	758
	Risk-based strategy	Annual 30-74	35293	663	29.9	83	3678
FH+ ages 40-49		biennial 50-74	9365	268	15.9	40	778
	Risk-based strategy	Annual 35-74	31196	587	28.2	82	3135
FH+ ages 50-64		biennial 50-74	9744	250	15.3	43	809
	Risk-based strategy	Annual 35-74	32308	501	25.4	79	3233
FH+ ages 65+		biennial 50-74	9999	215	13.4	41	829
	Risk-based strategy	Annual 35-74	32925	444	22.7	73	3282
PRS group 5 (2.0 < RR < 3.0)							
FH+ ages 30-39		biennial 50-74	9812	204	12.4	33	814
	Risk-based strategy	Annual 30-74	37225	496	23.1	68	3845
FH+ ages 40-49		biennial 50-74	10005	207	12.6	34	829
	Risk-based strategy	Annual 40-74	28112	411	20.9	63	2671
FH+ ages 50-64		biennial 50-74	10301	191	11.8	36	854
	Risk-based strategy	an40-50,bi50-74	19055	269	14.2	44	1922
FH+ ages 65+		biennial 50-74	10507	161	10.2	33	870
	Risk-based strategy	an40-50,bi50-74	19275	235	12.4	40	1940
PRS group 4 (1.5 < RR < 2.0)							
FH+ ages 30-39		biennial 50-74	10319	158	9.7	27	854
	Risk-based strategy	an30-50,bi50-74	28744	309	14.1	42	3146
FH+ ages 40-49		biennial 50-74	10466	159	9.7	28	866
	Risk-based strategy	an40-50,bi50-74	19196	246	12.4	37	1933
FH+ ages 50-64		biennial 50-74	10694	145	9.0	29	885
	Risk-based strategy	Biennial 40-74	15501	184	10.4	34	1459

FH+ ages 65+		biennial 50-74	10855	121	7.7	26	897
	Risk-based strategy	Biennial 40-74	15672	159	9.0	30	1473
PRS group 3 (1.0 < RR < 1.5)							
FH+ ages 30-39		biennial 50-74	10319	158	9.7	27	854
	Risk-based strategy	Biennial 30-74	19926	263	13.0	38	2037
FH+ ages 40-49		biennial 50-74	10466	159	9.7	28	866
	Risk-based strategy	Biennial 40-74	15212	220	11.8	35	1435
FH+ ages 50-64		biennial 50-74	10694	145	9.0	29	885
	Risk-based strategy	Biennial 45-74	13370	171	10.2	33	1199
FH+ ages 65+		biennial 50-74	10855	121	7.7	26	897
	Risk-based strategy	Biennial 45-74	13551	145	8.8	30	1213
PRS group 2 (0.5 < RR < 1.0)							
FH+ ages 30-39		biennial 50-74	11041	85	5.2	16	911
	Risk-based strategy	Biennial 35-74	18652	132	6.9	21	1833
FH+ ages 40-49		biennial 50-74	11117	84	5.2	16	918
	Risk-based strategy	Biennial 40-74	15964	113	6.2	19	1496
FH+ ages 50-64		biennial 50-74	11238	75	4.7	16	927
	Risk-based strategy	biennial 50-74	11238	75	4.7	16	927
FH+ ages 65+		biennial 50-74	11325	62	4.0	14	934
	Risk-based strategy	biennial 50-74	11325	62	4.0	14	934
PRS group 1 (0.0 < RR < 0.5)							
FH+ ages 30-39		biennial 50-74	11447	39	2.4	8	944
	Risk-based strategy	Triennial 30-74	14742	47	2.5	8	1724
FH+ ages 40-49		biennial 50-74	11483	39	2.4	8	946
	Risk-based strategy	Triennial 40-74	10983	38	2.1	7	1218
FH+ ages 50-64		biennial 50-74	11538	35	2.2	7	951
	Risk-based strategy	Triennial 50-74	7987	26	1.7	6	756
FH+ ages 65+		biennial 50-74	11578	28	1.8	7	954
	Risk-based strategy	Triennial 50-74	8012	22	1.4	5	758

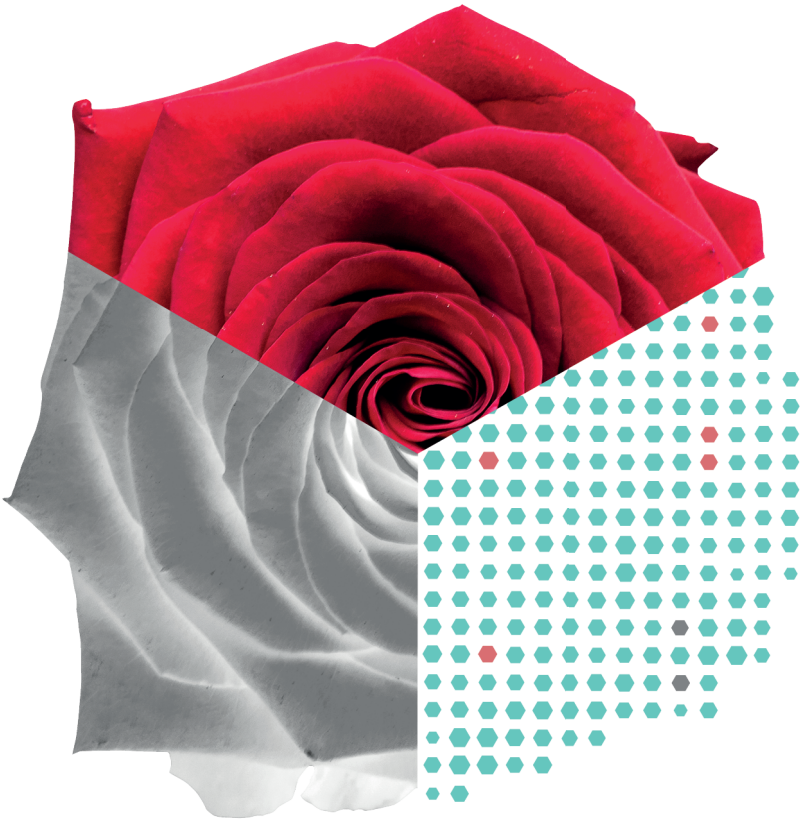
Benefits and harms projections of mammography screening based on polygenic risk scores and breast cancer family history combined. Outcomes for model GE (Georgetown-Einstein) per 1,000 women screened.

Risk group based on polygenic risk score (PRS) and breast cancer family history (FH)		Screening strategy	Number of screens	Life years gained	Breast cancer deaths averted	Over diagnoses	False positives
PRS group 7 (5.0 < RR < 10.0)							
FH+ ages 30-39		biennial 50-74	7417	774	38.6	11	589
	Risk-based strategy	Annual 30-74	31822	1550	62.5	14	3293
FH+ ages 40-49		biennial 50-74	7716	805	40.2	11	612
	Risk-based strategy	Annual 30-74	32912	1513	62.7	14	3407
FH+ ages 50-64		biennial 50-74	8289	865	43.2	12	658
	Risk-based strategy	Annual 30-74	34391	1381	60.4	15	3550
FH+ ages 65+		biennial 50-74	8904	702	37.2	13	713
	Risk-based strategy	Annual 30-74	35681	1155	52.1	16	3663
PRS group 6 (3.0 < RR < 5.0)							
FH+ ages 30-39		biennial 50-74	8979	545	27.6	11	723
	Risk-based strategy	Annual 30-74	35662	1030	42.8	14	3653
FH+ ages 40-49		biennial 50-74	9188	557	28.2	11	740
	Risk-based strategy	Annual 35-74	31419	975	41.9	14	3102
FH+ ages 50-64		biennial 50-74	9579	580	29.4	12	772
	Risk-based strategy	Annual 35-74	32409	875	39.7	15	3198
FH+ ages 65+		biennial 50-74	9996	462	24.9	12	809
	Risk-based strategy	Annual 35-74	33282	717	33.6	15	3274
PRS group 5 (2.0 < RR < 3.0)							
FH+ ages 30-39		biennial 50-74	9849	392	20.0	11	799
	Risk-based strategy	Annual 30-74	37764	722	30.4	15	3848
FH+ ages 40-49		biennial 50-74	9999	398	20.3	12	811
	Risk-based strategy	Annual 40-74	28395	647	28.9	14	2658
FH+ ages 50-64		biennial 50-74	10277	409	20.9	12	834
	Risk-based strategy	an40-50,bi50-74	19772	546	25.3	14	1959
FH+ ages 65+		biennial 50-74	10572	323	17.5	12	860
	Risk-based strategy	an40-50,bi50-74	20084	440	21.1	14	1988
PRS group 4 (1.5 < RR < 2.0)							
FH+ ages 30-39		biennial 50-74	10412	285	14.7	12	848
	Risk-based strategy	an30-50,bi50-74	29630	492	20.6	14	3196
FH+ ages 40-49		biennial 50-74	10521	288	14.8	12	857
	Risk-based strategy	an40-50,bi50-74	19997	438	19.4	14	1979
FH+ ages 50-64		biennial 50-74	10722	295	15.2	12	873
	Risk-based strategy	Biennial 40-74	15530	372	17.6	14	1443

FH+ ages 65+		biennial 50-74	10935	231	12.5	12	892
	Risk-based strategy	Biennial 40-74	15745	296	14.6	14	1463
PRS group 3 (1.0 < RR < 1.5)							
FH+ ages 30-39		biennial 50-74	10778	213	11.0	12	880
	Risk-based strategy	Biennial 30-74	20478	327	14.3	15	2065
FH+ ages 40-49		biennial 50-74	10859	216	11.1	12	887
	Risk-based strategy	Biennial 40-74	15656	304	13.8	14	1455
FH+ ages 50-64		biennial 50-74	11009	218	11.2	12	899
	Risk-based strategy	Biennial 45-74	13063	253	12.2	13	1167
FH+ ages 65+		biennial 50-74	11166	171	9.3	12	913
	Risk-based strategy	Biennial 45-74	13217	196	9.9	13	1181
PRS group 2 (0.5 < RR < 1.0)							
FH+ ages 30-39		biennial 50-74	11133	141	7.3	12	911
	Risk-based strategy	Biennial 35-74	18051	205	9.1	15	1777
FH+ ages 40-49		biennial 50-74	11187	141	7.3	12	916
	Risk-based strategy	Biennial 40-74	16017	198	9.0	14	1489
FH+ ages 50-64		biennial 50-74	11285	143	7.4	12	924
	Risk-based strategy	biennial 50-74	11285	143	7.4	12	924
FH+ ages 65+		biennial 50-74	11389	111	6.1	12	933
	Risk-based strategy	biennial 50-74	11389	111	6.1	12	933
PRS group 1 (0.0 < RR < 0.5)							
FH+ ages 30-39		biennial 50-74	11505	62	3.2	12	944
	Risk-based strategy	Triennial 30-74	14019	76	3.4	15	1431
FH+ ages 40-49		biennial 50-74	11529	62	3.2	12	946
	Risk-based strategy	Triennial 40-74	10986	74	3.4	14	1044
FH+ ages 50-64		biennial 50-74	11572	63	3.3	12	950
	Risk-based strategy	Triennial 50-74	8012	56	2.9	12	649
FH+ ages 65+		biennial 50-74	11618	49	2.7	12	954
	Risk-based strategy	Triennial 50-74	8042	43	2.4	12	652

Chapter 11

Discussion



DISCUSSION

In developed countries, most women within the 50-70 age range have been regularly screened for breast cancer in the last decades. (1, 2) However, measuring the public health impact of breast cancer screening has been challenging for several reasons. First, it is unknown how many breast cancers diagnoses and breast cancer deaths would have occurred had there been no screening. The lack of a control group of women who are not screened makes it difficult to quantify the impact of screening. Second, the simultaneous improvements in breast cancer screening and treatment make it difficult to quantify the contributions of either. These are areas where models come into play. (3, 4) Models can simulate a population of women in the presence and in the absence of various screening and treatment strategies. Further, models can extrapolate the findings from randomized controlled trials by synthesizing data on breast cancer epidemiology, demographics, screening accuracy, and treatment effectiveness to estimate the magnitude of harms and benefits associated with many different screening strategies. The predictions by the Cancer Intervention and Surveillance Modeling Network (CISNET) models have been used to support the current United States Preventive Services Task Force (USPSTF) screening guidelines. (5) Overall, there are numerous reasons why models can contribute to a better understanding of trends in breast cancer incidence and mortality. Nevertheless, breast cancer microsimulation models can also be perceived as complex and be challenging to fully understand.

Research question 1: How can model description, comparison, and validation contribute to a better understanding of model predictions?

Microsimulation model MISCAN-Fadia

One way to improve the understanding of model predictions is to provide a detailed description of the model. The tumor size-oriented MIcrosimulation SCreening ANalyses (MISCAN) model is characterized by exponential continuous tumor growth based on the tumor volume doubling time concept. The tumor FAtal DIAmeter (FADIA) concept represents distant metastasis of breast cancer. These concepts form an intuitive biological entry to modeling breast cancer natural history. One advantage is that tumor size can be observed at diagnosis and if real data on tumor progression rates becomes available in the future this can be used directly in the model. A challenge however, is that trials evaluating the performance of screening modalities often only report test sensitivity, and have to be recalibrated to tumor sizes in order to be applicable in the model. Logically, newer and more sensitive screening tests are able to detect tumors of smaller diameter sizes than less sensitive (older) screening modalities such single view film mammography. Similarly, the efficacy of breast cancer treatment found in studies (6) is translated into a tumor size that can be cured by a specific treatment.

In randomized controlled trials, randomization of participants is a key step to reduce the chance of systematic differences between study participants in the intervention and control groups. In the model this is imitated by simulating a target population twice with the exact same characteristics, except the screening strategy. In general, describing the demography, breast cancer natural history, screening and treatment part of a model and including the model inputs, should contribute to a better understanding of the model. In 2018, a special issue in Medical Decision Making was dedicated to providing a detailed description of all CISNET breast cancer models.(7)

Comparison of DCIS models

One of the most important harms of routinely screening asymptomatic women for breast cancer, that has profound implications for quality of life, is overdiagnosis and overtreatment. The magnitude of overdiagnoses has been a matter of extensive debate because the standard of care is that all tumors are treated immediately upon diagnosis. Moreover, overdiagnoses is difficult to measure as it not observable in individual women and estimates vary widely. (8) The CISNET models project that 34-72% of DCIS diagnoses are overdiagnosed in a biennial 50 to 74 screening scenario.(9) The comparison of multiple approaches to modeling DCIS (in chapter 3) showed that models assuming a stable background trend in breast cancer incidence predicted the highest rates of overdiagnoses of DCIS. The stable background trend implied that the majority of the increase in breast cancer diagnoses due to screening were overdiagnoses. Models with a relatively long pre-clinical duration of DCIS and therefore a relatively long period to detect DCIS by screening, also predicted a high percentage of DCIS overdiagnoses. Models including invasive breast cancer which can be non-progressive, predicted relatively low levels of DCIS overdiagnoses. Overall, and similar to what other studies have found, the comparative modeling outcomes showed that even though there is uncertainty about DCIS natural history, the amount of overdiagnoses among DCIS cases is substantial and greater than the amount of overdiagnoses among invasive breast cancers. (10)

Evidently, the quality of model inputs is related to the quality of model outputs. Since the information about DCIS natural history is still limited, the model projections for DCIS overdiagnoses may therefore not be sufficiently accurate yet to inform clinical practice. A key step in the improvement of our understanding of DCIS natural history and the associated value of modeling DCIS is using observed data from DCIS trials. The COMET(11), LORD, and LORIS (12) trials monitor women with DCIS with the intent of only offering treatment when needed and thereby reduce the risk of overtreatment. Future steps that modeling groups have to make are including new trial information and predictors for disease progression. Predictors for progression include cytologic grade, younger age at diagnosis, ethnicity, or DCIS tumor size. (13, 14)

External model validation

There is a complex interplay between multiple factors that contribute to the effects of screening and treatment on cancer incidence and mortality. These factors include, sensitivity and specificity of screening, screening frequency, attendance to screening, treatment effectiveness, treatment adherence, disease risk and natural history of the disease. Models can synthesize data from various sources to simulate the interplay between such factors and make predictions for the impact of screening and treatment. If collaborative modeling outcomes point to similar conclusions by different models, this should improve the credibility of the conclusion. To formally assess a model´s predictive ability, model predictions should be compared to observed clinical trial outcomes. This is called model validation. The comparison of model predictions to observed event data not used in model development, is called external validation and is seen as one of the strongest forms of model validation. (15)

The effectiveness of screening below age 50 is an important issue in breast cancer screening. While young women (< age 50) are at lower risk to develop breast cancer than older women, tumors grow faster and mammography performs less well due to the prevalence of dense breasts in younger women. (16) The different screening guidelines reflect the uncertainty about screening in this age group. The U.K. Age trial was specifically designed to address the question about the effectiveness of screening in women in the 40 to 49 age range. (17) In chapter 4, Five CISNET models, primarily built for making predictions of screening and treatment in the United States, made predictions for breast cancer screening in the United Kingdom. Predictions were compared to the findings of the U.K. Age trial that compared annual mammography screening of women ages 40 to 49 years with no screening in this age group. The models underestimated the effect of screening on breast cancer mortality at 10-year follow-up. On average, the modeled breast cancer mortality reduction due to screening was 15% (range across models, 13% to 17%) vs. 25% (95% CI, 3% to 42%) observed in the Age trial. (18) At 17-year follow-up, the models predicted 13% (range across models, 10% to 17%) vs. the non-significant 12% (95% CI, -4% to 26%) observed in the trial.

On closer inspection and comparison of model outcomes, we observed that models with slower tumor progression on average predicted a slight increase in breast cancer mortality reduction between 10 and 17-year follow-up. The models with faster tumor progression, and thus a shorter time to breast cancer metastases, on average showed a decline or stable trend in breast cancer mortality reduction. Given that the underestimation at 10-year follow-up was present across all models, it might be explained by a common model input not related to screening. Specifically, no treatment information has been reported in the trial. The models used a derived treatment dissemination based on U.K. surgical oncology reports that may have been different from the actual treatments received by women diagnosed with breast cancer in the trial.

It is known that if screening is first introduced there is a delay in the impact on cancer mortality. The Age trial is one example that shows that lifetime follow-up is important when measuring the impact of screening and treatment. If an extension of the U.K. screening program to women under age 50 was based on the conclusions of the trial at 10-year follow up, one could argue that based on the breast cancer mortality reduction at 17-year follow up this should be reversed. A different challenge of the Age trial was that an ongoing national screening program was in place for women aged 50 and older, and for justified ethical reasons women in both arms of the trial were invited to participate in this program. To assess the effectiveness of screening on breast cancer mortality, the trial restricted their analyses to breast cancers diagnosed during the intervention phase. With regard to screening quality in the trial, the models and the trial itself showed more breast cancer diagnoses due to symptoms (interval cancers) than from early detection by screening in the intervention group. We attributed this finding to the relatively low sensitivity of single view mammography at the time.

Overall we conclude that the models captured the observed long-term effect at 17-year follow-up of screening from age 40 to 49 years on breast cancer incidence and mortality in the UK Age trial, suggesting that the model structures, input parameters, and assumptions about breast cancer natural history are reasonable for estimating the impact of screening on mortality in this age group. It can be noted that it is quite common to have relatively wide confidence intervals in randomized trials on cancer screening. The wide confidence intervals are partly due to the limited number of women included and absolute number of breast cancer deaths. In modeling studies, the outcomes and simulations are not limited to a certain number of women, but models are ultimately informed by these observed data as well.

Which model aspects drive model predictions (MCLIR method)

A necessary step in the interpretation of collaborative model results is to understand how model structure and assumptions contribute to variations in cancer incidence and mortality predictions. However, explaining differences in model predictions is not always straightforward for reasons related to the nature of the disease. Modeling breast cancer involves the representation of unobservable processes such as tumor onset and tumor progression, upon which interventions are overlaid. To model breast cancer, models must make assumptions about the timing of tumor inception, tumor progression, and progression variability among tumors. These assumptions, in conjunction with model structure, impact 3 key determinants of screening effectiveness: 1) pre-clinical duration of breast cancer in which cancers could be detected by screening; 2) the sensitivity of the screening test; and 3) the improvement in prognosis from treatment, e.g., to what extent (earlier) treatment actually reduces (more) breast cancer mortality. The maximum clinical incidence reduction (MCLIR) method was used to isolate the effects of tumor onset,

tumor progression, screening test sensitivity, and breast cancer treatment by comparing model results before and after imposing a one-time screening intervention at age 62 under varying assumptions about screening performance and treatment effectiveness.

Even though different models may use the same data on screening sensitivity and breast cancer treatment effectiveness, the implementation of screening and treatment varies because model structures are different. The MCLIR method was designed to gain insight into how model structure and assumptions influence model predictions. The rationale behind this method is that in the absence of screening, breast cancers will only be diagnosed because of clinical symptoms; referred to as clinical incidence and defined as breast cancers diagnosed due to symptoms. Screening is assumed to detect some of these cancers before symptomatic diagnosis, thereby reducing clinical incidence, and possibly cancer mortality. Differences in 'clinical incidence reduction' reflect differences in how models portray the pre-clinical detectable phase of breast cancer (tumor onset and progression) and mechanisms of screen detection (incorporation of sensitivity). On the other hand, differences in breast cancer mortality are expected to capture model-specific assumptions about implementation of treatment as well as the impact of tumor onset and progression on breast cancer natural history.

The hypothetical 'perfect screening test' scenario showed that some models have relatively large numbers of tumors in existence at screening. On closer inspection, these models have in common a model structure that simulates tumor inception long before the start of the sojourn time (the screen-detectable phase). Moreover, the outcomes also indicated that the tumors in these models are, on average, slowly progressing with longer survival times. On the other hand, models with few cancers in existence at screening, were models with structures that simulated tumors at the start of the sojourn time and with assumptions of relatively fast tumor progression that resulted in shorter survival times on average. Overall, models may be perceived as complex, however the interplay between screening and treatment interventions with unobservable disease natural history is also complex in itself. The MCLIR method can isolate model parts and provide more insight into the factors that drive incidence and mortality predictions. Overall we conclude that in models, the timing of tumor inception and its effect on the length of the pre-clinical phase of breast cancer can have substantial impact on their predictions for breast cancer incidence and mortality reduction.

PART 2: QUANTIFYING THE HARMS AND BENEFITS OF AGE-BASED BREAST CANCER SCREENING IN THE UNITED STATES.

The evidence obtained from randomized controlled trials on the effectiveness of breast cancer screening in the past 30 years led to the widespread use of mammography screen-

ing. Despite this body of evidence, the magnitude of the harms and benefits of breast cancer screening has been debated extensively and the lack of consensus is reflected in the current screening guidelines. This debate has been fueled by the increase in harms such as false-positives and overdiagnoses. Also, the simultaneous improvements in breast cancer screening and treatment over time make it difficult to disentangle the contributions of either to the overall harms and benefits.

Research question 2: What are the benefits and harms of current age-based breast cancer screening in the United States?

Explaining the decline in U.S. breast cancer mortality

Advances in breast cancer screening and treatment have both contributed to the decline in U.S. breast cancer mortality in the last 30 years. In 2005, the CISNET models estimated that screening and treatment contributed about equally to the decline in breast cancer mortality between 1975 and 2000.(3) After the year 2000, two important developments have emerged: digital mammography screening and improvements in molecularly targeted treatments. To further reduce breast cancer mortality, it is useful to assess the relative contributions of screening and treatment to breast cancer mortality in the first decade of the 21st century. No single cancer registry in the U.S., nor any randomized trial, collected sufficient long-term information about ER/ERBB specific treatment to quantify the contributions of screening and treatment by molecular subtype at the population level.

We used 6 different CISNET models to simulate US breast cancer mortality from 2000 to 2012 for multiple birth cohorts using national data on plain-film and digital mammography patterns and performance, dissemination and efficacy of ER/ERBB2(HER2)-specific treatment, and competing mortality. In 2000, the contribution of screening to overall breast cancer mortality reduction was 44% and 56% of the reduction associated with treatment. In 2012 this changed; screening was estimated to be responsible for 37% and treatment for 63% of the total breast cancer mortality reduction in that year. Improvements in chemotherapy and hormone therapy were mainly responsible for this increase in the contribution of treatment. Molecular subtype tumors ER+/ERBB+ were mainly treated with Trastuzumab in 2012 and showed the largest relative contributions associated with treatment vs screening: 69% vs 31%. The ER-/ERBB- tumor group saw the lowest breast cancer mortality reduction (37%) and did not benefit from improvements in hormone therapy nor Trastuzumab. Overall, all models conclude there has been a shift in in the relative contributions associated with screening and treatment to U.S. breast cancer mortality. Advances in screening from film to digital mammography have contributed to the overall decline in breast cancer mortality. Even so, the dissemination of new molecularly targeted therapies and the improved delivery of standard treatment

regimens has had a stronger impact on breast cancer mortality than screening between 2000 and 2012.

Our analyses focused on explaining the decline in breast cancer mortality and did not investigate the harms associated with screening and treatment. However, in future perspective, one possible long-term implication of our findings could be that, if cancer treatments become more and more effective, more targeted, and less burdensome, early detection by screening could become less important. In such scenario, improved treatments could indirectly lead to a reduction in the number of screens and thereby a reduction in false-positives and overdiagnoses. It will be important to continuously evaluate the contributions of screening and treatment in light of new developments. In the meantime, improving the sensitivity and specificity of screening is the most direct way to reduce false positives and recall rates. The use of prognostic factors for invasive breast cancer or watchful waiting strategies in non-invasive cases could potentially reduce overdiagnoses.

Model predictions informing screening guidelines

One of the lessons learned in decades of breast cancer screening is that the harms do not always outweigh the benefits. In 2009, the United States Preventive Services Task Force used collaborative modeling outcomes to support the revision of their recommendations from annual screening beginning at age 40 years to biennial screening beginning at age 50. (19) In 2016, the CISNET models updated the model inputs to account for improvements in screening and systemic treatment. We estimated the magnitude of harms (false-positive mammograms, benign biopsies, overdiagnosis) and benefits (breast cancer mortality reduction, life-years gained, quality-adjusted life-years) of eight different screening strategies. Screening strategies varying in start age of screening (40, 45, 50) and screening interval (annual, biennial, and hybrid), where hybrid strategies consist of annual screening before age 50 followed by biennial screening, were evaluated. All models showed that, when considering the average-risk population, screening starting at age 40 led to substantially more false-positives and overdiagnoses among women in their forties than screening starting at age 50. Starting biennial screening at age 40 vs. 50 modestly lowered breast cancer mortality, and QALYs gained increased by 22% from 86 to 105 per 1.000 women screened. Overall, biennial screening strategies were efficient and preferred over annual strategies for average-risk women. Efficient strategies are strategies that result in the greatest gain in benefits per mammogram. Women at 2-to 4-fold average risk could consider annual screening at ages 40 or 50. Sensitivity analyses of screening cessation at older ages showed that comorbidity levels could be used to tailor stopping age of screening.

Overall, these results suggest that screening starting at age 40 has some benefits, but increases the harms substantially. From a public health perspective considering the ratio

between harms and benefits, extending the 50 to 74 biennial screening recommendations to include women aged 40 to 49 is not favorable for average risk women. However, from a woman's perspective the choice to start screening at age 40 may depend on the value she attaches to the potential benefits and harms of screening.

Radiation induced breast cancer

The ionizing radiation associated with repeated mammography may increase breast cancer risk and could lead to radiation induced cancer. To date, radiation induced breast cancer risk was based on exposure from routine screening only and assumed 4 views per screening. We considered radiation from routine screening for different subgroups of women, diagnostic work-up following an abnormal screening result, false-positive recalls, breast biopsies, and follow-up screening examinations. Variation in radiation dose was taken into account as some women receive more than the mean radiation dose for reasons related to breast thickness, breast augmentation, or breast movement during screening. Annual screening including diagnostic work-up among women aged 40 to 74 years induced 125 breast cancers and 16 breast cancer deaths per 100.000 women screened. Biennial screening from ages 50 to 74 resulted in 27 breast cancers and only 4 breast cancer deaths. Screening and diagnostic work-up among women with large breasts lead to 2.3 times more radiation exposure and were consequently at approximately two times greater risk of radiation induced breast cancer and breast cancer death than women with small or average-sized breasts. Overall, our estimates show that it is important to account for variation in radiation dose when quantifying the number of radiation induced breast cancer and breast cancer deaths

Previous analyses showed that the harms of annual compared to biennial screening greatly increased in terms of false-positives and overdiagnoses. We now showed that, especially when considering annual screening or screening initiation before age 50, the risk of radiation induced breast cancer and breast cancer death is substantial and should be taken into account by policy makers, healthcare providers, and ideally women themselves. Moreover, among women with large breasts who undergo more views on average for a complete screening examination, the radiation induced harms are even greater and approximately doubled. In light of the rapid adoption of digital 3-dimensional tomosynthesis in the United States and elsewhere, it is important to keep in mind that the radiation dose is similar or slightly greater than of digital mammography. It goes without saying that combining digital mammography with tomosynthesis doubles the amount of radiation exposure and risk for inducing breast cancer.

PART 3: PROJECTING THE HARMS AND BENEFITS OF RISK-BASED BREAST CANCER SCREENING IN THE UNITED STATES.

In developed countries, the majority of women adhere to breast cancer screening guidelines. Whilst all guidelines recommend women to be screened regularly, there are differences in the start and stop age of screening as well as in screening interval. (19-21) The guidelines have in common the age-based approach to recommend screening. The logic behind this approach is that age is the strongest risk-factor for most women and ethically all women should have the same rights to potential benefits of screening. However, there is also a downside to an age-based approach to screening. For instance, a screening guideline of biennial screening from ages 50 to 74 essentially treats all women between ages 50 and 74 as being at equal risk for developing breast cancer. It is known that breast cancer risk varies among women of the same age.

Research question 3: To what extent can risk-based breast cancer screening improve the harm-benefit ratio of current age-based screening guidelines?

Risk-stratified screening implies that women are screened in a way that is based on their risk level. A prerequisite is that ahead of screening some sort of risk-assessment has to be made. This could for instance be assessed by asking about their personal or family history of breast cancer, measuring their breast density, or testing for genetic risk factors such as SNPs or rare variants.

Tailoring breast cancer screening intervals by breast density and risk

Despite the consensus about screening women aged 50 and older that is reflected in the various age-based guidelines, it remains challenging to incorporate information on breast cancer risk into screening routines beyond age. Breast density is a risk factor for breast cancer, may change as women age, and affects mammography performance. (22, 23) We estimated the outcomes for screening strategies in the U.S. varying interval of screening (annual, biennial, and triennial) tailored to women aged 50 years or older with various combinations of breast density and relative risk. Four density levels, in line with the American College of Radiology's Breast Imaging reporting were considered: 1) almost entirely fat, 2) scattered fibroglandular density, 3) heterogeneously dense, and 4) extremely dense. Additionally, increased risk levels 1.3, 2.0, and 4.0 that represent for example post-menopausal obesity, history of a benign breast biopsy, or personal history of breast cancer were included. The results showed that screening, regardless of interval and age group, yielded more breast cancer deaths averted, life-years gained, and quality adjusted life-years among women with dense breast and among women at increased relative risk within each density group. In other words, higher breast cancer risk was

associated with more benefits of screening. The number of false-positives and benign biopsies decreased with increasing risk and density, while overdiagnoses increased by risk. When considering a cost-effectiveness threshold of $100,000 per QALY, triennial screening was the only effective strategy for women with low breast density at average risk. Biennial screening was cost-effective among women at increased risk regardless of density, and annual screening was only cost-effective across subgroups at the highest (4.0) risk level and breast density categories 3 and 4 (extremely dense).

Overall, we conclude that breast density and risk level can be used to guide screening intervals. Across women with varying levels of risk and breast density, those with dense breasts at increased risk are most likely to benefit from the current USPSTF guidelines of biennial screening from ages 50 to 74. From a policy maker perspective, the results suggest that only women with extremely dense breasts at the highest risk levels should consider annual screening. Otherwise, annual screening is not cost-effective. Triennial screening was cost-effective for a relatively large group of women with low breast density and average risk. In international perspective, triennial screening is standard practice in the U.K. while in the U.S. this interval is not considered in any guidelines. The modeling results show that triennial screening has a similar balance between harms and benefits compared to biennial screening. In absolute numbers, the benefits, but also the harms are greater for biennial screening, but if relative measures or harm-benefit ratios are leading, triennial screening could be considered for low density, average-risk women. It remains difficult to extend this analysis to younger (<50) women as breast density is unknown until the first mammogram. Incorporating changes in breast cancer risk over time or by age could potentially increase the benefits and reduce the harms of risk-stratified screening.

Personalizing breast cancer screening based on polygenic risk and family history

A first-degree family member diagnosed with breast cancer is a risk factor to develop breast cancer and relatively easy to assess. Polygenic risk can be assessed by a SNP test using blood or saliva and polygenic risk is presumed to remain unchanged during life. These characteristics are the rationale behind our study assessing risk-stratified screening approaches using first-degree family history (FH) and polygenic risk scores (PRS). The models established risk groups based on first-degree family history and risk groups based on a 77-and 167 SNP polygenic risk score. Annual, hybrid, biennial, and triennial digital mammography screening strategies starting at ages 30, 35, 40, 45, and 50 were evaluated for each risk group. Women at high risk due to a first degree family history of breast cancer and/or high polygenic risk could initiate screening before age 50. Women with below-average polygenic risk could consider triennial screening. We projected greater benefits (breast cancer deaths averted, life years gained) when targeted screening was based on polygenic risk scores rather than family history. The screening

approach combining risk from polygenic risk and family history resulted in the maximum improvement in benefits compared to current age-based screening guidelines.

Sensitivity analyses including additional, more recently identified SNP only modestly improved the benefits and harms. If the discriminatory performance of polygenic risk scores improves in the future, different screening scenarios may be optimal from a public health perspective. From an individual perspective, the attitude towards the harms and benefits of polygenic risk-based screening may result in a different preferred screening strategy. We noticed that quite some screening strategies were associated with more intense screening than the current biennial 50-74 screening guidelines. To remove this aspect and quantify the benefit from just the risk-stratification, we redistributed the guideline-concordant number of screens across all women. The outcomes showed that life-years gained and breast cancer deaths averted still increased modestly. Conversely, this showed that a considerable part of the projected increase in benefits was explained by the increase in cancer detection following from more screening examinations.

Increasing number of guidelines advise women to discuss individual breast cancer risk with their healthcare providers. Ongoing trials such as the WISDOM trial (24) and My-PEBS just started to investigate screening approaches based on genetic markers. Until results become available, the model estimates provide specific screening strategies based on genetic risk factors that could be considered in practice. Combining multiple risk factors such as polygenic risk, breast density, and reproductive, lifestyle, and hormonal factors is likely to improve risk prediction and the harm-benefit ratio for stratified screening. In all scenarios, obtaining genetic information should be done with utmost care and ethical approval. Other ethical aspects of genetic testing such as patient autonomy, accessibility to polygenic risk testing, and differential effects across ancestries should be considered before the implementation or recommendation of polygenic risk-based screening.

DIRECTIONS FOR FUTURE RESEARCH BY BREAST CANCER SIMULATION MODELS

Microsimulation models are commonly used to evaluate and quantify the benefits and harms, i.e. cost and effects of health care policies and interventions. Several applications and topics for future research related to breast cancer screening modalities, breast cancer detection, risk-based screening, and treatment are listed here.

Breast cancer screening modalities

- Estimate the potential impact of screening strategies combining multiple modalities such as mammography, tomosynthesis, magnetic resonance imaging, and/or liquid biopsies.

- Estimate the impact of breast self-examination strategies in developing countries.
- Evaluate active surveillance screening strategies using liquid biopsies to monitor disease activity and possible treatment response.

Breast cancer detection

- Estimate the harms and benefits of currently available blood-based liquid biopsies in detecting circulating tumor DNA and confirming healthy tissue.
- Estimate the required test performance for liquid biopsies to be cost effective.
- Estimate the current and future potential of computer aided detection reducing the harms of screening including false positives, overdiagnoses, and false reassurances.

Risk-based screening

- Estimate the cost and effects of screening targeted to individual, age-specific, breast cancer risk based on a combination of risk factors including polygenic risk (SNPs), breast density, rare variants, and lifestyle factors.
- Find the optimal screening strategies for mutation carriers who are at increased risk to develop breast cancer with distinct natural history.
- Estimate the potential of combining breast cancer risk (e.g. subtype-specific risk) with assumptions about tumor progression rates to inform screening strategies.

Breast cancer treatment

- Assess the impact of a new treatment or vaccine discovery that can prevent or treat metastatic breast cancer.
- Estimate 'watchful waiting' strategies for the treatment of DCIS.

Model development / methodology

- Develop models for the interaction between breast cancer risk and tumor progression.
- Develop models to predict local, regional, and distant breast cancer recurrence.
- Extend the current DCIS models by including prognostic factors for DCIS.
- Further develop the Maximum Clinical Incidence Reduction method to explore the effects of model structure and assumptions on predictions about the harms of screening.

REFERENCES

1. Altobelli E, Lattanzi A. Breast cancer in European Union: an update of screening programmes as of March 2014 (review). Int J Oncol. 2014;45(5):1785-92.
2. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. Br J Cancer. 2013;108(11):2205-40.
3. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005;353(17):1784-92.
4. Plevritis SK, Munoz D, Kurian AW, Stout NK, Alagoz O, Near AM, et al. Association of Screening and Treatment With Breast Cancer Mortality by Molecular Subtype in US Women, 2000-2012. JAMA. 2018;319(2):154-64.
5. Mandelblatt JS, Stout NK, Schechter CB, van den Broek JJ, Miglioretti DL, Krapcho M, et al. Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies. Ann Intern Med. 2016;164(4):215-25.
6. Early Breast Cancer Trialists' Collaborative G, Peto R, Davies C, Godwin J, Gray R, Pan HC, 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.
7. Alagoz O, Berry DA, de Koning HJ, Feuer EJ, Lee SJ, Plevritis SK, et al. Introduction to the Cancer Intervention and Surveillance Modeling Network (CISNET) Breast Cancer Models. Med Decis Making. 2018;38(1_suppl):3S-8S.
8. Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. J Med Screen. 2012;19 Suppl 1:42-56.
9. van Ravesteyn NT, van den Broek JJ, Li X, Weedon-Fekjaer H, Schechter CB, Alagoz O, et al. Modeling Ductal Carcinoma In Situ (DCIS): An Overview of CISNET Model Approaches. Med Decis Making. 2018;38(1_suppl):126S-39S.
10. Myers ER, Moorman P, Gierisch JM, Havrilesky LJ, Grimm LJ, Ghate S, et al. Benefits and Harms of Breast Cancer Screening: A Systematic Review. JAMA. 2015;314(15):1615-34.
11. Hwang ES, Hyslop T, Lynch T, Frank E, Pinto D, Basila D, 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.
12. Francis A, Thomas J, Fallowfield L, Wallis M, Bartlett JM, Brookes C, et al. Addressing overtreatment of screen detected DCIS; the LORIS trial. Eur J Cancer. 2015;51(16):2296-303.
13. Maxwell AJ, Clements K, Hilton B, Dodwell DJ, Evans A, Kearins O, et al. Risk factors for the development of invasive cancer in unresected ductal carcinoma in situ. Eur J Surg Oncol. 2018;44(4):429-35.
14. Gorringe KL, Fox SB. Ductal Carcinoma In Situ Biology, Biomarkers, and Diagnosis. Front Oncol. 2017;7:248.
15. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Med Decis Making. 2012;32(5):733-43.
16. Nelson HD, Zakher B, Cantor A, Fu R, Griffin J, O'Meara ES, et al. Risk factors for breast cancer for women aged

40 to 49 years: a systematic review and meta-analysis. Ann Intern Med. 2012;156(9):635-48.
17. Moss S. A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40. Trial Steering Group. J Med Screen. 1999;6(3):144-8.
18. Moss SM, Wale C, Smith R, Evans A, Cuckle H, Duffy SW. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. Lancet Oncol. 2015;16(9):1123-32.
19. Siu AL, On behalf of the U. S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2016;164(4):279-96.
20. Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. JAMA. 2015;314(15):1599-614.
21. Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, et al. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med. 2015;372(24):2353-8.
22. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2006;15(6):1159-69.
23. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med. 2003;138(3):168-75.
24. Esserman LJ, Study W, Athena I. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. NPJ Breast Cancer. 2017;3:34.

Conclusions

CONCLUSIONS

Research question 1: How can model description, comparison, and validation contribute to a better understanding of model predictions?

Describing the breast cancer natural history, screening, treatment, and demography component of micro simulation model MISCAN-Fadia provided necessary information to understand the workings of the model. The most important and distinct characteristics of the model are continuous tumor growth, the fatal diameter concept representing metastasized "fatal" breast cancer, and the use of tumor size as a proxy for screen detection and treatment effects. We concluded that the model is quite flexible and can synthesize data from different sources, but also requires recalibration of several inputs before these can be used in this tumor-size oriented model. In this detailed model description, we justified modeling choices, and listed considerations as well as limitations that should improve transparency.

The comparison of model predictions of overdiagnoses among screen detected DCIS was 34% to 72% and 2% to 12% among invasive breast cancers in a biennial 50-74 screening scenario. We concluded that regardless of differences in model structure and assumptions about breast cancer natural history, overdiagnoses among DCIS is extensive and as long as the standard of care is treatment of DCIS upon diagnosis, many women are overtreated. Convergence of overdiagnoses predictions can be achieved when data on, currently unobservable, DCIS progression rates becomes available from active surveillance trials.

The models´ predictive ability was formally assessed by the comparison of breast cancer incidence and mortality predictions of annual screening from ages 40 to 49 to observed outcomes in the Age trial. The models reproduced the patterns in breast cancer incidence, but underestimated breast cancer mortality reduction at 10- and were more accurate at 17-year follow-up. We concluded that the model structures, existing input parameters, and assumptions about breast cancer natural history are reasonable for estimating the impact of screening on mortality in the 40-49 age group.

The maximum clinical incidence reduction (MCLIR) method was used to compare models and disentangle the interplay between screening and treatment interventions with model-specific assumptions about unobservable breast cancer natural history. Overall, we concluded that in models, the timing of tumor inception and its effect on the length of the pre-clinical phase of breast cancer had substantial impact on predictions for breast cancer incidence and mortality reduction.

Research question 2: What are the benefits and harms of current age-based breast cancer screening in the United States?

The models consistently showed that biennial screening starting at age 40 instead of 50 lead to disproportionately more false-positives and overdiagnoses among average-risk women. Breast cancer mortality was only modestly lowered, but QALYs gained increased by 22%. Compared to annual screening strategies, biennial screening resulted in the greatest gain in benefits per mammogram and dominated annual strategies for average-risk women. Only for women at 2-to 4-fold average risk could consider annual screening at ages 40 or 50. Overall, we concluded that screening starting at age 40 has some benefits, but increased the harms substantially.

In light of the simultaneous improvements in breast cancer screening and treatment in the last decade, the models incorporated the transition from film to digital mammography and included molecular subtype specific breast cancer treatments to separate the contributions of either to breast cancer mortality reduction. In 2000, the contribution of screening to overall breast cancer mortality reduction was 44% vs. 56% explained by treatment. We showed that between 2000 and 2012 there has been a shift in relative contributions, screening was estimated to be responsible for 37% and treatment for 63% of the total breast cancer mortality reduction in 2012. The models concluded that dissemination and improved delivery of new molecularly targeted therapies has had a stronger impact than screening improvements on breast cancer mortality between 2000 and 2012.

The ionizing radiation associated with repeated mammography may increase breast cancer risk and could lead to radiation induced cancer. Annual screening including diagnostic work-up among women aged 40 to 74 years induced 125 breast cancers and 16 breast cancer deaths per 1.000 women screened. Biennial screening from ages 50 to 74 resulted in 27 breast cancers and only 4 breast cancer deaths. Overall, we concluded that it is important to account for variation in radiation amount caused by diagnostic work-up following an abnormal screening result, false-positive recalls, breast thickness, breast augmentation, breast biopsies, and follow-up screening examinations when quantifying the number of radiation induced breast cancer and breast cancer deaths

Research question 3: To what extent can risk-based breast cancer screening improve the harm-benefit ratio of current age-based screening guidelines?

The results of screening based on breast density and risk-level showed that increased breast cancer risk from either source was associated with more benefits of screening. Conversely, the number of false-positives and benign biopsies decreased with increasing risk and density while the number of overdiagnoses increased by risk. When considering a cost-effectiveness threshold of $100,000 per QALY, triennial screening was the only

effective strategy for women with low breast density at average risk. Biennial screening was cost-effective among women at increased risk regardless of density, and annual screening was only cost-effective across subgroups at the highest (4.0) risk level and breast density categories 3 and 4 (extremely dense). Overall, we concluded that breast density and risk level can be used to guide screening intervals.

We projected greater benefits (breast cancer deaths averted, life years gained) when screening was based on polygenic risk scores rather than family history. The screening approach combining risk from polygenic risk and family history resulted in the maximum improvement in benefits compared to current age-based screening guidelines. Women at high risk due to a first degree family history of breast cancer and/or high polygenic risk could initiate screening before age 50. Women with below-average polygenic risk could consider triennial screening. A large part of the projected increase in benefits was explained by the increase in cancer detection following from more screening examinations. Nevertheless, the benefits would still modestly increase at equal number of screens.

Summary (EN)

SUMMARY

Chapter 1 introduces the motivation and overall topic of this thesis: breast cancer, and breast cancer screening. The causes, risk-factors, incidence, survival, and breast cancer mortality are described. Potential benefits of early detection through screening are: life-years gained, improved quality of life, breast cancer deaths prevented, correct reassurance among women without breast cancer. Potential harms of screening include false reassurance, overdiagnosis, overtreatment, false-positive screening test results, and to some extent the temporary uncertainty after screening, and exposure to radiation that can induce breast cancer. The scientific body of evidence on breast cancer screening that has been gathered in the past decennia, has led to the widespread use of mammography worldwide. However, there is no consensus about which screening strategy is optimal. This is the area where simulation models are used to make projections about the effects of various different screening strategies. In this thesis, we investigate how model predictions can be better understood and to what extent risk-stratification can increase the benefits of breast cancer screening.

Part 1: Breast cancer micro-simulation: methods, comparative modeling, and model validation.

Chapter 2 provides an overview of microsimulation screening analysis – Fatial diameters (MISCAN-Fadia) model. The four main components of the model: demography, breast cancer natural history, screening , and treatment are described in detail. The MISCAN-Fadia model distinguishes itself from many other models by using a biological entry such as tumor growth and tumor size to modeling the natural history of breast cancer. The effectiveness of treatment and the sensitivity of screening are both linked to tumor size. The model adopts a 'fatal diameter' concept which implies that a cancer is not curable anymore and basically represents distant metastasis of breast cancer. The model is able to simulate many screening and treatment strategies in a short amount of time with varying adherence to screening and treatment. In each simulation, women differing in risk, birthyear, and life expectancy can be included. The model produces estimates of lifeyears gained, breast cancer eaths prevented, stage distributions, overdiagnoses, and interval cancer. Recent model developments include radiation induces breast cancers, breast density, and cancer by molecular subtype.

Chapter 3 shows how the CISNET breast cancer models simulate DCIS. Since the introduction of mammpgraphy screening in the 1980's in the United States, the incidence of DCIS, which is seen as a precursor of breast cancer, increased substantially. Uncertainty remains about the natural history of DCIS because tumor onset and tumor progression cannot be observed. In the 5 CISNET models, invasive breast cancer can develop from preclinical screen-detectable DCIS, or preclinical undetectable DCIS. A part of preclinical

DCIS may also regress. The models estimate that a large part of screen detected DCIS are overdiagnoses: 34%-72% in a biennial 50 to 74 screening strategy. Overdiagnosis is defined as a screen detected tumor that in the absence of screening would not have been found. The model predictions show no association between the amount of DCIS- and invasive overdiagnoses. The large differences in predictions of overdiagnosed DCIS cases, which is also found in other scientific literature, reflects the uncertainty around the natural history of DCIS. This underscores the importance of active surveillance trials such as the LORD, LORIS, and COMET trial that can provide more observed data on DCIS natural history.

Chapter 4 presents an external validation that compares model predictions to observed data from the 'U.K. Age trial'. The trial compared annual mammography screening in women aged 40 to 49 to a control group who were offered usual care, which is no screening in this age group. The 5 CISNET models used demography, screening attendance, and mammographic sensitivity from the trial in combination with extant assumptions about the onset en natural history of breast cancer to predict the incidence and mortality in the intervention and control arm. The results show that the effect of annual screening on breast cancer incidence is reproduced quite well. The average breast cancer mortality reduction after 10 years of follow-up was underestimated by the models 15% (range: 13% to 17%) compared to 25% (95% CI, 3% to 42%) in the trial. After 17 years of follow-up, the trial showed a 12% (95% CI, -4% to 26%) non-significant reduction and the models 13% (10% - 17%) on average. We conclude that the models reproduced the long term effects of the age trial reasonably well. This suggests that the existing model structures, model input parameters, and assumptions are suitable for estimating the effect of screening on breast cancer mortality in this age group.

In chapter 5, investigates how model structures and assumpations about the preclinical duration of breast cancer influence model predictions. The Maximum Clinical Incidence Reduction (MCLIR) method is used and extended to disentangle the effects of tumor growth rate, timing of tumor onset, screening sensitivity, and treatment effectiveness. The models do this in a simplified setting of a single screen at age 62 with varying assumptions about test sensitivity and treatment effectiveness. The MCLIR method compares changes in the number of breast cancer cases and deaths in 4 scenarios: 1. no screening, 2. a screen with perfect (100%) sensitivity and perfect treatment (100% cure), 3. a screen with sensitivity of digital mammography and perfect treatment, 4. a screen with sensitivity of digital mammography and realistic (observed) effectiveness of treatment. The models predict a 19% to 71% reduction in clinical incidence and 33% to 67% reduction in breast cancer mortality as a result of a perfect screening test and perfect treatment. In the scenario with sensitivity of digital mammography and realistic treatment effectiveness, the prediction converge: 11% to 24% clinical incidence reduction and 8% to 18% breast cancer mortality reduction. The timing of tumor onset and its

influence of the preclinical duration had the largest impact on model predictions. Models with relatively fast progressing tumors also had a shorter preclinical duration. The MCLIR method can shed light on the root of the differences between model predictions and can be applied in other disease settings where the effects of screening are modeled.

Part 2 – Quantifying the harms and benefits of breast cancer screening among women in the United States

In chapter 6, the harms and benefits of eight screening strategies varying in starting age and interval are estimated by 6 cisnet models. The target population was average risk women and women at increased breast cancer risk due to their breast density or co-morbidity. Biennial screening from ages 50 to 74 prevented 7 breast cancer deaths on average compared to no screening. Annual screening in the same age ranges would prevent 3 additional breastr cancer deaths, but would increase false-positives by almost 2.000 per 1.000 women screened over a lifetime. Starting annual screening at age 40 showed similar harms and benefits among women with 2 to 4 times the average risk. Women with moderate to severe comorbidity could stop screening at age 66 or 68. All 6 models conclude that starting screening at age 40 leads to a small benefit in terms of life years gained and breast cancer deaths prevented, but the increase in false positives and overdiagnoses is substantial. This quantitative analyses shows that biennial screening among average risk women is most efficient. Policy makers cab use this information to inform their decision about breast cancer screening policy.

In chapter 7, six breast cancer simulation models are used to assess the relative contributions of screening and treatment to the trend in breast cancer mortality between 2000 and 2012. Given the improvements in treatment and new adjuvant therapies which were given to breast cancer patients in these periods, the analysis focuses on combination of breast cancer subtypes estrogen receptor (ER) positive and human epidermal growth factor receptor (HER) 2. In 2000, the models estimate a 37% (27%-42%) breast cancer mortality reduction vs. no screening. 56% of this reduction is explained by breast cancer treatment and 44% by screening. However, in 2012 the total breast cancer mortality reduction is estimated at 49% (39%-58%) of which 37% is explained by screening and 63% by improvemens in treatment. For 3 out of the 4 subtypes it holds that treatment has made a larger contribution to the decline in breast cancer mortality, except for the ER-/HER2- tumors where the contributions of screening and treatment are estimated as approximately equal. The models conclude that in 2000 to 2012 the continued decline in overall breast cancer mortality can be explained for a larger part by new and improved treatments than by screening in this period.

Chapter 8 investigates to what extent the exposure to the ionizing radiation of repeated mammography screening contributes to breast cancer and breast cancer death. Prior research was based on 4 views per screening and did not account for breast size

and breast thickness, nor false-positives, diagnostic work-up, and variations in radiation dose caused by breast augmentation or breast positioning during screening. This study accounted for these factors because of their impact on the overall radiation dose and consequent radiation induced breast cancers. We estimated the radiation induced breast cancer in 8 screening strategies varying starting age (40, 45, 50) and screening interval (annual, biennial, hybrid). The benefits of annual screening of 100.000 women between ages 50 and 74 are estimated at 968 breast cancers prevented, but would also induce 125 breast cancer and 16 breast cancer deaths through radiation. Among women with large breasts, 8% receives more radiation doses during screening and this was estimated to cause 266 breast cancers of which 35 would lead to breast cancer death per 100.000 women screened. The results in this study show that it is important to account for variations in radiation dose from and after screening when determining the number of radiation induced breast cancers and breast cancer deaths.

Part 3: Projecting the lifetime harms and benefits of risk-based breast cancer screening.

In chapter 9, the effects of screening among women with varying breast density and risk are quantified. Three CISNET models simulate the effect of annual, biennial, and hybrid screening between ages 50 and 74 or 65 and 74. We distinguished four density groups spanning between almost entirely fat to extremely dense breast tissue. Increased breast cancer risk caused by other factors was modeled by including 4 relative risk groups: 1.0 (average risk), 1.3, 2.0, 4.0. The results show that in all screening intervals, the breast cancer deaths prevented and life years gained increased with breast density as well as increases in relative risk from other factors. At the same time, false positives and unnecessary biopsies decreased while overdiagnoses increased. The results in this study show that breast density and increased risk due to other factors can be useful in the formation of risk-based screening guidelines.

In chapter 10 we investigated the effects of screening based on breast cancer family history and small DNA variations and how these relate to the results of age-based screening guidelines. Two CISNET models estimated the effects of screening strategies with starting ages (30, 35, 40, 45, 50) and stopping age 74, and screening intervals (annual, biennial, triennial, hybrid). Among women younger than age 50 with a first-degree family member diagnosed with breast cancer; about 9% of the population, starting screening before age 50 would gain 44% life years and avert 24% breast cancer deaths compared to starting screening at age 50. However, the increase in the number of mammograms among these women also led to 25% more overdiagnoses and false positives would double. Screening based on polygenic risk gained 19% additional life years, prevented 11% breast cancer deaths, and overdiagnoses and false positives increased by 10% and 26%. Screening based on breast cancer family history and polygenic risk resulted in the

largest increase in benefits compared to current USPSTF guideline screening. This study showed that risk stratified screening can lead to fewer breast cancer deaths and more life years gained among women who are at increased risk of breast cancer due to polygenic risk and family history.

Samenvatting (NL)

SAMENVATTING

Hoofdstuk 1 introduceert zowel de aanleiding als het overkoepelende thema van dit proefschrift: borstkanker en borstkankerscreening. Naast de oorzaken en risicofactoren worden ook de incidentie, overleving, en sterfte aan borstkanker besproken. De potentiele voordelen van vroege opsporing door middel van borstkankerscreening zijn: gewonnen levensjaren, verhoogde levenskwaliteit, voorkomen borstkanker sterfgevallen, en terechte geruststelling bij gezonde vrouwen zonder borstkanker. De potentiele nadelen van screening zijn: onterechte geruststelling, overdiagnose en overbehandeling, fout-positieve uitslagen, en in mindere mate de tijdelijke onzekerheid na het screenen, en blootstelling aan straling wat kan bijdragen aan de ontwikkeling van borstkanker. Het wetenschappelijk bewijs voor borstkankerscreening wat in de afgelopen decennia is vergaard, heeft ertoe geleid dat wereldwijd vrouwen regelmatig worden gescreend. Echter is er geen consensus over welke screeningstrategie optimaal is. Dit is het gebied waarin simulatiemodellen gebruikt worden om voorspellingen te maken over de effecten van verschillende screeningstrategien. In dit proefschrift wordt onderzocht hoe we voorspellingen van meerdere modellen beter kunnen begrijpen en in welke mate risico-stratificatie de voordelen van borstkankerscreening kan vergroten.

Part 1: Breast cancer micro-simulation: methods, comparative modeling, and model validation.

Hoofdstuk 2 geeft een overzicht van het Microsimulation Screening Analysis – Fatal diameter (MISCAN-Fadia) model. De vier belangrijkste componenten van het model: demografie, natuurlijk beloop van borstkanker, screening en behandeling worden uitvoerig besproken. Het MISCAN-Fadia model onderscheidt zich van veel andere modellen door een biologische grondslag zoals tumor groei en tumor grootte te nemen om het ontstaan en verloop van borstkanker te simuleren. Zowel de effectiviteit van behandelingen in het genezen van borstkanker en de gevoeligheid van de screening test wordt in MISCAN-Fadia gelinkt aan tumor grootte. Het model hanteert het 'fatale diameter' concept wat de niet meer geneesbare staat van kanker, metastase op afstand, symboliseert in de vorm van een fatale tumor grootte. Het model is in staat om in een kort tijdsbestek vele screening strategieën en behandelschema's te simuleren met variërende screeningopkomst en therapietrouw . In deze simulatie kunnen vrouwen van verschillend geboortejaar, risico en levensverwachting tegelijkertijd worden gesimuleerd. Het model produceert schattingen voor onder andere gewonnen levensjaren, voorkomen borstkanker sterfgevallen, stadium verdelingen, overdiagnoses, en interval kankers. Recente ontwikkelingen omvatten het simuleren van vrouwen met verschillende borstdichtheid, de impact door straling geïnduceerde borstkankers, en trendanalyses in borstkanker naar moleculair subtypes.

Hoofdstuk 3 laat zien hoe de CISNET borstkankermodellen DCIS simuleren. Sinds de introductie van mammografie screening in de jaren '80 in de Verenigde Staten is de incidentie van DCIS, wat gezien wordt als voorstadium van en risicofactor voor borstkanker, sterk toegenomen. Er is onzekerheid over het natuurlijk beloop van DCIS omdat het ontstaan van een tumor en progressie snelheid niet observeerbaar zijn. In de 5 CISNET modellen kan invasieve borstkanker ontstaan uit pre-klinische screen-detecteerbare DCIS, of uit pre-klinische nog niet screen-detecteerbare DCIS. Daarnaast kan een deel van de pre-klinische DCIS ook regresseren. De modellen schatten dat een groot deel van de screen gedetecteerde DCIS overdiagnoses zijn: 34% - 72% bij tweejarig screenen tussen leeftijden 50 en 74. Overdiagnose is hier gedefinieerd als een screen gedecteerde tumor die in de afweigheid van screening niet gevonden zou zijn geworden. De model voorspellingen laten geen verband zien tussen de hoeveelheid DCIS-en invasieve overdiagnoses. De grote verschillen tussen voorspelde over-gediagnosticeerde DCIS gevallen, wat ook in andere wetenschappelijke literatuur wordt gevonden, geeft de onzekerheid over het natuurlijk beloop van DCIS weer. Dit onderschrijft het belang van active surveillance trials zoals de LORD, LORIS en COMET trial die meer geobserveerde data kunnen verschaffen over het natuurlijk beloop van DCIS.

Hoofdstuk 4 is een externe validatie waarbij model voorspellingen worden vergeleken met geobserveerde data uit de 'U.K. Age trial'. Deze trial vergeleek jaarlijkse mammografie screening in vrouwen tussen de 40 en 49 jaar met een controle groep die de gebruikelijke zorg ontvingen, wat geen screening is in deze leeftijdsgroep. De vijf CISNET modellen gebruikten demografie, screeningsdeelname, en mammografie sensitiviteit van de U.K. Age trial in combinatie met bestaande aannames over het ontstaan en natuurlijk beloop borstkanker om de incidentie en moraliteit in de interventie en controle arm te voorspellen. De resultaten laten zien dat het effect van jaarlijks screenen op borstkanker incidentie goed wordt gereproduceerd door de modellen. De gemiddelde voorspelde borstkanker mortaliteitsreductie na tien jaar follow-up was 15% (range: 13% tot 17%) en daarmee onderschat vergeleken met de 25% (95% CI, 3% to 42%) gevonden in de trial. Na 17 jaar vond de trial een 12% (95% CI, -4% to 26%) niet-significante reductie en de modellen gemiddeld 13% (10% - 17%). In deze externe validatie concluderen we dat de modellen de lange termijn effecten van de U.K. Age trial redelijk goed reproduceren. Dit suggereert dat de model structuren, model input parameters, en aannames geschikt zijn om het effect van screenen op borstkanker mortaliteit in deze leeftijdsgroep te schatten.

In hoofdstuk 5, wordt onderzocht hoe model structuren en aannames over de preklinische duur van borstkanker invloed hebben op voorspellingen. De 'Maximum Clinical Incidence Reduction' (MCLIR) methode wordt gebruikt en uitgebreid om de effecten van tumor groeisnelheid, de timing van het ontstaan van de tumor, screening test sensitiviteit en behandelingseffectiviteit te ontrafelen. De modellen doen dit door de effecten van één screen op leeftijd 62 te simuleren met variërende aannames over screening sensiti-

viteit en behandelingseffect. De MCLIR methode vergelijkt veranderingen in het aantal borstkankers (niet screen gedetecteerd) en borstkanker sterfgevallen in vier scenario's: 1. geen screening, 2. één screening test op leeftijd 62 met perfect (100%) sensitiviteit en perfecte (gegarandeerd genezende) behandeling, 3. één screening test met sensitiviteit van digitale mammografie op leeftijd 62 met en perfecte behandeling, 4. één screening test met sensitiviteit van digitale mammografie op leeftijd 62 met geobserveerde behandelingseffectiviteit. De modellen voorspellen een 19% tot 71% reductie in klinische incidentie als gevolg van een perfecte screen met perfecte behandeling en 33% tot 67% borstkanker mortaliteit reductie. In het scenario met sensitiviteit van digitale mammografie en realistische behandelingseffectiviteit liggen de voorspellingen een stuk dichterbij elkaar: 11% tot 24% reductie in klinische incidentie en 8% tot 18% borstkanker mortaliteit reductie. Het moment van tumor ontwikkeling en de invloed op de preklinische duur had het grootste effect op de model voorspellingen voor incidentie en mortaliteit. Modellen met een relatief snellere tumorgroei lieten gemiddeld genomen ook een kortere preklinische duur zien. De MCLIR methode kan licht schijnen op de oorsprong tussen verschillen in modelvoorspellingen en kan ook worden toegepast in andere ziektes waar het effect van screenen gemodelleerd wordt.

Part 2 – Quantifying the harms and benefits of breast cancer screening among women in the United States

In hoofdstuk 6 worden de voor-en nadelen van acht screening strategieën varierend in start leeftijd en screening interval door 6 modellen geschat. De doelgroepen zijn gemiddeld risico vrouwen en vrouwen met een verhoogd borstkanker risico door borstdichtheid of co-morbiditeit. Tweejaarlijks screening van leeftijd 50 tot 74 voorkwam gemiddeld 7 borstkanker sterfgevallen vergeleken met helemaal niet screenen. Jaarlijks screenen in dezelfde leeftijden zou 3 extra sterfgevallen voorkomen, maar wel tot bijna 2000 extra fout-positieven leiden per 1000 gescreende vrouwen. Jaarlijks screenen vanaf leeftijd 40 heeft vergelijkbare voor-en nadelen voor vrouwen met een 2 tot 4-keer zo hoog risico als gemiddeld risico vrouwen. Vrouwen met matig tot ernstige co-morbiditeit zouden kunnen stoppen met screenen op leeftijd 66 of 68. Alle 6 modellen concluderen dat starten met screenen op leeftijd 40 leidt tot een kleine toename in gewonnen levensjaren en voorkomen borstkanker sterfgevallen, maar de toename in nadelen zoals in fout-positieven en overdiagnosen is groter. Deze kwantitatieve analyse laat zien dat tweejaarlijks screening voor gemiddeld risico vrouwen het meest efficient is. Beleidsmakers kunnen met deze uitkomsten hun overwegingen of beslissingen omtrent borstkanker screenen ondersteunen.

In hoofdstuk 7 wordt, gebruikmakend van 6 borstkanker simulatiemodellen, onderzocht wat de relatieve bijdrage van screening en behandeling is aan de trends in borstkankersterfte tussen 2000 en 2012. Gezien de verbeteringen in behandeling en nieuwe

adjuvante therapieën die in deze periode aan borstkanker patiënten zijn gegeven, concentreert deze analyse zich op combinaties van borstkanker subtypes oestrogeen receptor (ER) positief en human epidermal growth factor receptor (HER) 2. In 2000 schatten de modellen de borstkanker mortaliteitsreductie op 37% (27%-42%) vs. geen screening. 56% van deze reductie verklaren de modellen door borstkanker behandeling en 44% is toe te wijzen aan screening. Echter, in 2012 wordt de totale borstkankersterfte reductie geschat op 49% (39%-58%) waarvan 37% door screening en 63% door verbeterde behandelingen. Voor drie van de vier subtypes geldt dat behandeling een grotere bijdrage heeft gehad in de afgenomen borstkankersterfte, behalve voor ER-/HER2- schatten de modellen de bijdrage van screening en behandeling als praktisch even groot. Daarmee concluderen de modellen dat in de periode 2000 tot 2012 de verdere afname in totale borstkankersterfte voor een groter deel te verklaren is door de verbeterde en nieuwe behandelingen dan door screening in deze periode.

Hoofdstuk 8 onderzoekt in welke mate herhaaldelijke mammografie screening leidt tot borstkanker of zelfs borstkankersterfte door blootstelling aan straling. Voorgaand onderzoek was gebaseerd op 4 röntgen foto's per screening en hield geen rekening met borstgrootte en dikte, fout-positieve screening resultaten, diagnostisch vervolgonderzoek na een abnormaal screenings resultaat, of variaties in stralingsdosering door een borstvergroting of verkeerde borstpositionering. Deze studie houdt wel rekening met deze factoren aangezien deze allemaal invloed hebben op het aantal door straling geïnduceerde borstkankers. We schatten de straling geïnduceerde borstkanker incidentie en sterfte voor 8 verschillende screening strategieën die variëren naar startleeftijd (40, 45, 50) en screening interval (jaarlijks, tweejaarlijks, hybride). De voordelen van het jaarlijks screenen van 100,000 vrouwen tussen leeftijd 40 en 74 worden geschat op 968 voorkomen borstkanker doden, maar dit leidt ook tot 125 borstkankers en 16 borstkanker doden die ontstaan door de straling. Voor 8% van de vrouwen met grote borsten die meer straling ontvangen bij screenen worden het aantal borstkankers en borstkanker doden geschat op 266 en 35 per 100,000 vrouwen. De resultaten uit dit onderzoek laten zien dat het belangrijk is om de variaties in de hoeveelheid schadelijke straling tijdens en na het screenen mee te nemen bij het bepalen van het aantal straling geïnduceerde borstkankers en borstkanker doden.

Part 3: Projecting the lifetime harms and benefits of risk-based breast cancer screening.

In hoofdstuk 9 wordt het effect van verschillende screening strategieën op vrouwen met variërende borstdichtheid en borstkankerrisico gekwantificeerd. Drie CISNET modellen simuleren het effect van jaarlijks, tweejaarlijks, en hybrid screenen tussen leeftijden 50 en 74 of 65 en 74. We onderscheiden voor borstdichtheid 4 groepen van bijna volledig vetweefsel tot extreem dicht borstweefsel. Voor een verhoogt borstkankerrisico door

andere factoren worden ook 4 risicogroepen gemodelleerd met een relatief risico van: 1.0 (gemiddeld risico), 1.3, 2.0, en 4.0. De resultaten laten zien dat voor alle screening intervallen de voorkomen borstkanker doden en gewonnen levensjaren toenemen naarmate zowel borstdichtheid als het borstkankerrisico toenemen. Tegelijkertijd nemen fout positieve screenings en onnodige biopsieën af, echter neemt het aantal over diagnoses wel toe. De resultaten uit deze studie laten zien dat borstdichtheid en een verhoogd risico bruikbaar kunnen zijn bij het opstellen van risico-gebaseerde screening richtlijnen.

In hoofdstuk 10 wordt onderzocht hoe de effecten van screening gebaseerd op familie historie van borstkanker en kleine afwijkingen in het DNA zich verhouden tot die van de op leeftijd gebaseerde screening richtlijnen. Twee CISNET modellen schatten de effecten van screening strategieën met verschillende startleeftijd (30, 35, 40, 45, 50) met stopleeftijd 74 en screening interval (jaarlijks, tweejaarlijks, driejaarlijks, hybride). Voor jonge vrouwen (<50) met een eerstegraads familielid met borstkanker; ongeveer 9% in de populatie, levert eerder beginnen met tweejaarlijks screenen 44% extra gewonnen levensjaren en 24% minder borstkanker doden ten opzichte van beginnen met tweejaarlijks screenen op leeftijd 50. Echter, de toename in mammogrammen leidt in deze groep ook tot naar schatting 26% meer overdiagnoses en het aantal fout positieven verdubbelt. Screenen gebaseerd op polygenetisch risico, levert op populatie niveau 19% extra gewonnen levensjaren, 11% minder borstkanker doden, overdiagnoses stijgen met 10% en fout positieven met 26%. Screenen gebaseerd op familie historie van borstkanker en polygenetisch risico levert de meeste voordelen op ten opzichte van de huidige USPSTF richtlijnen. Dit onderzoek laat zien dat risico gestratificeerd screenen voor een aantal groepen vrouwen met een verhoogt risico op borstkanker door een familie historie met borstkanker en afwijkingen in hun DNA minder borstkanker doden en meer levensjaren kan opleveren.

CURRICULUM VITAE

Jeroen van den Broek was born on August 26th 1987, in Eindhoven, the Netherlands. In 2006, he completed his secondary education ´Atheneum´ at scholengemeenschap Augustinianum in Eindhoven. In the beginning of 2007, he auditioned at conservatory Berklee College of Music for jazz drumming, Boston, USA. Later that year, he started studying 'Econometrics and Operational Research' at the Erasmus University in Rotterdam. He obtained his Bachelor of Science degree in 2010 and Master of Science degree with specialization 'Quantitative Finance' in 2012. For his bachelor- and master thesis, Jeroen investigated patterns of synchronization in international stock markets and estimated the profitability of high-frequency pairs trading strategies. In the final stages of his Bachelor education, he started his own company specializing in low-cost advertising for customer acquisition online and created a tool to automate advertisement-creation based on search engine keywords. In September 2013, he started working at the department of public health at Erasmus University Medical center in Rotterdam. His research focused on the quantification of the benefits and harms of breast cancer screening strategies by means of microsimulation modeling. In March 2017, Jeroen temporarily left the department of public health in Rotterdam to work as a Research Scholar at the Harvard University, School of Public Health, Boston, USA. This research used a polygenic risk score, consisting of many variations in the DNA sequence associated with breast cancer, to personalize breast cancer screening strategies. In August 2017, Jeroen returned at Erasmus Medical Center to continue his work on comparative modeling and risk-based screening. The results of this research are described in this thesis.

LIST OF PUBLICATIONS

Publications in this thesis

1. **Jeroen J. van den Broek**, Nicolien T. van Ravesteyn, Eveline A. Heijnsdijk, Harry J. de Koning.
   Simulating the impact of risk-based screening and treatment on breast cancer outcomes with MISCAN-Fadia. Medical Decision Making. 2018;38(1S):54–65.
2. **Jeroen J. van den Broek**, Nicolien T. van Ravesteyn, Jeanne S. Mandelblatt, Hui Huang, Mehmet Ali Ergun, Elizabeth S. Burnside, Helen C. Xu, Yisheng Li, Oguzhan Alagoz, Sandra J. Lee, Natasha K. Stout, Juhee Song, Amy Trentham-Dietz, Sylvia K. Plevritis, Sue M. Moss, Harry J. de Koning.
   Comparing CISNET Breast Cancer Incidence and Mortality Predictions to Observed Clinical Trial Results of Mammography Screening from Ages 40 to 49. Medical Decision Making 2018;38:140S-50S
3. **Jeroen J. van den Broek**, Nicolien T. van Ravesteyn, Mucahit Cevik, Clyde B. Schechter, Sandra J. Lee, Hui Huang, Yisheng Li, Diego F. Munoz, Sylvia Plevritis, Jeanne Mandelblatt, Harry J. de Koning, Natasha K. Stout*, Marjolein van Ballegooijen*.
   Comparing CISNET breast cancer models using the Maximum Clinical Incidence Reduction methodology. Medical Decision Making 2018;38:112S-25S.
4. Nicolien T. van Ravesteyn, **Jeroen J. van den Broek**, Xiaoxue Li, Harald Weedon-Fekjær, Clyde B. Schechter, Oguzhan Alagoz, Xuelin Huang, Don Weaver, Elizabeth S. Burnside, Rinaa S. Punglia, Harry J. de Koning, Sandra J. Lee.
   Modeling ductal carcinoma in situ (DCIS) – an overview of CISNET model approaches. Medical Decision Making. 2018;38:126S-39S.
5. Sylvia Plevritis, Diego Munoz, Allison Kurian, Oguzhan Alagoz, Aimee M. Near, Sandra J. Lee, **Jeroen J. van den Broek**, Xuelin Huang, Natasha K. Stout, Clyde B. Schechter , Brain L. Sprague, Harry J. de Koning, Amy Trentham-Dietz, Nicolien T. van Ravesteyn, Ronald Gagnon, Yaojen Chang , Yisheng Li, Mehmet Ali Ergun, Hui Huang, Donald A. Berry, Jeanne S. Mandelblatt.
   Association of Screening and Treatment With Breast Cancer Mortality by Molecular Subtype in US Women, 58 2000-2012. JAMA. 2018;319(2):154-64.
6. Mandelblatt JS, Stout NK, Schechter CB, **van den Broek JJ**, Miglioretti DL, Krapcho M, Trentham-Dietz A, Munoz D, Lee SJ, Berry DA, van Ravesteyn NT, Alagoz O, Kerlikowske K, Tosteson AN, Near AM, Hoeffken A, Chang Y, Heijnsdijk EA, Chisholm G, Huang X, Huang H, Ergun MA, Gangnon R, Sprague BL, Plevritis S, Feuer E, de Koning HJ, Cronin KA.

Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies. Annals of Internal Medicine. 2016 Feb 16;164(4):215-25.

7. Miglioretti DL, Lange J, **van den Broek JJ**, Lee CI, van Ravesteyn NT, Ritley D, Kerlikowske K, Fenton JJ, Melnikow J, de Koning HJ, Hubbard RA.
   Radiation-Induced Breast Cancer Incidence and Mortality From Digital Mammography Screening: A Modeling Study. Annals of Internal Medicine. 2016 Feb 16;164(4):205-14. doi: 10.7326/M15-1241.
8. Trentham-Dietz A, Kerlikowske K, Stout NK, Miglioretti DL, Schechter CB, Ergun MA, **van den Broek JJ,** Alagoz O, Sprague BL, van Ravesteyn NT, Near AM, Gangnon RE, Hampton JM, Chandler Y, de Koning HJ, Mandelblatt JS, Tosteson AN; Breast Cancer Surveillance Consortium and the Cancer Intervention and Surveillance Modeling Network.
   Tailoring Breast Cancer Screening Intervals by Breast Density and Risk for Women Aged 50 Years or Older: Collaborative Modeling of Screening Outcomes. Annals of Internal Medicine. 2016 Nov 15;165(10):700-712. doi: 10.7326/M16-0476.
9. **Jeroen J. van den Broek**, Clyde B. Schechter, Nicolien T. van Ravesteyn, A. Cecile J.W. Janssens, Michael C. Wolfson, Amy Trentham-Dietz, Jacques Simard, Douglas F. Easton, Jeanne S. Mandelblatt, Peter Kraft, Harry J. de Koning
   Personalizing Breast Cancer Screening Based on Polygenic Risk and Family History (submitted)

Publications not in this thesis

10. Sprague BL, Stout NK, Schechter C, van Ravesteyn NT, Cevik M, Alagoz O, Lee CI, **van den Broek JJ**, Miglioretti DL, Mandelblatt JS, de Koning HJ, Kerlikowske K, Lehman CD, Tosteson AN.
    Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. Annals of Internal Medicine. 2015 Feb 3;162(3):157-66. doi: 10.7326/M14-0692.
11. Burnside Elizabeth, Lee Sandra, Bennette Carrie, Near Aimee, Alagoz Oguzhan, Huang Hui, **van den Broek Jeroen**, Kim Joo Yeon , Ergun Mehmet , van Ravesteyn Nicolien, Stout Natasha, de Koning Harry, Mandelblatt Jeanne.
    Using Collaborative Simulation Modeling to Develop a Web-Based Tool to Support Policy-Level Decision Making About Breast Cancer Screening Initiation Age. MDM Policy Pract. 2017;2.

PHD PORTFOLIO

Seminars, Conferences, Courses	Period	Workload
Genetic Risk Prediction, UVA Amsterdam, the Netherlands.	2013	0.2 ECTS
Endnote for researchers, Erasmus MC, department of Public Health, Rotterdam, the Netherlands.	2013	0.1 ECTS
Planning and Evaluation of screening (NIHES course), Rotterdam, the Netherlands.	2014	1.4 ECTS
Statistical Computing in R, Erasmus MC, department of Public Health, Rotterdam, the Netherlands.	2016	1.0 ECTS
Patient Oriented Research; Design, Conduct, Analysis & clinical Implications, Erasmus MC, Rotterdam, the Netherlands.	2015	0.3 ECTS
Scientific integrity, Erasmus MC, Rotterdam, the Netherlands.	2015	0.3 ECTS
International Cancer Screening Network Meeting, Rotterdam, the Netherlands	2015	0.7 ECTS
Teach the Teacher (BKO workshop), Erasmus MC Desiderius School, Rotterdam, the Netherlands.	2015	0.7 ECTS
Scientific Career Orientation, Erasmus MC, Rotterdam, the Netherlands	2015	0.1 ECTS
Biomedical English Writing, Erasmus MC, Rotterdam, the Netherlands	2015	1.4 ECTS
Time management course, Erasmus MC, Rotterdam, the Netherlands	2015	0.2 ECTS
Providing feedback (BKO workshop), Erasmus MC, Rotterdam, the Netherlands	2016	0.2 ECTS
Presenting for scientists, Erasmus MC, Rotterdam, the Netherlands	2016	0.2 ECTS
Basic Human Genetics, Molecular Medicine, Erasmus MC, Rotterdam, the Netherlands	2016	0.5 ECTS
Research Seminars Harvard School of Public Health Program in Genetic Epidemiology and Statistical Genetics, Boston, MA. United States.	2017	1.0 ECTS
Genetic Epidemiology of Cancer, Harvard School of Public Health Program in Genetic Epidemiology and Statistical Genetics, Boston, MA. United States.	2017	1.4 ECTS
Cancer Intervention & Surveillance Modeling Network Meeting Breast working Group meeting, Washington DC, United states.	2013	0.7 ECTS
Breast Cancer Surveillance Consortium Working Group Meeting, Bethesda, MD, United states.	2014	0.7 ECTS
Research Seminars, Erasmus MC, department of Public Health, Rotterdam, the Netherlands.	2013-2018	3.7 ECTS
International Cancer Screening Network Meeting, Rotterdam, the Netherlands	2019	0.7 ECTS
	Total	**15.5 ECTS**

Presentations, posters & workshops	Period	Workload
Cancer Intervention & Surveillance Modeling Network Meeting, 2 Oral presentations at the Breast Working Group. Minneapolis, MN, & Bethesda MD, United states.	2014	2.0 ECTS
International Cancer Screening Network Conference, Rotterdam, the Netherlands. Oral presentation.	2015	1.0 ECTS

Breast working Group, Cancer Intervention & Surveillance Modeling Network Meeting, Seattle, United states. Two Oral Presentations.	2015	2.0 ECTS
Research Seminar department of Public Health: Oral Presentation: "Validating CISNET breast cancer models on the outcomes of the Age Trial". Erasmus MC, Rotterdam, the Netherlands	2016	0.5 ECTS
Breast working Group (2x), Cancer Intervention & Surveillance Modeling Network Meeting, Boston MA, & Bethesda MD, United states. 2 Oral Presentations.	2016	2.0 ECTS
Breast cancer screening symposium, National University of Singapore, Singapore.	2016	1.0 ECTS
Breast working Group (2x), Cancer Intervention & Surveillance Modeling Network Meeting, Stanford University, Palo Alto, CA, & Bethesda MD, United states. 3 Oral Presentations.	2017	3.0 ECTS
Breast working Group, Cancer Intervention & Surveillance Modeling Network Meeting, Ann Arbor MI, United states. Oral Presentation.	2018	1.0 ECTS
Poster at the International Cancer Screening Network Meeting, Rotterdam, the Netherlands	2019	1.0 ECTS
	Total	**13.5 ECTS**

Teaching activities	**Period**	**Workload**
Community Project Mentor, Assisting two groups of 3rd year Bachelor of Medicine students with a project on "Health Literacy", Erasmus MC, Rotterdam, the Netherlands	2014 - 2015	1.4 ECTS
Reviewing Bachelor Essays 3rd year Bachelor of Medicine students, Erasmus MC, Rotterdam, the Netherlands	2014	0.7 ECTS
Breast cancer screening symposium, National University of Singapore, Singapore. Workshop: "Interactive hands-on workshop on Microsimulation model MISCAN-Fadia".	2016	1.0 ECTS
Mentor in Mentoring program for junior researchers, Erasmus MC, Rotterdam, the Netherlands.	2017	0.5 ECTS
	Total	**3.6 ECTS**

Other	**Period**	**Workload**
Harvard School of Public Health international exchange at the department of Genetic Epidemiology, Boston, MA. Collaboration with Peter Kraft Phd, Professor of Epidemiology on Breast cancer screening based on genetic risk profile (chapter 9 of this thesis). Attended several seminars, presentations, including in-person meetings at the Dana-Farber Cancer Institute, Harvard School of Public Health and Harvard Medical School.	2017 (4 months)	4.0 ECTS
Reviewing manuscript for Eastern Mediterranean Health Journal.	2014	0.3 ECTS
	Total	**4.3 ECTS**
	Total Total	**36.9 ECTS**

DANKWOORD

Veel collega's, familie, en vrienden hebben direct of indirect een bijdrage geleverd aan de totstandkoming van dit proefschrift. Een aantal mensen wil ik hier in het bijzonder bedanken voor hun bijdrage. Als eerste mijn promotor and copromoter, Harry en Nicolien. Ik ben jullie erg dankbaar voor het vertrouwen dat jullie in mij hebben gehad. Ik heb de ruimte gekregen om te groeien, projecten te leiden, en ons werk internationaal te presenteren, dat waardeer ik enorm.

Harry, jouw specifieke kennis over borstkankerscreening, strategisch inzicht, en internationale netwerk heeft een grote bijdrage geleverd aan dit proefschrift. Ik heb je ervaren als een promotor met een kritische wetenschappelijke visie én humor. Daarnaast ben je samen met Nicolien een mentor geweest in alles wat bij het internationaal samenwerken komt kijken. In de afgelopen jaren hebben we ook veel leuke niet-werkgerelateerde momenten gehad in binnen-en buitenland en daarbij is onze gedeelde interesse in muziek en goed voetbal ((PSV) natuurlijk) een mooie bijkomstigheid. Ik kijk ernaar uit om na mijn promotie mijn onderzoek bij het Erasmus MC voort te zetten en daarmee bij te kunnen dragen aan het verder terugdringen van de kankersterfte.

Nicolien, in het begin moest ik met mijn econometrie achtergrond behoorlijk wennen want ik was niet thuis in de medische wereld en zeker niet in het screenen op borstkanker. Er is niemand waar ik zoveel van heb geleerd in de afgelopen jaren als jij; je inhoudelijke kennis van het modelleren, analyzeren en verbanden leggen heeft een zeer grote directe en indirecte bijdrage gehad aan dit proefschrift en mijn ontwikkeling als wetenschapper. In de afgelopen jaren heb je een stuk meer verantwoordelijkheden gekregen, zowel op werkgebied als met het moederschap en ondanks dat stond je altijd voor mij klaar, was je relaxt én flexibel! Bedankt voor je fijne begeleiding en de prettige samenwerking

Mijn dank gaat ook uit naar Marjolein, de vele discussies over de MCLIR waren zeer nuttig, daardoor heb ik veel geleerd over het natuurlijke beloop van kanker op modelniveau. Collega's van de screensectie, oud-kamergenoten, paranimfen Kevin en Lisa, en in het bijzonder het 'breast cancer team', bedankt voor de inhoudelijke discussies en gezelligheid.

Ik wil ook de leden van de kleine commissie hartelijk danken voor het beoordelen van mijn proefschrift, maar ook de andere leden van de promotiecommissie voor hun de interesse en bereidheid om met mij hierover van gedachten te wisselen.

Very important, this work is to a great extent the result of many collaborations within CISNET and the Breast Cancer Surveillance Consortium. In particular Jeanne, Tasha, Oguz, Clyde, Amy, Sandra, Don, Sylvia, and Diana, our in-person discussions, shared work and your contributions have been invaluable over the years. I would also like to thank all other members of cisnet breast working group and coauthors for their contribution and inspiration.

Peter I'm very grateful for the opportunity that you provided to extend my research at the Harvard School of Public Health. Your expertise and guidance in the polygenetics of breast cancer has been profound. I count myself lucky for having had your mentorship. Esther and Chris, you were wonderful colleagues to spend time with during and after work in Boston.

Lieve pap en mam, (Andre en Eveline), de onvoorwaardelijke liefde en mogelijkheden die ik heb gekregen van jongs af aan maken jullie echt geweldige ouders en tegelijkertijd goede vrienden die ik zonder enige twijfel iedereen kan toewensen. Stijn, ondanks dat we verschillend zijn ben je altijd jezelf en daarmee een voorbeeld. Lieve mooie Lisa, bedankt dat je mij al die jaren gevoed hebt met liefde, lekker eten, gezelschap en vertier; ook dankzij jou is mijn proefschrift afgekomen.