EVALUATION AND OPTIMIZATION OF BREAST CANCER SCREENING PROGRAMS:
THROUGH THE LENS OF EUROPE

Nadine Zielonke

Evaluation and Optimization of Breast Cancer Screening Programs: Through the Lens of Europe

The studies reported in this thesis were primarily funded by the European Unions' Horizon 2020 Programme project EU-TOPIA (Towards improved screening for breast, cervical and colorectal cancer in all of Europe) under grant agreement no 634753.

ISBN: 978-94-6423-501-2
Cover based on drawing of: Jasmine Lalta
Cover and inside layout by: Bregje Jaspers | www.proefschriftontwerp.nl
Print: ProefschriftMaken | www.proefschriftmaken.nl

The thesis was printed with financial support of the Department of Public Health, Erasmus MC University Medical Center Rotterdam and the Erasmus University Rotterdam.

Evaluation and Optimization of Breast Cancer Screening Programs:
Through the Lens of Europe

Evaluatie en optimalisatie van borstkankerscreeningprogramma's:
door de lens van Europa

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Thursday 9 December 2021 at 10.30hrs

By
Nadine Zielonke
born in Wismar, Germany

Erasmus University Rotterdam

Doctoral Committee:

Promotor:	prof. dr. H.J. de Koning
Other members:	prof. E. de Bekker-Grob prof.dr. M. Broeders prof. Z. Voko
Copromotor:	dr. N.T. van Ravesteyn

CONTENT

Chapter 1

General introduction

BREAST CANCER IN EUROPE

Breast cancer is a major public health problem in Europe. It is by far the most frequently diagnosed neoplasm in European women. At present, women in Europe have a 1:7 chance of developing breast cancer during their lifetime[1]. It is estimated that in 2020 there were 531,086 new diagnoses (26% of all new cancer diagnoses in women) in Europe, corresponding to an age-standardized incidence rate of 74.3 per 100,000 women.

There is substantial regional variation in breast cancer incidence rates throughout Europe. In 2020, rates were lowest in Eastern Europe (average of 57.1 per 100,000) and highest in Northern Europe (90.7 per 100,000)[2], Figure 1.
Despite important progress in detection, major advances in technology and shifts in treatment paradigms and therefore survival, there were 141,765 breast cancer deaths (16% of all cancer deaths in women) estimated for 2020[3]. The age-standardized mortality rate in Europe in 2020 was 14.8 per 100,000 women. In many countries, breast cancer mortality has decreased by 0.8%-1.5% per year since 2006. However, the decline started a decade later in Central and Eastern Europe and age-standardised mortality rates of breast cancer are considerably higher in Eastern European countries than in Western European countries (Figure 1). Great inequity thus persists in cancer mortality across Europe. The lowest mortality rates per 100,000 were estimated in Southern Europe (average of 13.3 per 100,000) and the highest in Eastern Europe (15.6 per 100,000).

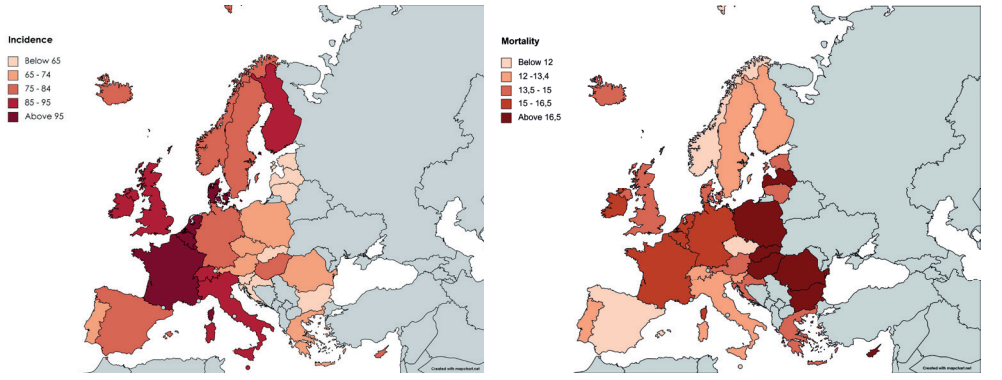


Figure 1. Age-standardized rate per 100,000 women, per country in 2020 (Source: Cancer Today[2])

Survival of breast cancer continues to vary across Europe. The five-year survival of breast cancer patients diagnosed in 2000-2007 is highest in Northern and Western Europe (85% and 84%, respectively) and lowest in Eastern Europe (75%)[4], mostly due to more advanced stage at diagnosis[5] and suboptimum access to adequate care[6].

Besides female sex, advancing age is the biggest breast cancer risk factor (Figure 2). There is a steep age gradient, with over 80% of all new cases occurring after the age of 50[4].

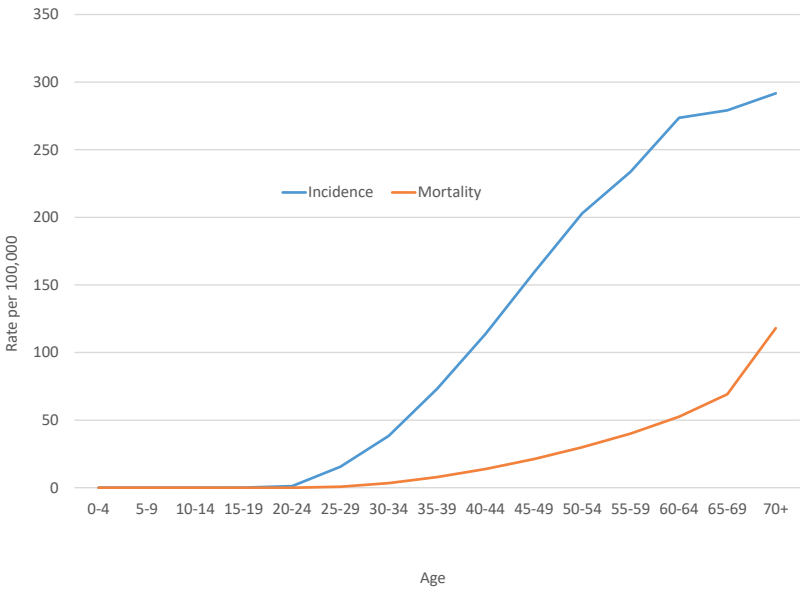


Figure 2. European breast cancer incidence and mortality by age in 2020 (Source: Cancer Today[2])

The classification of cancer by anatomic disease extent, i.e. stage, is often based on the TNM classification system, which defines the size of the tumor (T), possible regional lymph node involvement (N) and possible distant metastases (M). The stage of the breast cancer is the major determinant for prognosis and used for treatment selection. The larger the size, the less likely it is that the tumor can be cured[7 8]. Stage is an increasingly important component of cancer surveillance and cancer control and is used for the evaluation of screening and early detection efforts[9].

BREAST CANCER SCREENING

Through early detection in asymptomatic women, breast cancer screening aims to reduce morbidity associated with advanced stages of the disease, as well as cancer-specific mortality. Breast cancer can be detected earlier during the screen detectable phase and the diagnosis is brought forward (Figure 3).

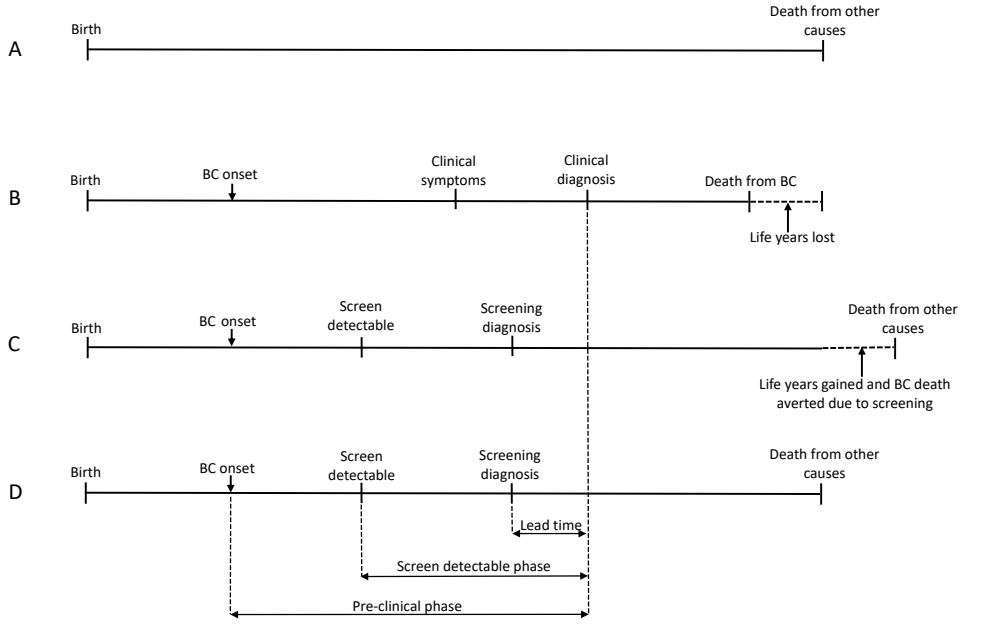


Figure 3. Four possible life-history scenarios

A: woman without breast cancer. B: woman with breast cancer, without screening. C: Woman with breast cancer, where screening brings the cancer diagnosis forward and she gains life years. D: Woman with breast cancer, where screening brings the cancer diagnosis forward, but her death is not postponed and no life years are gained. Screening is beneficial if life years are gained and the time of death is postponed (C). But screening is also considered beneficial if breast cancer is detected earlier, treatment is therefore less intensive and the quality of life might improve, although no additional life years are gained (D). The *pre-clinical phase* is the period of time between tumor inception and clinical diagnosis in the absence of screening. The *screen detectable phase* (also termed sojourn time*)* for a screening test, e.g., mammography, is the period that a cancer can be screen-detected. 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 the onset of breast cancer. The time that the diagnosis is brought forward by mammography is called *lead time*. Because screening brings the date of diagnosis forward, the period between diagnosis and death is generally longer with screening, even if breast cancer death is not postponed or prevented (D). BC: breast cancer

Randomised controlled trials and several observational studies have demonstrated that systematic screening of eligible women through quality-assured population-based programs for breast cancer reduces mortality from this disease[10-19]. Organized screening contributes to better care and earlier diagnosis: the introduction of screening is accompanied by a decrease in cases diagnosed at an advanced stage[20], likely also due to increased awareness of the disease by women and general practitioners[21], and by enabling facilities to diagnose and treat the disease[11]. According to the IARC Handbook of Cancer Prevention (2015)[22], organized screening programs are defined as organized at the national or regional level, with an explicit policy, including an active invitation of the entire target population and monitoring of cancer occurrence in the

target population. Breast cancer screening programs in Europe make use of (mostly digital) mammography, which is an X-ray image taken of the breasts called a mammogram, and has a relatively high sensitivity and specificity[23].

Along with the benefits of mortality reduction and life-years gained, breast cancer screening can also cause harm. The most important harms of breast cancer screening relate to overdiagnosis and overtreatment. Overdiagnosed cancers are cancers that would not have been detected in the absence of screening during the lifetime. Other important harms are false-positive results of the screening test and the possible risk of radiation-induced breast cancer[12]. Effectiveness and the balances of benefits and harms can vary greatly between cancer screening programs in different countries and settings[10].

BREAST CANCER SCREENING IN EUROPE

Based on the evidence of the beneficial effect of breast cancer screening, in 2003 the European Council published their first guidelines for organized mammography screening programs for early detection of breast cancer in asymptomatic women with a strong recommendation to inviting women ages 50-69, every two years[24 25] .

At present, breast cancer screening programs are well established in most European countries and all have some form of screening for breast cancer. Following the European recommendations and implementing organized screening programs was instrumental in ensuring that the vast majority of women in the chosen target age range in the EU member states have access to organized screening for breast cancer. But despite intensified efforts by the European Council since 2003, the implementation of organized, population-based mammography screening is not uniform across Europe and depends greatly on the policies in place in different countries, the organization of health care, and available resources[26 27].

Disparities exist in terms of the status of implementation, the extent to which screening programs are organized, the invitation coverage and the attendance to screening[27]. Some countries, particularly those in Eastern Europe, have less well developed programs or have not yet implemented (organized) screening[22]. Examination coverage (i.e., the proportion of the target population screened after invitation) varies across Europe, from less than 17% in Cyprus to more than 83% in the United Kingdom[10].

Disparities also exist with respect to the practiced screening strategy. Most European countries adopted the target age range for breast cancer screening as recommended by the European Commission for which there is a strong recommendation (50-69). Only a few countries adopted

a different age range and either invite women younger than 50 or they invite women beyond the age of 69. The screening interval is two years in all countries except for Malta and the United Kingdom (UK) where three yearly screening is practiced[27].

Beside these differences, European countries also differ with respect to the coexistence of organized and opportunistic screening activity. Opportunistic or non-organised screening refers to all other breast cancer screening activity where individual invitations are not sent to the women in the eligible population or when women undergo a mammography outside or additionally to the (existing) screening programme. Opportunistic screening could under certain circumstances be an alternative to programmatic population based screening, but is generally far more expensive and less efficient[28-30].

EVALUATION OF SCREENING

For screening to be effective, the screening program should be well organized and ensure high coverage of the eligible population; must incorporate a screening test that is accurate, feasible, affordable, culturally acceptable, and safe; should provide prompt diagnostic, treatment, and follow-up services to those with positive results from screening tests; and should ensure the quality at every step. It is very important to ensure a good balance between the harms and benefits of breast cancer screening. Regular evaluation provides an opportunity to consider whether crucial performance indicators of the program (e.g. recall rate, interval cancer rate or detection rate[31]) are changing or whether, in the light of new research findings, a screening program should be modified or even ended. Unfortunately, there are only few screening programs that have accomplished long-term evaluations on the balance between harms and benefits[10]. Often only short-term indicators for benefits and harms are available.

THE ROLE OF MICROSIMULATION MODELING

The sheer complexity of population screening makes it difficult to evaluate. For example, outcomes in situations with and without screening or with and without a certain change to the screening program are not known at the same time for the same population. In addition, changes to screening programs often accumulate over time, making it impossible to disentangle the impact of one change from the other. Besides, a program's benefits and harms can be estimated only after a sufficiently long follow-up time. A simulation model is a helpful tool to estimate the effect of each of the listed factors on cancer incidence and mortality. The MIcrosimulation SCreening ANalysis (MISCAN) model[32] is such a tool. To inform screening policy, MISCAN can

be used to evaluate and compare a wide range of different screening strategies, while holding selected conditions (e.g. the effect of treatment) constant[33 34].

MISCAN simulates individual life histories from birth to death, with and without breast cancer, in the presence and in the absence of screening. The model consists of four main components: population demographics, natural history of breast cancer, screening, and treatment. The model simulates a population consisting of individual life histories (Figure 3), based on a country's life-table. Subsequently, the natural history of breast cancer (in the absence of screening) is simulated resulting in the onset of breast cancer in a subset of women, which may eventually lead to breast cancer death. First, the model estimates outcomes for a situation without screening. Following, mammography screening and survival improvements after screen-detection are modelled. In the presence of screening, tumors can be screen-detected during the pre-clinical phase, before clinical symptoms are present. Therefore, screening may alter the simulated life histories, detecting cancer at an earlier stage and effective earlier treatment may prevent dying from the disease. Besides, screening might result in overdiagnosis and false positive results. MISCAN quantifies the long-term effectiveness of screening comparing all life histories in the absence of screening with the corresponding life histories in the presence of screening.

MISCAN–Breast has been developed to analyze trends in breast cancer incidence and mortality due to changes in lifestyle, improvement of treatment and implementation of screening strategies. In addition, the model can improve the understanding of the natural history of breast cancer, screening interventions and treatment and their effects on population trends in incidence and mortality. The model can be used to guide public health research and priorities, and can aid in the development of optimal cancer control strategies, e.g. by extending the age-range of eligible women optimising the attendance rate of the program.

THE EU-TOPIA PROJECT

The burden of breast cancer as well as breast cancer screening programs vary considerably throughout Europe and the long-term effectiveness of screening has only been assessed in a few countries[27]. These substantial differences may result in inappropriate interventions, excessive screening, and overtreatment on the one hand, or under-screening, delayed provision of appropriate treatment on the other. In an effort to identify opportunities to improve cancer outcomes across Europe, EU-TOPIA (Towards improved screening for breast, cervical and colorectal cancer in all of Europe), a five year project (2015-2020) funded by the European Commission's Horizon 2020 programme, was initiated.

In EU-TOPIA, we aimed to improve screening outcomes and to build capacity for future self-evaluation of the three existing cancer screening programs (breast, cervical and colorectal) throughout Europe by following the concept of the health policy cycle (Figure 4).

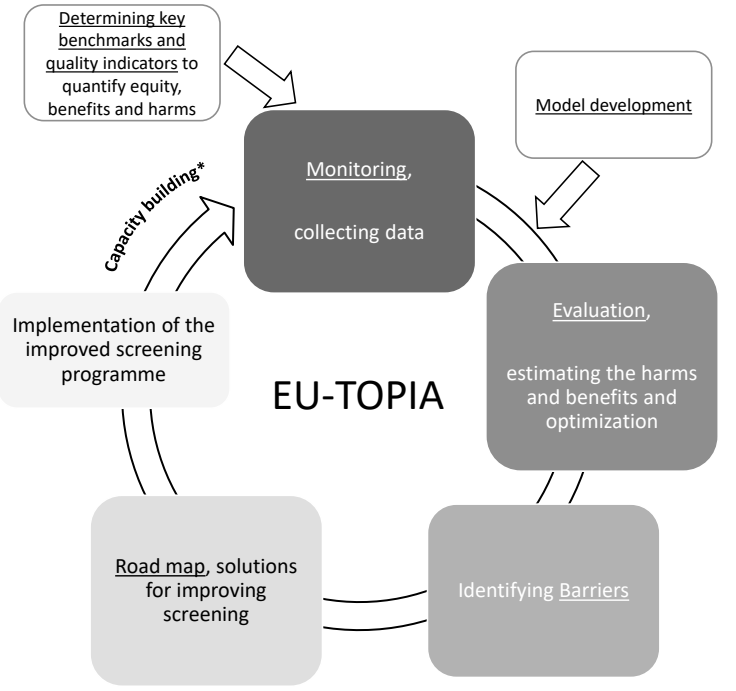


Figure 4. The EU-TOPIA Health policy cycle

Throughout the project, European policymakers were guided to develop their own road maps and actions towards improved screening programs. Road maps are defined as a strategic plan that includes a policy guide for improving cancer screening. They contain feasible changes, e.g., to extend or reduce the program, to change the screen test used, or to incorporate new developments in screening, and provide policymakers with evidence for increased, decreased or optimized use of screening. The EU-TOPIA research team provided tools to perform the consecutive steps that were defined as part of the health policy cycle leading to country-specific roadmaps:

1) Monitoring of current effectiveness of the screening program (EU-TOPIA Monitoring Tool)
2) Evaluation of future harms and benefits (EU-TOPIA Evaluation Tool)

3) Identification of barriers that hinder the implementation of optimal screening programs (EU-TOPIA Barrier Assessment Tool)
4) Scenario analyses of changes in cancer screening practices (EU-TOPIA Evaluation Tool)
5) Identification of key stakeholders (EU-TOPIA Stakeholder Analysis Tool)
6) Providing policy recommendations (EU-TOPIA Action Plans).

Almost all studies included in this thesis were performed within this project. For more detailed information about EU-TOPIA, the tools and workshops, consult the webpage www.eu-topia.org.

RESEARCH QUESTIONS AND THESIS OUTLINE

In this thesis, the consequences of variations of breast cancer screening practices and potential ways to (further) optimize screening programs across Europe are investigated.

This thesis is broadly split into two parts. Part 1 (Chapters 2-4) describes the effectiveness of breast cancer screening in Europe. Part 2 (Chapters 5-7) of this thesis explores different interventions and their impact on harms and benefits of breast cancer screening.

Part 1: The effectiveness of breast cancer screening: How well is it really working?
In Chapter 2, we summarize current evidence of breast cancer mortality reduction due to mammography screening for each European region. We include different types of studies in this systematic review, using a methodologically sound quality appraisal.

Chapter 3 presents an analysis of systematic reviews on benefits and harms of breast cancer screening. We will summarize data on different screening approaches and synthesize the results.

The following Chapter 4 illustrates how breast cancer screening in Europe has already impacted breast cancer mortality and how, through further optimising screening coverage, the number of breast cancer deaths of European women could be further reduced.

Part 2: Modelling the impact of different interventions on the harms and benefits of breast cancer screening: How can harmful screening effects be reduced and positive effects enhanced? And are there better screening practices possible?

In Chapter 5, we assess how harm-to benefit ratios vary if breast cancer screening would be extended to younger and/or older age groups compared to the age group 50-69. The hypothetical effects of adapting the age-range (i.e. 45-69, 45-74 or 50-74) is simulated for four European countries: The Netherlands, Finland, Italy and Slovenia, countries with high quality observational data which were selected to be representative of the four European regions.

The objective of Chapter 6 is to identify feasible changes leading to better long-term outcomes of breast cancer screening programs in Italy. Based on certain practical or cultural barriers, we aim to quantify the harms, benefits and costs of several screening strategies, considering overcoming said barriers.

With Chapter 7, we aim to assess whether the use of disability-adjusted life years (DALYs) averted or quality-adjusted life years (QALYs) gained in cost-effectiveness analyses of 24 screening strategies leads to differences in estimates of optimal strategies for breast cancer screening in The Netherlands.

Finally, this thesis is concluded with a general discussion in Chapter 8, which will answer the research questions and present implications and opportunities for future research.

REFERENCES

1. The European Commission. ECIS - European Cancer Information System. 2019;Available at: https://ecis.jrc.ec.europa.eu/.
2. Ferlay J, Ervik M, Lam F, et al. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.fr/today, accessed [14/05/2021].
3. Sung H, Ferlay J, Siegel R, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021
4. ECIS - European Cancer Information System: European Commission; [Available from: https://ecis.jrc.ec.europa.eu/.
5. Minicozzi P, Walsh PM, Sanchez MJ, et al. Is low survival for cancer in Eastern Europe due principally to late stage at diagnosis? Eur J Cancer 2018;93:127-37.
6. Sant M, Chirlaque Lopez MD, Agresti R, et al. Survival of women with cancers of breast and genital organs in Europe 1999-2007: Results of the EUROCARE-5 study. Eur J Cancer 2015;51(15):2191-205.
7. Saadatmand S, Bretveld R, Siesling S, et al. Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients. BMJ 2015;351:h4901.
8. Sant M, Allemani C, Capocaccia R, et al. Stage at diagnosis is a key explanation of differences in breast cancer survival across Europe. International Journal of Cancer 2003;106(3):416-22.
9. The Union for International Cancer Control. https://www.uicc.org/.
10. IARC. Cancer Screening in the European Union. Report on the implementation of the Council Recommendation on cancer screening. IARC, Lyon 2017;Available at: 2017_cancerscreening_2ndreportimplementation_en.pdf.
11. Marmot MG, Altman DG, Cameron DA, et al. The benefits and harms of breast cancer screening: an independent review. Br J Cancer 2013;108(11):2205-40.
12. Lauby-Secretan B, Scoccianti C, Loomis D, et al. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med 2015;372(24):2353-8.
13. Moss SM, Nyström L, Jonsson H, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of trend studies. J Med Screen 2012;19 Suppl 1:26-32.
14. Broeders M, Moss S, Nyström L, 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.
15. de Koning HJ. Mammographic screening: evidence from randomised controlled trials. Ann Oncol 2003;14(8):1185-9.
16. Nelson HD, Fu R, Cantor A, et al. Effectiveness of Breast Cancer Screening: Systematic Review and Meta-analysis to Update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med 2016;164(4):244-55.
17. Heinavaara S, Sarkeala T, Anttila A. Impact of organised mammography screening on breast cancer mortality in a case-control and cohort study. Br J Cancer 2016.
18. Tabár L, Vitak B, Chen THH, et al. Swedish two-county trial: Impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011;260(3):658-63.

19. Puliti D, Miccinesi G, Collina N, et al. Effectiveness of service screening: A case-control study to assess breast cancer mortality reduction. Br J Cancer 2008;99(3):423-27.
20. Canelo-Aybar C, Ferreira DS, Ballesteros M, et al. Benefits and harms of breast cancer mammography screening for women at average risk of breast cancer: A systematic review for the European Commission Initiative on Breast Cancer. J Med Screen 2021:969141321993866.
21. Mandrik O, Tolma E, Zielonke N, et al. Systematic reviews as a "lens of evidence": Determinants of participation in breast cancer screening. J Med Screen 2020:969141320930743.
22. IARC Working Group. IARC Handbooks of Cancer Prevention, Volume 15: Breast Cancer Screening. Lyon, 2016.
23. Pisano ED, Hendrick RE, Yaffe MJ, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology 2008;246(2):376-83.
24. Council of the European Union. Council Recommendation of 2 December 2003 on Cancer Screening (2003/878/EC). OJ L 327: 34-38. 2003.
25. Commission E. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth Edition. 2006.
26. Deandrea S, Molina-Barcelo A, Uluturk A, et al. Presence, characteristics and equity of access to breast cancer screening programmes in 27 European countries in 2010 and 2014. Results from an international survey. Prev Med 2016;91:250-63.
27. Basu P, Ponti A, Anttila A, et al. Status of implementation and organization of cancer screening in The European Union Member States-Summary results from the second European screening report. Int J Cancer 2018;142(1):44-56.
28. Mantellini P, Lippi G, . I costi dello screening. Un'analisi in logica activity based. Firenze Zadig Editore, 2011.
29. Bulliard JL, Ducros C, Jemelin C, et al. Effectiveness of organised versus opportunistic mammography screening. Ann Oncol 2009;20(7):1199-202.
30. de Gelder R, Bulliard JL, de Wolf C, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. Eur J Cancer 2009;45(1):127-38.
31. Muratov S, Canelo-Aybar C, Tarride JE, et al. Monitoring and evaluation of breast cancer screening programmes: selecting candidate performance indicators. BMC Cancer 2020;20(1):795.
32. Sankatsing VDV, Heijnsdijk EAM, van Luijt PA, et al. Cost-effectiveness of digital mammography screening before the age of 50 in The Netherlands. Int J Cancer 2015;137(8):1990-9.
33. de Gelder R, Heijnsdijk EA, Fracheboud J, et al. 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):165-72.
34. Plevritis SK, Munoz D, Kurian AW, 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.

Part 1:

THE EFFECTIVENESS OF BREAST CANCER SCREENING

Chapter 2

Evidence for reducing cancer specific mortality due to screening for breast cancer in Europe: a systematic review.

Zielonke N, Gini A, Jansen EEL, Anttila A, Segnan N, Ponti A, Veerus P, de Koning HJ, van Ravesteyn NT, Heijnsdijk EAM; EU-TOPIA consortium.

Eur J Cancer 2020; 127: 191-206.

ABSTRACT

Background

The aim of this study was to quantify the impact of organized mammography screening on breast cancer mortality across European regions. Therefore, a systematic review was performed including different types of studies from all European regions and stringently used clearly defined quality appraisal to summarize the best evidence.

Methods

Six databases were searched including Embase, Medline and Web of Science from inception to March 2018. To identify all eligible studies which assessed the effect of organized screening on breast cancer mortality, two reviewers independently applied predefined inclusion and exclusion criteria. Original studies in English with a minimum follow-up of five years that were randomized controlled trials (RCTs) or observational studies were included. The Cochrane risk of bias instrument and the Newcastle-Ottawa Scale were used to assess the risk of bias.

Results

Of the 5,015 references initially retrieved, 60 were included in the final analysis. Those comprised 36 cohort studies, 17 case-control studies and 7 RCTs. None were from Eastern Europe. The quality of the included studies varied: 19 were of very good or good quality. Of those, the reduction in breast cancer mortality in attenders versus non-attenders ranged between 33%-43% (Northern Europe), 43%-45% (Southern Europe) and 12%-58% (Western Europe). The estimates ranged between 4%-31% in invited versus non-invited.

Conclusion

This systematic review provides evidence that organized screening reduces breast cancer mortality in all European regions where screening was implemented and monitored, while quantification is still lacking for Eastern Europe. The wide range of estimates indicates large differences in the evaluation designs between studies, rather than in the effectiveness of screening.

INTRODUCTION

Breast cancer (BC) has become the most common cancer in women worldwide in both developed and developing countries[1 2]. Through early detection in asymptomatic women, screening aims to reduce morbidity associated with advanced stages of the disease, as well as cancer-specific mortality However, the benefits and harms of mammography screening have been debated heatedly in the last decades[3].

It is thirty-five years since randomized controlled trials (RCT) showed that mammography screening leads to a reduction in BC mortality[4], which resulted in various policy recommendations[5]. More recently, the effect of running mammography screening outside the experimental setting has been assessed. Several observational studies have demonstrated that BC screening reduces BC mortality[6-12]. However, screening has also harms. After careful evaluation of the balance between the benefits and adverse effects of mammography screening, the most recent review by the International Agency for Research on Cancer (IARC) concluded that there is a net benefit from inviting women 50 to 69 years as well as sufficient evidence for women up to 74 years of age to receive screening[3].

At present, population-based BC screening programs are ongoing, piloted or planned in 25 out of 28 EU member states for nearly 95% of women in the age group of 50-69 years[13]. BC screening is delivered mainly by organized programs encouraged by the European Commission, which has published quality assurance guidelines[14], which are currently being updated[15]. There is wide agreement on different aspects of the screening policy, such as the screening test (mammography), the minimum target age range (50–69 years) and the screening interval (two years). On the other hand, there are substantial differences within the European Union (EU) in the extent to which target populations are actually exposed to screening[16]. Currently there is nearly a two-fold difference among the EU-countries in the coverage by invitations and a more than five-fold difference in the participation rate reported[17].

A considerable number of systematic reviews have estimated the effectiveness of mammography screening in terms of a reduction in BC mortality. Some of these reviews included only RCTs[18-21], whereas others focused exclusively on observational studies[8 11 12]. Several reviews did not follow a standardized quality appraisal protocol[8 19 22 23]. These reviews demonstrated high variability in estimates which led to different conclusions and recommendations on the most appropriate screening strategy. The probably most extensive and recent review was done for the IARC handbook. Their average estimate was 40% reduction in the risk of death from BC for women attending mammographic screening[1]. However, to our knowledge, no review has summarized the current evidence for all European regions, including different types of studies, using a methodologically sound quality appraisal. The aim of this systematic review, therefore

was to systematically evaluate and quantify the impact of organized screening on BC-specific mortality across Europe.

The objective of this review is to answer the following questions:

1. What is the impact of organized BC cancer screening on BC mortality across Europe?
2. What are the differences between regions in Europe with regards to BC mortality reduction due to screening?

METHODS

This systematic review was done in accordance to a peer-reviewed protocol that is published and registered with PROSPERO (CRD42016042433)[24]. We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement and checklist when reporting our findings[25].

All methodological steps were performed by two independent reviewers (N.Z. and A.G.). Disagreement between the two investigators were solved by consensus or by consulting a third independent reviewer (E.E.L.J.).

Data sources and search

Study design, Table 1) served to define specific keywords used in our comprehensive bibliographic searches. Systematic bibliographic searches were conducted on the Embase, Medline Ovid, Web of Science, PubMed publisher, Google Scholar, and Cochrane Library. All databases were searched from inception to March 2018. The computer-assisted searches were designed and performed by a research librarian using controlled keywords to assess the concepts related to mammographic screening, BC and mortality among European countries. In Appendix 1, the detailed search strategies performed for every source are listed. To augment the search and to improve the likelihood of identifying studies that are only indexed in local journals, experts were asked to suggest additional articles that were not retrieved through the above-mentioned search strategy. Additional potentially eligible articles were identified by hand searching the reference lists of all included studies. The search was limited to articles written in English conducted in any European country and the authors only considered studies that included data from RCTs or observational studies such as prospective and retrospective controlled cohort or case-control studies. All references were managed in Thomson Reuters Endnote X7.1 and duplicates were removed.

Table 1. PICOS criteria for inclusion and exclusion of studies

Parameter	Inclusion criteria	Exclusion criteria
Population	People invited to / participating in organized[1] mass screening for breast cancer in a European country[2]	People from non-European countries.
Intervention	Organized screening for breast, cancer.	Other screening interventions (e.g. breast self-examination)
Control	People not invited to/not attending organized screening or people participating in opportunistic screening only.	No control group (everybody is screened)
Outcome	Change in breast cancer mortality due to screening.	No direct estimation of breast cancer mortality reduction due to screening[3]
Study design	Randomized controlled trials, retrospective and prospective observational (cohort or case-control) studies	Non-original research studies (e.g. editorials, letters, and conference abstracts), modeling/simulation studies, ecological studies[4].

[1]Based on the IARC Handbook of Cancer Prevention (2015) we defined organized screening as screening programs organized at national or regional level, with an explicit policy [2]Western Europe: Austria, Belgium, France, Germany, Ireland, Luxembourg, the Netherlands, United Kingdom and Switzerland. Northern Europe: Denmark, Estonia, Faroe Islands, Finland, Iceland, Latvia, Lithuania, Norway and Sweden. Southern Europe: Cyprus, Gibralta, Greece, Italy, Malta, Portugal, Spain. Eastern Europe: Bulgaria, Czech Republic, Croatia, Hungary, Poland, Romania, Slovakia, Slovenia. [3]Studies that only provide estimates on changes of survival rates, were excluded. [4]Ecological studies that simply compare trends between unmatched regions or single regions over time without statistical adjustments for e.g. baseline risk, excluded.

Study Selection

Two investigators independently reviewed the titles and abstracts of all references identified by the literature search by using the PICOS criteria displayed in Table 1. Then all potentially suitable articles were reviewed in depth and additional exclusions have been made applying eligibility criteria proposed by Elmunzer and colleagues[26]: i) studies in which data or patients were duplicated in other manuscripts; ii) studies in which data were not reported for at least 5 years of follow-up; iii) studies in which the total number of events and participants were not reported for each study group. If multiple studies compared the same region, period or population, or reported on the same trial, the study with the longest follow-up was retained. The full texts of all included publications were screened for eligibility. Relevant outcome data and study details such as first author; year of publication; country where the study was conducted; study design; screening target population; follow-up information; sample size of the study; assessment of confounding factors (such as adjustment for self-selection bias), and the reported estimates (with corresponding 95% confidence intervals [95%CI]) of the screening effect on BC mortality, were extracted. Furthermore, eligible articles were grouped according to European regions

(Northern, Western, Southern and Eastern Europe) following the classification provided by the EUROVOC Multilingual Thesaurus of the European Union[27].

Quality appraisal

We used the Cochrane risk of bias instrument[28] to assess the quality of the included RCTs. This tool helps to indicate the validity of the studies' results and the presence of any systematic error leading to an overestimation or underestimation of the true intervention effect. The tool considers the risk of bias within five domains, including randomized sequence generation, allocation concealment, masking outcome assessors, incomplete outcome data reporting, and selective outcome reporting. However, the sixth domain from the original tool, masking of participants and personnel, was not applied in this systematic review as it was deemed not applicable to screening. The reviewers judged each of the five domains and agreed in an overall judgement for each study as low, moderate, or high risk of bias.

To judge the quality of observational studies we used the Newcastle-Ottawa Scale (NOS)[29]. Using the tool with its specific questions for cohort studies and case-control studies respectively, each study is judged on several items, categorized into three areas: the selection of the study groups, the comparability of the groups, and the ascertainment of either the exposure or outcome of interest. Originally, the NOS does not award a point if the case definition of case-control studies was exclusively based on record linkage. However, many studies on cancer screening are based on data from cancer registries. As advocated in Anttila[30] cancer registries can be hold co-responsible for the quality and impact assessment of screening programmes when mandated and resourced adequately. Thus, when the percentage of histologically verified cases of the respective cancer register was known to be above 95% according to the International Agency for Research on Cancer[31 32], we qualified the case definition as independent validation and award a point on this question.

The highest quality studies are awarded with a score of nine.

The risk of bias of studies included in this review has been categorized as follows:

(I) Low risk RCTs,

(II) Moderate risk RCTs or score of 8 or 9 in observational studies,

(III) High risk RCTs or score 5 to 7 in observational studies and

(IV) Observational studies with a score from 0 to 4.

RESULTS

The PRISMA flowchart (Figure 1) presents the number of articles found and excluded in each stage. The initial search retrieved a total of 6,691 citations. The augmenting bibliographic search provided 153 additional references. After removing duplicates, 5,015 citations were identified of which 150

potentially eligible articles were selected for detailed full-text evaluation. After the preliminary full-text analysis using the eligibility criteria mentioned above, 89 references were excluded from our review. A detailed overview of the reasons for exclusion are presented in Appendix 3.

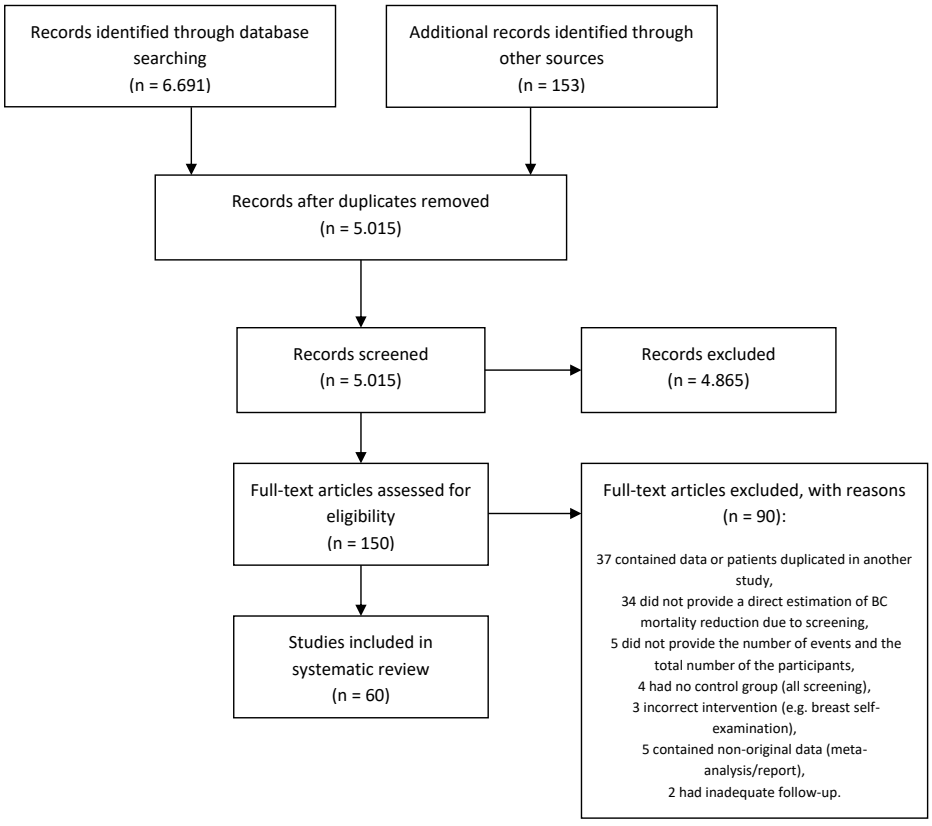


Figure 1. PRISMA flow chart of records through the review searching and inclusion process

Of the 153 additional references from experts, 31 were already identified by our initial literature search. The remaining 122 studies were excluded mostly because they did not provide a direct estimation of the impact of BC screening on cancer specific mortality.

In total 60 studies were included in the final in-depth analysis. Those included 38 cohort studies[33-70], 17 case-control studies[71-87] and 7 randomized controlled trials[88-94]. Details of these studies and their main characteristics – sorted by European region – are reported in Table 2 a-c. Thirty studies were included for Northern Europe[33-55 68 74 75 89-92 94]; 9 for Southern Europe[56-61 81 82 86]; and 22 studies for Western Europe[62-67 69-73 76-80 83-85 87 88 93]. None of the included studies came from Eastern Europe. The majority of studies (51/60) covered the age group 50-69.

Table 2a. Characteristics, risk of bias and results on breast cancer mortality of included studies, by quality score. Northern Europe

Study	Country	Study type	Participants	Attendance	Target age (years)	Follow-up (years)	Correction for self-selection bias	Quality score[a]	Effect sizes for breast cancer mortality, RR/ HR/OR (95%CI)	
									Invited vs not invited	Attenders vs not attenders
Andersson I, 1988	Sweden	RCT	I:21.088, C: 21.195	85%	45-79	9	n/a	A	RR = 0.96 (0.68-1.35)	
Tabar L., 2011	Sweden	RCT	I: 77.080, C: 55.985	85%	40-74	29	n/a	A	RR = 0.69 (0.56-0.84)	
Andersson I, 1997	Sweden	RCT	I: 13.528, C: 12.242	NA	<50	10	n/a	B	RR = 0.64 (0.45-0.89)	
Heinävaara S, 2016	Finland	Case-control	Cases: 1.907 Controls: 18.978	86%	50-69	7.4	yes	8/9		HR = 0.67 (0.49-0.90)
Olsen AH, 2007	Denmark	Cohort	Participants: 430.823 pyr, Non-participants: 634.224 pyr†	NA	50-69	10	n/a	8/9	RR = 0.80 (0.68–0.94)	
Olsen AH, 2013	Norway	Cohort	Participants: 1.182747 pyr, Non-participants: 1.152.755 pyr†	NA	50-69	6	n/a	8/9	RR = 0.93 (0.77-1.12)	
Tabar L, 2003	Sweden	Cohort	Participants: 2.399.000 pyr, Non-participants: 2.416.000 pyr†	85%	40-69	20	yes	8/9		RR = 0.59 (0.53-0.66)
Weedon-Fekjaer H, 2014	Norway	Cohort	Participants: 2.407.709 pyr, Non-participants: 12.785.325 pyr†	76%	50-69	15	n/a	8/9	RR = 0.72 (0.64-0.79)	
Hofvind S, 2013	Norway	Cohort	Participants: 4.814.060 pyr, Non-participants: 988.641 pyr†	NA	50-69	15	yes	8/9		RR = 0.57 (0.51-0.64)

Bjurtsam N, 2016	Sweden	RCT	I: 21.904, C: 30.318	84%	39-59	14	n/a	C	RR = 0.70 (0.53-0.93)	
Frisell J, 1997	Sweden	RCT	I: 40.318, C: 19.943	N/A	40-64	11.4	n/a	C	RR = 0.74 (0.50-1.10)	
Anttila A, 2008	Finland	Cohort s	Participants: 89.893, Non-participants: 68.862†	90%	50-59	15	n/a	7/9	Mortality rate -11.1% (-19.4-2.1)	
Anttinen A, 2006	Finland	Cohort	Participants: 552, Non-participants: 341†	71%	50-69	8.0-12.5	n/a	7/9	HR = 0.82 (0.59-1.12)	
Njor SH, 2015	Denmark	Cohort	Participants: 870.465 pyr, Non-participants: 828.508 pyr†	NA	50-69	14	yes	7/9		RR = 0.72 (0.59–0.87)
Anttila A, 2002	Finland	Cohort	Participants: 161.400 Wy, Non-participants: 155.400 Wy†	81.8%	50-59	8.5-10.5	n/a	6/9	RR = 0.81 (0.62-1.05)	
Duffy SW, 2006	Sweden	Cohort	Participants: 566.423, Non-participants: 542.187†	75%	40-69	>20	yes	6/9		RR = 0.57 (0.53-0.62)
Hakama M, 1995	Finland	Cohort	Participants: 3.708, Non-participants: 6.223†	86%	40-47	9	no	6/9		RR = 0.11 (0.00-0.71)
Hakama M, 1997	Finland	Cohort	Participants: 89.893, Non-participants: 68.862†	85%	45-69	6	n/a	6/9	RR = 0.76 (0.53-1.09)	
Jonsson H, 2001	Sweden	Cohort	Participants: 162.986 Non-participants: 98.608	NA	50-69	10.6	n/a	6/9	RR = 0.84 (0.67-1.05)	
Jonsson H, 2003	Sweden	Cohort	Participants: 43.749 Non-participants: 618.342	NA	40-64	22	n/a	6/9	RR = 0.84 (0.67-1.05)	
Jonsson H, 2003	Sweden	Cohort	Participants: 83.830, Non-participants: 41.608	NA	70-74	10	n/a	6/9	RR = 0.82 (0.57-1.19)	

Table 2a. Continued

Kalager M, 2010	Norway	Cohort	Participants: 2.337.323 pyr, Non-participants: 1.866.741 pyr†	NA	50-69	8	n/a	6/9	RR = 0.90 (0.76-1.04)[b]	
Parvinen I, 2006	Finland	Cohort	Participants: 963.362 pyr, Non-participants: 1.016.664 pyr†	NA	55-69	15	n/a	6/9	Helsinki: RR = 1.11 (0.95-1.29), Tampere: RR = 0.86 (0.65-1.12), Turku: RR = 0.64 (0.47-0.88)	
Sarkeala T, 2008	Finland	Cohort	Participants: 1.439.753 pyr, Non-participants: 34.803.524 pyr	NA	50-69	10	yes	6/9		RR = 0.62 (0.43-0.85)
Tabar L, 2001	Sweden	Cohort	Participants: 1.100.931 pyr, Non-participants: 1.213.136 pyr†	85%	40-69	30	yes	6/9	RR = 0.50 (0.41-0.60)	RR = 0.52 (0.43-0.63)
Gabe R, 2007	Iceland	Case-control	Cases: 226, Controls†: 902	61-68%	40-69	N/A	yes	5/9		OR = 0.65 (0.39-1.09)
Hellquist BN, 2011	Sweden	Cohort	Participants: 7.261.415, Non-participants: 8.843.852	80-90%	40-49	16	no	5/9	RR = 0.74 (0.66-0.83)	RR = 0.71 (0.62-0.80)
Jonsson H, 2000	Sweden	Cohort	Participants: 202.152 Non-participants: 237.279	NA	40-49	8	n/a	5/9	RR = 0.91 (0.72-1.15)	
Jonsson H, 2007	Sweden	Cohort	Participants: 109.000, Non-participants: 77.000	NA	40-74	11	yes	5/9	RR = 0.74 0.62-0.88)	
Parvinen I, 2015	Finland	Cohort	Participants: 1.439.753 pyr, Non-participants: 34.803.524 pyr	86.7	40-84	>10	n/a	5/9	TKU vs. RoF: RR = 0.85 (0.66-1.1) TKU vs. HEL: RR = 0.75 (0.57-1.00)	

BC = Breast cancer, pyr = person years, Wy = Women years, OR = Odds ratio; RR = Relative risk; HR = Hazard ratio; I: Intervention group; C: Control group; NA: not available; n/a: not applicable; TKU: Turku, RoF: Rest of Finland, HEL: Helsinki, Target age: Ages targeted by the screening program; Follow-up: Follow-up after initiation of the screening program. [a] Quality assessment made according to the Newcastle-Ottawa scale and the Cochrane risk of bias instrument. Risk of bias for RCT was categorized as follow: A (Low risk), B (Moderate risk) and C (High risk).; [†] Controls were drawn from the same population as the intervention group; [b] this value was recomputed as RR from the results provided in the original article

Table 2b. Characteristics, risk of bias and results on breast cancer mortality of included studies, by quality score. Southern Europe

Study	Country	Study type	Participants	Attendance	Target age (years)	Follow-up (years)	Correction for self-selection bias	Quality score[a]	Effect sizes for breast cancer mortality, RR/HR/OR (95%CI)	
									Invited vs not invited	Attenders vs not attenders
Palli D, 1986	Italy	Case-control	Cases: 57, Controls: 257†	NA	40-70		yes	8/9		BC Mortality: OR = 0.57 (95%CI: 0.35-0.89)
Puliti D, 2012	Italy	Cohort study	Participants: 32.544, Non-participants: 18.552†	56%	50-69	16.5	yes	8/9		50-59: RR = 0.55 (0.41-0.75), 60-69: RR = 0.49 (0.38-0.64)
Barco I, 2015	Spain	Cohort study	Participants: 496, Non-participants: 1.325†	NA	50-69	6	no	7/9		HR = 0.33 (0.18-0.63)
Puliti D, 2008	Italy	Case-control	Cases: 2.371 (Exp: 297), Controls: 9.484 (Exp: 1.718)†	n/a	50-74	n/a	yes	7/9	OR = 0.75 (0.62-0.92)	OR = 0.50 (0.42-0.60)
Paci E, 2002	Italy	Cohort study	Participants: 254.890 pyr, Non-participants: 164.742 pyr†	NA	50-69	8	n/a	6/9	RR = 0.75 (0.54-1.04)	
Palli D, 1989	Italy	Case-control	Cases: 103 (Exp: 55), Controls: 515(Exp: 355) †	n/a	40-49, 50+	n/a		6/9		40-49: OR = 0.63 (0.24-1.64), 50+: OR = 0.51 (0.29-0.89)
Ascunce EN, 2007	Spain	Cohort study	Participants: 185, Non-participants: 123†	85%	50-69	14	n/a	5/9	RR = 0.65 (0.51-0.82)	
Paci E, 2005	Italy	Cohort study	Participants: 2.105, Non-participants: 2.339†	NA	50-69	5	n/a	5/9	BC Mortality: RR = 0.73 (95%CI: 0.61-0.87)	

BC = Breast cancer, pyr = person years, Wy = Women years, OR = Odds ratio; RR = Relative risk; HR = Hazard ratio; I: Intervention group; C: Control group; NA: not available; n/a: not applicable; Target age: Ages targeted by the screening program; Follow-up: Follow-up after initiation of the screening program. [a] Quality assessment made according to the Newcastle-Ottawa scale and the Cochrane risk of bias instrument; [†] Controls were drawn from the same population as the intervention group

Table 2c. Characteristics, risk of bias and results on breast cancer mortality of included studies, by quality score. Western Europe

Study	Country	Study type	Participants	Attendance	Target age (years)	Follow-up (years)	Correction for self-selection bias	Quality score[a]	Effect sizes for breast cancer mortality, RR/HR/OR (95%CI)	
									Invited vs not invited	Attenders vs not attenders
Moss S, 2015	UK	RCT	I: 53.883 C: 106.953	81%	39-41	17	yes	A		RR = 088 (0.74-1.04)
Johns LE, 2017	UK	Cohort study	Participants: 2.407.709 pyr, Non-participants: 12.785.325 pyr†	74%	49-64	15	yes	9/9	RR = 0.79 (0.73-0.84)	RR = 0.68 (0.63-0.73)
Johns LE, 2017	UK	Case-control[b]	Cases: 11.754 (Exp: 5.109) Controls: 37.601 (Exp: 20.545) †	n/a	49-64		yes	9/9	OR = 0.79 (0.71-0.88)	OR = 0.53 (0.46-0.62)
Allgood PC, 2008	United Kingdom	Case-control	Cases: 284 (Exp: 208), Controls: 568 (Exp: 505) †	n/a	50-70		yes	8/9		OR = 0.65 (0.48-0.88)
Massat NJ, 2015	UK	Case-control	Cases: 391, Controls: 417†	61.7%	47-89		yes	8/9		OR = 0.69 (0.50-0.94)
Massat NJ, 2015	UK	Case-control	Cases: 869, Controls: 1.642†	70.5-62.8%	47-89		yes	8/9		OR = 0.61 (0.44-0.85)
Otto S, 2012	Netherlands	Case-control	Cases: 755, Controls: 3.739†	79%	50-75	n/a	yes	8/9		OR = 0.51 (0.40-0.66)
Paap E, 2014	Netherlands	Case-control	Cases: 1.233, Controls: 2.090	81.3%	50-75	n/a	yes	8/9		OR = 0.42 (0.33-0.53)
Alexander FE, 1999	United Kingdom	RCT	I: 28.628 C: 26.026	NA	45-65	14	no	C		RR = 0.79(0.60-1.02)
Broeders MJM, 2002	Netherlands	Case-control	Cases: 157 (Exp: 157), Controls: 758 (Exp: 758) †	n/a	40-80		no	7/9		40-49: OR = 0.90 (0.38-2.14), 50-59: OR = 0.71 (0.35-1.46), 60-69: OR = 0.80 (0.42-1.54)
Ernst M, 2004	Netherlands	Cohort study	Participants: 419, Non-participants: 250†	NA	50-69	8	n/a	6/9	HR = 0.75 (0.57-1.01)	

Fielder HM, 2004	UK	Case-control	Cases: 419 (Exp: 275), Controls: 717 (Exp: 535) †	n/a	50-75		yes	6/9		OR = 0.75 (0.49-1.14)
Mook S, 2011	Netherlands	Cohort study	Participants: 958, Non-participants: 1.634†	70-80%	50-69	10	no	6/9		HR = 0.62 (0.50-0.86)
van Dijk JAAM, 1996	Netherlands	Case-control	Cases: 82 (Exp: 15), Controls: 410 (Exp: 101) †	n/a	65+	n/a	no	6/9		RR = 0.56 (0.28-1.13)
Miltenburg GAJ, 1998	Netherlands	Case-control	Cases: 177 (Exp: 51), Controls: 531 (Exp: 64) †	n/a	50-69		no	5/9		OR = 0.54 (0.37-0.79)
Moss S, 1999	UK	Cohort study	Participants: 45.607, Non-participants: 190.496	65-70%	45-64	16	no	5/9		RR = 0.74 (0.63-0.86)
Sankatsing V, 2017	Netherlands	Cohort study	Participants: NA, Non-participants: NA†	80%	50-74	13-20	n/a	5/9	% rate change 2010 compared with 1980: -30	
Otto SJ, 2003	Netherlands	Cohort study	Participants: 8.414, Non-participants: 14.971†	NA	55-74	>10	n/a	4/9	% rate change 2001 compared with 1986-88: -19.9 (-26.6 to -14.2)	
Peer PM, 1995	Netherlands	Cohort study	Participants: 166.307 Wy, Non-participants: 154.103 Wy	87%	35-49	16	n/a	4/9	RR = 0.94 (0.68-1.29)	
van Dijk JAAM, 1997	Netherlands	Cohort study	Participants: 16.383Wy, Non-participants: 17.487 Wy	46%	65+	8-13	n/a	4/9	RR = 0.53 (0.27-1.04)	
van Schoor G, 2010	Netherlands	Case-control	Cases: 76 (Exp: 50), Controls: 750 (Exp: 596) †	n/a	40-49	n/a	no	4/9		OR = 0.50 (0.30-0.82)
van Schoor G, 2011	Netherlands	Case-control	Cases: 282, Controls: 1.410†	NA	50-69		no	4/9		OR = 0.65 (0.49-0.87)

BC = Breast cancer, pyr = person years, Wy = Women years, OR = Odds ratio; RR = Relative risk; HR = Hazard ratio; I: Intervention group; C: Control group; NA: not available; n/a: not applicable; Target age: Ages targeted by the screening program; Follow-up: Follow-up after initiation of the screening program. [a] Quality assessment made according to the Newcastle-Ottawa scale and the Cochrane risk of bias instrument. Risk of bias for RCT was categorized as follow: A (Low risk), B (Moderate risk) and C (High risk).; [b] Nested case-control study; [†] Controls were drawn from the same population as the intervention group

Considering the results of all 60 studies included in this review, BC mortality reduction estimates for invited vs. non-invited women varied from 4%[89] to 36%[90] in Northern Europe, from 25%[59 82] to 35%[56] in Southern Europe and from 6%[66] to 47%[67] in Western Europe. When comparing BC mortality of screened vs. non-screened women, estimates varied from 2%[50] to 89%[38] in Northern Europe, from 43%[86]to 67%[57] in Southern Europe and 12%[64] to 58%[80] in Western Europe. Of the 60 included studies, 40 had statistically significant results.

European Regions

The quality of the included studies was miscellaneous. Among the 60 included reported results, 5% (3/60) fell into quality category I, 27% (17/60) were graded as quality category II, 60% (37/60) fell into category III and 8% (5/60) into category IV. Due to the numerousness of included studies, we will only highlight the results from those 19 studies from group I and II. The estimated effect of organized mammographic screening on BC specific mortality from these studies, by European regions, is described in Figure 2. The entirety of the risk of bias assessment of all included cohort studies, case-control studies, and RCTs is displayed in Appendix 4-6, respectively. Additionally, Appendix 7 is a summary of the risk of bias assessment for the RCTs used in this review.

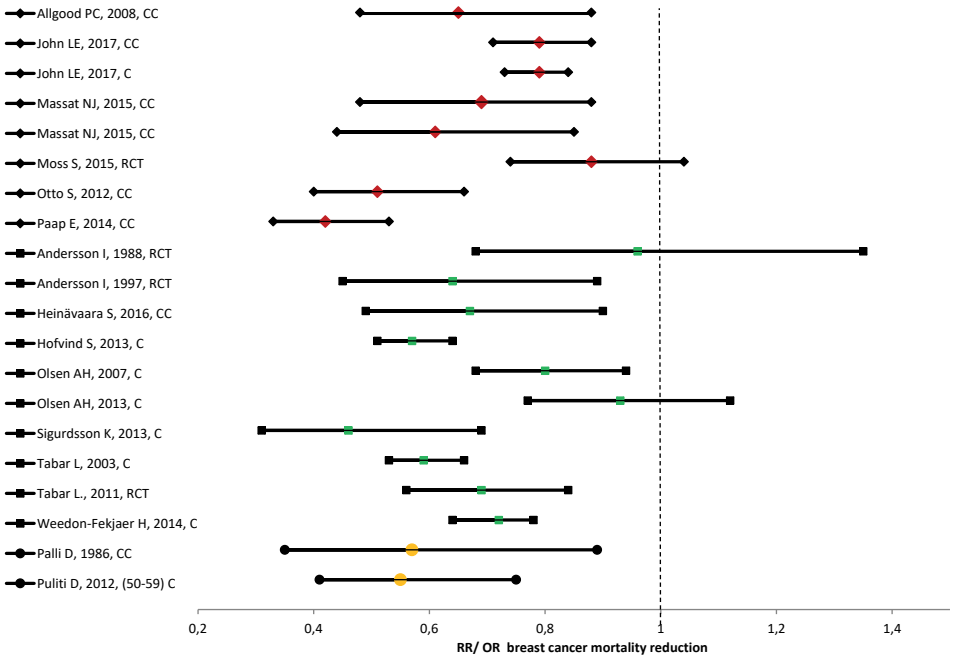


Figure 2. Forest plot displaying the effectiveness of organized mammographic screening on breast cancer specific mortality, of studies with (very) good quality (group I and II), by European regions (Western Europe: diamond, Northern Europe: square, Southern Europe: circle). The 95% confidence intervals for individual studies are represented by a horizontal line.

North

A total of 30 studies were selected and reported for Northern European countries, including Denmark, Finland, Sweden, Norway and Iceland (Table 2a).

Five of these references were randomized controlled trials, all from Sweden, of which two were of low risk of bias and one of moderate risk of bias. Two of the (very) good studies establish the protective effect of being invited to mammographic screening. Tabar (2011)[94] reports the long-term effect of mammographic screening in the Swedish Two-County trial. They found a highly significant reduction in BC mortality in women invited to mammographic screening (RR = 0.69 [95% CI: 0.56-0.84]), whereas the Malmö mammographic screening trial[89] initially did not find a statistically significant effect of screening (RR = 0.96 [95% CI: 0.68-1.35]) after a little less than 9 years of follow-up. Andersson[90] provided additional follow-up data of the two cohorts from the Malmö Mammographic Screening Trial, particularly for women between 45 and 50 years of age who were followed for an average of 10 years. They conclude that being invited to screening lowers the BC mortality significantly (RR = 0.64 [95% CI: 0.45-0.89]).

The reviewers appraised five of the 30 cohort studies from northern Europe to be of good quality. Three come from Norway. The Norwegian BC screening program was initiated in 1996, when it began as a pilot study in four of the 19 Norwegian counties. The program targets women ages 50 to 69 who are invited every two years. Olsen (2013)[48] observed the change in BC mortality due to screening comparing it to historical control groups in the four pilot study countries, using an incidence-based approach. The cohort study has a short follow-up of only 6 years and a reported RR of 0.93 (0.77-1.12). Two other reports included 15 years of follow-up. Weedon-Fekjaer's prospective cohort study estimated that invitation to mammographic screening was associated with a 28% reduced risk of death from BC compared with not being invited (RR = 0.72 [0.64-0.79])[55]. Hofvind (2013)[40] compared BC mortality of women attending screening with that of a non-screened cohort, considering incidence based mortality (IBM). Fifteen years after the start of the program, the reduction was estimated to be 43% (RR = 0.57 [0.51-0.64]). For Copenhagen (Denmark), Olsen (2007)[49] analyzed IBM of women invited to the routine mammography by linking screening registry, cancer registry, cause of death registry, and population registry data for individual women age 50-69. Using historical comparison groups, the effect of invitation to mammography screening every two years was as BC mortality reduction of 20% (RR = 0.80 [0.68–0.94]). Tabar et al (2003)[54] assessed the long-term effects (20 years) of mammographic service screening on BC mortality in two Swedish counties for women aged 40-69 years. Taking potential biases (e.g. age and self-selection bias) into account, BC mortality of screened women was 41% lower than that of unscreened women (RR = 0.59 [0.53-0.66]).

Heinävaara (2016)[75] evaluated the long-term effect of organized mammography screening on IBM in Finland in 1992–2011 among 50–84-year-old women using a case–control design. The effect of screening, corrected for self-selection bias, was 33% (HR = 0.67 [0.49–0.90]).

South

The characteristics of the nine included articles from Southern European countries are reported in Table 2b. All selected studies were performed in Italy and Spain. One case-control study and one cohort study, both from Florence/Italy, were judged to be of good quality (category II). In a rural area near Florence a population-based screening program for BC was started in 1970. The case-control study by Palli (1986)[86] showed that women who have been screened at least once had a BC mortality reduction of 43% compared to women never screened (OR = 0.57 [0.35-0.89]). Puliti (2012)[61] followed up women invited to the Florentine screening program every two years at age 50-69. Using an incidence-based approach, the estimated mortality reduction was 45% among 50 to 59 year-old women (RR = 0.55 [0.41-0.75]) and 51% among 60 to 69 year-old women (RR = 0.49 [0.38-0.64]) after 16 years of follow-up.

West

From Western European countries, the reviewers included 22 studies which exclusively came from From Western European countries, the reviewers included 22 studies which exclusively came from the Netherlands and United Kingdom (Table 2c).

The UK Age Trial (Moss, 2015)[93] was the only RCT from this region that was judged to carry a low risk of bias. However, it only refers to the specific group of women aged 40-49 after 17 years of follow-up. Annual mammography screening below age 50 leads to a rate ratio (RR) for BC mortality of 0.88 (0.74-1.04).

Six case-control studies reached a score of 8 or 9 (of 9), all with fairly similar results: Allgood (2008)[71] performed a study in the East Anglia region after the initiation of the breast screening program in 1989. The odds ratio (OR) for death from BC in women who attend at least one routine screen compared to those who did not attend was 0.65 (0.48-0.88). Massat (2016)[76] assessed the impact of the NHS BC Screening program 20 years after the inception and showed a BC mortality reduction of 39% among attenders (OR = 0.61 [0.44-0.85]) In a companion case-control study, Massat (2016)[76] [77]reported that breast screening attendance reduces the fatality risk by 31% (OR= 0.69 [0.50-0.94]). A 47% BC mortality reduction for attending women was found in a nested case-control study by Johns (2017)[87] (OR = 0.53 [0.46-0.62]), who evaluated the effectiveness of the NHS breast screening program in England and Wales. All of the British observational study results were corrected for self-selection bias. For the Netherlands, Paap (2014)[80] estimated the benefit of the population-based screening program to be as high as 58% (OR = 0.42 [0.33-0.53]) for screened compared to unscreened women. Otto's (2012)[79] assessment of the effectiveness of mammography screening of Dutch women indicated a significant association between attending mammography screening and risk of breast cancer death (OR = 0.51 [0.40–0.66]). Johns (2017)[70] conducted the first individual-based cohort evaluation of population breast screening in the UK, to estimate the impact of the NHS breast screening program (NHSBSP)

on BC mortality. After adjustment for self-selection bias, the mortality reduction was 32% (RR = 0.68 [0.63–0.73]).

East

No studies from Eastern Europe met the inclusion criteria.

DISCUSSION

To the best of the authors' knowledge, this review is the first that comprises evidence from RCTs as well as observational studies and stringently uses transparent grading tools to appraise the quality of each included reference and then highlights only those which provide the most valid information. The results fortify that mammography screening leads to reduced mortality from BC and the evaluation studies conducted in the three European regions where screening was implemented are confirming this conclusion.

The large number of possibly eligible studies for this review as well as the number of other (systematic) reviews on this topic reflect the long history of evaluations regarding the benefits of mammographic screening, including some contrasting views.

In 2012, the Independent UK Panel on BC screening relied mainly on findings from RCTs in order to provide estimates of the level of benefits and harms. Based on 11 trials with 13 years of follow-up they concluded that the relative risk reduction was 20% in women invited for screening[19]. Gøtzsche and Jørgensen[21], who included only RCTs in their review, found that the trials with adequate randomization did not find a statistically significant effect of screening on BC mortality. Nevertheless, in the past decade concerns have been raised about the applicability of RCTs in times of growing availability of service screening and about the validity of these trials. More recently, the evaluation of screening benefits has shifted to population-based screening services, and observational studies became the main contributors of new information on the impact of BC screening on BC mortality reduction[8]. Prerequisites for methodologically sound results therefore are individual data on screening exposure that is sufficiently long (>5 years), reliable information on the vital status as well as cancer data which can be directly linked to a women's screening history and to her cause of death. The susceptibility to bias can furthermore be limited when studies use incidence-based mortality (IBM) and adjust for self-selection bias. By using standardized tools to judge on the presence of all of those methodological components, we were able to identify those observational studies that are qualitatively consistent with well conducted RCTs. Of the 38 cohort studies included in this systematic review, 24 considered IBM and therewith only observed BC deaths in women diagnosed after their first invitation to (or attendance in) mammographic screening. In that way, these studies only account for

a risk of BC death at a time, when it could have been affected by service screening. BC mortality reductions were consistently greater when the analysis compared screened vs. unscreened women rather than women who were invited vs not invited to screening. All of the 17 case-control studies included in this review compared women attending in screening to non-attending ones. The attractiveness of the case–control approach is that it uses observed mortality and it requires fewer participants than cohort studies. Thus it is a very efficient tool to evaluate (new) organized screening programs[6 95]. However, non-compliers, those women who did not accept the invitation to screening within organized programs, can potentially have a different risk of death from BC than the general population. Therefore, one major disadvantage of this study design is the tendency to selection bias. Duffy et al.[9] provided a method of adjustment for potential confounders. The majority of the included case-control studies adjusted for self-selection bias.

While most researchers agree that the combination of both screening and treatment leads to a reduction of BC mortality, some claim that the reduction of BC mortality observed in Europe since the 1990s is mostly due to changes in cancer treatment[96]. Changes in treatment over time – in Norway, for example, multidisciplinary breast care centers have been introduced parallel with the organized screening program[40 46 97] - make the results difficult to interpret. Both, case-control or IBM studies implicitly imply a treatment effect though. In order to disentangle the synergistic effect of screening with better treatment modeling analysis under different assumptions are needed. In their simulation modeling study, Plevritis et al (2018)[98] evaluated the contributions associated with screening and treatment to BC mortality reductions for US women. The estimated reduction in BC mortality rate between 2000 and 2012 was 49%, of which 37% were associated with screening and 63% with treatment, although the associations varied by BC molecular subtype.

We discovered a lack of eligible studies from Eastern Europe on mortality reduction due to screening. One main explanation could be serious (financial) barriers to organizing and/or evaluating screening services[17]. Among the regions included in this study, some populations had long-established screening programs running since the end of the 1980s (e.g. Finland) and since the beginning of the 1990s (the Netherlands, Norway, Tuscany and Turin (Italy)) with complete coverage of populations at screening age, but potentially different age groups covered across these areas. Particularly for eastern European countries, opportunistic screening has been offered to women since the early 1990s[99] and still plays an important role in explaining low participation rates in the organized programs[17]. In most eastern European countries breast screening programs started more recently: Hungary in 2001, Estonia in 2005, Lithuania in 2006, Cracow (Poland) in 2007 and Slovenia in 2008. Hence a long running monitoring and evaluation system is either still missing or does not provide sound results yet.

This qualitative review is based on well-defined a priori criteria and a rigorous systematic methodology. Nonetheless, we note four potential limitations. First, non-English-language studies were excluded. Second, the large number of included studies and their methodological designs led to a wide range of estimates of mortality reduction due to screening. Therefore we did not aim for a meta-analysis and to synthesize the results, but rather to highlight the reported evidence. Three, we used very strict PICOS criteria during the selection process following an in-depth quality assessment through the Cochrane and NOS tools to limit the risk of bias. While these choices may limit the number of references that will be included in this systematic review, it guarantees the best available evidence on which we base the conclusions. Lastly, this review did not include grey literature and thus solely relies on published studies. Therefore it might be affected by publication bias, as published literature appears to be predominantly biased towards positive results[100].

The variation in the point estimates from individual studies indicates differences in evaluation designs, e.g., in ages of follow-up of breast cancer incidence or mortality, duration of follow-up since first invitation, comparison group and assessment methods of self-selection bias, rather than variability of the effectiveness of screening. It would have been very important to describe the patterns in more detail according to the above factors, but it was often not possible yet, due to lack of information provided in many studies. Recent studies suggest that the impact can be highly variable, depending e.g. if breast cancers during screening age only, or also after the last invitation round would be included[75 101]. It would be very important to assess the screening impacts after the whole life span since the first invitation, and describe the variable effects in the various follow-up windows of relevance.

We prove that there are several methodologically appropriate approaches that are able to capture the true beneficial effect of mammographic screening. However, in order to assess the validity of these results, future reviews would strongly profit from quality appraisal tools which are specifically developed to judge the impact of screening, as well as the quality of European record linkage practice.

Acknowledgments

We like to thank Wichor Bramer, biomedical information specialist at the medical library at Erasmus MC (Rotterdam, the Netherlands) for his valuable help, expert inputs and solutions in the making of this systematic review.

Funding

This systematic review is part of the EU-TOPIA project, funded by the EU-Framework Programme (Horizon 2020) of the European Commission, project reference 634753. The authors alone are responsible for the views expressed in this manuscript.

REFERENCES

1. IARC Working Group, IARC Handbooks of Cancer Prevention, Volume 15: Breast Cancer Screening. Lyon, 2016.
2. Bray, F., et al., Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 2018. 68(6): p. 394-424.
3. Lauby-Secretan, B., et al., Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med, 2015. 372(24): p. 2353-8.
4. Shapiro, S., et al., Ten- to fourteen-year effect of screening on breast cancer mortality. J Natl Cancer Inst, 1982. 69(2): p. 349-55.
5. de Koning, H.J., Mammographic screening: evidence from randomised controlled trials. Ann Oncol, 2003. 14(8): p. 1185-9.
6. Autier, P. and M. Boniol, Breast cancer screening: evidence of benefit depends on the method used. BMC Med, 2012. 10: p. 163.
7. Autier, P., et al., Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. BMJ, 2011. 343: p. d4411.
8. Broeders, M., 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: p. 14-25.
9. Duffy, S.W., et al., Correcting for non-compliance bias in case-control studies to evaluate cancer screening programmes. Journal of the Royal Statistical Society Series C-Applied Statistics, 2002. 51: p. 235-243.
10. Loberg, M., et al., Benefits and harms of mammography screening. Breast Cancer Res, 2015. 17: p. 63.
11. Moss, S.M., et al., The impact of mammographic screening on breast cancer mortality in Europe: a review of trend studies. J Med Screen, 2012. 19 Suppl 1: p. 26-32.
12. Njor, S., et al., Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. J Med Screen, 2012. 19 Suppl 1: p. 33-41.
13. Basu, P., et al., Status of implementation and organization of cancer screening in The European Union Member States-Summary results from the second European screening report. Int J Cancer, 2018. 142(1): p. 44-56.
14. Commission, E., European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth Edition. 2006.
15. European Commission Initiative on Breast Cancer, Recommendations from the European Breast Guidlines. 2021. Available at: https://ecibc.jrc.ec.europa.eu/recommendations/.
16. Giordano, L., et al., Mammographic screening programmes in Europe: organization, coverage and participation. J Med Screen, 2012. 19 Suppl 1: p. 72-82.

17. IARC, Cancer Screening in the European Union. Report on the implementation of the Council Recommendation on cancer screening. IARC, Lyon, 2017. Available at: 2017_cancerscreening_2ndreportimplementation_en.pdf.
18. Gøtzsche, P.C., Relation between breast cancer mortality and screening effectiveness: systematic review of the mammography trials. Dan Med Bull, 2011. 58(3): p. A4246.
19. Marmot, M.G., et al., The benefits and harms of breast cancer screening: an independent review. Br J Cancer, 2013. 108(11): p. 2205-40.
20. Nelson, H.D., et al., Effectiveness of Breast Cancer Screening: Systematic Review and Meta-analysis to Update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med, 2016. 164(4): p. 244-55.
21. Gøtzsche, P.C. and K.J. Jørgensen, Screening for breast cancer with mammography. Cochrane Database Syst Rev, 2013(6): p. CD001877.
22. Irvin, V.L. and R.M. Kaplan, Screening mammography & breast cancer mortality: meta-analysis of quasi-experimental studies. PLoS One, 2014. 9(6): p. e98105.
23. Schopper, D. and C. de Wolf, How effective are breast cancer screening programmes by mammography? Review of the current evidence. Eur J Cancer, 2009. 45(11): p. 1916-23.
24. Zielonke, N., et al., Evidence for cancer specific and all-cause mortality reduction due to screening for breast, cervical and colorectal cancer: a systematic review. PROSPERO 2016: CRD42016042433 2016.
25. Moher, D., et al., Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev, 2015. 4: p. 1.
26. Elmunzer, B.J., et al., Comparing the effectiveness of competing tests for reducing colorectal cancer mortality: a network meta-analysis. Gastrointest Endosc, 2015. 81(3): p. 700-709 e3.
27. Multilingual Thesaurus of the European Union (EUROVOC). Available at: http://eurovoc.europa.eu/drupal/?q=request&uri=http://eurovoc.europa.eu/100277 .
28. Higgins JP, G.S., Cochrane Handbook for Systematic Review of Interventions v5.1.0. Available at: http://training.cochrane.org/handbook. Accessed April, 2016.
29. Wells GA, Shea B, O'Conell D, et al. Ottawa Hospital Research Institute. The Newcastle-Ottawa scale (NOS) for assessing the quality of nonrandomized studies in meta-analysis. Available at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed April, 2016.
30. Anttila, A., et al., Towards better implementation of cancer screening in Europe through improved monitoring and evaluation and greater engagement of cancer registries. Eur J Cancer, 2015. 51(2): p. 241-51.
31. Forman D, B.F., Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M, Steliarova-Foucher E, Swaminathan R Ferlay J, Cancer Incidence in Five Continents, Vol. X. IARC Scientific Publication No. 164. Lyon: International Agency for Research on Cancer., 2014.
32. Bray, F., et al., Cancer Incidence in Five Continents, Vol. XI (electronic version). Lyon: International Agency for Research on Cancer. Available from: http://ci5.iarc.fr, 2017.

33. Anttila, A., J. Koskela, and M. Hakama, Programme sensitivity and effectiveness of mammography service screening in Helsinki, Finland. J Med Screen, 2002. 9(4): p. 153-158.
34. Anttila, A., et al., Impacts of the Finnish service screening programme on breast cancer rates. BMC Public Health, 2008. 8.
35. Anttinen, J., H. Kautiainen, and T. Kuopio, Role of mammography screening as a predictor of survival in postmenopausal breast cancer patients. Br J Cancer, 2006. 94(1): p. 147-151.
36. Duffy, S.W., Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. Cancer Epidemiol Biomarkers Prev, 2006. 15(1): p. 45-51.
37. Hakama, M., et al., Effectiveness of the public health policy for breast cancer screening in Finland: Population based cohort study. BR MED J, 1997. 314(7084): p. 864-867.
38. Hakama, M., et al., Effectiveness of screening for breast cancer in women under 50 years at entry: The Kotka pilot project in Finland. Int J Cancer, 1995. 63(1): p. 55-57.
39. Hellquist, B.N., et al., Effectiveness of population-based service screening with mammography for women ages 40 to 49 years. Cancer, 2011. 117(4): p. 714-722.
40. Hofvind, S., et al., Breast cancer mortality in participants of the Norwegian Breast Cancer Screening Program. Cancer, 2013. 119(17): p. 3106-3112.
41. Jonsson, H., et al., Service screening with mammography in Northern Sweden: Effects on breast cancer mortality - An update. J Med Screen, 2007. 14(2): p. 87-93.
42. Jonsson, H., et al., Service screening with mammography of women aged 50-69 years in sweden: Effects on mortality from breast cancer. J Med Screen, 2001. 8(3): p. 152-160.
43. Jonsson, H., et al., Service screening with mammography. Long-term effects on breast cancer mortality in the county of Gävleborg, Sweden. Breast, 2003. 12(3): p. 183-193.
44. Jonsson, H., et al., Service screening with mammography in Sweden: Evaluation of effects of screening on breast cancer mortality in age group 40-49 years. Acta Oncol, 2000. 39(5): p. 617-623.
45. Jonsson, H., et al., Service screening with mammography of women aged 70-74 years in Sweden: Effects on breast cancer mortality. Cancer Detect Prev, 2003. 27(5): p. 360-369.
46. Kalager, M., et al., Effect of screening mammography on breast-cancer mortality in Norway. New Engl J Med, 2010. 363(13): p. 1203-1210.
47. Njor, S.H., et al., Decline in breast cancer mortality: How much is attributable to screening? J. Med. Screen., 2015. 22(1): p. 20-27.
48. Olsen, A.H., et al., Breast cancer mortality in Norway after the introduction of mammography screening. Int J Cancer, 2013. 132(1): p. 208-214.
49. Olsen, A.H., S.H. Njor, and E. Lynge, Estimating the benefits of mammography screening: The impact of study design. Epidemiology, 2007. 18(4): p. 487-492.
50. Parvinen, I., et al., Mammography screening in three Finnish residential areas: Comprehensive population-based study of breast cancer incidence and incidence-based mortality 1976-2009. Br J Cancer, 2015. 112(5): p. 918-924.
51. Parvinen, I., et al., Service screening mammography reduces breast cancer mortality among elderly women in Turku. J Med Screen, 2006. 13(1): p. 34-40.

52. Sarkeala, T., S. Heinävaara, and A. Anttila, Breast cancer mortality with varying invitational policies in organised mammography. Br J Cancer, 2008. 98(3): p. 641-645.
53. Sigurdsson, K. and E.J. Ólafsdóttir, Population-based service mammography screening: The Icelandic experience. Breast Cancer: Target Ther, 2013. 5: p. 17-25.
54. Tabar, L., et al., Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. Lancet, 2003. 361(9367): p. 1405-1410.
55. Weedon-Fekjræ, H., P.R. Romundstad, and L.J. Vatten, Modern mammography screening and breast cancer mortality: Population study. BMJ (Online), 2014. 348.
56. Ascunce, E.N., et al., Changes in breast cancer mortality in Navarre (Spain) after introduction of a screening programme. J Med Screen, 2007. 14(1): p. 14-20.
57. Barco, I., et al., Comparison of Screened and Nonscreened Breast Cancer Patients in Relation to Age: A 2-Institution Study. Clin Breast Cancer, 2015.
58. García Fernández, A., et al., Mortality and recurrence patterns of breast cancer patients diagnosed under a screening programme versus comparable non-screened breast cancer patients from the same population: Analytical survey from 2002 to 2012. Tumor Biol, 2014. 35(3): p. 1945-1953.
59. Paci, E., et al., Quantification of the effect of mammographic screening on fatal breast cancers: The Florence Programme 1990-96. Br J Cancer, 2002. 87(1): p. 65-69.
60. Paci, E., et al., Early diagnosis, not differential treatment, explains better survival in service screening. Eur J Cancer, 2005. 41(17): p. 2728-2734.
61. Puliti, D., et al., Balancing harms and benefits of service mammography screening programs: A cohort study. Breast Cancer Res, 2012. 14(1).
62. Ernst, M.F., et al., Breast Carcinoma Diagnosis, Treatment, and Prognosis before and after the Introduction of Mass Mammographic Screening. Cancer, 2004. 100(7): p. 1337-1344.
63. Mook, S., et al., Independent prognostic value of screen detection in invasive breast cancer. J Natl Cancer Inst, 2011. 103(7): p. 585-597.
64. Moss, S.M., 16-Year mortality from breast cancer in the UK trial of early detection of breast cancer. Lancet, 1999. 353(9168): p. 1909-1914.
65. Otto, S.J., et al., Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: A systematic review. Lancet, 2003. 361(9367): p. 1411-1417.
66. Peer, P.G.M., et al., Effect on breast cancer mortality of biennial mammographic screening of women under age 50. Int J Cancer, 1995. 60(6): p. 808-811.
67. Van Dijck, J.A.A.M., et al., Breast-cancer mortality in a non-randomized trial on mammographic screening in women over age 65. Int J Cancer, 1997. 70(2): p. 164-168.
68. Tabár, L., et al., Beyond randomized controlled trials: Organized mammographic screening substantially reduces breast carcinoma mortality. Cancer, 2001. 91(9): p. 1724-1731.
69. Sankatsing, V.D.V., et al., The effect of population-based mammography screening in Dutch municipalities on breast cancer mortality: 20 years of follow-up. Int J Cancer, 2017. 141(4): p. 671-677.
70. Johns, L.E., et al., Effect of population breast screening on breast cancer mortality up to 2005 in England and Wales: An individual-level cohort study. Br J Cancer, 2017. 116(2): p. 246-252.

71. Allgood, P.C., et al., A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality. Br J Cancer, 2008. 98(1): p. 206-209.
72. Broeders, M.J.M., et al., Repeated mammographic screening reduces breast cancer mortality along the continuum of age. J Med Screen, 2002. 9(4): p. 163-167.
73. Fielder, H.M., et al., A case-control study to estimate the impact on breast cancer death of the breast screening programme in Wales. J Med Screen, 2004. 11(4): p. 194-198.
74. Gabe, R., et al., A case-control study to estimate the impact of the Icelandic population-based mammography screening program on breast cancer death. Acta Radiol, 2007. 48(9): p. 948-955.
75. Heinavaara, S., T. Sarkeala, and A. Anttila, Impact of organised mammography screening on breast cancer mortality in a case-control and cohort study. Br J Cancer, 2016.
76. Massat, N.J., et al., Impact of screening on breast cancer mortality: The UK program 20 years on. Cancer Epidemiol Biomarkers Prev, 2016. 25(3): p. 455-462.
77. Massat, N.J., et al., Explaining the better prognosis of screening-exposed breast cancers: Influence of tumor characteristics and treatment. Cancer Epidemiol Biomarkers Prev, 2016. 25(3): p. 479-487.
78. Miltenberg, G.A.J., et al., Seventeen-year evaluation of breast cancer screening: The DOM project, The Netherlands. Br J Cancer, 1998. 78(7): p. 962-965.
79. Otto, S.J., et al., Mammography screening and risk of breast cancer death: A population-based case - Control study. Cancer Epidemiol Biomarkers Prev, 2012. 21(1): p. 66-73.
80. Paap, E., et al., Breast cancer screening halves the risk of breast cancer death: A case-referent study. Breast, 2014. 23(4): p. 439-444.
81. Palli, D., et al., Time interval since last test in a breast cancer screening programme: A case-control study in Italy. J EPIDEMIOL COMMUNITY HEALTH, 1989. 43(3): p. 241-248.
82. Puliti, D., et al., Effectiveness of service screening: A case-control study to assess breast cancer mortality reduction. Br J Cancer, 2008. 99(3): p. 423-427.
83. Van Dijck, J.A.A.M., et al., Mammographic screening after the age of 65 years: Evidence for a reduction in breast cancer mortality. Int J Cancer, 1996. 66(6): p. 727-731.
84. Van Schoor, G., et al., Effective biennial mammographic screening in women aged 40-49. Eur J Cancer, 2010. 46(18): p. 3137-3140.
85. Van Schoor, G., et al., Increasingly strong reduction in breast cancer mortality due to screening. Br J Cancer, 2011. 104(6): p. 910-914.
86. Palli, D., et al., A case-control study of the efficacy of a non-randomized breast cancer screening program in florence (Italy). Int J Cancer, 1986.
87. Johns, L.E., A.J. Swerdlow, and S.M. Moss, Effect of population breast screening on breast cancer mortality to 2005 in England and Wales: A nested case-control study within a cohort of one million women. J Med Screen, 2017: p. 969141317713232.
88. Alexander, F.E., et al., 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. Lancet, 1999. 353(9168): p. 1903-1908.
89. Andersson, I., et al., Mammographic screening and mortality from breast cancer: The Malmo mammographic screening trial. BR MED J, 1988. 297(6654): p. 943-948.

90. Andersson, I. and L. Janzon, Reduced breast cancer mortality in women under age 50: updated results from the Malmö Mammographic Screening Program. J Natl Cancer Inst Monographs, 1997(22): p. 63-67.
91. Bjurstam, N., et al., The Gothenburg breast screening trial. Cancer, 2003. 97(10): p. 2387-2396.
92. Frisell, J., et al., Followup after 11 years - Update of mortality results in the Stockholm mammographic screening trial. BREAST CANCER RES TREAT, 1997. 45(3): p. 263-270.
93. Moss, S.M., et al., 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): p. 1123-1132.
94. Tabár, L., et al., Swedish two-county trial: Impact of mammographic screening on breast cancer mortality during 3 decades. Radiology, 2011. 260(3): p. 658-663.
95. van der Waal, D., et al., Case-control Studies on the Effectiveness of Breast Cancer Screening: Insights from the UK Age Trial. Epidemiology, 2015. 26(4): p. 590-6.
96. Bleyer, A., C. Baines, and A.B. Miller, Impact of screening mammography on breast cancer mortality. Int J Cancer, 2016. 138(8): p. 2003-12.
97. Kalager, M., et al., Comparative analysis of breast cancer mortality following mammography screening in Denmark and Norway. Ann Oncol, 2014. 25(6): p. 1137-1143.
98. Plevritis, S.K., et al., Association of Screening and Treatment With Breast Cancer Mortality by Molecular Subtype in US Women, 2000-2012. JAMA, 2018. 319(2): p. 154-164.
99. Rosso, S., et al., Up-to-date estimates of breast cancer survival for the years 2000-2004 in 11 European countries: the role of screening and a comparison with data from the United States. Eur J Cancer, 2010. 46(18): p. 3351-7.
100. Joober, R., et al., Publication bias: what are the challenges and can they be overcome? J Psychiatry Neurosci, 2012. 37(3): p. 149-52.
101. Beau, A.B., et al., Limitations in the Effect of Screening on Breast Cancer Mortality. J Clin Oncol, 2018. 36(30): p. 2988-2994.

SUPPLEMENTARY MATERIALS

Collaborators

Piret Veerus, Ahti Anttila, Sirpa Heinävaara, Tytti Sarkeala, Marcell Csanadi, Janos Pitter, György Széles, Zoltan Voko, Silvia Minozzi, Nereo Segnan, Carlo Senore, Marjolein van Ballegooijen, Inge Driesprong - de Kok, Andrea Gini, Eveline Heijnsdijk, Erik Jansen, Harry de Koning, Iris Lansdorp – Vogelaar, Nicolien van Ravesteyn, Nadine Zielonke, Urska Ivanus, Katja Jarm, Dominika Novak Mlakar, Maja Primic-Žakelj, Martin McKee, Jennifer Priaulx

Appendix 1. Computer-assisted search code by reference databases.

Source	Selection code
Embase	('breast tumor'/exp OR mammography/exp OR (((breast OR mamma*) NEAR/10 (cancer* OR neoplas* OR tumo* OR carcino* OR adenocarcin*)) OR mammogra*):ab,ti) AND (screening/ exp OR (screen* OR ((annual* OR periodic*) NEAR/3 examination*))) AND (mortality/de OR 'cancer mortality'/de OR (mortalit* OR (death NEXT/1 rate*)):ab,ti) NOT ([Conference Abstract]/lim OR [Letter]/lim OR [Note]/lim OR [Editorial]/lim) AND [english]/lim AND (europe/exp OR (europe* OR Andorra* OR Austria* OR Balkan* OR Belgi* OR Albania* OR Baltic-State* OR Bosnia* OR Herzegovina* OR Bulgaria* OR Croatia* OR Czech* OR Hungar* OR Kosovo* OR Macedonia* OR Moldova* OR Montenegr* OR Poland* OR polish* OR Belarus* OR Romania* OR Russia* OR Serbia* OR Slovakia* OR Slovenia* OR Ukraine* OR France* OR french OR German* OR Gibraltar* OR Great-Brit* OR uk OR united-kingdom* OR England* OR Scotland* OR Wales* OR welsh OR Greece* OR Ireland* OR Italy OR Italian OR Liechtenstein* OR Luxembourg* OR Monaco* OR Netherlands* OR dutch OR holland OR Portug* OR San-Marino* OR Scandinavia* OR Nordic* OR Denmark* OR danish OR Finland* OR finnish OR Iceland* OR Norwa* OR norwegian OR Sweden* OR swedish OR Spain* OR spanish OR Switzerland* OR swiss):ab,ti,ca,ta,cy,ad) AND ('observational study'/exp OR 'cohort analysis'/exp OR 'longitudinal study'/exp OR 'retrospective study'/exp OR 'prospective study'/exp OR 'health survey'/de OR 'health care survey'/de OR 'epidemiological data'/de OR 'case control study'/de OR 'cross-sectional study'/de OR 'correlational study'/de OR 'population research'/de OR 'family study'/de OR 'major clinical study'/de OR 'multicenter study'/de OR 'comparative study'/de OR 'follow up'/de OR 'clinical study'/de OR 'clinical article'/de OR 'clinical trial'/exp OR 'randomization'/exp OR 'intervention study'/de OR 'open study'/de OR 'community trial'/ de OR 'review'/exp OR 'systematic review'/exp OR (((observation* OR epidemiolog* OR famil* OR comparativ* OR communit*) NEAR/6 (stud* OR data OR research)) OR cohort* OR longitudinal* OR retrospectiv* OR prospectiv* OR population* OR (national* NEAR/3 (stud* OR survey)) OR (health* NEAR/3 survey*) OR ((case OR cases OR match*) NEAR/3 control*) OR (cross NEXT/1 section*) OR correlation* OR multicenter* OR (multi* NEXT/1 center*) OR 'follow up' OR followup* OR clinical* OR trial OR random* OR review*):ab,ti)
Medline Ovid	(exp "Breast Neoplasms"/ OR exp Mammography/ OR (((breast OR mamma*) ADJ10 (cancer* OR neoplas* OR tumo* OR carcino* OR adenocarcin*)) OR mammogra*).ab,ti.) AND ("Mass Screening"/ OR exp "Early Diagnosis"/ OR (screen* OR ((annual* OR periodic*) ADJ3 examination*)) OR (early ADJ3 (diagnos* OR detect*))) AND (exp mortality/ OR (mortalit* OR (death ADJ rate*)). ab,ti.) NOT (letter OR news OR comment OR editorial OR congresses OR abstracts).pt. AND english. la. AND (exp europe/ OR (europe* OR Andorra* OR Austria* OR Balkan* OR Belgi* OR Albania* OR Baltic-State* OR Bosnia* OR Herzegovina* OR Bulgaria* OR Croatia* OR Czech* OR Hungar* OR Kosovo* OR Macedonia* OR Moldova* OR Montenegr* OR Poland* OR polish* OR Belarus* OR Romania* OR Russia* OR Serbia* OR Slovakia* OR Slovenia* OR Ukraine* OR France* OR french OR German* OR Gibraltar* OR Great-Brit* OR uk OR united-kingdom* OR England* OR Scotland* OR Wales* OR welsh OR Greece* OR Ireland* OR Italy OR Italian OR Liechtenstein* OR Luxembourg* OR Monaco* OR Netherlands* OR dutch OR holland OR Portug* OR San-Marino* OR Scandinavia* OR Nordic* OR Denmark* OR danish OR Finland* OR finnish OR Iceland* OR Norwa* OR norwegian OR Sweden* OR swedish OR Spain* OR spanish OR Switzerland* OR swiss). ab,ti,jn,cp,in.) AND ("observational study"/ OR exp "Cohort Studies"/ OR "Health Surveys"/ OR "Epidemiologic Studies"/ OR "Case-Control Studies"/ OR "Cross-Sectional Studies"/ OR "multicenter study"/ OR "comparative study"/ OR "clinical study"/ OR exp "clinical trials"/ OR "Random Allocation"/ OR "review"/ OR (((observation* OR epidemiolog*) ADJ6 (stud* OR data OR research)) OR cohort* OR longitudinal* OR retrospectiv* OR prospectiv* OR population* OR (national* ADJ3 (stud* OR survey)) OR (health* ADJ3 survey*) OR ((case OR cases OR match*) ADJ3 control*) OR (cross ADJ section*) OR correlation* OR multicenter* OR (multi* ADJ center*) OR "follow up" OR followup* OR clinical* OR trial OR random* OR review*).ab,ti.)

Cochrane	((((breast OR mamma*) NEAR/10 (cancer* OR neoplas* OR tumo* OR carcino* OR adenocarcin*)) OR mammogra*):ab,ti) AND ((screen* OR ((annual* OR periodic*) NEAR/3 examination*)) OR (early NEAR/3 (diagnos* OR detect*))) AND ((mortalit* OR (death NEXT/1 rate*)):ab,ti) AND ((europe* OR Andorra* OR Austria* OR Balkan* OR Belgi* OR Albania* OR Baltic-State* OR Bosnia* OR Herzegovina* OR Bulgaria* OR Croatia* OR Czech* OR Hungar* OR Kosovo* OR Macedonia* OR Moldova* OR Montenegr* OR Poland* OR polish* OR Belarus* OR Romania* OR Russia* OR Serbia* OR Slovakia* OR Slovenia* OR Ukraine* OR France* OR french OR German* OR Gibraltar* OR Great-Brit* OR uk OR united-kingdom* OR England* OR Scotland* OR Wales* OR welsh OR Greece* OR Ireland* OR Italy OR Italian OR Liechtenstein* OR Luxembourg* OR Monaco* OR Netherlands* OR dutch OR holland OR Portug* OR San-Marino* OR Scandinavia* OR Nordic* OR Denmark* OR danish OR Finland* OR finnish OR Iceland* OR Norwa* OR norwegian OR Sweden* OR swedish OR Spain* OR spanish OR Switzerland* OR swiss))
Web-of-science	TS=(((((breast OR mamma*) NEAR/10 (cancer* OR neoplas* OR tumo* OR carcino* OR adenocarcin*)) OR mammogra*)) AND ((screen* OR ((annual* OR periodic*) NEAR/2 examination*)) OR (early NEAR/2 (diagnos* OR detect*))) AND ((mortalit* OR (death NEAR/1 rate*))) AND ((europe* OR Andorra* OR Austria* OR Balkan* OR Belgi* OR Albania* OR Baltic-State* OR Bosnia* OR Herzegovina* OR Bulgaria* OR Croatia* OR Czech* OR Hungar* OR Kosovo* OR Macedonia* OR Moldova* OR Montenegr* OR Poland* OR polish* OR Belarus* OR Romania* OR Russia* OR Serbia* OR Slovakia* OR Slovenia* OR Ukraine* OR France* OR french OR German* OR Gibraltar* OR Great-Brit* OR uk OR united-kingdom* OR England* OR Scotland* OR Wales* OR welsh OR Greece* OR Ireland* OR Italy OR Italian OR Liechtenstein* OR Luxembourg* OR Monaco* OR Netherlands* OR dutch OR holland OR Portug* OR San-Marino* OR Scandinavia* OR Nordic* OR Denmark* OR danish OR Finland* OR finnish OR Iceland* OR Norwa* OR norwegian OR Sweden* OR swedish OR Spain* OR spanish OR Switzerland* OR swiss)) AND (((observation* OR epidemiolog* OR famil* OR comparativ* OR communit*) NEAR/5 (stud* OR data OR research)) OR cohort* OR longitudinal* OR retrospectiv* OR prospectiv* OR population* OR (national* NEAR/2 (stud* OR survey)) OR (health* NEAR/2 survey*) OR ((case OR cases OR match*) NEAR/2 control*) OR (cross NEAR/1 section*) OR correlation* OR multicenter* OR (multi* NEAR/1 center*) OR"follow up" OR followup* OR clinical* OR trial OR random* OR review*)) AND DT=(article) AND la=(english)
Pubmed publisher	("Breast Neoplasms"[mh] OR Mammography[mh] OR (((breast OR mamma*[tiab]) AND (cancer*[tiab] OR neoplas*[tiab] OR tumor*[tiab] OR tumors[tiab] OR tumour*[tiab] OR carcino*[tiab] OR adenocarcin*[tiab])) OR mammogra*[tiab])) AND ("Mass Screening"[mh] OR "Early Diagnosis"[mh] OR (screen*[tiab] OR ((annual*[tiab] OR periodic*[tiab]) AND examination*[tiab])) OR (early AND (diagnos*[tiab] OR detect*[tiab])))) AND (mortality[mh] OR (mortalit*[tiab] OR (death rate*[tiab])))) NOT (letter[pt] OR news[pt] OR comment[pt] OR editorial[pt] OR congresses[pt] OR abstracts[pt]) AND english[la] AND (europe[mh] OR (europe* OR Andorra* OR Austria* OR Balkan* OR Belgi* OR Albania* OR Baltic-State* OR Bosnia* OR Herzegovina* OR Bulgaria* OR Croatia* OR Czech* OR Hungar* OR Kosovo* OR Macedonia* OR Moldova* OR Montenegr* OR Poland* OR polish* OR Belarus* OR Romania* OR Russia* OR Serbia* OR Slovakia* OR Slovenia* OR Ukraine* OR France* OR french OR German* OR Gibraltar* OR Great-Brit* OR uk OR united-kingdom* OR England* OR Scotland* OR Wales* OR welsh OR Greece* OR Ireland* OR Italy OR Italian OR Liechtenstein* OR Luxembourg* OR Monaco* OR Netherlands* OR dutch OR holland OR Portug* OR San-Marino* OR Scandinavia* OR Nordic* OR Denmark* OR danish OR Finland* OR finnish OR Iceland* OR Norwa* OR norwegian OR Sweden* OR swedish OR Spain* OR spanish OR Switzerland* OR swiss)) AND ("observational study"[pt] OR "Cohort Studies"[mh] OR "Health Surveys"[mh] OR "Epidemiologic Studies"[mh] OR "Case-Control Studies"[mh] OR "Cross-Sectional Studies"[mh] OR "multicenter study"[pt] OR "comparative study"[pt] OR "clinical study"[pt] OR "clinical trials"[pt] OR "Random Allocation"[mh] OR "review"[pt] OR (((observation*[tiab] OR epidemiolog*[tiab]) AND (stud*[tiab] OR data OR research)) OR cohort*[tiab] OR longitudinal*[tiab] OR retrospectiv*[tiab] OR prospectiv*[tiab] OR population*[tiab] OR (national*[tiab] AND (stud*[tiab] OR survey)) OR (health*[tiab] AND survey*[tiab]) OR ((case OR cases OR match*[tiab]) AND control*[tiab]) OR (cross section*[tiab]) OR correlation*[tiab] OR multicenter*[tiab] OR (multi center*[tiab]) OR "follow up" OR followup*[tiab] OR clinical*[tiab] OR trial OR random*[tiab] OR review*[tiab])) AND publisher[sb]
Google scholar	"breast | mammary cancer | neoplasm | tumor | carcinoma | adenocarcinoma" screening | "annual | periodic examination" | "early diagnosis | detection" mortality | "death rate" europe cohort | longitudinal | prospective | retrospective | trial | epidemiological | epidemiologic

Appendix 2. Number of potentially relevant citations for source.

Source	Number citations	Number distinct citations
Initial searches		
Embase.com	2.169	2.132
Medline Ovid	1.891	814
Web-of-science	1.171	553
Cochrane	182	51
Google scholar	200	121
Total	5.613	3.671
Expert opinion	153	122
Total references	5.766	3.793

Appendix 3. Characteristics of excluded studies.

Study	Reason for exclusion
Alexander F E, 1994	Contained data or patients duplicated in another study
Alexander F E, 1997	Contained data or patients duplicated in another study
Autier P, 2012	Study did not provide a direct estimation of BC mortality reduction due to screening.
Baker SG, et al	Study did not provide a direct estimation of BC mortality reduction due to screening.
Barchielli A, 2001	Study did not provide a direct estimation of BC mortality reduction due to screening.
Bastos, 2017	Study did not provide a direct estimation of BC mortality reduction due to screening.
Beau, 2017	Contained data or patients duplicated in another study
Bjurstam N, 1997	Contained data or patients duplicated in another study
Bjurstam N, 1997	Contained data or patients duplicated in another study
Bjurstam, 2003	Contained data or patients duplicated in another study
Blamey R W, 2002	No control group, all screening
Blanks RG, 2000	The study does not reports the number of events and the total number of the participants
Broeders MJM, 2001	Study did not provide a direct estimation of BC mortality reduction due to screening.
Celko A, 1996	Study did not provide a direct estimation of BC mortality reduction due to screening.
Chamberlain J, 1988	Contained data or patients duplicated in another study
Chen H H, 1995	Contained data or patients duplicated in another study
Christensen L H, 2006	Study did not provide a direct estimation of BC mortality reduction due to screening.
Collette C, 1992	Contained data or patients duplicated in another study
Collette HJA, 1985	Contained data or patients duplicated in another study
Collette HJA, 1992	Contained data or patients duplicated in another study
Day N,1995	No control group, all screening
De Koning H, 1995	Follow up not long enough
De Waard F, 1988	Contained data or patients duplicated in another study
Domingo L, 2012	Study did not provide a direct estimation of BC mortality reduction due to screening.
Duffy S W, 2003	Contained data or patients duplicated in another study
Duffy SW 2, 2006	Contained data or patients duplicated in another study
Duffy SW, 2002	Contained data or patients duplicated in another study
Duffy SW, 2010	Study did not provide a direct estimation of BC mortality reduction due to screening.
Ellman R, 1993	Contained data or patients duplicated in another study
Fracheboud J, 2007	Study did not provide a direct estimation of BC mortality reduction due to screening.
Frisell J, 1991	Contained data or patients duplicated in another study
Frisell J, 1997	Contained data or patients duplicated in another study
Garcia Fernandez A, 2014	Study did not provide a direct estimation of BC mortality reduction due to screening.
García-Fernández A,2015	Study did not provide a direct estimation of BC mortality reduction due to screening.
Garne JP, 1997	Study did not provide a direct estimation of BC mortality reduction due to screening.
Gastrin G, 1993	No mammography but BSE

Gastrin G, 1994	No mammography but BSE
Gorini G, 2004	Study did not provide a direct estimation of BC mortality reduction due to screening.
Hakama M, 1999	No control group, all screening
Hanley, 2017	Study did not provide a direct estimation of BC mortality reduction due to screening.
Kalager M, 2009	Study did not provide a direct estimation of BC mortality reduction due to screening.
Kalager M, 2012	Study did not provide a direct estimation of BC mortality reduction due to screening.
Kalager M, 2014	Non-original data (meta-analysis)
Kauhava L, 2006	Study did not provide a direct estimation of BC mortality reduction due to screening.
Kolozsvári LR, 2013	Study did not provide a direct estimation of BC mortality reduction due to screening.
Larsson LG, 1996	Contained data or patients duplicated in another study
Larsson LG, 1997	Contained data or patients duplicated in another study
Lenner P, 1997	Follow up not long enough
Lind H, 2010	Study did not provide a direct estimation of BC mortality reduction due to screening.
Májek O, 2010	Study did not provide a direct estimation of BC mortality reduction due to screening.
Majek O, 2011	Study did not provide a direct estimation of BC mortality reduction due to screening.
McCann, 1999	Non-original data (meta-analysis/report)
McCann, 1999	Non-original data (meta-analysis/report)
Minelli L, 2007	No numbers of screening population
Moss S M, 1992	Contained data or patients duplicated in another study
Moss S M, 2006	Contained data or patients duplicated in another study
Mukhtar T K, 2013	No numbers of screening population
Nyström L, 2002	Duplicates numbers from original studies.
Nyström, 2017	Contained data or patients duplicated in another study
Olsen AH, 2005	Contained data or patients duplicated in another study
Olsson Å. 2011	Study did not provide a direct estimation of BC mortality reduction due to screening.
Ondrusova M, 2012	Study did not provide a direct estimation of BC mortality reduction due to screening.
Otten JDM, 2008	Study did not provide a direct estimation of BC mortality reduction due to screening.
Paap E, 2010	Contained data or patients duplicated in another study
Paci E, 2002 a	Contained data or patients duplicated in another study
Paci E, 2007	Contained data or patients duplicated in another study
Paci E, 2008	Study did not provide a direct estimation of BC mortality reduction due to screening.
Paci, 2011	Contained data or patients duplicated in another study
Parvinen I, 2011	No control group, all screening
Pons-Vigués, M, 2008	Study did not provide a direct estimation of BC mortality reduction due to screening.
Roberts M M, 1990	Contained data or patients duplicated in another study
Sarkeala T, 2008	Control group was modelled.
Sarkeala T, 2014	Study did not provide a direct estimation of BC mortality reduction due to screening.
Sasieni P, 2003	Study did not provide a direct estimation of BC mortality reduction due to screening.

Appendix 3. Continued

Sigurdsson K, 2013	Study did not provide a direct estimation of BC mortality reduction due to screening.
Szynglarewicz B, 2008	Follow up not long enough
Tabár L, 1985	Contained data or patients duplicated in another study
Tabár L, 1992	Contained data or patients duplicated in another study
Tabár L, 1993	Contained data or patients duplicated in another study
Tabár L, 1995	Contained data or patients duplicated in another study
Tabár L, 1995	Contained data or patients duplicated in another study
Tabár L, 1997	Contained data or patients duplicated in another study
Tabár L, 2000	Contained data or patients duplicated in another study
Tabár L, 2002	Contained data or patients duplicated in another study
Tabar, 2017	Study did not provide a direct estimation of BC mortality reduction due to screening.
Törnberg S, 1994	The study did not report the total number of the participant in each study group.
Van Dijk JAAM, 1994	Contained data or patients duplicated in another study
Verbeek ALM, 1884	Contained data or patients duplicated in another study
Verbeek ALM, 2003	Non-original data (meta-analysis/report)
Wärnberg F, 1999	Study did not provide a direct estimation of BC mortality reduction due to screening.

Appendix 4. Risk of bias in Cohort studies according to Newcastle-Ottawa scale.

	Selection				Comparability		Outcome			
Study	Representativeness of the exposed cohort	Selection of the non-exposed	Ascertainment of exposure	Absence of interest outcome at start of study	Study controls for age	Any additional factors	Assessment of outcome	Follow-up Length	Adequacy of follow-up	**Final result**
Anttila A, 2002	Yes. The cohort represents women eligible for inclusion in the BC screening program in Helsinki. (*)	Drawn from the same community. (*)	no individual data on exposure.	Yes. "[...] incidence based mortality, which included only those breast cancer deaths in which the incident breast cancer case had been diagnosed in the given age and period window with or without screening."(*)	Not controlled for age.	No additional factors	BC deaths were retrieved through record linkage with National Cause of Death Register. (*)	Yes. 8.5-10.5 years. (*)	No statement	6/9
Anttila A, 2008	Yes. The cohort included all women eligible for invitation Finnish BC screening program. (*)	Drawn from the same community. (*)	Secure record. (*)	Yes. "Only those deaths were included, where the diagnosis of breast cancer took place in the given calendar period and age group."(*)	Controls matched for age. (*)	No additional factors	BC deaths were retrieved through record linkage with National Cause of Death Register. (*)	Yes. 12 years. (*)	No statement	7/9
Anttinen A, 2006	Yes. The cohort is representative for women aged 50-69 years in the Health Care District of Central Finland. (*)	Drawn from the same community. (*)	Secure record. (*)	Yes, all individuals are alive at the start of the study. (*)	Not controlled for age.	treatment, tumor characteristics (*)	Yes. Survival data was collected from files of Finnish Cancer Registry and Statistics Finland. (*)	Yes. 8.0-12.5 years. (*)	No statement	7/9
Ascunce EN, 2007	Yes. The cohort included all women eligible for invitation to the screening programs in Navarre/Spain. (*)	Drawn from the same community - pre-screening period. (*)	no individual data on exposure.	Yes. "We excluded women with breast cancer diagnose before 1991." (*)	Controls matched for age. (*)	No additional factors	BC deaths were retrieved through record linkage with Navarre Mortality Register. (*)	Yes. We assume at least 5 years of follow-up. (*)	No statement	6/9

Appendix 4. Continued

Barco I, 2015	Selected group. BC patients in a reference hospital in Barcelona. Not representative for individuals in the community.	Drawn from the same community. (*)	Secure record. (*)	Yes, all individuals are alive at the start of the study. (*)	Controls matched for age. (*)	Tumor characteristics. (*)	BC deaths were retrieved through record linkage with National Population Register. (*)	Yes. 6 years (*)	No statement	7/9
Duffy SW, 2006	Yes. The cohort included all women eligible for invitation to the screening programs in 13 areas in Sweden. (*)	Drawn from the same community, pre-screening period. (*)	Secure record on individual exposure from screening centers. (*)	Yes. "Cancer cases exposed to screening are defined as those attending their last schedule screening appointment before diagnosis."(*)	Not controlled for age.	No additional factors	BC deaths were retrieved through record linkage with Swedish Cause of Death Register. (*)	Yes. >10 years. (*)	No statement	6/9
Ernst M, 2004	Selected group. BC patient in a hospital cohort without description of inclusion criteria. Not representative for individuals in the community.	Drawn from the same community . (*)	Secure record. (*)	All have been BC patients in the same hospital. No bias expected. (*)	Controls matched for age. (*)	prognostic factors (*)	No description.	Yes. 8 years. (*)	3 patients lost to follow up. Unlikely to cause bias. (*)	7/9
Hakama M, 1995	Yes. The cohort included all women eligible for invitation to the screening program in Kotka. (*)	Drawn from the same community. (*)	no individual data on exposure.	Yes, "refined mortality" refers to BC diagnosed after the start of screening. (*)	Controls matched for age. (*)	No additional factors	BC deaths were retrieved through record linkage with National Population Register. (*)	Yes. 9 years (*)	No statement	6/9
Hakama M, 1997	Representative of the average population in the community. The cohort included all women eligible for invitation to the Finnish screening program. (*)	Drawn from the same community. (*)	Secure record. (*)	Yes, "refined mortality" refers to BC diagnosed after the start of screening. (*)	Controls matched for age. (*)	No additional factors	BC deaths were retrieved through record linkage with National Population Register. (*)	No information on mean follow-up.	No statement	6/9

Hellquist BN, 2011	Yes. The cohort is representative for all women living in Swedish counties where the lower age limit for mammographic screening is 40. (*)	Drawn from different community (counties with different age limits).	No. Information on screening exposure only for BC deaths.	Yes. "Only invitation and attendance before diagnosis were of interest."(*)	Controls matched for age. (*)	No additional factors	BC deaths were retrieved through record linkage with Swedish Cancer registry. (*)	Yes. 16 years. (*)	No statement	5/9
Hofvind S, 2013	Yes. The cohort included all women eligible for invitation to the NBCSP. (*)	Drawn from the same community. (*)	Secure record. (*)	Yes. "Women who were diagnosed with breast cancer before the postal date of the invitation […] were excluded."(*)	Controls matched for age. (*)	County, calendar period, time since inclusion in the cohort (*)	BC deaths were retrieved through record linkage with Statistic Norway. (*)	Yes. 15 years. (*)	No statement	8/9
Johns LE, 2017	Yes. "Study area covered around one third of England and the whole of Wales."*	Drawn from the same community. (*)	Secure record. Individual-level screening exposure and mortality outcome data from NHSBSP (*)	Yes. Cohort was free of breast cancer on 1 January 1991. *	Controls matched for age. (*)	Socioeconomic factors *	BC deaths were retrieved through record linkage with Office for National Statistics (ONS). (*)	Yes. 15 years *	4% lost to follow up. Unlikely to cause bias. (*)	9/9
Jonsson H, 2000	Yes. The cohort is representative for all women living in Swedish counties where the lower age limit for mammographic screening is 40. (*)	Drawn from different community.	no individual data on exposure (counties where the lower age limit is 50).	Yes. "A breast cancer case was defined as a case of invasive breast cancer […] diagnosed at age 40-49 years during the reference or the study periods." (*)	Not controlled for age.	Region, time period (*)	BC deaths were retrieved through record linkage with Swedish Cancer Register. (*)	Yes. 8 years. (*)	No statement	5/9

Appendix 4. Continued

Jonsson H, 2001	Yes. The cohort is representative for all women living in Swedish counties where mammographic screening started in 1986-87. (*)	Drawn from different community (counties where mammographic screening started in 1993 or later).	No individual data on exposure used for the analysis.	Yes. "A breast cancer case was defined as a case of invasive breast cancer […] diagnosed at age 50-69 during the reference or the study period." (*)	Controls matched for age. (*)	Region, start of screening, year of follow-up (*)	BC deaths were retrieved through record linkage with Swedish Cause of Death Register. (*)	Yes. 10,1 years. (*)	No statement	6/9
Jonsson H, 2003	Yes. The cohort is representative for all women living in Swedish counties where the upper age limit for mammographic screening is 74. (*)	Drawn from different community (counties with lower age limits).	No individual data on exposure used for the analysis.	Yes. "A breast cancer was […] diagnosed at age 70-74 years during the reference or the study period." (*)	Controls matched for age. (*)	Region (*)	BC deaths were retrieved through record linkage with Swedish Cause of Death Register. (*)	Yes. 10.1 years. (*)	No statement	6/9
Jonsson H, 2003	Yes. The cohort included all women eligible for invitation to the screening program in Gävleborg. (*)	Drawn from different community.	No individual data on exposure used for the analysis.	Yes. "Only deaths from breast cancer diagnosed after the first invitation to screening were analyzed." (*)	Controls matched for age. (*)	Other causes of death, region, reference period (*)	BC deaths were retrieved through record linkage with Swedish Cause of Death Register. (*)	Yes. 22 years. (*)	No statement	6/9
Jonsson H, 2007	Yes. "The female population in Västnorrland and Norbotten constituted the study population."(*)	Drawn from different community.	No individual data on exposure used for the analysis.	Yes. "Cohort in the study groups were defined to include only breast cancer cases diagnosed after their first invitation to screening."(*)	Not controlled for age.	Lead time (*)	BC deaths were retrieved through record linkage with Swedish Cause of Death Register. (*)	Yes. 11 years. (*)	No statement	5/9
Kalager M, 2010	Yes. The cohort included all women eligible for invitation to the screening program in specific Norwegian regions. (*)	Historical controls drawn from the same community. (*)	Secure record. *	Yes. "Women diagnosed with breast cancer prior to first invitation date were excluded." (*)	Not controlled for age.	No additional factors	BC deaths were retrieved through record linkage with National Cause of Death Register. (*)	No. 2,2 years average.	Yes, complete follow-up (*)	6/9

Mook S, 2011	No. Cohort (hospital) includes BC patients that had to meet highly specific inclusion criteria.	Drawn from the same community .(*)	Secure record. (*)	All have been BC patients in the same hospital. No bias expected. (*)	Controls matched for age. (*)	prognostic factors (*)	Study uses mix of methods to assess outcome. Hence possible bias due to misclassification.	Yes. > 10 years. (*)	No statement	6/9
Moss S, 1999	Yes. The cohort is representative for British women aged 45-64. (*)	Different geographical area.	no description if info on (individual) screening exposure is available.	Yes. Only breast cancer deaths in women with BC diagnosed after entry to the trial were considered. (*)	Controls matched for age. (*)		BC deaths were retrieved through record linkage ("flagging") with UK National Health service Central Register. (*)	Yes. 16 years. (*)	No statement	5/9
Njor SH, 2015	Yes. The cohort included all women eligible for invitation to the Funen screening program. (*)	Drawn from the same community. Funen pre-screening period. (*)	Secure record. (*)	Yes. "Women with breast cancer diagnosis prior to the date of first invitation to mammography screening were excluded."(*)	Controls matched for age. (*)	No additional factors	BC deaths were retrieved through record linkage with Danish Central Population Register. (*)	Yes. 14 years. (*)	No statement	7/9
Olsen AH, 2007	Yes. The cohort included all women eligible for invitation to the screening program in Copenhagen. (*)	Drawn from the same community. Copenhagen pre-screening period. (*)	Secure record. (*)	Yes. "We […] included only deaths from breast cancer diagnosed during the observation period."(*)	Controls matched for age. (*)	Calendar period, region, screening exposure (*)	BC deaths were retrieved through record linkage with Danish Cause of Death Register. (*)	Yes. 10 years. (*)	No statement	8/9
Olsen AH, 2013	Yes. The cohort included all women eligible for invitation to the screening program in 4 Norwegian regions. (*)	Historical controls drawn from the same community. (*)	Secure record. Individual screening histories. (*)	Yes. "Women diagnosed with breast cancer prior to first invitation date were excluded." (*)	Controls matched for age. (*)	Region, time period, age at death (*)	BC deaths were retrieved through record linkage with Norw. Cause of Death Register. (*)	Yes. 5,9 years (*)	No statement	8/9
Otto SJ, 2003	Yes. The cohort included all women eligible for invitation to the screening programs in the Netherlands. (*)	Drawn from the same community .(*)	no individual data on exposure.	No demonstration.	Not controlled for age.	No additional factors	BC deaths were retrieved through record linkage with Cause of Death Register at CBS. (*)	Yes. >10 years. (*)	No statement	4/9

Appendix 4. Continued

Paci E, 2002	Yes. "The target population was the resident female population aged 50-69 years, who were invited over the period from 1990-96." (*)	Drawn from the same community . (*)	Secure record. (*)	Yes. Only breast cancer deaths after first invitation were included. (*)	Controls matched for age. (*)	tumor characteristics (*)	BC deaths were retrieved through record linkage with Tuscany tumor register. (*)	Yes. 8 years. (*)	No statement	**8/9**
Paci E, 2005	Yes. The cohort included all breast cancers diagnosed in women aged 50-69 in Turin and Florence. (*)	Drawn from the same community . (*)	written self-report. "Screening file".	Yes. This study uses a retrospective cohort where all participants have BC at the start of study. No risk of selection bias. (*)	Controls matched for age. (*)	City, tumor characteristics (*)	No statement.	No. Median follow up 4,5 years.	No statement	**5/9**
Parvinen I, 2006	Yes. The cohort is representative for women aged 55-69 years in the cities of Tuku, Tampere and Helsinki. (*)	Drawn from the same community. (*)	no individual data on exposure.	Yes. Mortality analysis only considered new breast cancer cases during the 11-year incidence period. (*)	Controls matched for age. (*)	No additional factors	Yes. "Data on […] breast cancer deaths were obtained from the Finnish Cancer Registry."(*)	Yes. 15 years. (*)	No statement	**6/9**
Parvinen I, 2015	Yes. The cohort is representative for women aged 40-84 years in the cities of Tuku, Helsinki and RoF. (*)	Drawn from a different region.	no individual data on exposure.	Yes. Results refer to incidence based mortality. (*)	Controls matched for age. (*)	No additional factors	Yes. Breast cancer deaths were obtained from the Finnish Cancer Registry. (*)	Yes. 10 years (*)	No statement	**5/9**
Peer PM, 1995	Yes. "The present study is restricted to women [in Nijmegen] born in 1925-1939, aged under 50 at January 1, 1975."(*)	Drawn from different geographical area.	no description if info on (individual) screening exposure is available.	Yes. Only BC diagnoses between 1975 until the end of 1990 were considered. (*)	Not controlled for age.	No additional factors	BC deaths were retrieved through record linkage with local registrar's office. (*)	Yes. 16 years. (*)	No statement	**4/9**

Puliti D, 2012	Representative of the average population in the community. The cohort included all women eligible for invitation to the Florentine screening program. (*)	Drawn from the same community. (*)	Secure record. (*)	Yes. Individuals with previous BC diagnosis were excluded. (*)	Controls matched for age. (*)	marital staus, deprivation index (*)	BC deaths were retrieved through record linkage with regional mortality registry. (*)	Yes. 16.5 years (*)	No statement	**8/9**
Sankatsing V, 2017	Yes. The cohort included all women eligible for invitation to the screening programs in the Netherlands. (*)	Drawn from the same community . (*)	no individual data on exposure.	No demonstration.	Controls matched for age. (*)	No additional factors	BC deaths were retrieved through record linkage with Cause of Death Register at CBS. (*)	Yes. >10 years. (*)	No statement	**5/9**
Sarkeala T, 2008	Yes. The present study is representative for the whole female population in the age group 60-79 years at death in 1992-2003."(*)	Modelled controls.	Secure record. (*)	No demonstration.	Controls matched for age. (*)	Policy category, calendar time (*)	BC deaths were retrieved through record linkage with National Population Registry. (*)	Yes. 10 years. (*)	No statement	**6/9**
Tabar L, 2001	Yes. The cohort included all women in Dalarna or Östegötland at the ages 40-69. (*)	Drawn from the same community .(*)	Secure record. (*)	Yes. Refined mortality. "Only deaths resulting from incident tumors diagnosed during each given time period were used in this primary analysis."(*)	Not controlled for age.	No additional factors	BC deaths were retrieved through record linkage with National Cause of Death register. (*)	Yes. 30 years. (*)	No statement	**6/9**
Tabar L, 2003	Yes. The cohort included all women in Dalarna or Linköping at the ages 40-69. (*)	Drawn from the same community .(*)	Secure record. (*)	Yes. Refined mortality. Both, diagnosis and death took place in one of the two periods o be included in the analysis.(*)	Controls matched for age. (*)	Self-selection bias and changes in cancer incidence (*)	BC deaths were retrieved through record linkage with Regional oncology centers and National Cause of Death register. (*)	Yes. 20 years (*)	No statement	**8/9**

Appendix 4. Continued

van Dijk JAAM, 1997	Yes. Study included all Nijmegen women born between 1995 and 1909 who had been invited to screening before the end of 1990.(*)	Drawn from different geographical area.	no description if info on (individual) screening exposure is available.	Yes. Refined mortality. "Patients diagnosed with breast cancer before this date [Jan 1 1978, start of screening pogram in Nijmegen] were excluded from the analysis."(*)	Not controlled for age.	No additional factors	Blind assessment based on medical records. (*)	Yes. > 10 years. (*)	No statement	**4/9**
Weedon-Fekjaer H, 2014	Yes. "We included all Norwegian women aged 50 to 79 years between 1986 and 2009." (*)	Drawn from the same community. (*)	Secure record. (*)	Yes. "We carefully separated breast cancers diagnosed in women before invitation to first screening from those diagnosed after invitation to avoid misclassification of breast cancer death according to exposure status."(*)	Controls matched for age. (*)	County, time period, cohort (*)	BC deaths were retrieved through record linkage with National Cause of Death Register. (*)	Yes. 24 years (*)	No statement	**8/9**

(*) The presence of this symbol means the study fitted the selected criteria and it was accounted in the final result. BC = Breast Cancer

Appendix 5. Risk of bias in Case-Control studies according to Newcastle-Ottawa scale.

	Selection				Comparability		Exposure			
Study	**Case definition**	**Representativeness of the cases**	**Control selection**	**Control definition**	**Study controls for age**	**Any additional factors**	**Ascertainment**	**Same method for case and control**	**Non-Response rate**	**Final result**
Allgood PC, 2008	adequate definition. "Cases were deaths from female breast cancer in women aged 50-70 years at diagnosis." (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	No BC death. "The controls were alive at the time of death of their case and were also known to the data source before the diagnosis of the case" (*)	Cases and controls matched according to age. (*)	Place of residence, deprivation. (*)	Secure record. Complete screening history from individual screening units. (*)	Yes (*)	No respondents described.	**8/9**
Broeders MJM, 2002	adequate definition with elaborate histological confirmation. (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	Yes. Referents were alive […] at the time of death of the case, […], and were free of BC at their index invitation. (*)	Cases and controls matched according to age. (*)	No additional factors	Secure record. Complete screening history from Screening Registry Nijmegen. (*)	Yes (*)	No respondents described.	**7/9**
Fielder HM, 2004	adequate definition. "cases were deaths from breast cancer in women aged 50-7 at diagnosis and who were diagnosed after he instigation of screening in 1991." (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	No control for history of BC.	Cases and controls matched according to age. (*)	No additional factors	Secure record. Complete screening history from health care database. (*)	Yes (*)	No respondents described.	**6/9**

Appendix 5. Continued

Gabe R, 2007	adequate definition. "The cases in this study were defined as women who died from breast cancer after the minimum age for screening and during the epoch of the organised mammography based program." (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	No description of history of BC. Controls were alive at the time the case died.	Cases and controls matched according to age. (*)	screening area, tumor characteristics (*)	No description	Yes (*)	No respondents described.	**6/9**
Heinävaara S, 2016	Adequate definition. Based on record linkage with Finnish Cancer Register with high %MV (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	No BC diagnosis. "Those (controls) diagnosed with breast cancer […] or died before the cases' index invitation date were excluded." (*)	Cases and controls matched according to age. (*)	year of index invitation, municipality (*)	Secure record. Complete screening history from Mass Screening Registry. (*)	Yes (*)	No respondents described.	**8/9**
Johns LE, 2017	adequate definition. Based on record linkage with Cancer Registry and Death Certificates, but cases should be histologically verified.(*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	"cohort members with no diagnosis of breast cancer at the date of diagnosis of the case and who were alive at the time of death of the case" *	Cases and controls matched according to age. (*)	Socio economic status *	Individual screening history from the NHS screening program *	Yes (*)	Same rate for both groups (0%) (*)	**9/9**

Massat NJ, 2015	adequate definition. "All women who died of primary breast cancer […] ages 47 to 89 […] and who had been first diagnosed with primary breast cancer (invasive) ages 47 to 89 and since 1990 were selected cases" (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	No BC death. "The controls were alive at the case date of death […] and had a first diagnosis pr primary breast cancer (invasive) within 6 month." (*)	Cases and controls matched according to age. (*)	Place of residence, deprivation.(*)	Secure record. Complete screening history from health care database. (*)	Yes (*)	No respondents described.	**8/9**
Massat NJ, 2015	adequate definition. "All women registered as having primary breast cancer as the leading cause of death, as having died ages 47-89 years […], and having been first diagnosed with breast cancer ages 47 to 89 years and since 1990 were selected cases." (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	No BC death. "Each control was alive at the case's date of death and had not been diagnosed with breast cancer prior to the case's date of first diagnosis." (*)	Cases and controls matched according to age. (*)	Place of residence, deprivation, number of invitations.(*)	Secure record. Complete screening history from health care database. (*)	Yes (*)	No respondents described.	**8/9**
Miltenburg GAJ, 1998	adequate definition. Based on record linkage with Death Certificates and Cancer Registry with high %MV (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	No description of history of BC. Controls were alive at the time the case died.	Cases and controls matched according to age. (*)	No additional factors	Secure record. Complete screening history from Screening Registry Utrecht. (*)	Yes (*)	No respondents described.	**6/9**

Appendix 5. Continued

Otto S, 2012	adequate definition. Based on record linkage with Death Certificates and Cancer Registry with high %MV (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	Yes. Women with BC diagnosis before first invitation were excluded. (*)	Cases and controls matched according to age. (*)	Calendar year of invitation (*)	Secure record. Complete screening history from Comprehensive Cancer Center Rotterdam. (*)	Yes (*)	No respondents described.	**8/9**
Paap E, 2014	adequate definition. Based on record linkage with Death Certificates and Cancer Registry with high %MV (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	Yes. Had to be free of BC at the moment of index invitation and had to be alive at time of death of case. (*)	Cases and controls matched according to age/ year of birth. (*)	area of residence (*)	Secure record. Complete screening history from regional screening organizations. (*)	Yes (*)	No respondents described.	**8/9**
Palli D, 1986	adequate definition with elaborate histological confirmation. (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	"Three controls had had breast cancer diagnosed in the previous year"	Cases and controls matched according to age. (*)	Municipality *	Secure record. Complete screening history from C.S.P.O. database. (*)	Yes (*)	Same rate for both groups (0%) (*)	**8/9**
Palli D, 1989	adequate definition. But only based Death Certificates.	Reviewers doubt series of cases is complete. Potential for selection bias.	Community controls. Controls were drawn from the same source. (*)	No description of history of BC. Controls were alive at the time the case died.	Cases and controls matched according to age. (*)	Place of residence, number of children, family history (*)	Secure record. Complete screening history from Center for Study and Prevention of Oncological disease, Florence. (*)	Yes (*)	Same rate for both groups (2%) (*)	**6/9**
Puliti D, 2008	Adequate definition. But based only on death certificates, no further review.	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	No BC diagnosis. "All controls had to be free of breast cancer up to the date of diagnosis of the matched case." (*)	Cases and controls matched according to age. (*)	Place of residence (*)	Secure record. Complete screening history from screening database. (*)	Yes (*)	No respondents described.	**7/9**

van Dijk JAAM, 1996	Adequate definition. With independent validation by panel of physicians. (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	Controls were alive and free of BC at their index invitation and residing in Nijmegen at the time of death of the case. (*)	Cases and controls matched according to age. (*)	No additional factors	Written self-report	Yes (*)	No respondents described.	**6/9**
van Schoor G, 2010	Adequate definition. With independent validation from committee of physicians. (*)	Consecutive series of cases (*)	Community controls. Controls were drawn from the same source. (*)	No description of history of BC. Referents were eligible for screening and alive at the time the case died.	Cases and controls matched according to age. (*)	No additional factors	No description	No description	No respondents described.	**4/9**
van Schoor G, 2011	Adequate definition. With independent validation from committee of physicians. (*)	No description.	Community controls. Controls were drawn from the same source. (*)	"Referents had to be eligible for screening, they did not have BC at the time of invitation and were living in Nijmegen at the time of death of the case."(*)	Cases and controls matched according to age. (*)	No additional factors	No description	No description	No respondents described.	**4/9**

(*) The presence of this symbol means the study fitted the selected criteria and it was accounted in the final result. BC = Breast Cancer

Appendix 6. Risk of bias for randomized control trials according to Cochrane risk of bias instrument.

Study	Entry[a]	Judgement	Support for judgement	Study	Entry[a]	Judgement	Support for judgement
Alexander FE, 1999	Random sequence generation (selection bias)	High risk	Sequences generates in clusters based on GP practices (high risk for SES bias)	Andersson I, 1988	Random sequence generation (selection bias)	Low risk	Quote: "Half of the women in each birth cohort were randomly selected." Probably done.
	Allocation concealment (selection bias)	Unclear risk	Method of concealment is not described. Probably high risk.		Allocation concealment (selection bias)	Unclear risk	Method of concealment is not described.
	Blinding of outcome Assessment (detection bias)	Unclear risk	The process is described as "flagging" which seems to be standard for the UK.		Blinding of outcome Assessment (detection bias)	Low risk	Independent and blinded endpoint committee.
	Incomplete outcome data Addressed (attrition bias)	Low risk	"Flagging" fails in 2% of all cases. Attrition failure seems unrelated to allocation and outcome.		Incomplete outcome data Addressed (attrition bias)	Unclear risk	The study did not address this outcome.
	Selective reporting (reporting bias)	Low risk	The study protocol is not available but it is clear that the published reports include all expected outcomes.		Selective reporting (reporting bias)	Low risk	study protocol not available but published results include all expected outcomes measures.
	Final judgement[b]	High Risk	C		Final judgment	Low risk	A
Study	Entry[a]	Judgement	Support for judgement	Study	Entry[a]	Judgement	Support for judgement
Andersson I, 1997	Random sequence generation (selection bias)	Unclear risk	Quote: "Women were randomly allocated". Probably done.	Bjurtsam N, 2016	Random sequence generation (selection bias)	High risk	study conducted cluster randomization based on date of birth
	Allocation concealment (selection bias)	Unclear risk	Method of concealment is not described.		Allocation concealment (selection bias)	Unclear risk	Method of concealment is not described.
	Blinding of outcome Assessment (detection bias)	Low Risk	Quote: "BC mortality was assessed by record linkage to the Swedish cause o death register."		Blinding of outcome Assessment (detection bias)	Low risk	BC death were classified through blinded endpoint commitees and linked to Swedish Cause of Death register.
	Incomplete outcome data Addressed (attrition bias)	Unclear risk	The study did not address this outcome.		Incomplete outcome data Addressed (attrition bias)	Unclear risk	The study did not address this outcome.
	Selective reporting (reporting bias)	Low Risk	Published results include all expected outcomes.		Selective reporting (reporting bias)	Low risk	The study protocol is not available but it is clear that the published reports include all expected outcomes.
	Final judgment	Moderate risk	B		Final judgment	High Risk	C

Study	Entry[a]	Judgement	Support for judgement	Study	Entry[a]	Judgement	Support for judgement
Frisell J, 1997	Random sequence generation (selection bias)	high risk	selection was done based on birth dates.	Moss S, 2015	Random sequence generation (selection bias)	Low risk	Quote: "randomization done by health authority computer system with specifically written software"
	Allocation concealment (selection bias)	high risk	Risk of crossover. Quote: "The use of mammography in the control group in 1981 when the first trial stared was around 8 percent.		Allocation concealment (selection bias)	Low risk	Quote: "The uninvited control group were unaware of their inclusion in the trial."
	Blinding of outcome Assessment (detection bias)	high risk	No blinded assessment. The responsible clinicians themselves determined cause of death of study population.		Blinding of outcome Assessment (detection bias)	Low risk	Individuals were followed up through record linkage with NHS Central Register (for vital status and cause of death)
	Incomplete outcome data Addressed (attrition bias)	Unclear risk	The study did not address this outcome.		Incomplete outcome data Addressed (attrition bias)	Low risk	Individuals lost to follow up were reported, same reasons across groups.
	Selective reporting (reporting bias)	Low risk	The study protocol is not available but it is clear that the published reports include all expected outcomes.		Selective reporting (reporting bias)	Low risk	study protocol not available but published results include all expected outcomes measures.
	Final judgment	High risk	C		Final judgment	Low risk	A
Study	Entry[a]	Judgement	Support for judgement				
Tabar L, 2011	Random sequence generation (selection bias)	Low risk	Randomization on community level but broken down into small clusters. Study protocol in Tabar L, 1985				
	Allocation concealment (selection bias)	Unclear risk	Method of concealment is not described.				
	Blinding of outcome Assessment (detection bias)	Low risk	Blinded local endpoint committees determined case status and cause of death.				
	Incomplete outcome data Addressed (attrition bias)	Unclear risk	The study did not address this outcome.				
	Selective reporting (reporting bias)	Low risk	all expected major outcome measures reported.				
	Final judgment	Low risk	A				

[a] Blinding of participant and personnel, Blinding of outcome assessment (patient-reported outcomes), and incomplete outcome data addressed (Short-term outcomes, 2-6 weeks) were not considered in the risk assessment of this analysis. [b] The summary assessment of the overall risk of bias is based on the judgment of the individual domains, with special emphasis on selection bias.

Appendix 7. Summary of possible bias in randomized control trials.

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of outcome Assessment (detection bias)	Incomplete outcome data Addressed (attrition bias)	Selective reporting (reporting bias)	Final judgment
Alexander FE, 1999	-	?	?	+	+	High risk
Andersson I, 1988	+	?	+	?	+	Low Risk
Andersson I, 1997	?	?	+	?	+	moderate risk
Bjurtsam N, 2016	-	?	+	?	+	High risk
Frisell J, 1997	-	-	-	?	+	High risk
Moss S, 2015	+	+	+	+	+	Low Risk
Tabar L, 2011	+	?	+	?	+	Low Risk

-: High risk bias, ?: unclear risk of bias, +: low risk of bias.

Appendix 8. Conflict of interest and/or funding statements of all included studies.

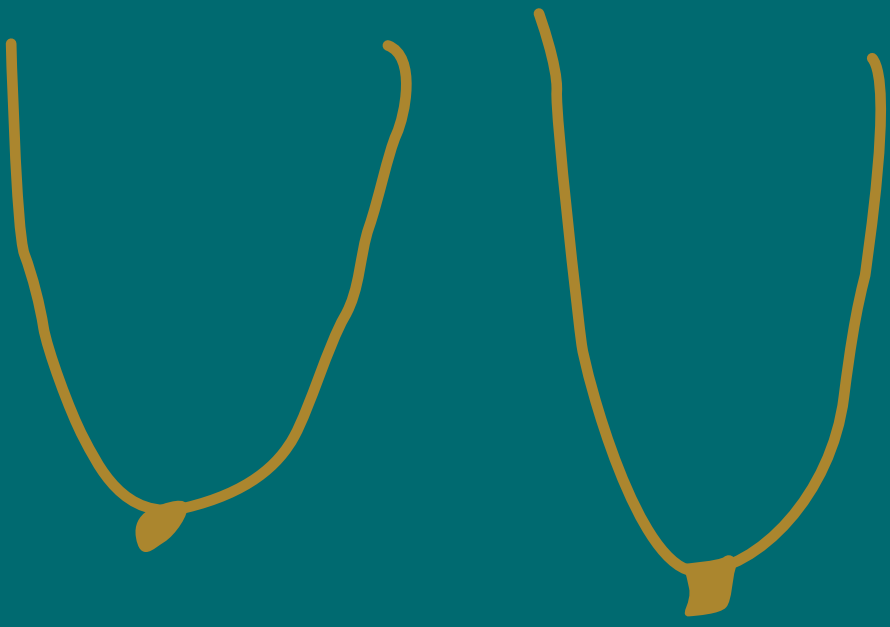
Study	Conflict of interest and/or funding statement
Alexander FE, 1999	The Edinburgh randomised trial was funded by the Cancer Research Campaign and the Chief Scientist's Office (CSO) of the Scottish Home and Health Department.
Allgood PC, 2008	No statement
Andersson I, 1988	This study was supported by the Swedish Cancer Society, project No 2357- B87-O1XA.
Andersson I, 1997	No statement
Anttila A, 2002	No statement
Anttila A, 2008	The author(s) declare that they have no competing interests.
Anttinen A, 2006	No statement
Ascunce EN, 2007	No statement
Barco I, 2015	The authors have stated that they have no conflicts of interest.
Bjurtsam N, 2016	The Gothenburg trial was funded by the City of Gothenburg board of health care. Conflict of interest: None.
Broeders MJM, 2002	No statement
Duffy SW, 2006	This work was supported by the American Cancer Society, through a grant from the Longaberger Company, and a grant from the European Commission through the European Breast Cancer Network, project 3.3.
Ernst M, 2004	No statement
Fielder HM, 2004	The European Commission Public Health Programme, within the European Breast Cancer Network, provided partial funding support.
Frisell J, 1997	No statement
Gabe R, 2007	Funding was provided for Rhian Gabe by the International Agency for Research on Cancer, Lyon.
Hakama M, 1995	The evaluation of the project was supported by the Finnish Slot Machine Association and the Finnish Cancer Institute.
Hakama M, 1997	Funding: Cancer Society of Finland, the Finnish Slot Machine Association, and the Finnish Cancer Institute. Conflict of interest: None.
Heinävaara S, 2016	The authors declare no conflict of interest.
Hellquist BN, 2011	The authors made no disclosures.
Hofvind S, 2013	All authors are employed at the Cancer Registry of Norway, which is administering the Norwegian Breast Cancer Screening Program.
Johns LE, 2017	The authors declare no conflict of interest.
Johns LE, 2017	This work was funded by the Policy Research Programme of the Department of Health for England (grant no. 0040064)
Jonsson H, 2000	This study was supported by the Swedish Cancer Society and the European Commission.
Jonsson H, 2001	This study was supported by the Swedish Cancer Society and the European Commission.
Jonsson H, 2003	This study was supported by the Gävle Cancer Foundation, Swedish Cancer Society and the European Commission.
Jonsson H, 2003	The present study was supported by the Swedish Cancer Society and the European Commission.
Jonsson H, 2007	The study was financially supported by the county councils in Västerbotten, Norrbotten, Västernorrland and Jämtland.

Appendix 8. Continued

Kalager M, 2010	Supported by the Cancer Registry of Norway and the Research Council of Norway. No potential conflict of interest relevant to this article was reported.
Massat NJ, 2015	No potential conflicts of interest were disclosed. This study was funded by a grant from the UK Department of Health.
Massat NJ, 2015	No potential conflicts of interest were disclosed. This study was funded by a grant from the UK Department of Health.
Miltenburg GAJ, 1998	This study was funded by the Praeventiefonds and the National Health Insurance Board.
Mook S, 2011	Dutch Cancer Society (NKI 2009-4363 to M.K.S.); Dutch National Genomics Initiative-Cancer Genomics Center (NKI CGC 2008-2012 to L.J.V.t.V.). The funding sources had no role in study design, collection, analysis, or interpretation of data, writing of the article, or in decisions relating to publication.
Moss S, 1999	This study was undertaken by the Cancer Screening Evaluation Unit, which receives support from the UK Department of Health.
Moss S, 2015	SMM received funding from the National Institute for Health Research Health Technology Assessment, American Cancer Society, Cancer Research UK, Department of Health, Medical Research Council, and the US National Cancer Institute, during the conduct of the study. All other authors declare no competing interests.
Njor SH, 2015	This study was financially supported by the Esper and Olga Boel Foundation. The funding source had no role in the study design, the analysis, interpretation of data, writing of the manuscript, or the decision to submit it for publication.
Olsen AH, 2007	Supported by Danish Cancer Society.
Olsen AH, 2013	Grant sponsor: Norwegian Research Council 189505/V50.
Otto SJ, 2003	None declared.
Otto S, 2012	This work was supported by the Dutch Health Care Insurance Council (CVZ) and the National Institute for Public Health and the Environment (RIVM).
Paap E, 2014	H.J.K. has received funding from SCOR Global Life SE (SGL). G.J.H. received payment from Philips Health care for an invitational lecture in Dubai. The authors reported no other financial interests related to this research. This study was supported by a research grant from the Dutch Cancer Society (KUN 2006-3571).
Paci E, 2002	We thank the American Cancer Society for financial support. Eugenio Paci has been partially supported by a UICC-ICCRETT grant. The Florence Screening Programme is a pilot project of the European Network for Breast Cancer Screening (contract no. S12.307923 (2000 CVG2-031)).
Paci E, 2005	No statement
Palli D, 1986	This work was supported by grants from the O.E.R. (Osservatorio Epidemiologico Regionale - Regione Toscana) and the A.I.R.C. (Associazione Italiana Ricerca sul Cancro).
Palli D, 1989	This work was supported by a grant from O.E.R. (Osservatorio Epidemiologico Regionale-Regione Toscana).
Parvinen I, 2006	No statement
Parvinen I, 2015	The authors declare no conflict of interest.
Peer PM, 1995	This ongoing study is supported by the Nationale Ziekenfondsraad.
Puliti D, 2008	This study was supported by the partial contribution of a research grant of the Italian League against cancer (Rome) and of Italian Ministry of Health.
Puliti D, 2012	The authors declare that they have no competing interests.
Sankatsing V, 2017	This work was supported by the National Institute for Public Health and the Environment (RIVM).

Appendix 8. Continued

Sarkeala T, 2008	No statement
Tabar L, 2001	Financial support received from the American Cancer Society.
Tabar L, 2003	COI: None declared. The American Cancer Society provided financial support for this study through a gift from the Longaberger company.
Tabar L., 2011	No potential conflicts of interest to disclose.
van Dijk JAAM, 1996	This study was supported financially by the Dutch Health Insurance Fonds Council "Ziekenfondsraad".
van Dijk JAAM, 1997	We are grateful to the Dutch ''Nationale Ziekenfondsraad'' for financial support.
van Schoor G, 2010	COI: None declared. This study was funded by an internal grant from the Radboud University Nijmegen Medical Centre. The funding source had no involvement in the study.
van Schoor G, 2011	The authors declare no conflict of interest. This study was funded by an internal grant from the Radboud University Nijmegen Medical Centre. The funding source had no involvement in the study.
Weedon-Fekjaer H, 2014	Funding: This study was supported by the Norwegian Research Council (reference No 189503). Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.



Chapter 3

Systematic reviews as a 'lens of evidence': Determinants of benefits and harms of breast cancer screening

Mandrik O, Zielonke N, Meheus F, Severens JLH, Guha N, Herrero Acosta R, Murillo R.

Int J Cancer 2019; 145: 994-1006.

ABSTRACT

Background

This systematic review, stimulated by inconsistency in secondary evidence, reports the benefits and harms of breast cancer (BC) screening and their determinants according to systematic reviews.

Methods

A systematic search, which identified 9,976 abstracts, led to the inclusion of 58 reviews.

Results

BC mortality reduction with screening mammography was 15–25% in trials and 28–56% in observational studies in all age groups, and the risk of stage III+ cancers was reduced for women older than 49 years. Overdiagnosis due to mammography was 1–60% in trials and 1–12% in studies with a low risk of bias, and cumulative false-positive rates were lower with biennial than annual screening (3–17% vs 0.01–41%). There is no consistency in the reviews' conclusions about the magnitude of BC mortality reduction among women younger than 50 years or older than 69 years, or determinants of benefits and harms of mammography, including the type of mammography (digital vs screen-film), the number of views, and the screening interval. Similarly, there was no solid evidence on determinants of benefits and harms or BC mortality reduction with screening by ultrasonography or clinical breast examination (sensitivity ranges, 54–84% and 47–69%, respectively), and strong evidence of unfavourable benefit-to-harm ratio with breast self-examination.

Conclusion

The reviews' conclusions were not dependent on the quality of the reviews or publication date. Systematic reviews on mammography screening, mainly from high-income countries, systematically disagree on the interpretation of the benefit-to-harm ratio. Future reviews are unlikely to clarify the discrepancies unless new original studies are published.

INTRODUCTION

The traditional evidence-based medicine pyramid places systematic reviews with meta-synthesis on the pinnacle of a hierarchy of evidence. The recently proposed update of the pyramid applies systematic reviews as a lens through which other types of studies should be appraised, considering synthesised evidence as a tool for stakeholders [1]. But does this lens always provide the same image, and if not, what can affect the conclusions of systematic reviews?

Many reviews on benefits and harms of breast cancer screening (BCS) have been published over several years. Some of these reviews were used as a basis for developing national or international guidelines, leading to inconsistent recommendations. In a set of systematic reviews, we summarise the data from reviews on four screening approaches – screening mammography, ultrasonography, clinical breast examination (CBE), and breast self-examination (BSE) – or their combinations, among the general population. To our knowledge, no study has previously synthesised the results from systematic reviews on determinants of benefits and harms (Part 1), participation rate (Part 2), or cost-effectiveness (Part 3) of BCS approaches or explored the possible differences in the conclusions of systematic reviews on this topic.

In this first review, we aim to report:

(1) Variability in the outcomes of the reviews (mortality reduction, overdiagnosis, false-positive rates (FPR), mortality induced, and intermediate outcomes of BCS);
(2) Variability in the determinants of benefits and harms;
(3) Review characteristics that explain the variability in the outcomes and derived conclusions.

METHODS

The design of this study was reported in the published protocol[2], and registered with the International prospective register of systematic reviews (PROSPERO, #CRD42016050764). We systematically searched the PubMed via Medline, Scopus, Embase, and Cochrane databases in August 2016 and conducted updates and searches for grey literature in February 2017 and again in April 2018 (Appendix 1).

Following the protocol, we excluded reviews not using a systematic (reproducible) literature search. Deviating from the protocol, we included two reviews on which consensus was not reached after two rounds of discussions. For each of the included reviews, we tabulated the outcomes, the score by the Assessing the Methodological Quality of Systematic Reviews (AMSTAR) checklist[3], the limitations of the reviews, and the limitations of the original studies (if

their quality was assessed by the reviews and considered in the conclusions). We also narratively summarised the outcomes of the reviews that scored two or higher on the AMSTAR checklist, considering the reviews with lower scores as non-systematic. For the reviews with updates, we synthesised the evidence from the most recent publication, separately reporting the conclusions of the previous versions.

The uni- and multinomial regressions were run in RStudio to assess an impact of factors on the AMSTAR quality score and conclusions of the reviews regarding mammography screening.

RESULTS

We identified 9,976 abstracts through our systematic search and 228 additional reviews through a non-systematic search (Figure 1). The inter-rater reliability between two reviewers for decisions on full-text inclusion was 85% (Cohen's kappa = 0.63; substantial agreement). The excluded reviews are indicated in Appendix 2.

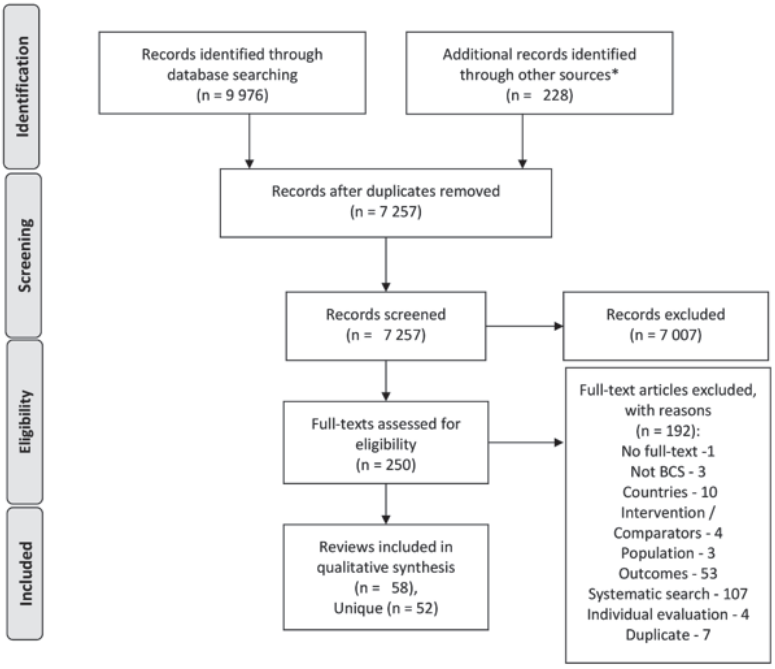


Figure 1. Prisma Flow Diagram

The 58 included reviews, of which 52 were without updates (Appendix 3), reported data on benefits (n = 30), harms (n = 9), or both (n = 19). Most reviews on benefits and harms of BCS were not limited to a particular geographical region or setting; the others searched for studies comparable to the target countries, such as the UK[4-6], the USA[7-13], Canada[14-17], Australia[18], the Republic of Korea[19], or Japan[20], or limited the literature search to a specific region (Asia in one systematic [21] and Europe in five narrative reviews [22-26]).

We did not identify systematic reviews reporting the benefits or harms of mammography screening in low- and middle-income countries (LMICs)[27]; BSE outcomes were reported for China and the Philippines, and CBE outcomes for India. Trials reporting final outcomes of mammography screening, cited in the reviews, were conducted only in, and observational evidence was mainly from, high-income jurisdictions (Appendix 3). A fixed-effects model was used in some of the reviews assessing the clinical outcomes of BCS programmes, including the cluster of Cochrane reviews[20, 28-33], which may signify an assumption of no cross-population differences in the interventions and outcomes.

The structure of the identified outcomes reported in the reviews on benefits and harms of BCS by screening modality is presented in Figure 2.

Screening mammography

Benefits of screening mammography among all age groups

There is consistency on breast cancer mortality (BCM) reduction among meta-analyses (Figure 3a) and reviews without meta-synthesis (Appendix 4), but no consistency in the interpretation of the size of the effect, the importance of the effect, and conclusions on screening with the observed risk or odds ratios being justified. The mean size of effect pooled from randomised controlled trials (RCTs) is 15–25%[6, 9, 11, 16, 19, 20, 25, 30, 34-38], from models/estimates is 11–33%[34-36], and from observational/population evidence is 28–56%[9, 25, 37]. The Cochrane review reported statistically non-significant all-cancer mortality reduction. The all-cause mortality reduction was also statistically non-significant in all the included reviews (Appendix 4).

Overall, the reviews of screening mammography reported high variability of accuracy and intermediate outcomes including sensitivity, size and proportion of small and advanced tumours at diagnosis, proportional interval cancer rate, interval cancer ratio, and positive predictive value (PPV). The most frequently reported outcomes, sensitivity and PPV, had ranges of 51–97% and 2–22%, respectively (Appendix 5).

Although screen-detected tumours may be slow-growing[33] and thus lead to overdiagnosis[8], tumour size is considered one of the most potent predictors of tumour behaviour in breast cancer (BC)[39]. The reviews were not fully consistent in concluding that mammography resulted in stage

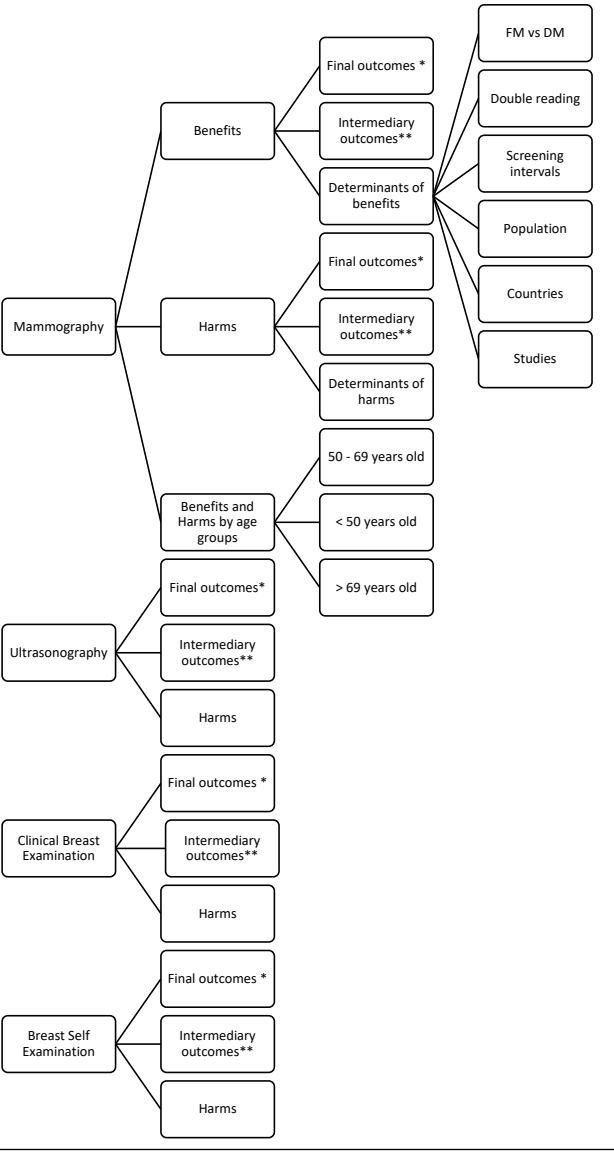


Figure 2. Structure of the outcomes of the reviews on benefits and harms of mammography, ultrasonography, clinical breast examination and breast self-examination

* Final outcomes for the benefits of screening: breast cancer mortality, cancer mortality, general mortality; final outcomes for harms of screening: overdiagnosis, overtreatment, false-positive diagnosis, and radiation-induced deaths.

** Intermediary outcomes for the benefits of screening: sensitivity, size and proportion of small and advanced of tumours at diagnosis, proportional interval cancer rate, interval cancer ratio, positive predictive value; Intermediary outcomes for the harms of screening: specificity, recall rates.

Abbreviations: FM – field mammography; DM – digital mammography.

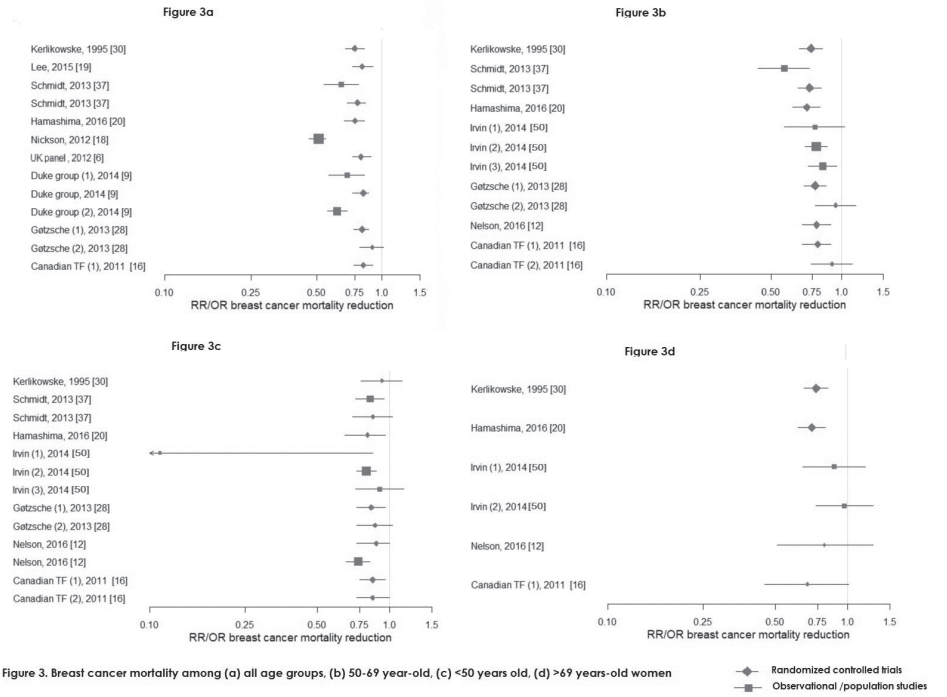


Figure 3. Breast cancer mortality reduction among (a) all age groups, (b) 50–69 year-old, (c)<50 year-old, (d)>69 year-old women.

Duke group (2014): (1) Case-control studies, (2) Incidence – based mortality studies; Gotzsche (2013): (1) All randomized trials; (2) Truly randomized trials; Canadian task force (2011): (1) All randomized trials; (2) Truly randomized trials; Irvin (2014): (1) Birth cohort comparison; (2) Geographical comparison; (3) Geographical-Historical Comparisons.

shift or detection of smaller tumours[8, 12, 39-41]. We observed that the difference in the conclusions was related to how the target stage shift was defined (stage II+ vs stage III+). No statistically significant relative risk (RR) reduction was observed for shift of stage II+ cancers (Appendix 5). Risk of stage III+ cancers was reduced with mammography screening for women older than 49 years (RR, 0.62; 95% confidence interval, 0.46–0.83) compared with no screening [12].

Determinants of benefits of screening mammography

There was no consistency in the reviews whether digital mammography has higher or lower accuracy than screen-film mammography (Appendix 6)[13, 15, 42-44]. The reviews suggested that digital mammography performs better in women younger than 50 years, premenopausal or perimenopausal, with heterogeneously or extremely dense breast tissue[15, 43]. Four reviews concluded on inconsistent evidence on recall rates[13, 15, 42, 44], and one on shorter examination times with digital mammography[15].

The included reviews also compared one- versus two-view mammography, double versus single reading, and screening with different intervals (from 12 to ≥36 months). The review by Kerlikowske et al. (1995) reported similar BCM reduction with one- and two-view mammography[30]. Posso et al. (2017)[45] summarising the evidence from studies where the recall decision was reached by consensus between two readers concluded on similar detection and FPR, while Dinnes et al. (2001) suggested that double reading can improve accuracy compared with single reading if a positive decision by any of the readers is sufficient for recall[5].

Because there were no head-to-head trials comparing effectiveness of BCS by screening intervals, the reviews based their conclusions on indirect comparisons. The conclusions of five reviews were inconsistent about the sufficiency of the evidence on BCM differences with annual versus biennial or triennial screening [9, 11, 12, 16, 30]. One review found that younger women (<50 years) may benefit more from annual screening, but this evidence was insufficient [9].

Besides organisational aspects of screening, the reviews also considered breast cancer incidence by age, because higher incidence defined a higher effect of screening, and consistency of effect by country. Humphrey et al. (2002) reported that the highest incidence occurred before menopause[46].

The review of Myers et al. (2015) suggested that inconsistency in screening outcomes may be higher in the USA, where there is no single provider for BCS programmes, due to variability between patients, clinicians, and insurers[11]. Meta-regression analysis of the pooled odds ratios of BCM from case–control studies on BCS did not vary significantly by country[18].

The reviews' conclusions were affected by the characteristics of the original evidence included (trials or observational studies), and by the way the original evidence was analysed and synthesised. There is no observed relationship between initiation dates of RCTs and the reported BCM reduction[12] . According to Kerlikowske et al. (1995), studies initiated before 1980 had lower RR than later studies; the reported confidence intervals of the pooled risk ratios are much wider for later studies than for earlier publications[30]. Reviews based on observational evidence report larger BCM reduction than the conclusions based on data from RCTs, with the lowest impact on BCM within the best-randomised trials (Appendix 4).

Harms of screening mammography among all age groups

The main harms reported in the systematic reviews were overdiagnosis, overtreatment related to overdiagnosis, FPR, false-positive biopsies, and deaths attributable to radiation induced breast cancer (Appendix 7). The psychological impact of screening is not presented here, because this was not included in the search terms.

Definitions and measurements for overdiagnosis (ranged 0–84%) varied by: type of original evidence, source of cases for the denominator (unscreened, screened detected, entire follow-

up, etc.), duration of follow up, accounting for ductal carcinoma in situ (DCIS) and other in situ lesions, adjustment for breast cancer risk and lead time (Appendix 7). In general, studies using unscreened population in the denominator report higher overdiagnosis and lower rates of overdiagnosis were reported among the pooled values from RCTs and studies with a low risk of bias (Figure 4)[6, 9-12, 16, 19, 20, 28, 47]: 1–12% (2 of 3 reviews).

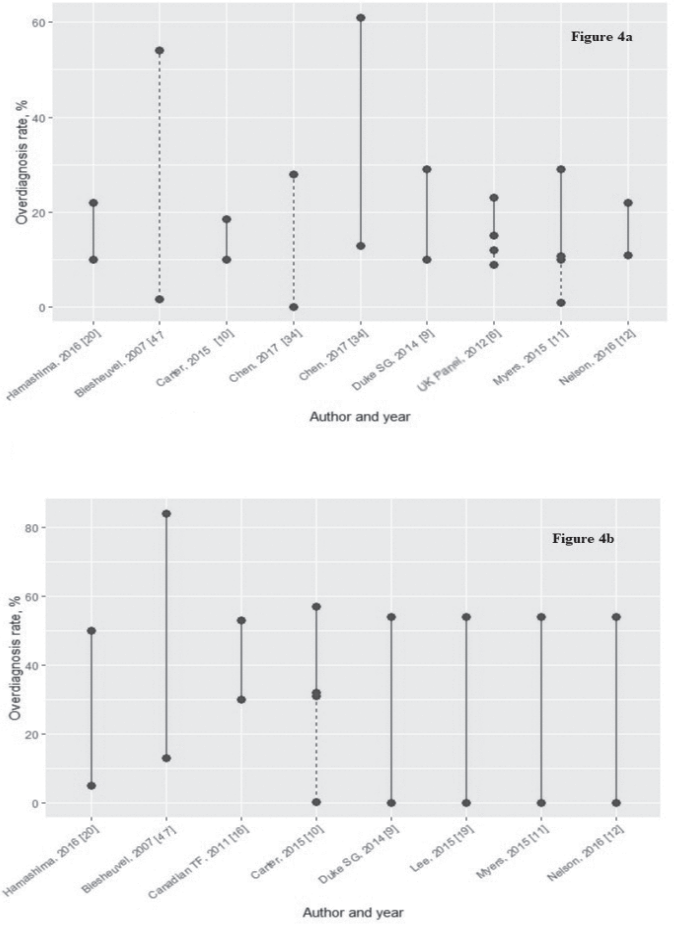


Figure 4. Overdiagnosis rate reported in systematic reviews of (a) randomised controlled trials and (b) observational studies.

Dashed line: Figure 4a. - Low risk of bias studies, Figure 4b. - Models.

Source of cases in denominators: Hamashima, 2016 [20] - not described; Biesheuvel, 2007 [47] - unscreened; Carter, 2015 – mixed (unscreened – 10%, screen expected – 4-76%, screen detected – 17-31%); Chen, 2017 - unscreened; Duke Synthesis Group, 2014 [9] – not described (unscreened -29%, screen detected -19%, entire follow up – 11%); The UK Panel, 2012 [6] – screen detected (16-19%) and entire follow up (10-11%); Myers, 2015 [11] - mixed (screen detected – 19%, entire follow up – 11%); Nelson, 2016 [12] – not described; Canadian Task Force, 2011 [16] - not described; Lee, 2015 [19] - not described.

Four reviews reported a higher risk of lumpectomies and mastectomies that could be related to a lead-time bias or overdiagnosis[12, 14, 28, 48]. Screen-detected breast cancers were more frequently treated with radiotherapy [12, 28, 48], but not with chemotherapy or hormone therapy[28, 48].

Similar to overdiagnosis, FPR and rate of false-positive biopsies varied significantly by screening interval, age of initiation, previous screening experience, and source of evidence (Appendix 7). The ranges of non-cumulative FPR were 6.5–8% with annual screening and 1–11% with biennial screening (Appendix 7)[9, 11, 48], and of cumulative FPR (after 10 years or lifetime) were 3–63% with annual and 7–60% with biennial screening [9, 12, 14, 16, 19, 20, 28, 38, 48] (Figure 5). Two reviews comparing these screening intervals concluded that FPR is higher with annual screening[11, 12]. The ranges for non-cumulative rate of false-positive biopsies were 2–12% with annual screening[11, 12] and 0.07–9% with biennial screening[9, 12, 14]; the cumulative rates (≥10 screenings) were 0.01–41% with annual screening [9, 11, 48] and 3–17% with biennial screening[9, 11, 48].

In contrast to the other harms, rate of deaths attributable to radiation was not significant in the reviews reporting on the topic[12,49,13]. Further, the most frequently reported intermediate outcomes of harms were specificity (>82% in all the reviews) and recall rate (3–14%) (Appendix 5).

Determinants of harms of screening mammography

Two reviews concluded on limited or no evidence whether overdiagnosis is higher with annual than biennial screening[9, 11]. FPR was considered to be higher with more frequent screenings [11, 12] and with longer duration of screening[11]. FPR was also higher for the first screen than for subsequent screens[11], in women with a family history of breast cancer and high breast density, and in women using hormone therapy[12]. The rate of false-positive biopsies per screen decreased with the availability of previous screening results[11]. Radiation-related harms increased with higher doses of exposure, younger age at exposure, and longer follow-up[48].

Similar to benefits, harms were not always consistent by country. Several reviews suggested that harms related to BCS may be higher in the USA[7, 11, 14, 20, 28], with possible explanations related to different screening and diagnostic guidelines, shorter screening interval, no national provider for screening services, and health-care provision through private centres.

Benefits and harms of screening mammography by age groups

Systematic reviews and meta-analyses of the RCTs show a positive effect (22–35%) of mammography screening on BCM reduction among women aged 50–69 years compared with no screening (Figure 3b, Appendix 4)[12, 13, 16, 20, 28, 30, 37]. All except one systematic review of observational evidence report BCM reduction of 17–49% in this age group[12, 18, 24, 26, 37, 50].

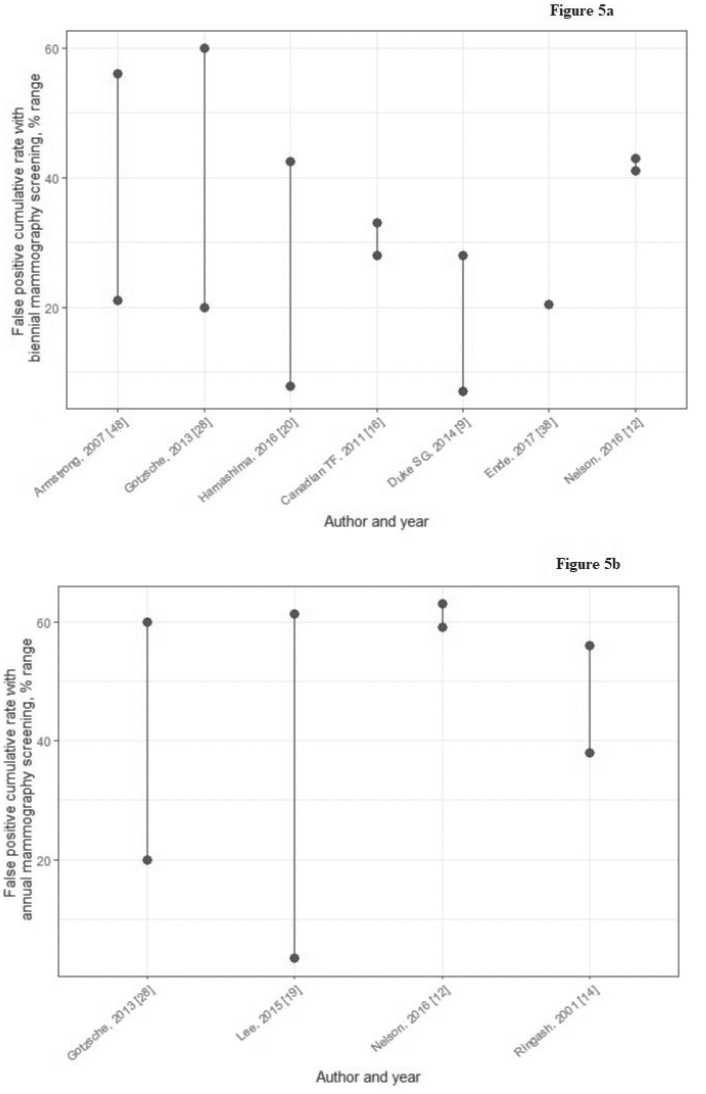


Figure 5. False positive cumulative rates with biennial (a) and annual (b) screening.

The conclusions and interpretations of the statistical findings of systematic reviews of either RCTs or observational studies reporting BCM reduction among women younger than 50 years[30, 33, 46] were and remain inconsistent (Figure 3c, Appendix 4)[9, 11, 12, 20, 28, 38, 50]. There was no review reporting all-cancer mortality reduction in this age group, and two meta-analyses concluding on statistically non-significant reduction in all-cause mortality[16, 38]. Seven reviews assumed that mammography screening has a higher benefit for women older than 50 years and a lower benefit for younger women[11-13, 20, 28, 38, 48], because of the lower test sensitivity of mammography due to higher breast density [12, 13, 29] and, possibly, faster-growing tumours. Myers et al. (2015) suggested that initiating screening at younger ages probably results in greater BCM reduction, but the magnitude of this incremental reduction is uncertain[11]. In the high-quality review by Nelson et al. (2016)[12] , the reduction in risk of advanced stage II+ or stage III+ breast cancers was not statistically significant for women younger than 50 years.

Two included reviews suggest that the rate of overdiagnosis may be larger among women aged 40–49 years[12, 38], with more than 25% of cases of breast cancer diagnosed among women in their 40s being low-grade DCIS, of which only 14% if left untreated could lead to invasive cancer after several decades[48].

Although FPR with a single examination was higher for older women [11, 12], the cumulative FPR was higher among women who initiated screening early (mainly <50 years)[12, 19, 46]. The reviews focusing on women younger than 50 years reported cumulative FPR of 20–56%[14, 38, 48]. The probability of receiving a certain diagnostic method was age-dependent: women aged 40–49 years experience the highest rate of additional imaging[12], and therefore may face higher radiation-related harms, whereas their rate of false-positive biopsies is lower than that of older women [12].

Regarding BCS-induced deaths, several reviews reported limited evidence for women screened annually for 10 years beginning at age 40 years. The estimated number of induced fatal breast cancers is small (8–25 per 100,000 women screened in 3 of 4 reviews)[12, 14, 16, 46] (Appendix 5), and is higher with earlier initiation of screening[12].

Similarly to the reviews on younger populations, systematic reviews report inconsistent BCM reduction among women older than 69 years (Figure 3d, Appendix 4)[8, 11, 12, 16, 20, 30, 50]. A review by Galit et al. (2007) concluded on lower BCM among women aged 75–84 years who underwent screening compared with those who did not [8], whereas other reviews concluded on no clear benefit for women older than 70 years[11, 12]. Regular mammography has been associated with smaller and earlier-stage tumours among women older than 74 years, which could also be clinically insignificant[8]. The reviews on BCS benefits and harms among women older than 69

years were based on limited evidence on BCM reduction from RCTs and harms specific to this age group, and did not report all- cancer or all-cause mortality.

Ultrasonography

No high-quality review (out of 6 included) identified studies reporting BCM reduction in BCS among the general population using ultrasonography alone or in combination with mammography (Appendix 8). The reviews targeting Asian populations reported high variability in sensitivity (54–84%), PPV (0.64–6.4%), and FPR (0.9–19.3%) of ultrasonography, with specificity of 96–98% and cancer detection rate of 2–3% per 1,000 screens. The highest-quality reviews [12, 29] concluded that ultrasonography is not justified as a supplementary tool for BCS, because of no solid evidence on its benefits. The reviews did not report transparently which factors can affect the accuracy of ultrasonography.

Clinical breast examination

The 10 included systematic reviews that assessed data on clinical breast examination agreed that the existing data on benefits of CBE are insufficient, because there is no solid evidence on a statistically significant impact of CBE on BCM (Appendix 8)[9, 13, 16, 20]. The range for sensitivity of CBE is 28–36% in the community [13] and 47–69% in RCTs in all except one review[13, 19, 20, 46]. The sensitivity of CBE was improved by spending more time on examination and by using a thorough technique[13]. The specificity of CBE was above 88% in all the reviews[13, 20, 46]. Compared with no screening, CBE was associated with a higher rate of false-positive biopsies [13, 46] and FPR [9, 12]. No solid evidence was identified on an impact of CBE on life expectancy and overdiagnosis [9].

Five reviews report no solid evidence on benefits of CBE combined with screening mammography versus mammography alone[9, 11, 12, 20, 30] (Appendix 8). The reviews' conclusions varied from "insufficient evidence on effects of CBE" to "no benefits of CBE in terms of mortality reduction"; the review by Lee et al. (2015) reported an incremental sensitivity of CBE added to mammography of 4–6%, with a decrement in specificity of 2%[19]. Limited data were available on harms of CBE added to mammography, with higher FPR and recall rates reported[11, 12]. Similarly to ultrasonography, the reviews did not report sufficiently on factors affecting the accuracy of CBE, besides an observation of lower sensitivity of screening in real-world versus trial settings.

Breast self-examination

Six reviews were consistent on no benefit of BSE on BCM (mainly referring to the 3 trials conducted)[12, 13, 31, 46], all-cause mortality[12], or number of cancers detected [31] (Appendix 8). The sensitivity of BSE was 20–41% in a real-world setting versus 40–89% on silicone models [13, 17]. The specificity of BSE on silicone models was 66–81%[17] . The reviews included reported harms related to FPR, including false-positive biopsies [12, 13, 31].

Quality of the reviews and factors affecting their conclusions

The quality of all of the included reviews varied from 1 to 10 on AMSTAR score (Appendix 9). The reviews were scored the highest on the attributes related to an adequate search approach, description of the included studies, and combining the results, and the lowest on reporting conflicts of interest, assessing publication bias, including grey literature, and reporting excluded studies (Figure 6). Multiple regression analysis was used to test if a year of publication, targeting high- income country (versus none), declaring funding, or including the evidence only from controlled trials significantly predicted AMSTAR score of the reviews. The results of the regression indicated that all four factors explained 22% of the variance (R^2=.22, F(6,45) = 2.09, p =.07) with funding and target country being not significant factors. The year of publication (β = .12, P<.05) and type of evidence included (β = -.82, P<.05) explained 16% of variance (R^2=.16, F(2,49) = 4.61, p =.01) with model being a better-fit than the univariate analyses.

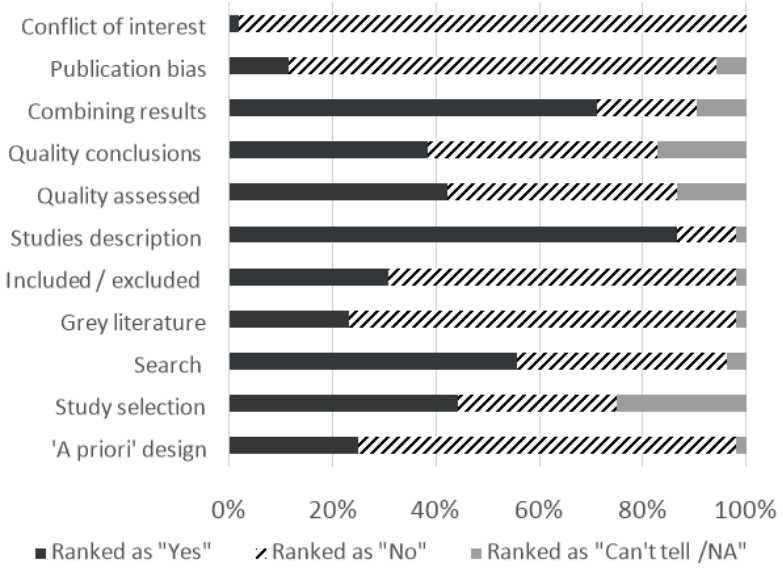


Figure 6. Quality of systematic reviews reporting benefits and/or harms of breast cancer screening.

The results of uni- and multi- variate regressions did not identify significance of such factors as AMSTAR score, date of publication, funding, using qualitative or meta-synthesis, or reporting benefits, harms, or both in the conclusions of the reviews on mammography screening (p >.05). The conclusions of the reviews reporting similar statistical results were not always identical and may be based on interpretation of statistics, choice of the main outcomes, rigorousness of inclusion criteria, and source of evidence. The conclusions of the reviews updated periodically

with the new evidence (Appendix 10) did not differ substantially from the previous versions. The publications from one cluster mainly reported similar values for outcomes.

While based on the same RCTs, reviews were inconsistent in the conclusions of trials' biases either in relation to benefits or harms estimation (Appendix 11). The reviews of observational evidence frequently included different studies; the quality of most of them was judged as fair or moderate and the selection bias was the main risk (Appendix 11).

DISCUSSION

Systematic reviews of BCS focus on mammography more than on the other screening approaches, and evaluate benefits of screening more frequently than harms. The available systematic reviews of either benefits or harms of BCS mainly target high-income countries; all RCTs and most of the observational studies on screening mammography were conducted in high-income jurisdictions, on ultrasonography in the USA and Asia, on CBE in North America and Asia, and on BSE in North America, Europe, the Russian Federation, and Asia.

The reviews' conclusions on any of the screening approaches were not seen to evolve with time, although some recent updates of the guidelines reported lower importance of mammography screening for younger women compared to earlier versions[12, 51]. We also did not observe a difference in the conclusions of the narrative and systematic reviews. The reviews with high AMSTAR scores and close publication or search date could reach contradictory conclusions on the benefit-to-harm ratio of mammography screening and the justification for its implementation. We found no evidence that variability in the reviews' conclusions was related to objective reasons (search date, rigorousness of inclusion criteria, choice of an outcome, source of evidence). The reviews of more rigorous evidence generally reported both lower benefits and lower harms. We did not see major additive value from the new reviews or updates of the previous reviews on BCS. We conclude that until new high-quality cohort or RCT results are published, additional reviews on BCS with mammography, ultrasonography, CBE, or BSE would not be of great value.

Summaries of evidence: mammography

The reviews are consistent in reduction in BCM among the general population and women aged 50–69 years, but not all-cancer or all-cause mortality. Both all-cancer and all-cause mortality may serve as the least biased outcomes of the efficacy of screening, avoiding possible mortality misclassifications. However, they may not be sensitive enough to detect the magnitudes in effects. Thus, disease-specific mortality may present the pure effect of the screening programme, while all-cancer and all-cause mortality may be considered in health-care resource allocation and

priority setting, enabling comparison of the relative value of screening mammography with other health-care innovations improving survival of the population.

The pure benefits and harms of mammography remain heterogeneous. BCS trials are highly diverse in their protocol designs, adherence, and evaluations; combining the outcomes of the RCTs into meta-analyses generates the expectations, but does not predict the outcomes of a specific program (which can either fail or succeed reaching higher effectiveness than meta-synthesised efficacy). Differences between reviews in quality assessment comprise not only identification of bias but also the assignment of overall quality scores, leading to variation in inclusion of RCTs. Subsequently, results of the reviews vary and conclusion were inconsistent. In general, the assessed reviews of RCTs have greater similarity in included studies but larger variability in quality assessment while reviews on observational studies show an opposite trend. If this overview will include only reviews incorporated the quality of studies in their conclusions, the disagreements among the reviews would remain. The impact of screening mammography on stage shift – the most potent intermediate predictor of screening efficacy – was positive for stage III+ breast cancer. BCM increases with progressing tumour stage[52], and therefore reduction of advanced tumours should improve patients' survival. Tabar et al. (2015) calculated that BCM reduction was reaching 28% in the trials achieving 20% or more reduction in advance cancers[53]. Since BCS programs are long-term planned and costly, detection of advanced cancers should serve as an early indicator of the possible success of the pilot BCS program.

The effectiveness of BCS relates to multiple parameters, including treatment access and efficacy. Regarding access, the health-care settings depicted in RCTs included in systematic reviews may reflect the current situation in LMICs, allowing an approximation of the expected benefits and harms for jurisdictions with limited resources. Furthermore, breast cancer survival also has improved dramatically through the decades due to treatment advances, with age-standardised 5-year survival reaching 85% or higher in 17 high-income countries and 80% or higher in 34 countries worldwide[54], which may diminish the benefits of mammography screening. If efficacy of late-stage treatments for breast cancer improves more, the clinical benefit of screening may decrease. Concurrently, the accuracy of mammography may also have improved through the years, favouring the benefit-to-harm ratio. Decisions on the rationale for screening should always be a balanced choice of the intervention able to offer the highest benefits with minimum harms, and preferably lower costs.

For women younger than 50 years or older than 69 years, the reviews were not consistent in their conclusions on BCM reduction, with no impact of screening on all-cancer mortality reported. For younger women, most reviews show no impact of mammography screening on early breast cancer detection. The harms may be also higher among younger women (radiation exposure and

FPR) and older women (overdiagnosis, because of shorter life expectancy); thus, the evidence collected by included reviews is not consistent on benefit-to-harm ratio for these age groups.

There was no consistency in determinants of higher benefits and lower harms of screening mammography, although double reading may improve sensitivity if the recall decision is based on at least one reader. DCIS is frequently detected and treated during mammography screening. Considering that relative survival with DCIS reached 100% even after 15 years of follow-up[52], the quality control system should advise on clear and non-aggressive management of screen-detected DCIS. The benefit-to-harm ratio may also be improved with availability of previous screening results. The guidelines on strict quality control and management of non-cancerous lesions could be more important in countries without a national screening provider, like the USA, where harms may be higher than in other countries.

Benefits to harm ratios of mammography screening among women 50 to 59 year old could not remain the same in all jurisdictions. As indicated, effective screening requires organized programs and may varying with disease incidence, population characteristics, and structures of financial and health-care systems. Considering the high variability in determinants of benefits and harms of screening, implementing BCS programmes without proper evaluation in these countries is risky, and so the results of the reviews should be extrapolated to LMICs with caution.

For LMICs with high breast cancer incidence and mortality, available early detection programmes, and sufficient capacity, piloting mammography screening among women aged 50–69 years should be combined with evaluation of implementation outcomes before programme scale-up.

Summaries of evidence: ultrasonography, clinical breast examination, and breast self-examination

The reviews agree on no solid evidence of mortality reduction with ultrasonography and CBE, and evidence of no effect and higher harms with BSE. Although our review could not summarise evidence from the reviews on reduction in advance breast cancers with CBE, the IARC Handbook on BCS concluded on sufficient strength of evidence regarding shifts in the stage distribution of tumours detected[55]. Because mortality reduction with ultrasonography and CBE screening is not confirmed while evidence of potential harms exists, r population programmes applying these approaches in countries without access to mammography are questionable. The sensitivity of both methods varies significantly, and real-world implementation may not reach the accuracy reported in trials. The accuracy of these screening approaches is provider-dependent; although CBE is perceived as a low-cost modality, its implementation in communities may entail substantial expenses related to quality assurance, invitations, and opportunity costs.

Because of the lack of solid evidence, the benefits and harms of ultrasonography and CBE should be explored further within pilot studies. We consider that appropriate implementation studies on these interventions are necessary even in countries with limited resources, because opportunistic benefits and costs may affect the functioning of the other health programmes.

Research and information gaps

We consider that additional reviews should be discouraged until new original evidence is available. The quality of reviews could be better standardised if the authors were systematically required to apply quality grading instruments to their submitted manuscripts.

More original research on benefits and harms of CBE, ultrasonography, and mammography screening among older women is required, which is especially important considering increasing life expectancy. Research targeted at improving the benefit-to-harm ratio of BCS should be encouraged.

The lack of primary and secondary research in LMICs does not enable extrapolation of the evidence to these settings. Because all screening approaches are operator-dependent, high-quality studies are required to gather effectiveness and implementation outcomes of the piloted BCS programmes.

Limitation

Considering the large scope of this systematic review, it is possible that we missed some of the important information despite the comprehensive approach to the evidence search and data extraction. We noted the limitations of using AMSTAR for judging the quality of reviews on cancer screening; some questions on AMSTAR may not be important for reviews of screening studies (such as conflicts of interest of the included studies), low AMSTAR scores may be related to journals' editorial policies on reporting, and high AMSTAR scores may not always mean the absence of biases.

Conclusion

Mammography screening for women aged 50 to 69 years results to decrease in BCM, but not all-cancer and all-cause mortality. It also causes harms, such as overdiagnosis and FPR, which are higher with more frequent screening. The conclusions of the reviews on benefits and harms of mammography were not consistent for the other age groups. No clear determinants of benefits and harms of mammography screening were identified. The other BCS approaches, such as US, CBE, BSE, cause harms but do not have sufficient evidence on mortality decrease.

Systematic reviews of mammography screening, mainly targeting high-income countries, are discordant in their interpretation of benefits and harms of screening, and their ratio. Their

conclusions are not related to their AMSTAR quality score, funding, objectives or the year of publication.

Acknowledgements

The authors are grateful to Taras Vereschak and Kostyantyn Dmitriev, who helped to screen the abstracts, Dr Maribel Almonte, Dr Sabina Rinaldi, Dr Beatrice Lauby-Secretan, Dr Robert Smith, and anonymous reviewers who provided valued comments and information regarding the manuscript content, Dr Armando Baena, who advised with graphical presentations, Dr Jin Young Park, Dr Bochen Cao, and Dr Chunqing Lin, who helped with translations, and Dr. Karen Muller who helped to stylistically edit the manuscript.

Definitions

Accuracy - ability of a test to discriminate between the target condition and health, such as sensitivity, specificity, and test predictive values;

Ductal carcinoma in situ - non-invasive or pre-invasive breast cancer;

False-positive rate – proportion or percentage of screening tests in which a test result improperly indicates presence of breast cancer when in reality it is not present;

Overdiagnosis - the diagnosis of a tumour that would not go on to cause symptoms or death in the woman's lifetime;

Positive Predictive Value - probability that a woman with a positive screening test truly has cancer.

REFERENCES

1. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evidence-based medicine 2016;21: 125-7.
2. Mandrik O, Ekwunife OI, Zielonke N, Meheus F, Severens JL, Lhachimi SK, Murillo R. What determines the effects and costs of breast cancer screening? A protocol of a systematic review of reviews. Systematic reviews 2017;6: 122.
3. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. Journal of clinical epidemiology 2009;62: 1013-20.
4. Petticrew MP, Sowden AJ, Lister-Sharp D, Wright K. False-negative results in screening programmes: systematic review of impact and implications. Health technology assessment (Winchester, England) 2000;4: 1-120.
5. Dinnes J, Moss S, Melia J, Blanks R, Song F, Kleijnen J. Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: findings of a systematic review (Structured abstract) Breast, vol. 10, 2001:455-63.
6. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. Lancet 2012;380: 1778-86.
7. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med 2007;146: 502-10.
8. Galit W, Green MS, Lital KB. Routine screening mammography in women older than 74 years: a review of the available data. Maturitas 2007;57: 109-19.
9. Havrilesky L, Gierisch JM, Moorman P, McCrory D, Ghate S, Williams J, Montgomery RC, Crowley M, Kosinski A, Grimm L, Davidson B, Kendrick AS, Chobot M, Gray R, Sanders G, Myers E, Systematic Review of Cancer Screening Literature for Updating American Cancer Society Breast Cancer Screening Guidelines. Duke Evidence Synthesis Group for American Cancer Society, 2014.
10. Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. Bmj 2015;350: g7773.
11. Myers ER, Moorman P, Gierisch JM, Havrilesky LJ, Grimm LJ, Ghate S, Davidson B, Mongtomery RC, Crowley MJ, McCrory DC, Kendrick A, Sanders GD. Benefits and Harms of Breast Cancer Screening: A Systematic Review. Jama 2015;314: 1615-34.
12. Nelson HD, Cantor A, Humphrey L, Fu R, Pappas M, Daeges M, Griffin J. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews Screening for Breast Cancer: A Systematic Review to Update the 2009 US Preventive Services Task Force Recommendationed. Rockville (MD): Agency for Healthcare Research and Quality (US), 2016.
13. Elmore JG, Armstrong K, Lehman CD, Fletcher SW. Screening for breast cancer. Jama 2005;293: 1245-56.
14. Ringash J. Preventive health care, 2001 update: screening mammography among women aged 40 - 49 years at average risk of breast cancer. Canadian Medical Association Journal, vol. 164, 2001:469-76.

15. Medical Advisory Secretariat. Cancer screening with digital mammography for women at average risk for breast cancer, magnetic resonance imaging (MRI) for women at high risk: an evidence-based analysis. Ont Health Technol Assess Ser [Internet], 2010.
16. Fitzpatrick-Lewis D., Hodgson N, Ciliska D, Peirson P, Gauld M, Liu YY Breast Cancer Screening. Canadian Task Force on Preventive Health Care. McMaster University, 2011.
17. Baxter N. Preventive health care, 2001 update: should women be routinely taught breast self-examination to screen for breast cancer? CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne 2001;164: 1837-46.
18. Nickson C, Mason KE, English DR, Kavanagh AM. Mammographic screening and breast cancer mortality: a case-control study and meta-analysis. Cancer Epidemiology, Biomarkers and Prevention, vol. 21, 2012:1479-88.
19. Lee EH, Park P, Kim NS, Seo HJ, PhD, Ko KL, Min JW, Shin MH, Lee K, Lee S, Choi N, Hur MH, Kim DI, Kim MJ, Kim SY, Sunwoo S, Dang JY, Kim SY, Kim Y, Lee WC, Jeong J. The Korean guideline for breast cancer screening. J Korean Med Assoc 2015;58: 408-19.
20. Hamashima C, Hamashima CC, Hattori M, Honjo S, Kasahara Y, Katayama T, Nakai M, Nakayama T, Morita T, Ohta K, Ohnuki K, Sagawa M, et al. The Japanese Guidelines for Breast Cancer Screening. Jpn J Clin Oncol 2016;46: 482-92.
21. Huang Y, Pang Y, Wang Q, Li JY. [Evaluation on the accuracy of high-frequency ultrasound being used in the breast cancer screening program in women from Asian countries: a systematic review]. Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi 2010;31: 1296-9.
22. Hofvind S, Ponti A, Patnick J, Ascunce N, Njor S, Broeders M, Giordano L, Frigerio A, Tornberg S, Project E, Euroscreen Working G, Van Hal G, et al. False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. J Med Screen 2012;19 Suppl 1: 57-66.
23. Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, Paci E. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. J Med Screen 2012;19 Suppl 1: 42-56.
24. Broeders M, Moss S, Nystrom L, Njor S, Jonsson H, Paap E, Massat N, Duffy S, Lynge E, Paci E. 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.
25. Moss SM, Nystrom L, Jonsson H, Paci E, Lynge E, Njor S, Broeders M. The impact of mammographic screening on breast cancer mortality in Europe: a review of trend studies. J Med Screen 2012;19 Suppl 1: 26-32.
26. Njor S, Nystrom L, Moss S, Paci E, Broeders M, Segnan N, Lynge E. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. J Med Screen 2012;19 Suppl 1: 33-41.
27. World Bank Country and Lending Groups: The World Bank Group, 2018.
28. Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2013: Cd001877.

29. Gartlehner G, Thaler K, Chapman A, Kaminski-Hartenthaler A, Berzaczy D, Van Noord Megan G, Helbich Thomas H. Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk Cochrane Database of Systematic Reviews: John Wiley & Sons, Ltd, 2013.
30. Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography: A meta-analysis. Journal of the American Medical Association 1995;273: 149-54.
31. Kösters J, Gotzsche PC. Regular self-examination or clinical examination for early detection of breast cancer Cochrane Database of Systematic Reviews: John Wiley & Sons, Ltd, 2003.
32. Njor S, Nystrom L, Moss S, Paci E, Broeders M, Segnan N, Lynge E, Euroscreen Working G. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. J Med Screen 2012;19 Suppl 1: 33-41.
33. Olsen O, Gotzsche PC. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2001: Cd001877.
34. Chen TH, Yen AM, Fann JC, Gordon P, Chen SL, Chiu SY, Hsu CY, Chang KJ, Lee WC, Yeoh KG, Saito H, Promthet S, et al. Clarifying the debate on population-based screening for breast cancer with mammography: A systematic review of randomized controlled trials on mammography with Bayesian meta-analysis and causal model. Medicine 2017;96: e5684.
35. Koleva-Kolarova RG, Zhan Z, Greuter MJ, Feenstra TL, De Bock GH. Simulation models in population breast cancer screening: A systematic review. Breast 2015;24: 354-63.
36. Schiller-Fruhwirth IC, Jahn B, Arvandi M, Siebert U. Cost-Effectiveness Models in Breast Cancer Screening in the General Population: A Systematic Review. Applied health economics and health policy 2017;15: 333-51.
37. Schmidt AF, Rovers MM, Klungel OH, Hoes AW, Knol MJ, Nielen M, de Boer A, Groenwold RH. Differences in interaction and subgroup-specific effects were observed between randomized and nonrandomized studies in three empirical examples. Journal of clinical epidemiology 2013;66: 599-607.
38. van den Ende C, Oordt-Speets AM, Vroling H, van Agt HME. Benefits and harms of breast cancer screening with mammography in women aged 40-49 years: A systematic review. Int J Cancer 2017;141: 1295-306.
39. Nagtegaal ID, Duffy SW. Reduction in rate of node metastases with breast screening: consistency of association with tumor size. Breast Cancer Res Treat 2013;137: 653-63.
40. Gøtzsche PC. Relation between breast cancer mortality and screening effectiveness: Systematic review of the mammography trial. Danish Medical Bulletin 2011;58.
41. Autier P, Boniol M, Middleton R, Dore JF, Hery C, Zheng T, Gavin A. Advanced breast cancer incidence following population-based mammographic screening. Annals of oncology : official journal of the European Society for Medical Oncology 2011;22: 1726-35.
42. Ho C, Hailey D, Warburton R, MacGregor J, Pisano E, Joyce J. Digital mammography versus film-screen mammography: technical, clinical and economic assessments. Technology report no 30. Canadian Coordinating Office for Health Technology Assessment, 2002.

43. Rothenberg BM, Ziegler KM, Aronson N. Technology evaluation center assessment synopsis: full-field digital mammography. Journal of the American College of Radiology : JACR 2006;3: 586-8.
44. Iared W, Shigueoka DC, Torloni MR, Velloni FG, Ajzen SA, Atallah AN, Valente O. Comparative evaluation of digital mammography and film mammography: systematic review and meta-analysis. Sao Paulo medical journal = Revista paulista de medicina 2011;129: 250-60.
45. Posso M, Puig T, Carles M, Rue M, Canelo-Aybar C, Bonfill X. Effectiveness and cost-effectiveness of double reading in digital mammography screening: A systematic review and meta-analysis. Eur J Radiol 2017;96: 40-9.
46. Humphrey L, Chan BKS, Detlefsen S, Helfand M. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews Screening for Breast Cancered. Rockville (MD): Agency for Healthcare Research and Quality (US), 2002.
47. Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. The Lancet Oncology 2007;8: 1129-38.
48. Armstrong K, Moye E, Williams S, Berlin JA, Reynolds EE. Screening mammography in women 40 to 49 years of age: a systematic review for the American College of Physicians. Ann Intern Med 2007;146: 516-26.
49. Erpeldinger S, Fayolle L, Boussageon R, Flori M, Laine X, Moreau A, Gueyffier F. Is there excess mortality in women screened with mammography: a meta-analysis of non-breast cancer mortality. Trials, vol. 14, 2013:368.
50. Irvin VL, Kaplan RM. Screening mammography & breast cancer mortality: meta-analysis of quasi-experimental studies. Database of Abstracts of Reviews of Effects, 2014:e98105.
51. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan B, Nygren P, Humphrey L. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews Screening for Breast Cancer: Systematic Evidence Review Update for the US Preventive Services Task Forceed. Rockville (MD): Agency for Healthcare Research and Quality (US), 2009.
52. Saadatmand S, Bretveld R, Siesling S, Tilanus-Linthorst MM. Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients. Bmj 2015;351: h4901.
53. Tabar L, Yen AM, Wu WY, Chen SL, Chiu SY, Fann JC, Ku MM, Smith RA, Duffy SW, Chen TH. Insights from the breast cancer screening trials: how screening affects the natural history of breast cancer and implications for evaluating service screening programs. The breast journal 2015;21: 13-20.
54. Allemani C, Weir HK, Carreira H, Harewood R, Spika D, Wang XS, Bannon F, Ahn JV, Johnson CJ, Bonaventure A, Marcos-Gragera R, Stiller C, et al. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). Lancet 2015;385: 977-1010.
55. Breast cancer screening. IARC Working Group on the Evaluation of Cancer-Preventive Interventions, 2014.

SUPPLEMENTARY MATERIALS

The complete supplementary materials (Appendix 1- 11) can be found in the open access publication.

Chapter 4

The potential of breast cancer screening in Europe

Zielonke N, Kregting LM, Heijnsdijk EAM, Veerus P, Heinävaara S, McKee M, de Kok IMCM, de Koning HJ, van Ravesteyn NT; EU-TOPIA collaborators.

Int J Cancer. 2021 Jan 15;148(2):406-418.

ABSTRACT

Background

Currently, all European countries offer some form of breast cancer screening. Nevertheless, disparities exist in the status of implementation, attendance, and the extent of opportunistic screening. As a result, breast cancer screening has not yet reached its full potential. We examined how many breast cancer deaths could be prevented if all European countries would biennially screen all women aged 50-69 for breast cancer.

Methods

We calculated the number of breast cancer deaths already prevented due to screening as well as the number of breast cancer deaths which could be additionally prevented if the total examination coverage (organised plus opportunistic) would reach 100%. The calculations are based on total examination coverage in women aged 50-69, the annual number of breast cancer deaths for women aged 50-74, and the maximal possible mortality reduction from breast cancer, assuming similar effectiveness of organised and opportunistic screening,

Results

The total examination coverage ranged from 49% (East), 62% (West), 64% (North) to 69% (South). Yearly 21,680 breast cancer deaths have already been prevented due to mammography screening. If all countries would reach a 100% examination coverage, 12,434 additional breast cancer deaths could be prevented annually, with the biggest potential in Eastern Europe. With maximum coverage, 23% of their breast cancer deaths could be additionally prevented, while in Western Europe it could be 21%, in Southern Europe 15%, and in Northern Europe 9%.

Conclusion

This study illustrates that by further optimising screening coverage, the number of breast cancer deaths in Europe can be lowered substantially.

INTRODUCTION

Breast cancer is a major public health problem in Europe. It is by far the most frequently diagnosed neoplasm in European women and is responsible for nearly one third of all new cancer cases among women in 31 European countries in 2018[1]. Breast cancer is also the leading cause of death in European women[1 2].

Randomised trials and several observational studies have demonstrated that systematic screening of eligible women through quality-assured population-based programmes for breast cancer reduces mortality from this disease[3-15].

Based on this evidence, in 2003 the European Commission's Initiative on Breast cancer Guidelines Development Group (GDG) published their first guidelines for organised mammography screening programmes for early detection of breast cancer in asymptomatic women with a strong recommendation to inviting women ages 50-69, every two years[16 17]. The guidelines and recommendations have been updated and expanded regularly ever since based on updated evidence on efficacy or diagnostics, resulting in extending the recommendations to triennial or biennial screening the age-groups 45-49 and 70-74 in the context of an organised screening programme[17].

At present, breast cancer screening programmes are well established in most European countries and all have some form of screening for breast cancer. Nevertheless, disparities exist in terms of the status of implementation, the extent to which screening programmes are organised, the invitation coverage, the coexistence with opportunistic screening activity and the attendance to screening[18].

In order to know to which extent the European recommendations have been adopted, reports on the implementation have been published in 2007 and 2017[3 18]. It was shown here as well as in other studies that the coverage of (organised) screening is of key importance in order to tap the full public health potential in terms of reduction in mortality from breast cancer[19 20].

However, in most European countries, opportunistic and organised screening coexist. Thus, to expect mortality reductions only from population-based screening programmes would probably lead to an underestimation of the total effectiveness of screening.

The primary aim of this study was to investigate what the effect would be of an increased or even complete breast cancer screening coverage on breast cancer mortality for each European country and if this effect differs between the four European regions. Therefore, we estimate how many breast cancer deaths have already been prevented due to screening and how many

deaths could additionally be prevented if countries would screen all women in the age-group 50-69 years every two years for breast cancer with a hypothetical 100% coverage of screening in the advised target age groups. The secondary aim was to provide an overview of screening practice and the amount of organised as well as opportunistic screening in Europe.

METHODS

Data

Data providers

As part of the EU-TOPIA project (TOwards imProved screening for breast, cervical and colorectal cancer In All of Europe), we collected data (see indicators listed in this section) of a recent year from over 36 data providers from 31 countries (see list of collaborators). They were either European screening organisers, researchers and/or policymakers. The data providers were contacted to collect any missing data, to correct any apparent inconsistencies and to approve on the use of it. For only a few countries (Greece, Portugal and Romania) data was completely missing despite best efforts of the authors to involve potential data providers. By utilizing other data sources like published reports[3] or online databases (e.g. the Cancer Mortality Database of the WHO[21] or ECIS - European Cancer Information System[22]), we filled these data gaps.

While our focus was clearly on national data, those were not available for a few countries. In Belgium, Spain, Sweden, Switzerland and the UK, healthcare delivery is organised at regional level with effectively independent screening programmes. Therefore, the data for the Belgian regions as well as the data for Scotland, Northern Ireland, England and Wales are presented separately in this study, while the data providers from Spain, Sweden and Switzerland could provide national estimates.

Indicators

Examination coverage of organised screening

Based on the IARC Handbook of Cancer Prevention (2015)[23] we defined organised screening as screening programmes organised at the national or regional level, with an explicit policy, including an active invitation of the entire target population and monitoring of cancer occurrence in the target population. For this study, the examination coverage of organised screening was specified as the proportion (%) of the target population (here: 50-69-year-old women) screened in the chosen report year after invitation. For countries without a population-based programme, the proportion is zero.

Examination coverage of opportunistic screening

Opportunistic or non-organised screening refers to all other breast cancer screening activity where individual invitations are not sent to the women in the eligible population or when women undergo a mammography outside or additionally to the (existing) screening programme[3 23]. Mammograms for symptomatic women are not counted as opportunistic screening. Generally, opportunistic screening is not monitored and is thus difficult to quantify. We asked the data providers to estimate opportunistic breast cancer screening by utilizing insurance data, survey results or by providing their expert opinion. If that was not possible, we applied the mean examination coverage of opportunistic screening of the European region.

Total examination coverage

We based our calculations on the total examination coverage as the sum of both organised and opportunistic examination coverage. For countries without an organised breast cancer screening programme and no estimate of opportunistic screening, we applied the region-specific average of the total examination coverage.

Breast cancer deaths

We included the absolute number of breast cancer deaths in women aged 50 to 74 years in the report year for each country or region within a country. In addition to the recommended screening ages range 50-69 we included breast cancer deaths for five additional years in ages 70-74 to account for death occurring after the last screening round.

Mortality reduction

The maximal possible mortality reduction is taken from a recently published systematic review on breast cancer mortality reduction due to screening[7]. In this publication, the authors identified those studies among 61 included studies that provided best evidence for breast cancer mortality reduction due to screening for each European region, based on observed data.

The identified studies (Table 1) represent point estimates for breast cancer mortality reduction due to breast cancer screening for each European region. These point estimates were 33% in Finland (North), 50% in Italy (South) and 58% in the Netherlands (West). We assume those reductions to be the same across all screened age groups. No studies from Eastern Europe met the initial inclusion criteria and subsequently evidence for mortality reduction due to breast cancer screening was lacking. Consequently, for these countries we applied the point estimate from Southern Europe as it is the medium value and because these two regions may seem fairly comparable in terms of the extent of screening coverage and the role of opportunistic screening.

Table 1. Overview of point estimates of breast cancer mortality reduction due to breast cancer screening from best evidence studies, per European region

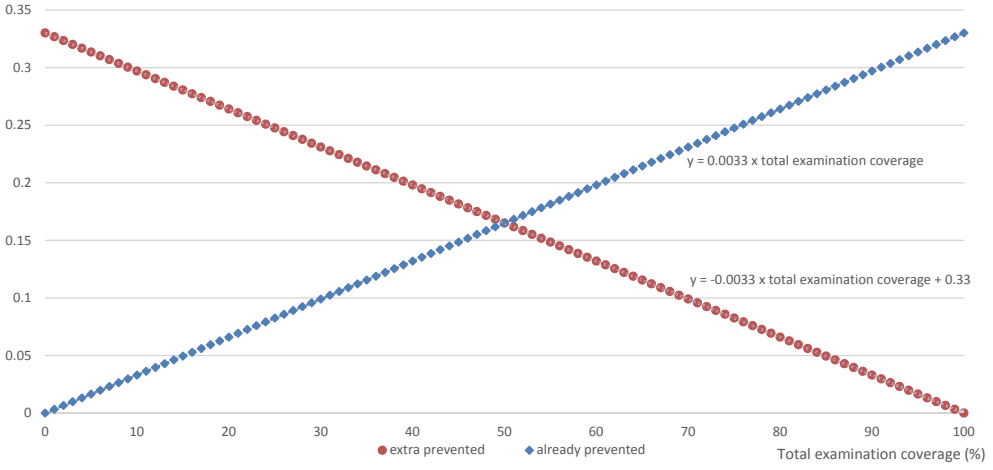
Study	Region	Country	Study type	Target age	Effect sizes for breast cancer mortality[a], (95%CI)
Heinävaara S, 2016[9]	North	Finland	Case-control	50-69	HR = 0.67 (0.49-0.90)[b]
Puliti D, 2008[33]	South	Italy	Case-control	50-74	OR = 0.50 (0.42-0.60)[b]
Paap E, 2014[12]	West	Netherlands	Case-control	50-75	OR = 0.42 (0.33-0.53)[b]

[a]Attenders/non-attenders. [b]Estimates corrected for self-selection bias.CI = Confidence interval, HR = Hazard Ratio, OR = Odds ratio

Calculations

We calculated for each country the number of breast cancer deaths which have already been prevented due to screening as well as the number of breast cancer deaths which could be additionally prevented if the total examination coverage (organised plus opportunistic) would reach 100%, assuming similar effectiveness of organised and opportunistic screening. We made four more assumptions to base our calculations on: first, that the underlying breast cancer mortality between current screening attenders and non-attenders is similar. Second, the maximal effect of breast cancer mortality reduction due to breast cancer screening differs across European regions, but is assumed to be the same in each of the region's countries, respectively. Third, the effects of breast cancer related therapy on the improvement of breast cancer specific mortality are implicitly accounted for in the level of reported breast cancer mortality and possible levels of breast cancer mortality reduction. They are also assumed to be the same in each region. And fourth, that the relationship between examination coverage and breast cancer mortality reduction is a linear one. Through linear interpolation of the point estimates from the best evidence studies for each European region, we were able to assign a potential breast cancer mortality reduction to any level of total screening coverage (calculation examples for each region are in Figure 1).

For example, based on the point estimates of breast cancer mortality reduction due to screening from the best evidence in each region (Table 1), the number of breast cancer deaths that were already prevented in a North European country would be calculated as 0.0033*total examination coverage*annual number of breast cancer deaths of women aged 50-74. For a South and East European country it would be 0.005*total examination coverage*annual number of breast cancer deaths of women aged 50-74 and for a West European country 0.0058*total examination coverage*annual number of breast cancer deaths of women aged 50-74.



Region	breast cancer deaths already prevented	Additionally preventable breast cancer deaths
North	y = 0.0033 * total examination coverage	y = -0.0033 * total examination coverage + 0.33
South	y = 0.005 * total examination coverage	y = -0.005 * total examination coverage + 0.5
West	y = 0.0058 * total examination coverage	y = -0.0058 * total examination coverage + 0.58
East	y = 0.005 * total examination coverage	y = -0.005 * total examination coverage + 0.5

Figure 1. (Potential) breast cancer mortality reduction, per total examination coverage (example region North)

By means of this graph, the number of already prevented breast cancer deaths and additionally preventable breast cancer deaths can be derived for any possible country.

The blue line (squares) represents the interpolated trend of the already prevented breast cancer deaths when the maximal possible breast cancer mortality reduction is 33% (Northern Europe). In a hypothetical Northern European country, the total examination coverage is 60% and 3,000 annual breast cancer deaths occur. These deaths need to be multiplied with the value on the y-axis resulting from the respective value on the x-axis (total examination coverage). Or alternatively. 0.0033*60=0.198 and 0.198*3,000=594. Thus, 594 women did not die of breast cancer due to current screening activity. To calculate the corresponding number of breast cancer deaths that could be additionally prevented if the examination coverage would increase to 100%, one needs to calculate the number of breast cancer deaths in the absence of screening first (i.e. the observed number of breast cancer deaths plus the breast cancer deaths that have already been prevented, thus 3,000 plus 594). Based on the total examination coverage, following the red line (circles), one can take the respective factor from the y-axis that these 3,594 deaths need to be multiplied with (or alternatively, y = -0.0033 * total examination coverage + 0.33). Hence, we calculated the factor on the y-axis to be 0.132 (-0.0033*60+0.33) and therefore 474 additional breast cancer deaths could be prevented. For the other three European regions, the calculations should be based on the respective regional values shown in the table above.

In contrast, the breast cancer deaths that could be additionally prevented if the screening coverage would increase to 100% is based on the number of breast cancer deaths in the absence of screening (i.e. the observed number of breast cancer deaths plus the breast cancer deaths that have already been prevented). In a North European country this number would be calculated as (-0.0033* total examination coverage +0.33)* annual number of breast cancer deaths of women aged 50-74 in the absence of screening. For a South and East European country it would be (-0.005* total examination coverage +0.5)* annual number of breast cancer deaths of women aged 50-74 in the absence of screening and for a West European country (-0.0058* total examination coverage +0.58)* annual number of breast cancer deaths of women aged 50-74 in the absence of screening (Figure 1).

Despite differences in target age range and frequency, for this study all calculations were based on the hypothetical situation of a uniform policy of screening women biennially between the ages 50 and 69. The observed coverage rates were adjusted accordingly.

Sensitivity analyses

Because of uncertainties around some assumptions made, the following sensitivity analyses were performed.

A sensitivity analysis was performed in which potential gains were calculated up to a maximal coverage of 84%, which is the highest screening coverage found in a European country (i.e. Denmark).

In addition, sensitivity analyses were performed in which the effectiveness of opportunistic screening was 10%, 20%, and 30% lower than organised screening. In these analyses, the percentages that could be gained to reach an examination coverage of 100%, were distributed over organised and opportunistic screening to the same distribution as was already present in the specific country (e.g. if present screening coverage was 40% organised and 20% opportunistic (ratio 2:1), the additional coverage was 27% organised and 13% opportunistic (2:1)).

To assess the impact of the regional point estimates on the maximal possible breast cancer mortality reduction on the regional results of this study, we performed a sensitivity analysis where we varied the point estimates across all European countries, i.e. we applied a 33% (North), a 50% (South) and a 58% (West) breast cancer mortality reduction due to screening irrespective of the location of the country.

RESULTS

Screening practice and Examination coverage

Most European countries adopted the target age range for breast cancer screening as recommended by the European Commission for which there is a strong recommendation (50-69). Only a few countries adopted a different age range and either invite women younger than 50 or they invite women beyond the age of 69, while a few stop inviting women at the age of 62 and 64, respectively. The screening interval was two years in all countries except for Malta and the United Kingdom (UK) where three yearly screening was practiced (Table 2).

The examination coverage of organised breast cancer screening was highest in Northern Europe and lowest in Eastern Europe (an average of 59% compared to 39%, Table 2). In contrast, the examination coverage of opportunistic screening was lowest in Northern Europe and highest in Southern Europe (5% compared to 32%). The total examination coverage ranged from 49% in Eastern Europe, 62% in Western Europe, 64% in Northern Europe to 69% in Southern Europe. With 84% and 25%, Denmark and Switzerland had the highest and the lowest total examination coverage, respectively.

Prevented breast cancer deaths

Based on the collected data, 42,051 women die of breast cancer in Europe every year. Due to the existence of breast cancer screening, 21,680 breast cancer deaths have already been prevented annually. Consequently, with no breast cancer screening activities, 63,731 women would have died of the cancer. Thus, 34% of breast cancer specific deaths have been prevented due to mammography screening across Europe. We calculated that 12,434 breast cancer deaths could additionally be prevented annually if breast cancer screening coverage would be extended to 100%. The regional results are presented in Figure 2 where Western Europe sticks out due to its population size as well as the biggest regional point estimate of breast cancer mortality reduction. In Western Europe, 22,031 women died of breast cancer in the reported year (red column). Due to the average total examination coverage of 61.5%, 13,147 breast cancer deaths were already averted. Hence, in the absence of screening, 35,178 women would have died annually of breast cancer (red striped column). If screening coverage would increase to 100%, only 14,742 breast cancer deaths would occur (grey striped column) as 7,298 additional breast cancer deaths could be averted annually. The respective numbers for all European countries and regions are presented in Table 3.

Table 2. Overview of national background data used as input

Country/region	Report year	breast cancer deaths 50-74	Examination coverage 50-69 (%)[1]		
			organised	opportunistic	total
North					
Denmark	2014	521	81.1	3.0	84.1
Estonia[2]	2016	121	37.4	8.0	45.4
Finland	2014	390	78.9	3.9	82.8
Iceland	2015	25	58.7	2.0	60.7
Latvia	2016	247	26.7	8.1	34.8
Lithuania	2016	265	44.2	5.0	49.2
Norway	2016	347	72.3	5.0	77.3
Sweden[3]	2016	605	76.5	1.0	77.5
Total North		*2,521*	*59.5*	*4.5*	*64.0*
West					
Austria[4]	2014	658	25.0	20.0	45.0
Wallonia (B)	2015	386	7.0	45.0	52.0
Brussel (B)	2015	69	11.6	42.0	53.6
Vlaanderen (B)	2015	736	51.0	18.2	69.2
France[3]	2015	5,043	51.6	13.5	65.1
Germany	2015	7,575	51.2	5.0	56.2
Ireland[5]	2015	335	53.3	3.9	57.2
Luxembourg	2013	29	56.0	5.7	61.7
Netherlands[3]	2015	1,628	75.8	5.0	80.8
Switzerland	2015	616	14.5	10.5	25.0
Scotland (UK)[6, 7]	2015	444	62.1	0	62.1
N. Ireland (UK)[6, 7]	2016	133	81.4	0	81.4
Wales (UK)[6, 7]	2016	264	76.6	0	76.6
England (UK)[6, 7]	4115	4,115	75.4	0	75.4
Total West		*21,972*	*49.0*	*12.1*	*61.5*
East					
Bulgaria	2015	711	-	49.0	49.0[8]
Croatia	2015	533	37.5	12.0	49.5
Czech Republic[4]	2016	823	57.6	3.0	60.6
Hungary[9•]	2015	1,197	22.5	19.5	42.0
Poland	2016	3,421	38.7	19.9	58.6
Romania[10]	2016	1,867	-	49.0	49.0[8]
Slovakia	2017	542	-	30.0	30.0
Slovenia	2015	177	40.1	13.0	53.1
Total East		*9,271*	*39.3*	*16.2*	*49.0*

South					
Cyprus	2017	58	35.1	32.4[11]	63.1
Greece[10]	2016	824	-	*68.9*	*68.9*[8]
Italy	2013	3,900	42.3	19	61.3
Malta[7]	2016	40	52.9	19.5	72.4
Portugal[3, 10]	2013	762	33.8	32.4[11]	66.2
Spain	2016	2,644	62	19.5	81.5
Total South		*8,228*	*45.2*	*32.4*	*68.9*

[1] The examination coverage of organised/ opportunistic screening was specified as the proportion (%) of the target population (here: 50-69 year old women) screened in the index year after invitation. [2] Screening ages 50-62. [3] Screening ages 50-74. [4] Screening ages 45-69. [5] Screening ages 50-64. [6] No opportunistic screening activity due to The Ionising Radiation (Medical Exposure) Regulations 2017. [7] 3-years screening interval. [8] Total screening is average or the region. [9] Screening ages 45-64. [10] Data from ECIS[4], Globocan[5] and the 2nd screening report[6]. [11] Opp. screening is average of the region.

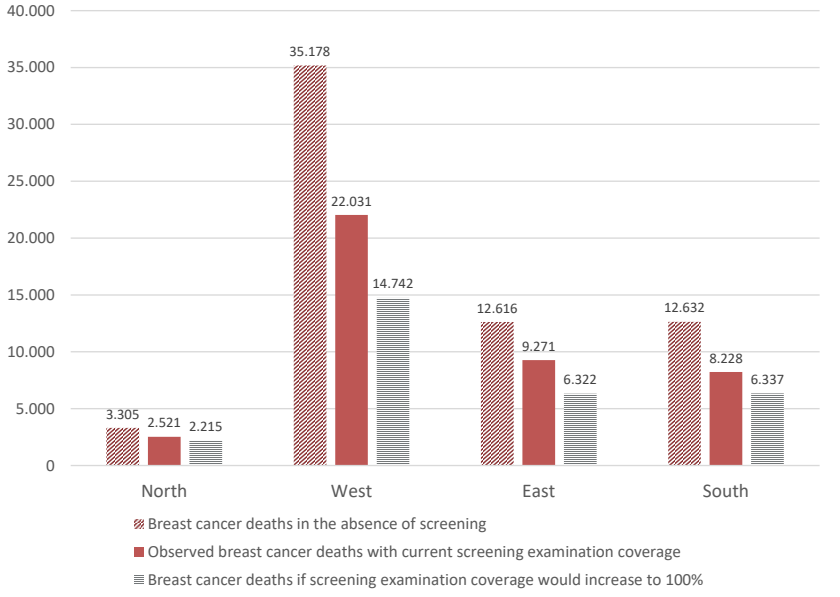


Figure 2. Annual number of observed and preventable breast cancer deaths, ages 50-74, per European region

Northern Europe: Denmark, Estonia, Finland Iceland, Latvia, Lithuania, Norway and Sweden.
Western Europe: Austria, Belgium, France, Germany, Ireland, Luxembourg, The Netherlands, United Kingdom and Switzerland.
Eastern Europe: Bulgaria, Czech Republic, Croatia, Hungary, Poland, Romania, Slovakia and Slovenia.
Southern Europe: Cyprus, Gibraltar, Greece, Italy, Malta, Portugal and Spain

Table 3. Number of (non-) preventable breast cancer deaths, and the results of the sensitivity analysis

	Prevented breast cancer deaths					Sensitivity analysis													
						max. European coverage		Sens -10%[1]		Sens -20%[1]		Sens -30%[1]		Max West[2]		Max North[2]		Max South[2]	
	A # BC deaths already prevented due to current screening coverage	B # BC deaths prevented if screening coverage would increase to 100	C # BC deaths in the absence of screening	A/C	B/C	# BC deaths already prevented due to current screening coverage	# BC deaths prevented if screening coverage would increase to 84%	# BC deaths already prevented due to current screening coverage	# BC deaths prevented if screening coverage would increase to 100%	# BC deaths already prevented due to current screening coverage	# BC deaths prevented if screening coverage would increase to 100%	# BC deaths already prevented due to current screening coverage	# BC deaths prevented if screening coverage would increase to 100%	# BC deaths already prevented due to current screening coverage	# BC deaths prevented if screening coverage would increase to 100%	# BC deaths already prevented due to current screening coverage	# BC deaths prevented if screening coverage would increase to 100%	# BC deaths already prevented due to current screening coverage	# BC deaths prevented if screening coverage would increase to 100%
North																			
Denmark	200	38	721	28%	5%	200	0	199	38	198	37	197	37	496	94	200	38	378	72
Estonia	21	26	142	15%	18%	21	18	21	25	20	25	20	24	43	52	21	26	36	43
Finland	147	30	537	27%	6%	147	2	146	30	146	30	146	30	360	74	147	30	276	57
Iceland	6	4	31	20%	13%	6	2	6	4	6	4	6	4	14	9	6	4	11	7
Latvia	32	60	279	11%	21%	32	45	31	58	31	57	31	56	62	117	32	60	52	97
Lithuania	51	53	316	16%	17%	51	42	51	53	50	52	49	51	106	110	51	53	86	90
Norway	119	35	466	26%	8%	119	11	117	34	116	33	116	33	282	84	119	35	219	65
Sweden	208	59	813	26%	7%	208	16	209	59	209	59	208	59	494	140	208	59	383	109
Total	784	306	3,305	24%	9%	784	136	780	301	777	297	773	294	1,858	680	784	306	1,440	539
Comp. base case							*45%*		*98%*		*97%*		*96%*		*223%*		*100%*		*176%*
West																			
Austria	232	284	890	26%	32%	232	201	216	266	200	250	185	234	232	284	115	140	191	181
Wallonia (B)	167	154	553	30%	28%	167	103	147	135	129	118	137	107	167	154	80	74	136	125

Brussel (B)	31	27	100	31%	27%	31	17	28	24	23	21	22	19	31	27	15	13	25	22
Vlaanderen (B)	493	221	1,229	40%	18%	493	107	472	212	454	203	438	195	493	221	218	98	389	174
France	3,059	1,645	8,102	38%	20%	3,059	893	3,002	1,600	2,711	1,511	2,665	1,471	3,059	1,645	1,380	742	2,434	1,308
Germany	3,663	2,868	11,238	33%	26%	3,663	1,825	3,604	2,827	3,562	2,790	3,523	2,755	3,663	2,868	1,725	1,350	2,960	2,318
Ireland	166	125	501	33%	25%	166	79	164	124	163	122	161	121	166	125	78	59	134	101
Luxembourg	16	10	45	36%	22%	16	6	16	10	16	10	16	10	16	10	7	5	13	8
the Netherlands	1,436	338	3,064	47%	11%	1,436	53	1,424	335	1,411	331	1,400	328	1,436	338	592	139	1,104	259
Switzerland	104	313	720	15%	44%	104	247	104	296	99	281	95	267	104	313	55	166	88	264
Scotland (UK)	250	153	694	36%	22%	249	89	250	138	250	122	250	107	250	153	114	70	200	122
N, Ireland (UK)	119	28	252	47%	11%	119	3	119	25	119	22	119	19	119	28	49	11	91	21
Wales (UK)	211	63	475	44%	13%	211	19	211	57	211	51	211	44	211	63	89	27	164	49
England (UK)	3,198	1,060	7,313	44%	15%	3,198	339	3,198	954	3,198	848	3,198	742	3,198	1,060	1,363	452	2,490	826
Total	13,147	7,289	35,178	37%	21%	13,146	3,981	12,954	7,003	12,545	6,682	12,421	6,420	13,147	7,289	5,880	3,345	10,419	5,779
Comp. base case							*55%*		*96%*		*92%*		*88%*		*100%*		*46%*		*79%*
East																			
Bulgaria	231	240	942	24%	26%	231	160	201	205	173	177	193	158	282	288	137	140	231	235
Croatia	175	177	708	25%	25%	175	120	172	172	166	166	162	161	215	217	104	105	175	177
Czech Republic	358	230	1,181	30%	20%	358	136	358	229	355	227	353	226	446	287	206	132	358	230
Hungary	318	439	1,515	21%	29%	318	318	307	416	304	395	301	374	385	532	193	266	318	439
Poland	1,418	992	4,839	29%	21%	1,418	605	1,436	962	1,370	915	1,309	870	1,761	1,232	820	574	1,418	992
Romania	605	630	2,472	24%	26%	605	420	650	566	543	482	448	405	741	756	360	367	605	618
Slovakia	176	183	718	24%	26%	176	194	96	201	83	175	70	150	114	263	60	137	96	220
Slovenia	64	57	241	27%	24%	64	14	74	56	71	54	69	52	79	70	38	33	64	57
Total	3,345	2,949	12,616	27%	23%	3,345	1,968	3,293	2,807	3,065	2,592	2,905	2,397	4,023	3,645	1,917	1,755	3,264	2,969
Comp. base case							*67%*		*95%*		*88%*		*81%*		*124%*		*60%*		*101%*
South																			
Cyprus	29	14	87	33%	17%	29	9	27	15	25	14	25	13	37	20	16	9	29	16
Greece	433	176	1,257	34%	14%	433	75	387	153	328	129	274	108	549	223	243	99	433	176
Italy	1,724	1,097	5,624	31%	20%	1,724	647	1,641	1,047	1,574	1,002	1,511	958	2,152	1,369	989	629	1,724	1,097
Malta	23	9	63	36%	14%	23	10	22	8	21	8	20	8	29	11	13	5	23	9
Portugal	377	194	1,139	33%	17%	377	103	312	173	293	161	275	150	475	244	213	109	377	194
Spain	1,818	402	4,462	41%	9%	1,818	45	1,239	342	1,205	331	1,171	320	2,370	523	973	215	1,818	402
	4,404	1,891	12,632	35%	15%	4,404	888	3,629	1,738	3,445	1,645	3,276	1,556	5,611	2,391	2,446	1,066	4,404	1,893
							47%		*92%*		*87%*		*82%*		*126%*		*56%*		*100%*
ALL	**21,680**	**12,434**	**63,731**	34%	20%	21,680	**6,973**	20,657	**11,849**	19,832	**11,215**	19,375	**10,667**	24,639	**14,005**	11,028	**6,472**	19,528	**11,180**
Comp. base case						*100%*	*56%*	*95%*	*95%*	*91%*	*90%*	*89%*	*86%*	*114%*	*113%*	*51%*	*52%*	*90%*	*90%*

BC = Breast cancer. [1] Effectiveness of opportunistic screening to lower cancer specific mortality was set to be 10%, 20%, and 30% lower than organised screening. In these analyses the gained percentages of screening coverage (up to 100%) were distributed over organised and opportunistic screening to the same distribution as was already present in the specific country (e.g. if present screening coverage was 40% organised and 20% opportunistic (ratio 2:1), the additional coverage was 27% organised and 13% opportunistic (2:1)). [2] Application of each of the regional point estimates across all European countries, i.e. we applied a 58% (West), a 33% (North), and a 50% (South) breast cancer mortality reduction due to screening irrespective of the location of the country.

Figure 3 presents the relative effect of a 100% total examination coverage for each country, i.e. showing the share of breast cancer deaths that could additionally be prevented when countries would screen all women 50 to 69 years of age every two years. Most countries could potentially avert additional 20%-29% of their breast cancer deaths. In contrast, all Nordic countries have consistently high coverage rates through their organised programmes and less additional breast cancer deaths could potentially be prevented when screening would be extended to 100%.

Sensitivity analyses

As shown in Table 3, assuming a maximal coverage of 84% instead of 100% led to a significant drop in prevented breast cancer deaths (6,975 averted deaths compared to 12,438). This cut is predominantly explained by countries who already have a comparably high screening coverage and lose the additional benefit of increasing up to 100% (e.g. the Netherlands, Spain or Denmark).

Assuming that opportunistic screening is 10% less effective as organised screening led to a 5% reduction of the additionally preventable breast cancer deaths. A 20% and 30% lowered effectiveness led to a 10% and 14% reduction, respectively. The effect was biggest in countries with a high percentage of opportunistic screening (e.g. Wallonia/Belgium).

Applying the Western European point estimate for mortality reduction across all of Europe, breast cancer deaths already prevented increased by 14% and breast cancer deaths that can additionally be prevented increased by 13%. This analysis has the biggest impact for Northern Europe (plus 223%), where the point estimate was the smallest in the base analysis. When the estimates from Northern and Southern Europe were applied, the number of breast cancer deaths prevented decreased by 49% and 10%, while the additionally preventable breast cancer deaths decreased by 48% and 10%, respectively, compared to the base calculation.

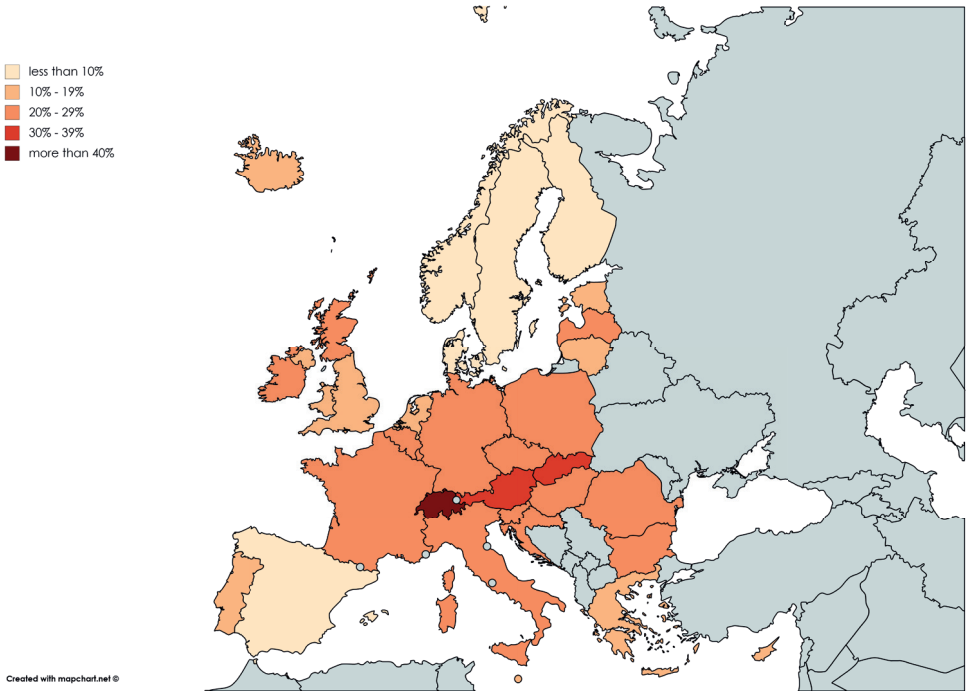


Figure 3: Percentage of breast cancer deaths that could be additionally be prevented if examination coverage would increase to 100%, per European country*

*Belgium is depicted as one country whereas in the calculation three highly autonomous regions Flanders, Wallonia and Brussels are included. These regions have very disparate screening programs for breast cancer (see Table 2) resulting in very different effects of an increased total examination coverage (Table 3).
Only 8 of the 26 Swiss cantons have organised breast cancer screening programmes which causes substantial variation in the distribution of organised versus opportunistic screening across regions. On a national level, total examination coverage was only 25% in 2015 (14% organized and 11% opportunistic) according to the national expert. Thus, a national examination coverage of 100% would further reduce breast cancer deaths by 44%.

DISCUSSION

This study illustrates how breast cancer screening in Europe already has a substantial impact by preventing nearly 21,700 breast cancer deaths per year. In addition, through further optimising screening coverage, the number of breast cancer deaths of European women could be further reduced significantly. The effect would be particularly notable in Eastern and Western Europe. Thus, rolling-out a breast cancer screening programme with complete coverage across the country is particularly favourable for Swiss women as it would further reduce breast cancer deaths by 44%. In contrast, all Nordic countries have consistently high coverage rates through their organised programmes (between 72% and 81%) plus a very low coverage of opportunistic screening for breast cancer (between 1% and 5%). When the total examination coverage for women aged 50-69 is already as high as 84%, not many additional breast cancer deaths could potentially be prevented if screening was extended to 100%.

Screening provides both harms and benefits, and therefore it is important to ensure a good balance between the two. Information on the balances of benefits and harms is needed to demonstrate that a chosen screening policy and programme with all its components and protocols is appropriate for any given country. In this paper, however, we focus solely on the primary aim of (organised) breast screening which is to reduce mortality from breast cancer through early detection[16 20].

The calculations for this present analysis are based on the assumption that opportunistic and organised breast cancer screening can lead to the same level of cancer specific mortality reduction. However, past studies resulted in slightly conflictive results. For example, a study in Denmark found that the sensitivity was twice as high for organised screening, while the specificity of organised and opportunistic screening was found to be similar[24]. Hofvind et al. compared opportunistic breast cancer screening in Vermont (USA) with organised breast cancer screening in Norway[25]. Both screening systems detected cancer at about the same rate and at the same prognostic stage. A study from Switzerland found that there was little difference in stage distribution and detection rates between cantons with only opportunistic screening and cantons with both organised and opportunistic screening[26], indicating that both are similarly effective. It was noted however that the quality of opportunistic screening in Switzerland probably benefitted from the training of radiographers, a higher reading volume of radiologists and the technical and quality-controlled procedures of the organised programme.
In summary, the main differences between organised and opportunistic screening can be seen in attendance[27], equity[27], and cost-effectiveness[28] which are all (much) better in organised screening. With regards to quality aspects, opportunistic screening might be quite similar to that of organised screening. Moreover, since opportunistic screening takes place next to organised screening in most countries (Bulgaria, Romania, Slovakia and Greece being the exception), it can profit from advantages of the organised system. Consequently, we are confident that by conflating opportunistic and organised screening for calculations and argumentations, we can increase the relevance of this paper.

The European guidelines for quality assurance in breast cancer screening and diagnosis consider participation rates above 70% as acceptable and above 75% as desirable[29]. In line with those guidelines, we do not actually propagate a screening coverage of 100% as this probably conflicts with informed choice[30]. However, by basing our calculations on a hypothetical goal of a screening coverage of 100% of eligible women, we assessed the maximum potential of breast cancer screening for each country.

This study focuses on screening women ages 50-69 as this is currently the practice in most European countries. Despite some exceptions (Table 2), women aged 70–74 are usually not eligible for mammography screening because there was insufficient evidence that screening would reduce mortality for women in this age group. Previous randomized controlled trials (RCTs) and observational studies on breast cancer screening have not generally included women aged 70 years and over. In their newest screening (conditional) recommendations, however, the European Commission Initiative on Breast Cancer suggests that average-risk and asymptomatic women between 45 and 49 as well as between 70 and 74 years old, have mammography screening for breast cancer.

Several further considerations inform the interpretation of our study. There is an ongoing debate as to which study design is the gold standard for estimating the true effect of screening on cancer specific mortality[23 31 32]. For this study, we considered that high-quality case control studies[7] provide the most informative data. RCTs were conducted more than 20 years ago when adherence to screening was less and the quality of screening programmes and breast cancer care were less advanced than today. In contrast, observational studies of screening are known to be prone to bias as there is no unselected unscreened group. Women who do not participate in screening might have a higher a priori risk of breast cancer mortality. If that was so, our assumption of a proportional relationship between screening coverage and reduction in breast cancer mortality would not hold. Therefore, it was of particular importance to base our analysis on estimates of mortality reduction that were not influenced by self-selection bias.

The regional point estimates from individual studies on mortality reduction due to breast cancer screening, which our calculations are based on, differ quite significantly. These differences indicate differences in evaluation designs, in target ages, in ages of follow-up of breast cancer incidence or mortality, in duration of follow-up since first invitation, in comparison groups, and in assessment methods of self-selection bias[9 12 33 7]. Therefore, the region-specific point estimates are not directly comparable with each other and they should not be used as a "quality indicator" for organized breast cancer screening in each region. Despite the different effect sizes, we are confident that our three regional estimates do not present an overestimation of the benefit of mammographic screening. They are well in the range of an analysis of Broeders at al. from 2012[5] who present a pooled breast cancer mortality reduction for women who actually participated in screening of 38% based on incidence based mortality studies (OR = 0.62 [0.56-0.69]) and 48% based on case control studies (OR = 0.52 [0.42-0.65], adjusted for self-selection). An analysis similar to this study has been published in 2013. Mackenbach and McKee[34] estimated there would be over 17,000 fewer breast cancer deaths each year if all countries in the EU could reduce death rates to those in the best performing country, Sweden. However, this study was based on cause- and age-specific death rates only rather than the combination of cause- and age-specific mortality and the extent of screening activity.

To our knowledge, there have been no other studies so far that have estimated the effect of breast cancer screening on cancer specific mortality when brought to its full potential based on the total extent of breast cancer screening activities in Europe. We were able to provide an extensive overview of the amount of organised as well as opportunistic screening in Europe by consulting national experts. Accordingly, some of the national estimates on screening uptake have never been published before. However, our study also has some potential limitations. The first limitation is the uncertainty regarding the coverage of opportunistic screening as these numbers are based on expert opinion or on national extrapolations of regional observations. Secondly, because the organised breast cancer screening in the UK as well as Malta is triennially rather than every two years, this led to a slight overestimation of the breast cancer death prevented. Third, our calculations probably led to an underestimation of the already prevented and additionally preventable deaths for the few countries which invite and screen women that are younger than 50 or older than 69. The fourth limitation is the fact that the number of breast cancer deaths and the estimates of examination coverage come from the same report year although the most recent breast cancer deaths rather reflect the past (e.g. 5-10 years ago) than current screening practice.

Our analysis paves the way for further research as it could potentially be applied to the other two cancer sites for which the European Council recommends screening: cervical and colorectal cancer.

This study illustrates that by further optimising screening coverage, the number of breast cancer deaths in Europe could be lowered substantially. Therefore, countries which do not yet offer organized screening for the target age range of 50 to 69 should strongly consider it based on our results. In addition, even when programmes to screen for breast cancer exist, much is still to be done. This includes increasing screening coverage through evidence-based interventions[35 36] and removing barriers to effective breast cancer screening[37 38].

Acknowledgments

We like to thank all data providers and EU-TOPIA workshop participants for their valuable help, expert inputs and critical feedback in the making of this study.

Funding

This study is part of the EU-TOPIA project, funded by the EU-Framework Programme (Horizon 2020) of the European Commission, project reference 634753. The authors alone are responsible for the views expressed in this manuscript.

Collaborators

Data providers[1] and EU-TOPIA consortium members (or both[2])

Austria	Gerald Gredinger[1]
Belgium - Wallonia	Michel Candeur[1]
Belgium - Brussel	Marc Arbyn[1], Cindy Simoens[1]
Belgium - Vlaanderen	Isabel de Brabander[1]
Bulgaria	Plamen Dimitrov[1], Zdravka Valerianova[1]
Croatia	Andrea Supe[1]
Czech Republic	Ondřej Ngo[1], Ondřej Májek[1]
Denmark	Elisabeth Lynge[1]
Estonia	Piret Veerus[2]
Finland	Sirpa Heinävaara[2], Ahti Anttila, Tytti Sarkeala
France	Agnes Rogel[1]
Germany	Vanessa Kääb-Sanyal[1], Klaus Kraywinkel[1]
Hungary	Marcell Csanadi[2], György Széles, Zoltan Voko
Italy	Carlo Senore[2], Nereo Segnan
Iceland	Rún Friðriksdóttir[1]
Ireland	Patricia Fitzpatrick[1]
Latvia	Inga Brokere[1]
Lithuania	Jurgita Grigariene[1]
Luxembourg	Diane Pivot[1]
Malta	Stephanie Xuereb[1]
The Netherlands	Linda de Munck[1], Inge de Kok, Andrea Gini, Eveline Heijnsdijk, Erik Jansen, Harry de Koning, Iris Lansdorp – Vogelaar, Nicolien van Ravesteyn
Norway	Solveig Hofvind[1]
Poland	Anna Macios[1]
Spain	Nieves Ascunce Elizaga[1]
Slovakia	Soňa Senderáková[1]
Slovenia	Katja Jarm[2], Urska Ivanus, Dominika Novak Mlakar
Sweden	Lennarth Nyström[1]
Switzerland	Jean-Luc Bulliard[1]
UK - Scotland	John Quinn[1]
UK - Northern Ireland	Jeni Rosborough[1]
UK - Wales	Ardiana Gjini[1]
UK - England	Radoslav Latinovic[1], Martin McKee

REFERENCES

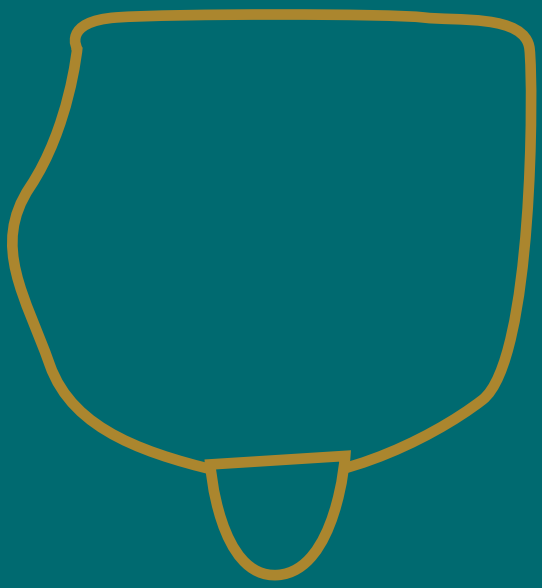
1. ECIS - European Cancer Information System: European Commission; [Available from: https://ecis.jrc.ec.europa.eu/.
2. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. Eur J Cancer 2018;103:356-87.
3. IARC. Cancer Screening in the European Union. Report on the implementation of the Council Recommendation on cancer screening. IARC, Lyon 2017;Available at: 2017_cancerscreening_2ndreportimplementation_en.pdf.
4. Nelson HD, Fu R, Cantor A, et al. Effectiveness of Breast Cancer Screening: Systematic Review and Meta-analysis to Update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med 2016;164(4):244-55.
5. Broeders M, Moss S, Nyström L, 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.
6. Moss SM, Nyström L, Jonsson H, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of trend studies. J Med Screen 2012;19 Suppl 1:26-32.
7. Zielonke N, Gini A, Jansen E, et al. Evidence for reducing cancer specific mortality due to screening for breast cancer in Europe: a systematic review. Eur J Cancer 2020;127:191-206.
8. Weedon-Fekjræ H, Romundstad PR, Vatten LJ. Modern mammography screening and breast cancer mortality: Population study. BMJ (Online) 2014;348.
9. Heinavaara S, Sarkeala T, Anttila A. Impact of organised mammography screening on breast cancer mortality in a case-control and cohort study. Br J Cancer 2016.
10. Olsen AH, Lynge E, Njor SH, et al. Breast cancer mortality in Norway after the introduction of mammography screening. Int J Cancer 2013;132(1):208-14.
11. Puliti D, Miccinesi G, Zappa M, et al. Balancing harms and benefits of service mammography screening programs: A cohort study. Breast Cancer Res 2012;14(1).
12. Paap E, Verbeek ALM, Botterweck AAM, et al. Breast cancer screening halves the risk of breast cancer death: A case-referent study. Breast 2014;23(4):439-44.
13. Massat NJ, Dibden A, Parmar D, et al. Impact of screening on breast cancer mortality: The UK program 20 years on. Cancer Epidemiol Biomarkers Prev 2016;25(3):455-62.
14. Johns LE, Coleman DA, Swerdlow AJ, et al. Effect of population breast screening on breast cancer mortality up to 2005 in England and Wales: An individual-level cohort study. Br J Cancer 2017;116(2):246-52.
15. Tabár L, Vitak B, Chen THH, et al. Swedish two-county trial: Impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011;260(3):658-63.
16. Council of the European Union. Council Recommendation of 2 December 2003 on Cancer Screening (2003/878/EC). OJ L 327: 34-38. Brussels, 2003.
17. European Commission Initiative on Breast Cancer. Recommendations from the European Breast Guidlines. 2020;Available at: https://ecibc.jrc.ec.europa.eu/recommendations/.

18. Basu P, Ponti A, Anttila A, et al. Status of implementation and organization of cancer screening in The European Union Member States-Summary results from the second European screening report. Int J Cancer 2018;142(1):44-56.
19. Chen TH, Yen AM, Fann JC, et al. Clarifying the debate on population-based screening for breast cancer with mammography: A systematic review of randomized controlled trials on mammography with Bayesian meta-analysis and causal model. Medicine (Baltimore) 2017;96(3):e5684.
20. Marmot MG, Altman DG, Cameron DA, et al. The benefits and harms of breast cancer screening: an independent review. Br J Cancer 2013;108(11):2205-40.
21. WHO. Cancer Mortality Database 2019;Available at: http://www-dep.iarc.fr/WHOdb/WHOdb.htm.
22. The European Commission. ECIS - European Cancer Information System. 2019;Available at: https://ecis.jrc.ec.europa.eu/.
23. IARC Working Group. IARC Handbooks of Cancer Prevention, Volume 15: Breast Cancer Screening. Lyon, 2016.
24. Bihrmann K, Jensen A, Olsen AH, et al. Performance of systematic and non-systematic ('opportunistic') screening mammography: a comparative study from Denmark. J Med Screen 2008;15(1):23-6.
25. Hofvind S, Vacek PM, Skelly J, et al. Comparing screening mammography for early breast cancer detection in Vermont and Norway. J Natl Cancer Inst 2008;100(15):1082-91.
26. Bulliard JL, Ducros C, Jemelin C, et al. Effectiveness of organised versus opportunistic mammography screening. Ann Oncol 2009;20(7):1199-202.
27. Espinas JA, Aliste L, Fernandez E, et al. Narrowing the equity gap: the impact of organized versus opportunistic cancer screening in Catalonia (Spain). J Med Screen 2011;18(2):87-90.
28. de Gelder R, Bulliard JL, de Wolf C, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. Eur J Cancer 2009;45(1):127-38.
29. Perry N, Broeders M, de Wolf C, et al. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition—summary document. Annals of Oncology 2008;19(4):614–22.
30. van Agt H, Fracheboud J, van der Steen A, et al. Do women make an informed choice about participating in breast cancer screening? A survey among women invited for a first mammography screening examination. Patient Educ Couns 2012;89(2):353-9.
31. Autier P, Boniol M. Breast cancer screening: evidence of benefit depends on the method used. BMC Med 2012;10:163.
32. Lauby-Secretan B, Scoccianti C, Loomis D, et al. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med 2015;372(24):2353-8.
33. Puliti D, Miccinesi G, Collina N, et al. Effectiveness of service screening: A case-control study to assess breast cancer mortality reduction. Br J Cancer 2008;99(3):423-27.
34. Mackenbach JP, Karanikolos M, McKee M. The unequal health of Europeans: successes and failures of policies. Lancet 2013;381(9872):1125-34.
35. Jepson R, Clegg A, Forbes C, et al. The determinants of screening uptake and interventions for increasing uptake: a systematic review. Health Technol Assess 2000;4(14):i-vii, 1-133.

36. Camilloni L, Ferroni E, Cendales BJ, et al. Methods to increase participation in organised screening programs: a systematic review. BMC Public Health 2013;13:464.
37. Turnbull E, Priaulx J, van Ravesteyn NT, et al. A health systems approach to identifying barriers to breast cancer screening programmes. Methodology and application in six European countries. Health Policy 2018;122(11):1198-205.
38. Priaulx J, de Koning HJ, de Kok I, et al. Identifying the barriers to effective breast, cervical and colorectal cancer screening in thirty one European countries using the Barriers to Effective Screening Tool (BEST). Health Policy 2018;122(11):1190-97.

Part 2:

MODELLING THE IMPACT OF DIFFERENT INTERVENTIONS ON BREAST CANCER SCREENING EFFECTIVENESS



Chapter 5

Extending age ranges in breast cancer screening in four European countries: Model estimations of harm-to-benefit-ratios

Nadine Zielonke, H. Amarens Geuzinge, Eveline A.M. Heijnsdijk, Sirpa Heinävaara, Carlo Senore, Katja Jarm, Harry J. de Koning, Nicolien T. van Ravesteyn, on behalf of the EU-TOPIA consortium.

Cancers 2021, 13(13), 3360;

ABSTRACT

Background

The main benefit of breast cancer (BC) screening is a reduction in mortality from BC. However, screening also causes harms such as overdiagnosis and false-positive results. The balance between benefits and harms varies by age. This study aims to assess how harm-to-benefit ratios of BC screening vary by age in the Netherlands, Finland, Italy and Slovenia.

Methods

Using microsimulation models, we simulated biennial screening with 100% attendance at varying ages for cohorts of women followed over a lifetime. The number of overdiagnoses, false-positive diagnoses, BC deaths averted and life-years gained (LYG) were calculated per 1,000 women. We compared four strategies (50–69, 45–69, 45–74 and 50–74) by calculating four harm-to-benefit ratios, respectively.

Results

Compared to the reference strategy 50–69, screening women at 45–74 or 50–74 years would be less beneficial in any of the four countries than screening women at 45–69, which would result in relatively fewer overdiagnoses per death averted or LYG. At the same time, false-positive results per death averted would increase substantially.

Conclusion

Adapting the age range of BC screening is an option to improve harm-to-benefit ratios in all four countries. Prioritization of considered harms and benefits affects the interpretation of results.

INTRODUCTION

The main benefit of breast cancer screening is a reduction in breast cancer mortality through early detection[1–6]. However, screening also causes harm. Important harms associated with breast cancer screening are overdiagnosis and false-positive results[5].

Based on evidence regarding the harms and benefits, the European Commission's Initiative on Breast Cancer Guidelines Development Group (GDG) strongly recommends inviting women ages 50–69 to mammography screening every two years[7]. Therefore, most European countries adopted biennial screening for breast cancer in this age range[8,9]. Updated evidence on efficacy resulted in extended (conditional) recommendations to triennial or biennial screening for age groups 45–49 and 70–74 in an organized screening programme[7].

Several factors influence the balance between benefits and harms of screening women younger than 50 and older than 69 years. The most important is that breast cancer incidence increases with age10 11. Furthermore, the sensitivity of mammography decreases with increasing breast density. Younger women have higher breast density, with lower test sensitivity and more false-positive results12–14. These two factors might result in smaller benefits and more harms of screening. In contrast, the benefits of screening women ages 70–74 might be limited due to the higher death rate from competing causes with advancing age, thus fewer life-years gained (LYG) and increases in overdiagnosis.

Unfortunately, there are only a few screening programmes that have accomplished long-term evaluations on the balance between harms and benefits[8]. Often only short-term indicators for benefits and harms are available. Despite several previous studies which assessed the harm-to-benefit-ratios of existing programs for breast cancer[12,15,16], there is no published analysis of the relationship between harms and benefits for varying age ranges and countries.

Therefore, the aim of this study is to assess harm-to-benefit ratios of breast cancer screening vary by age in four European countries. To this end, we calibrated and validated a microsimulation model for each of the four exemplary countries. This study was conducted within the scope of EU-TOPIA. In this project, one exemplary country with high-quality observational data was selected to be representative for each European region (the Netherlands for Western Europe, Finland for Northern Europe, Slovenia for Eastern Europe and Italy for Southern Europe). Using these country-specific models, we estimated the harms and benefits of various screening age ranges.

MATERIALS AND METHODS

Model Overview

The effects of screening for varying age groups were assessed using the Microsimulation Screening Analysis (MISCAN) model[17]. MISCAN simulates individual life histories and assesses the consequences of introducing a screening program on these life histories using the Monte Carlo method. Possible events in the life histories are birth and death of a person, onset of a pre-clinical ductal carcinoma in situ (DCIS), transitions between disease states, participation in screening and screen- or clinical detection of a cancer. (see Supplementary materials for more information on the MISCAN-Breast structure and underlying assumptions).

For each of the four countries, we adjusted and calibrated the MISCAN model to reflect differences in population demography (i.e., age distribution of the population and life expectancy), disease risk (i.e., breast cancer incidence and stage distribution) and potential differences in the natural history of breast cancer. In developing each model, we used a specific calibration process (Supplementary materials, chapter 6). The model optimized a set of unobservable parameters (e.g., stage-specific sensitivity) to match observed data (e.g., detection rates). Thus, we first validated the model versions replicating the data that were used in the calibration process (internal validation). Then, we externally validated the models against best evidence based on a recently published systematic review on breast cancer mortality reductions due to screening[4] (Supplementary materials, chapter 7).

Analysis

For each country, we simulated a cohort of 10 million women born in 1975 and followed all women from age 45 until death. First, we simulated the reference screening strategy with biennial screenings from age 50 to 69 years, assuming 100% examination coverage. We assumed 100% to achieve harm and benefit predictions of the tested screening strategies unaffected by external behavioural factors. We then determined the harms and benefits in comparison to no screening. Next, we determined the incremental harms and benefits of extending biennial breast cancer screening to start at age 45 and to stop at age 74.

Outcomes

Benefits were expressed as breast cancer deaths averted and LYG. Harms were ex-pressed as false positives and overdiagnoses, calculated as the difference in the number of diagnosed breast cancers in the presence of screening and in the absence of screening, using lifelong follow-up.

For each screening strategy, we determined the following harm-to-benefit ratios by dividing the harms by the benefits:

- Overdiagnosed breast cancer cases/averted breast cancer deaths;
- False-positive results/averted breast cancer deaths;
- Overdiagnosed breast cancer cases/LYG;
- False-positive results/LYG.

Compared to the reference strategy, an alternative screening strategy could be considered more optimal if one or more harm-to-benefit ratio is smaller.

Sensitivity Analysis

To evaluate how assumptions and parameter values influence the harm-to-benefit ratios and whether the relative differences between strategies change, we performed several sensitivity analyses. First, we assessed the influence of country-specific calibrated values for stage-specific sensitivity by using the highest and the lowest sensitivities and applied them across all countries. Second, we considered the highest and lowest observed referral rates and applied them across all countries. Third, we used observed examination coverage (Table 1) instead of 100%.

RESULTS

Model Calibration and Validation

The calibrated models for Slovenia, Finland, the Netherlands and Italy reproduced the country-specific trends in breast cancer incidence and mortality quite well (Supplementary materials, chapter 6, Table S2 and Figures S4–S11), that is, the simulated model predictions were mostly within the 95% confidence intervals of the corresponding observed outcomes. Subsequently, we validated our model predictions against observed breast cancer mortality reductions due to mammography screening in the Netherlands, Finland and Italy from a systematic (Supplementary materials, Table S3-S4). Due to a lack of studies from Eastern Europe, we validated the Slovenian model by comparing the modelled and observed interval cancer rates (Supplementary materials, Table S5-S8).

Outcomes of Different Screening Strategies

If 1,000 women underwent biennial mammography between the ages of 50 and 69 (10 screening rounds) and were followed over their lifetimes, the models predicted that around 9,000 screening tests would be performed. Compared to a situation without screening, 7 breast cancer deaths would be averted in Slovenia, 8 in Finland, 13 in the Netherlands and 11 in Italy (Table 2). These differences are largely driven by the differences in background incidence rates (chapter 6,

Table 1. Input values for the parametric sensitivity analysis, per country

	Slovenia	Finland	Netherlands	Italy
Examination coverage by per age-group [1]				
45-49	54.3%[2]	85.0%[2]	75.5%[2]	59.6%[2]
50-54	54.3%	85.0%	75.5%	59.6%
55-59	65.0%	85.9%	76.2%	63.2%
60-64	52.4%	86.8%	76.3%	63.9%
65-69	48.8%	73.0%	75.7%	61.5%
70-74	48.8%[2]	73.0%[2]	70.1%	61.5%[2]
Stage-specific sensitivity of digital mammography DCIS	0.726	0.596[3]	0.865[4]	0.821
Stage-specific sensitivity of digital mammography T1a	0.785	0.811	0.553[3]	1[4]
Stage-specific sensitivity of digital mammography T1b	0.656	0.761[4]	0.481[3]	0.717
Stage-specific sensitivity of digital mammography T1c	0.780[3]	0.946[4]	0.857	0.814
Stage-specific sensitivity of digital mammography T2+	1	1	1	1
Referral rate by age[5]				
< 50	0.040	0.030	0.030[3]	0.065[4]
> 50	0.034	0.028	0.023[3]	0.058[4]

[1] The examination coverage of (organized) screening is specified as the proportion (%) of the target population per age-group screened in the chosen report year after invitation. These observed parameters stem from the following years: Finland 2014, Netherlands & Italy 2015, Slovenia 2016. [2] For those countries that screen women within the age-range 50-69, we assume the same examination coverage for the age-group 45-49 and 70-74 as the nearest age-group which we have observed data for. [3] This country has the lowest calibrated sensitivity/observed referral rate for the respective cancer stages. [4] This country has the highest calibrated sensitivity/observed referral rate for the respective cancer stage. [5] The referral rate represents the percentage of participants with abnormal screening results who are referred for further diagnostic testing. This rate depends on the screening protocol adopted for referring women to assessment (i.e. positivity criteria, double vs. single reading), previous opportunistic screening, as well as the quality of screening tests.

Supplementary methods). The models also predicted that there would be 3 (range 2.5–3.3 across countries) overdiagnosed breast cancer cases per 1,000 women when screening between ages 50–69 (Table 2). The overdiagnosed breast cancer cases/breast cancer deaths averted ratio is estimated to range between 0.2 (Italy) and 0.5 (Slovenia). The false-positives/breast cancer deaths averted is estimated to range between 11.6 (the Netherlands) and 45.7 (Italy). Hence, 0.2–0.5 women would be overdiagnosed and 12–46 women would be confronted with a false-positive finding for every woman prevented from dying from breast cancer.

Table 2. (Incremental) Screening outcomes per country and screening strategy.

			Harms		Benefits		Harm-to-Benefit-ratios			
Country	Strategy[1]	Number of screening tests	Overdiagnosed BC cases	False Positives	BC deaths averted	LY gained	Overdiagnosed BC cases / BC deaths averted	False Positives / BC deaths averted	Overdiagnosed BC cases / LY gained	False Positives / LY gained
Slovenia	**50-69***	9,236	3.3	275.8	7.3	96.5	0.5	37.9	0.034	2.9
	45-74	13,723	+1.8	+220.8	+2.6	+32.7	0.5	50.3	0.040	3.8
	45-69	11,696	+0.1	+150.2	+0.8	+18.2	0.4	53.0	0.030	3.7
	50-74	11,264	+1.7	+58.4	+1.9	+14.8	0.5	36.6	0.045	3.0
Finland	**50-69***	9,170	2.6	212.3	7.7	105.3	0.3	27.6	0.025	2.0
	45-74	13,632	+1.5	+135.6	+3.2	+38.8	0.4	31.9	0.029	2.4
	45-69	12,034	+0.4	+96.7	+1.4	+24.2	0.3	34.2	0.023	2.4
	50-74	11,183	+1.4	+48.8	+2.4	+19.4	0.4	25.9	0.032	2.1
Netherlands	**50-69**	8,948	3.2	150.1	13.0	185.6	0.2	11.6	0.017	0.8
	45-74	13,288	+1.9	+172.5	+4.2	+59.5	0.3	18.8	0.021	1.3
	45-69	11,388	+0.2	+129.7	+1.8	+40	0.2	19.0	0.015	1.2
	50-74*	10,848	+1.7	+29.5	+2.5	+19.6	0.3	11.6	0.024	0.9
Italy	**50-69***	9,186	2.5	488.5	10.7	152.1	0.2	45.7	0.016	3.2
	45-74	13,657	+1.5	+338.8	+3.5	+49.2	0.3	58.2	0.020	4.1
	45-69	11,641	+0.1	+219.1	+1.4	+32.0	0.2	58.7	0.014	3.8
	50-74	11,203	+1.4	+105.5	+2.1	+17.2	0.3	46.3	0.023	3.5

Model projections for 2020-2075. Screening outcomes are presented per 1,000 women, aged 45 years followed over their lifetime. [1] Each strategy is compared to no screening. * Current screening strategy.

BC: breast cancer; LY: Life years. Number of screening rounds per strategy: 50-69: 10; 45-74: 15; 45-69: 12.5; 50-74: 12.5. We assumed 100% adherence to screening strategies including follow-up.

In all countries, adding screening below the age of 50 or after the age of 69 resulted in more life-years gained and more breast cancer deaths averted, but at the expense of increases in harms. For example, screening 1,000 women aged 50–74 in Finland is expected to avert 2.4 additional breast cancer deaths, but it would also yield 1.4 additional overdiagnosed cases (Table 2).

In all countries, the false-positive-related ratios are larger for the younger age ranges and smaller for the older ones compared to reference strategy 50–69. In contrast, the over-diagnosis-related ratios are larger for the older age ranges and tend to be smaller for the strategies where women are screened below the age of 50 (Table 2).

The percentage change in the harm-to-benefit ratios in comparison to the reference strategy is presented in Figure 1. In all countries, screening women between ages 45–69 would result in smaller overdiagnosis-related ratios. This is particularly pronounced for the ratio of overdiagnosed breast cancer cases to life-years gained. This ratio is 11% (Finland) to 13% (Italy) smaller for the strategy 45–69 than for the reference strategy. On the other hand, the false-positive-related harm-to-benefit ratios for adding screening before the age of 50 or after the age of 69 are less favourable than for screening women between ages 50 and 69.
Of the three alternative strategies, 45–74 is the least optimal age range for screening women in Slovenia, the Netherlands and Italy, as it would lead to an increase in all ratios. In Finland, the least optimal strategy for screening women appears to be 50–75, where the overdiagnosis-related ratios would result in substantial increases (51% and 67%, respectively, Figure 1).

Sensitivity Analysis

The overdiagnosis-related ratios were relatively insensitive to changing screening test characteristics (Supplementary materials, Table S9). However, the false-positive-related ratios were strongly affected by referral rates, leading to an average 14% reduction when applying the lowest age-specific referral rates vs. a two-fold increase when applying the highest age-specific referral rates across all countries. Applying the observed coverage in-stead of 100% increased the overdiagnosis-related ratios on average by 3% and diminished the false-positive-related ratios by 15%. Varying the values of our input parameters did not affect the magnitude of change of each of the harm-to-benefit ratios when com-pared to the reference strategy of ages 50–69 (Supplementary materials Figures S12–S16).

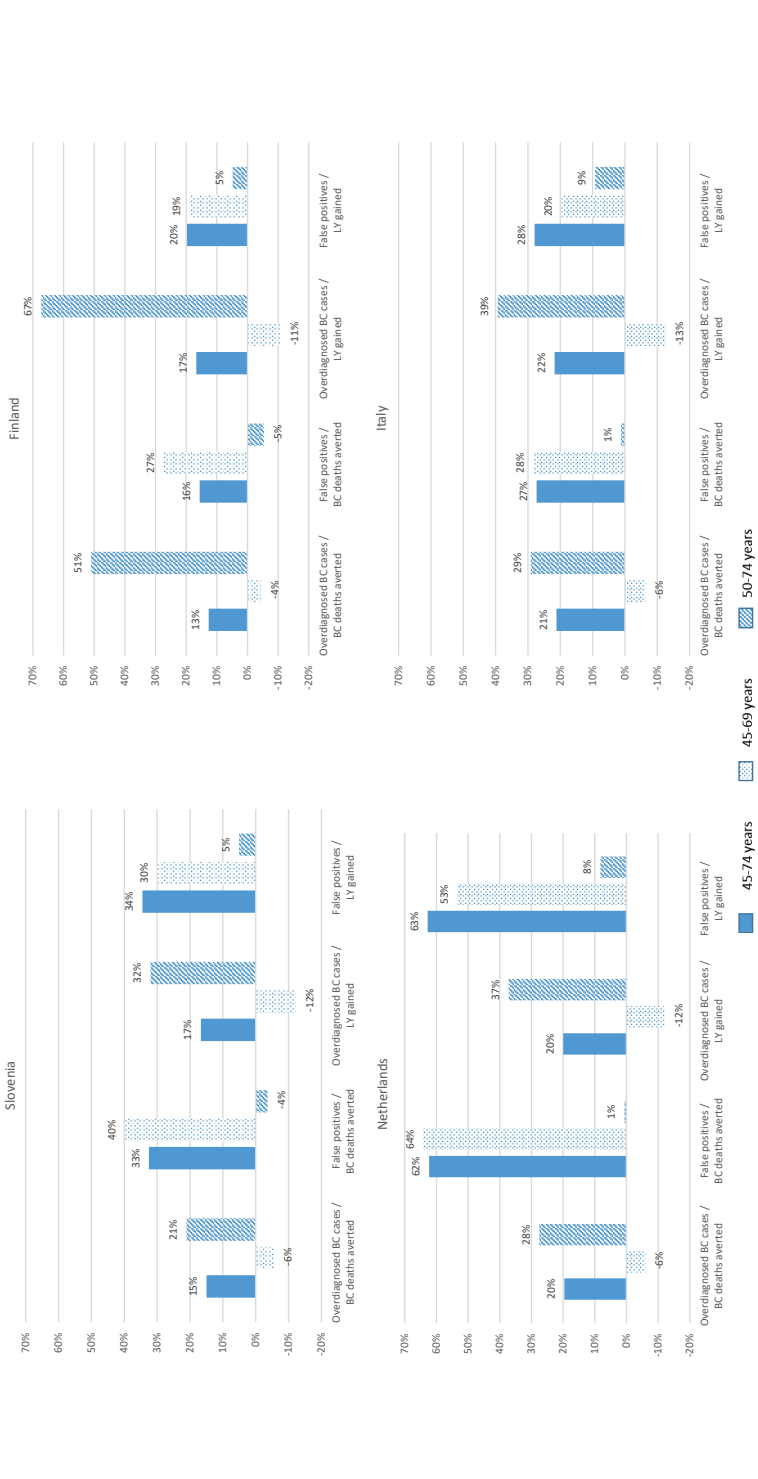


Figure 1. Percentage change in harm-to-benefit-ratios in comparison to the reference strategy 50-69, per alternative screening strategy and country

OD: Overdiagnosis; BC: Breast cancer; LY: Life years.

DISCUSSION

We were able to calibrate and validate four country-specific microsimulation models in order to investigate long-term outcomes of four breast cancer screening strategies for each European region. Therefore, our results are likely to be relevant to other European countries as well. We found that the ratio of overdiagnosed breast cancer/breast cancer deaths averted could be optimized if screening programs would screen women between ages 45 and 69. By extending the target age range, both the number of life-years gained and breast cancer deaths averted due to screening would increase. However, aside from benefits, extending the screening ages is also associated with additional harms. Of the three alternative strategies, 45–74 is the least optimal age range for screening women in Slovenia, the Netherlands and Italy, while the least optimal range is 50–75 in Finland.

The impact of the two harms used in our study is considerably different. False-positive results are the most frequent harm of mammography screening, leading to unnecessary testing and an increased benign biopsy rate. In contrast, overdiagnosis is less common, but has a substantial impact. The detection of overdiagnosed cancers turns women into patients, leading to surgery and treatments, which can cause harm and adversely affect quality of life[5]. Moreover, overdiagnosis leads to additional costs and use of healthcare resources. In contrast, false-positive results cause only short-term anxiety, and there is no measurable health utility decrement from this harm[18].

It can be debated whether the most serious harm (overdiagnosis) of screening should have equal priority to the most important benefit (the reduction in breast cancer mortality)[19]. However, we believe that the comparability of the two events should be considered. The value of a life saved versus an overdiagnosed case or their consequences are obviously of different magnitude[20]. Being overdiagnosed markedly influences the quality of life of women who experience it as it may cause suffering and anxiety, but it does not affect life expectancy. However, breast cancer screening extends lives[5,21], and therefore many women think overdiagnosis is worth the gain from the potential reduction in breast cancer mortality. In a discrete-choice experiment, Sicsic[22] estimated that women would be willing to accept on average 14.1 overdiagnosed cases and 47.8 false-positive results to avoid one breast-cancer-related death. These results indicate that women consider over-diagnosis 3.4 times as harmful as false-positive results. The ratios we found are well be-low these thresholds for overdiagnosis per death averted. In all modelled strategies and countries, there are more deaths averted (range 2–3) for every overdiagnosed case. In contrast, two strategies (45–69 and 45–74) in Slovenia and Italia, respectively, have false-positive results per averted breast cancer death above this threshold.

Our analysis was based on a cohort approach, where women 45 years of age were followed until death. While this approach still considers country-specific all-cause-mortality differences, it eliminates all other external factors such as differences in age structure and makes it possible to solely judge the effect of a change in screening strategy and to compare this effect between countries. However, in reality the differences in age structures between countries might actually play a role and thus affect the decision for a change in screening policy. Of the four countries in this analysis, the Italian population is relatively young, and the Finnish population is relatively old (Supplementary materials, Table S1).

To our knowledge, no previous studies analysed the relationship between harms and benefits for varying age ranges and countries. Some studies have specifically assessed the harm-to-benefit ratios for breast cancer screening, but only for the age range 50–69. The EUROSCREEN group estimated 4 overdiagnosed cases and 7 to 9 averted breast cancer deaths per 1,000 women, giving a ratio between 0.6 and 0.4[20,23]. An independent United Kingdom review found an overdiagnosis/breast cancer deaths averted ratio of three to be acceptable[5]. The variation in these results may represent methodological differences, for example in study design and length of follow-up[24]. Our findings for Southern Europe (Italian model) are in line with results of a modelling study for the Basque country, where Arrospide et al.[25] estimated an overdiagnosis/breast cancer deaths averted ratio of 0.3. Van Luijt[26] evaluated the Norwegian Breast Cancer Screening Program in a microsimulation study and estimated a harm-to-benefit ratio of 0.23, whereas we estimated the ratio to be 0.32 for Northern Europe (Finnish model). In a life table model analysis for the United Kingdom, Pashayan[27] assessed that woman who undergo age-based triennial screening between 50 and 69 have twice as many overdiagnosed cases than prevented breast cancer deaths. In contrast, we estimated four times more benefits than harms for Western Europe (Dutch model), despite a shorter screening frequency and higher assumed attendance.

Differences in model estimated ratios likely reflect differences of overdiagnosis estimates, which can vary due to factors such as contrasting definitions of the population at risk. Besides, differences in main model assumptions including the natural history of the disease, differences in length of follow-up and differences in goodness-of-fit of each model can also explain varying estimates[24,28].

Some limitations of this study have to be considered. First, the improvement of prognosis is based on trial data for women age 50–69 years[29,30]. We assumed the same improvement in survival for women outside this age range[31]. Second, our predictions are based on a cohort of women born in 1975. If life expectancy for older women continues to increase in the future, then we might have underestimated the benefits and overestimated the harms of screening for the strategy that screened beyond the age of 69. Third, we maintained the standard two-year

screening interval now adopted for the 50 to 69 age range for the alternative strategies, but there is uncertainty about the optimal screening interval for these age ranges, with recommendations ranging between 1 and 3 years. Future work could address different screening intervals by age.

We based our analysis on a comparison to the biennial screening from age 50 to 69 years irrespective of the actual screening policy in each of the four countries. However, the Dutch national breast cancer screening program invites women between 50 and 75 years of age. For the Netherlands, we found that when changing the reference strategy to the current strategy, our findings consistently show that starting screening 5 years earlier would lead to better overdiagnosis-related ratios. This is consistent with a previous microsimulation study based on the same Dutch model showing that digital mammography screening between age 40 and 49 in the Netherlands, in addition to the current screening strategy, is cost-effective [17].

The triad of benefits, harms and costs is a key element of health policy decision making. Future research should extend the harm-to-benefit ratios of breast cancer screening to a cost-effectiveness analysis. Such an analysis would consider additional screening effects, such as treatment-related advantages or quality of life, as well as costs.

CONCLUSIONS

Our study provides insight as to how harm-to-benefit ratios of breast screening pro-grams could be improved by adapting the age range of screened women. Assuming different strategies, this modelling study represents meaningful information on the magnitude of harms and benefits. However, the interpretation of our results depends on how the considered harms and benefits are prioritized by political decision makers.

REFERENCES

1. Council of the European Union. Council Recommendation of 2 December 2003 on Cancer Screening (2003/878/EC). Off. J. Eur. Union 2003, 327, 34–38.
2. Nelson, H.D.; Fu, R.; Cantor, A.; Pappas, M.; Daeges, M.; Humphrey, L.L. Effectiveness of Breast Cancer Screening: Systematic Review and Meta-analysis to Update the 2009 U.S. Preventive Services Task Force Recommendation. Ann. Intern. Med. 2016, 164, 244–255.
3. Broeders, M.; Moss, S.; Nyström, L.; Njor, S.; Jonsson, H.; Paap, E.; Massat, N.; Duffy, S.; Lynge, E.; Paci, E. 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.
4. Zielonke, N.; Gini, A.; Jansen, E.E.; Anttila, A.; Segnan, N.; Ponti, A.; Veerus, P.; de Koning, H.J.; van Ravesteyn, N.T.; Priaulx, J. Evidence for reducing cancer specific mortality due to screening for breast cancer in Europe: A systematic review. Eur. J. Cancer 2020, 127, 191–206.
5. Marmot, M.G.; Altman, D.G.; Cameron, D.A.; Dewar, J.A.; Thompson, S.G.; Wilcox, M. The benefits and harms of breast cancer screening: An independent review. Br. J. Cancer 2013, 108, 2205–2240.
6. Lauby-Secretan, B.; Scoccianti, C.; Loomis, D.; Benbrahim-Tallaa, L.; Bouvard, V.; Bianchini, F.; Straif, K. Breast-Cancer Screening—Viewpoint of the IARC Working Group. N. Engl. J. Med. 2015, 372, 2353–2358.
7. European Commission Initiative on Breast Cancer. Recommendations from the European Breast Guidlines. 2021. Available online: https://ecibc.jrc.ec.europa.eu/recommendations/ (accessed on 22 December 2020).
8. IARC. Cancer Screening in the European Union. Report on the Implementation of the Council Recommendation on Cancer Screening; IARC: Lyon, France, 2017. Available online: 2017_ cancerscreening_2ndreportimplementation_en.pdf (accessed on 08 November 2020).
9. Zielonke, N.; Kregting, L.M.; Heijnsdijk, E.A.M.; Veerus, P.; Heinävaara, S.; McKee, M.; de Kok, I.M.; de Koning, H.J.; van Ravesteyn, N.T.; Collaborators, T.E. The potential of breast cancer screening in Europe. Int. J. Cancer 2021, 148, 406–418.
10. IARC Working Group. IARC Handbooks of Cancer Prevention, Volume 15: Breast Cancer Screening; International Agency for Research on Cancer, Lyon, France, 2016.
11. Bray, F.; Colombet, M.; Mery, L.; Pineros, M.; Znaor, A.; Zanetti, R.; Ferlay, J. Cancer Incidence in Five Continents, Vol. XI (Electronic Version); International Agency for Research on Cancer: Lyon, France 2017. Available online: http://ci5iarcfr 2017 (accessed on 14 January 2021).
12. van den Ende, C.; Oordt-Speets, A.M.; Vroling, H.; van Agt, H.M. Benefits and harms of breast cancer screening with mammography in women aged 40–49 years: A systematic review. Int. J. Cancer 2017, 141, 1295–1306.
13. Mandrik, O.; Zielonke, N.; Meheus, F.; Severens, J. (Hans); Guha, N.; Acosta, R.H.; Murillo, R. Systematic reviews as a 'lens of evidence': Determinants of benefits and harms of breast cancer screening. Int. J. Cancer 2019, 145, 994–1006.

14. Wanders, J.O.; Holland, K.; Veldhuis, W.B.; Mann, R.M.; Pijnappel, R.M.; Peeters, P.H.; van Gils, C.H.; Karssemeijer, N. Volumetric breast density affects performance of digital screening mammography. Breast Cancer Res. Treat. 2017, 162, 95–103.
15. Beau, A.B.; Lynge, E.; Njor, S.H.; Vejborg, I.; Lophaven, S.N. Benefit-to-harm ratio of the Danish breast cancer screening programme. Int. J. Cancer 2017, 141, 512–518.
16. Van Ravesteyn, N.T.; Miglioretti, D.L.; Stout, N.K.; Lee, S.J.; Schechter, C.B.; Buist, D.S.; Huang, H.; Heijnsdijk, E.A.M.; Trentham-Dietz, A.; Alagoz, O.; et al. What level of risk tips the balance of benefits and harms to favor screening mammography starting at age 40? Ann. Intern. Med. 2012, 156, 609–617.
17. Sankatsing, V.D.; Heijnsdijk, E.A.; Van Luijt, P.A.; Van Ravesteyn, N.T.; Fracheboud, J.; De Koning, H.J. Cost-effectiveness of digital mammography screening before the age of 50 in The Netherlands. Int. J. Cancer 2015, 137, 1990–1999.
18. Tosteson, A.N.A.; Fryback, D.G.; Hammond, C.S.; Hanna, L.G.; Grove, M.R.; Brown, M.; Wang, Q.; Lindfors, K.; Pisano, E.D. Consequences of False-Positive Screening Mammograms. JAMA Intern. Med. 2014, 174, 954–961.
19. Jørgensen, K.J.; Klahn, A.; Gøtzsche, P.C. Are benefits and harms in mammography screening given equal attention in scientific articles? A cross-sectional study Review. BMC Med. 2007, 5, 12.
20. Hofvind, S.; Román, M.; Sebuødegård, S.; Falk, R.S. Balancing the benefits and detriments among women targeted by the Norwegian Breast Cancer Screening Program. J. Med Screen. 2016, 23, 203–209.
21. Goodwin, J.S.; Sheffield, K.; Li, S.; Tan, A. Receipt of Cancer Screening Is a Predictor of Life Expectancy. J. Gen. Intern. Med. 2016, 31, 1308–1314.
22. Sicsic, J.; Pelletier-Fleury, N.; Moumjid, N. Women's Benefits and Harms Trade-Offs in Breast Cancer Screening: Results from a Discrete-Choice Experiment. Value Health 2018, 21, 78–88.
23. Paci, E.; Group, E.W. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. J. Med. Screen. 2012, 19 (Suppl. 1), 5–13.
24. De Gelder, R.; Heijnsdijk, E.A.; Van Ravesteyn, N.T.; Fracheboud, J.; Draisma, G.; De Koning, H.J. Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiol. Rev. 2011, 33, 111–121.
25. Arrospide, A.; Rue, M.; Van Ravesteyn, N.T.; Comas, M.; Larrañaga, N.; Sarriugarte, G.; Mar, J. Evaluation of health benefits and harms of the breast cancer screening programme in the Basque Country using discrete event simulation. BMC Cancer 2015, 15, 671.
26. van Luijt, P.A.; Heijnsdijk, E.A.M.; de Koning, H.J. Cost-effectiveness of the Norwegian breast cancer screening program. Int. J. Cancer 2017, 140, 833–840.
27. Pashayan, N.; Morris, S.; Gilbert, F.J.; Pharoah, P.D. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. JAMA Oncol. 2018, 4, 1504–1510.
28. Jahn, B.; Sroczynski, G.; Bundo, M.; Mühlberger, N.; Puntscher, S.; Todorovic, J.; Rochau, U.; Oberaigner, W.; Koffijberg, H.; Fischer, T.; et al. Effectiveness, benefit harm and cost effectiveness of colorectal cancer screening in Austria. BMC Gastroenterol. 2019, 19, 209.

29. Tabár, L.; Vitak, B.; Chen, T.H.-H.; Yen, A.M.-F.; Cohen, A.; Tot, T.; Chiu, S.Y.-H.; Chen, S.L.-S.; Fann, J.C.-Y.; Rosell, J.; et al. Swedish two-county trial: Impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011, 260, 658–663.
30. Andersson, I.; Aspegren, K.; Janzon, L.; Landberg, T.; Lindholm, K.; Linell, F.; Ljungberg, O.; Ranstam, J.; Sigfusson, B. Mammographic screening and mortality from breast cancer: The Malmo mammographic screening trial. Br. Med. J. 1988, 297, 943–948.
31. van den Broek, J.J.; van Ravesteyn, N.T.; Mandelblatt, J.S.; Huang, H.; Ergun, M.A.; Burnside, E.S.; Xu, C.; Li, Y.; Alagoz, O.; Lee, S.J.; 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. Mak. 2018, 38 (Suppl. 1), 140S–150S.

SUPPLEMENTARY MATERIAL

A) EU-TOPIA MISCAN-BREAST MODEL DESCRIPTION

1. Model Purpose

Trends in breast cancer incidence and mortality and the (potential) impact of interventions depend on many kinds of factors related to the biology of breast cancer, the characteristics of the population, and the potential impact and usage of early detection and treatment. A simulation model is a helpful tool to estimate the effect of each of the listed factors on cancer incidence and mortality. MISCAN–Breast is developed to analyse trends in breast cancer due to changes in lifestyle, improvement of treatment and implementation of screening strategies. The purpose of MISCAN–Breast can be described in three specific aims:

1. To simulate breast cancer incidence and mortality according to observed figures.
2. To compare screening strategies, allowing the user to improve existing screening programmes as well advising countries on the effects of implementing a breast cancer screening programme.
3. To predict how breast cancer screening and treatment practices will impact future incidence and mortality.

2. Model Overview

2.1. General Model Structure

MISCAN-Breast is a stochastic, semi-Markov microsimulation model. In a microsimulation model, individuals are simulated one at a time instead of as proportions of a cohort. The advantage of this is that new events can be dependent on past events of that individual, giving the model a 'memory'. The model is stochastic, which means that sequences of events are simulated by drawing from distributions of probabilities and durations instead of using fixed values. Therefore, the outcomes of the model are subject to random variation.

MISCAN uses the Monte Carlo method to simulate all events in the program. Possible events are birth and death of a person, onset of a pre-clinical ductal carcinoma in situ (DCIS), transitions between disease states, participation in screening and screen- or clinical detection of a cancer. First, breast cancer incidence and breast cancer mortality are estimated in a situation without screening. Subsequently, screening and treatment related improvements in survival are simulated, in order to determine the impact of screening and treatment on the life histories.
MISCAN–Breast consists of four parts:

- Demography part
- Natural history part
- Treatment part
- Screening part

These parts are not physically separated in the program, but it is useful to consider them separately.

2.2. Demographic Part

MISCAN–Breast first generates a series of individual life histories in the demography part to form a population according to the Demography Parameters. Each woman in the population consists of a date of birth and a date of death from other causes than breast cancer. These dates are drawn from birth and life tables that are representative for the population under consideration. The maximum age that a person can reach in the model is set to 100 years.

2.3. Natural History Part

After individual life histories are simulated in the demography part of MISCAN-Breast, natural histories of breast cancer are simulated for a subset of these women in the Natural History Part (as only a few women will develop breast cancer). Breast cancer starts with the onset of a pre-clinical ductal carcinoma in situ (DCIS) and continues with its progression through the invasive successive states T1A, T1B, T1C and T2+. The development from a DCIS into cancer depends on lymph node status, age-specific transition probabilities and the duration distribution. At each stage, a tumour may become screen-detected if screening is present or clinically detected because of symptoms. The possible transitions between the different states are explained in Figure S1. The life history of each person is altered according to the breast cancer histories (natural history) that is simulated for that person. If a woman dies from breast cancer before she dies from other causes, her death age is adjusted accordingly.

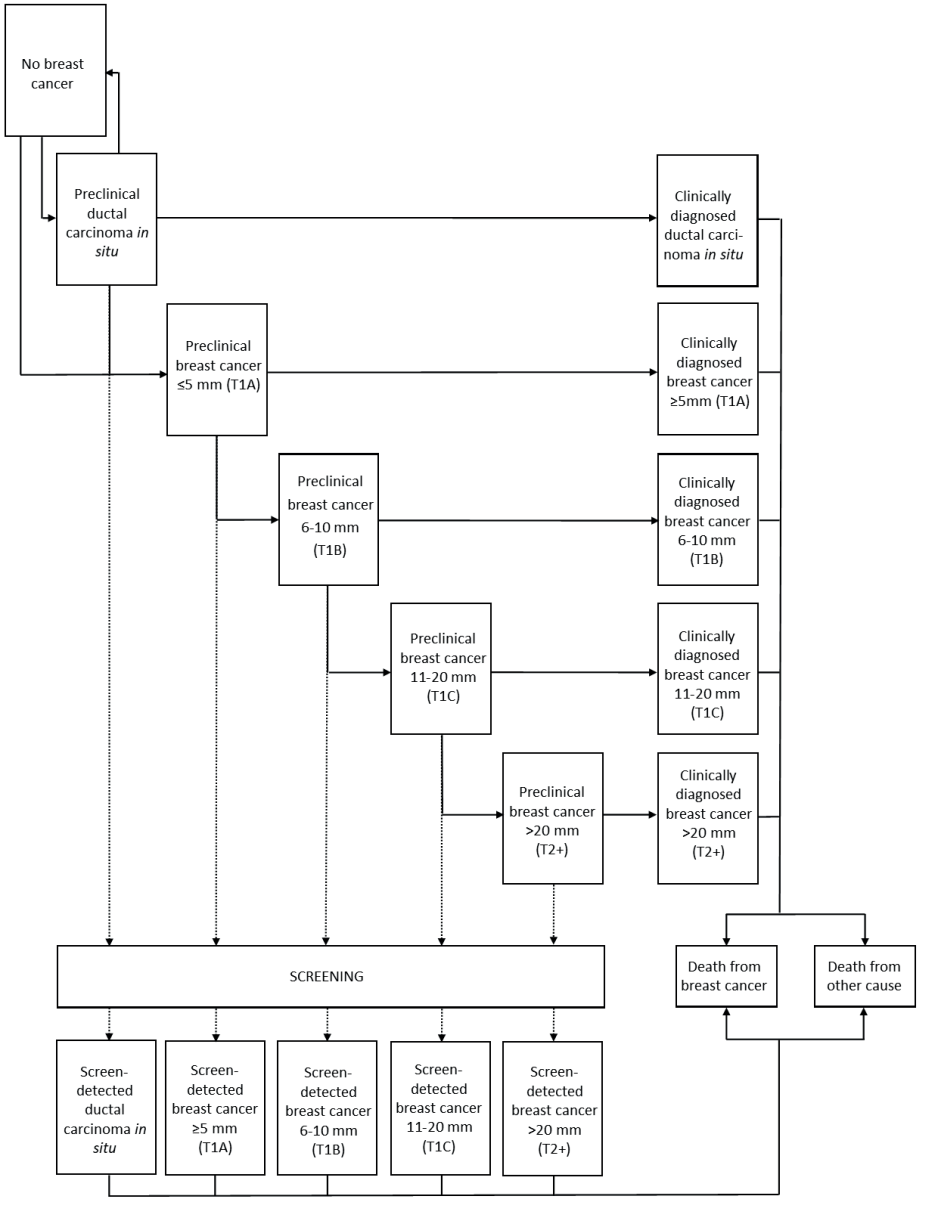


Figure S1. Transitions in the MISCAN-Breast model. The arrows represent the possible transitions.

2.4. Treatment Part

Any screen detected or clinically detected disease transits directly to one of the four treatment states: no therapy; chemotherapy; hormonal therapy; or a combination of chemotherapy and hormone therapy. The probability for a certain treatment is dependent on disease stage, age, calendar year, and detection mode.

A woman in a treatment state can die, either of breast cancer or from other causes. Other-cause mortality is determined by the life table.

The survival time is dependent on age- and stage- and treatment specific survival estimates.

2.5. Screening Part

In the third part of the model, screening for breast cancer is simulated. After the life history of a person is adjusted for breast cancer, the history will now be adjusted for the effects of screening. The screening part is simultaneously run with the natural history part, making detection of DCIS lesions or cancers in different states possible.

Persons can be invited to participate in screening at specified ages as defined in the screening policy. Depending on the test used and the tumour stage at the moment of the screening test, there is a probability of a positive test result. Screening may detect a DCIS lesion or invasive cancers. Women with a true positive result may receive (adjuvant) treatment. In the model, treatment starts immediately once a tumour is screen- or clinically detected. Screening leads to the detection of smaller tumours (in comparison to clinically detected tumours), which may improve survival after diagnosis (stage shift). The effectiveness of screening depends on the screening test characteristics (see Parameter Overview). The effect of screening on the life history of an individual is explained in Figure S2.

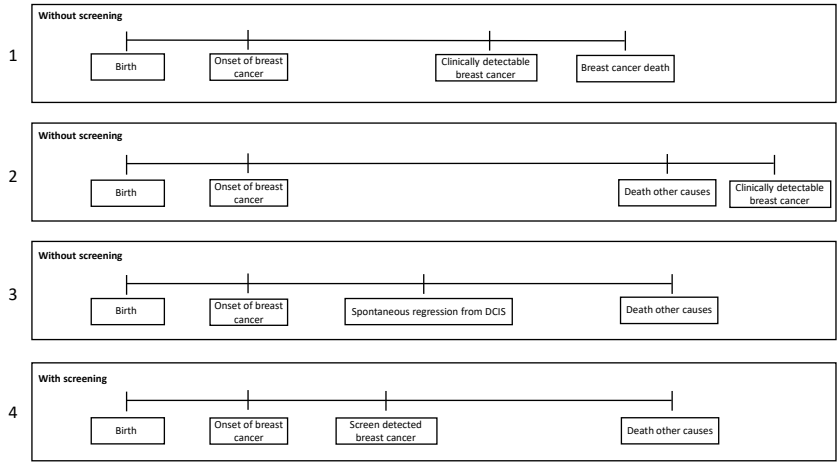


Figure S2. Effect of screening on life history.

Screening affects three women differently. The first box is the life history of a woman who has an onset of breast cancer, is diagnosed clinically, and dies of breast cancer. The second box is the life history of a woman who also has an onset of breast cancer, but who dies of other causes before this would be detected. The third box is the life history of a woman who has an onset of breast cancer, but also a spontaneous regression, this woman would not have been diagnosed without screening. The fourth box indicates the situation for these three women had screening been introduced. The woman in the first box no longer dies from breast cancer; the other two women do not benefit from screening. They have been overdiagnosed.

3. Model Output

All events (for example onset of disease, screen, diagnosis, death by cause) in the model are counted for each woman in two situations: a situation without breast cancer screening and a situation with breast cancer screening. The output of the model is flexible and can be for example: the number of invitations, tests, diagnosis (by mode of detection) and deaths, all by age and year, and when possible cancer stage.

The model can also provide "unobservable" events as lead time, overdiagnosis and breast cancer deaths prevented. In addition, costs and QALYs can be calculated.

4. Model Assumptions

As explained in the Section 'Model overview', the model consists of four parts; a demography, natural history, a screening and a treatment part. Since some inputs are not (directly) observable in the data and a model is a simplified version of a complex process. Therefore, several assumptions have to be made for the parts.

Because, in principle, all diagnosed breast cancers are treated, it is not exactly known how the disease would have developed without intervention. Hypotheses have to be made on the onset rate of preclinical cancer and the average sojourn time (the period in which a pre-clinical tumour can become detected by screening). The fraction of invasive breast cancers that is preceded by DCIS is uncertain, and it is unknown whether all preclinical DCIS cases progress, or if some are dormant or regress. Furthermore, assumptions need to be made on the age- and stage-specific fatality of cancer and the impact of a screening program. The role of underlying time trends, such as an increasing prevalence of certain risk factors for breast cancer and developments in breast cancer treatment and screening further complicates our analyses.

This does not imply that the natural history of breast cancer is completely unknown. Some indirect inferences can be made from randomized controlled trials and screening data. The average duration of pre-clinical cancer, for instance, is proportional to the ratio between the detection rate in the initial screening round and the clinical incidence rate without screening. Based on stage-specific incidence rates of clinically diagnosed and screen-detected breast cancer and rates of interval cancer, progression and regression rates might be assessed. Using detection rates per screening round and interval cancer rates, the sensitivity of mammography could be estimated. The survival after a breast cancer diagnosis and the influence of a screening program on these survival rates can be deducted from the randomized trials. In MISCAN, the effectiveness of screening was based on the Swedish Trials[1–4]. Despite the fact that these trials have been performed in the seventies and eighties, the observed breast cancer mortality in the Netherlands (and other countries) could be reasonably modelled using these data.

5. Model Parameters

5.1. Demographic Part

1. One-year birth cohorts
2. Proportion of the population in each birth cohort
3. Life table

5.2. Natural History Part

1. Mean duration of preclinical screen-detectable cancer by age and stage
2. The probability of a transition between the stages
3. Annual increase in background breast cancer incidence (without screening)
4. Long-term relative survival by clinical stage and age
   a. Without adjuvant treatment
   b. With hormonal treatment
   c. With chemotherapy
   d. With hormonal and chemotherapy
5. Reduction in risk of dying of breast cancer by age and preclinical stage after screen-detection

5.3. Treatment Part

Proportion by age and stage that are treated with adjuvant therapy

5.4. Screening Part

1. Screening attendance by age
2. Test sensitivity of digital mammography by age and preclinical stage
3. Improvement of survival
4. Proportion referral by age

6. Model Calibration

The values of some parameters in Section 5 can be observed directly, and will be calculated based on data in the EUTOPIA data template. Other values are available from literature.

However, some model parameters cannot be derived from observational data (e.g., mean duration of preclinical screen-detectable cancer by age and stage). Therefore, these parameter values need to be calibrated (fitted). Calibration involves estimating the parameter values in a way that the simulated outcomes fit the observed data.

6.1. Calibration Parameters

1. Mean duration of screen-detectable preclinical stage by age and stage
2. Test sensitivity of digital mammography by preclinical stage
3. Onset of the disease by age
4. Increasing incidence by year
5. Probabilities of immediate and slow progression from DCIS to T1A
6. Survival

6.2. Regional EU-TOPIA Models

Because the values of the calibrated parameters might differ across Europe, four different models were calibrated. From each European region, an exemplary country with high quality observational data, including the screening behaviour of that population, was selected to be representative for that region (the Netherlands for Western Europe, Finland for Northern Europe, Slovenia for Eastern Europe and Italy for Southern Europe).

6.2.1. Age structure of the female population

Table S1. Age-structure of the exemplary countries, women in 2018.

% of Total Female Population						
Country	**45 to 49**	**50 to 54**	**55 to 59**	**60 to 64**	**65 to 69**	**70 to 74**
Italy	7.9%	8.0%	7.1%	6.3%	6.0%	5.3%
Netherlands	7.3%	7.3%	7.0%	6.3%	5.8%	5.2%
Slovenia	6.8%	7.4%	7.1%	7.1%	6.2%	4.6%
Finland	5.8%	6.6%	6.6%	6.7%	6.9%	5.9%
% of female population 45–74						
Italy	19.4%	19.7%	17.5%	15.5%	14.7%	13.1%
Netherlands	18.8%	18.9%	18.0%	16.1%	15.0%	13.3%
Slovenia	17.3%	18.8%	18.1%	18.1%	16.0%	11.9%
Finland	15.1%	17.1%	17.1%	17.5%	17.9%	15.3%

Source: https://ec.europa.eu/eurostat/web/population-demography-migration-projections/data/database.

6.2.2. Breast-Cancer Screening Program Netherlands

In the Netherlands, a national population-based screening program is operational since 1990, inviting women 50–69 years of age for a biennial screening examination (Fracheboud et al. 2001). Women 70–75 years of age are invited since 1999. The Dutch Breast Cancer Screening Programme is carried out by five regional Cancer Screening Organisations (65 screening units) and coordinated, monitored and evaluated by the National Institute for Public Health and the Environment. In 2003, the first digital mammography unit was introduced and in 2010, all screening examinations were performed using digital mammography.

The invitation coverage for organized mammography screening was 100.3% in 2016, the participation rate was 77.3%, and the examination coverage was 77.5%. The detection rate per 1,000 screened women was 1.52 for DCIS and 5.30 for invasive BC in 2016.

6.2.3. Breast-Cancer Screening Program Finland

The nationwide Finnish breast cancer screening program started in 1987.

From 1992 until 2006, the national target population consisted of only women aged 50–59 invited biennially, and based on Government Decree on Screenings it widened up to 50–69 during 2007–2016. The coverage of quality assured screening registration has improved with time, and it reached complete coverage of all service providers in 2005. The Mass Screening Registry, a section of the Finnish Cancer Registry, maintains the individual level data on screening invitations to and participation in mammography screening.

The invitation coverage for organized mammography screening was 94.9% in 2014, the participation rate was 82.9%, and the examination coverage was 78.7%.

The detection rate per 1,000 screened women was 0.72 for DCIS and 5.36 for invasive BC in 2014.

6.2.4. Breast-Cancer Screening Program Italy

In Italy, screening by mammography is organized regionally and started in the City of Florence in 1990. The current Ministry of Health's guidelines recommend that women aged 50–69 are personally invited every two years. Several regions invite women from the age of 45 (annually) and/or up to age 74–75 (biennially). All 20 regions work under the umbrella of the National centre for screening monitoring (ONS). Together with the Italian group for mammography screening (GISMa), ONS is responsible for monitoring, performance evaluation, data collection and promotion of the organized breast screening programmes in Italy.

The invitation coverage for organized mammography screening was 81.0% in 2015, the participation rate was 54%, and the examination coverage was 43.8%. We assume that 19% of all woman age 50–69 were screened outside the programme.

The detection rate per 1,000 screened women was 0.7 for DCIS and 3.7 for invasive BC in 2013.

6.2.5. Breast-Cancer Screening Program Slovenia

The breast cancer screening programme in Slovenia (DORA) is organized, national and population based. The implementation started in 2008 and has been gradually expanded until rollout was completed in April 2018.

DORA invites women aged 50 to 69 years to screening mammography every two years. In 2016 screenings were carried out in 11 screening units; on 12 stationary screening mammographs and in 2 mobile screening units. Responsible institution: Institute of Oncology Ljubljana.

The invitation coverage for organized mammography screening was 53.4% in 2016, the participation rate was 75.7%, and the examination coverage was 40.4%. We assume that 13% of all woman age 50–69 were screened outside the programme.

The detection rate per 1,000 screened women was 1.31 for DCIS and 5.03 for invasive BC in 2016

6.3. Calibration Process

In calibrating each new country-specific MISCAN-Breast model, we used a specific calibration process composed of 4 steps (Figure S3):

1. The starting point is an existing model that was calibrated for the Netherlands (Model A)
2. All model inputs that are based on observed data, such as the demographic characteristics of the population, cancer survival and the implemented screening programme, are adjusted to represent the situation in country B (step II).
3. Observed data is collected that will be compared with the model outputs. These observed data are the so-called calibration targets: BC incidence and mortality rates, interval cancer rates, stage distribution and detection rates (Step III).
4. Based on the observed data the model optimizes a set of unobservable parameters to meet the observed data in the model output (step IV). MISCAN uses the Nelder and Mead simplex ("Amoeba") multivariate minimization routine, which has been adapted for optimizing random functions. The model runs repeatedly, and the simulation runs are compared to the counts in the observations. The total deviance is the sum of the individual deviances. A convergence criterion is set by a Kendall Tau test. The required significance level of the test can be specified.
5. After the model fitted a set of unobservable parameters, the outcomes of step IV are evaluated. First by evaluating the model fit with the calibration targets (by visual inspection including 95% confidence interval). If the fit is not satisfying, the model inputs are re-evaluated and step IV is repeated. Parameters of the natural history of the cancer, e.g. stage-specific mean duration of screen-detectable breast cancer and progression, are initially assumed to be equal between model country A and B. Only if the calibrated model for country B does not meet the calibration targets, these natural history parameters will be calibrated as well, together with the initial set of calibration parameters.
6. If the fit is satisfying, the model is validated by replicating mortality reduction due to mammography screening based on outcomes observed in the selected published studies (data not used in the calibration process, external validation). Briefly, we adjust the models to reflect specific demographics (e.g. year and age of BC deaths),and screening policy of this study (starting age, stopping age, screening modality, interval, participation) in accordance with the selected published study for country B. Then we compare the model predictions and the observed outcomes of the study. If the outcomes predicted by the model are within the 95% confidence interval of the corresponding study's observed outcomes, a calibrated and validated model for country B is ready to be used.

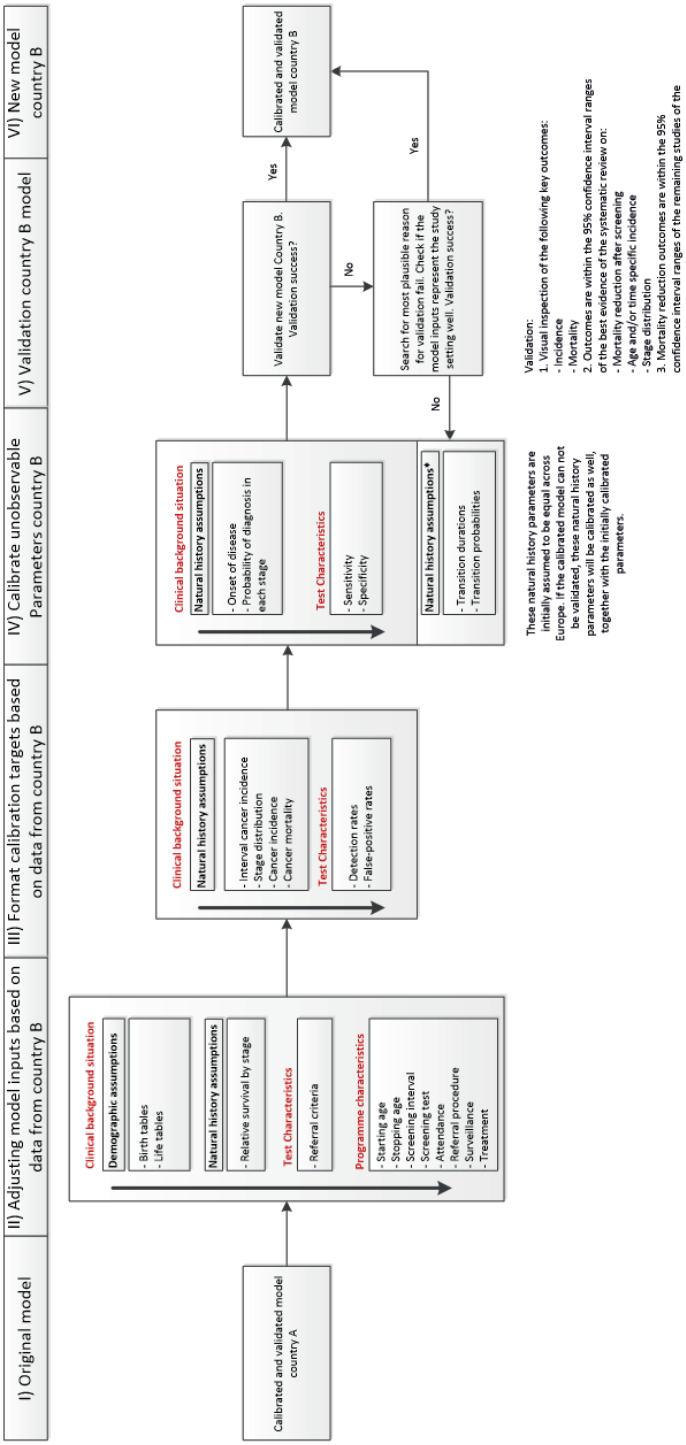


Figure S3. Calibration and validation process for development of MISCAN-Breast country specific models.

6.4. Calibration Results

6.4.1. Country Specific Calibration Parameters

Table S2. Model input parameters.

Input parameter	Netherlands	Finland	Italy	Slovenia
Stage-specific sensitivity of digital mammography DCIS	0.865	0.596	0.821	0.726
Stage-specific sensitivity of digital mammography T1a	0.553	0.811	1.000	0.785
Stage-specific sensitivity of digital mammography T1b	0.481	0.761	0.717	0.656
Stage-specific sensitivity of digital mammography T1c	0.857	0.946	0.814	0.780
Stage-specific sensitivity of digital mammography T2+	1	1	1	1
Breast cancer onset	0.291	0.279	0.267	0.267
Onset hazard age 30 years	0.0000705	0.0000135	0.000000302	0.0000201
Onset hazard age 50 years	0.0098	0.0067	0.0078	0.0060
Onset hazard age 70 years	0.0190	0.0135	0.0148	0.0060
Onset hazard age 100 years	0.0245	0.0000242	0.0018	0.0000158
Stage-specific duration (years) screen-detectable preclinical stage DCIS	0.669	0.921	0.669*	1.525
Stage-specific duration (years) screen-detectable preclinical stage T1a	0.720	1.079	0.720*	0.702
Stage-specific duration (years) screen-detectable preclinical stage T1b	1.106	1.550	1.106*	1.495
Stage-specific duration (years) screen-detectable preclinical stage T1c	1.492	1.200	1.492*	1.737
Stage-specific duration (years) screen-detectable preclinical stage T2+	1.171	1.269	1.171*	1.304
Mortality factor ages 0–72.5	2.2	0.8	2	1.2
Mortality factor ages 72.5–100	4.5	5	5	5
Referral rate <50	0.030	0.030	0.065	0.040
Referral rate >50	0.023	0.028	0.058	0.034

* This natural history parameter is assumed to be the same as in Model A (Netherlands) as the calibration results were satisfying

6.5. The Netherlands (West, model A)

The MISCAN model for breast cancer screening in the Netherlands has been well reported and validated in the past[5,6]. The parameters were calibrated to recent data from the Dutch screening organizations on interval cancers (between 2004 and 2011), screen-detected cancers (2004–2013) and stage distribution at detection for screen-detected cancers and interval cancers. Simultaneously, the parameters were calibrated to data on breast cancer incidence between 1975 and 2013 by five year age groups, from the National Cancer Registry and the Eindhoven Cancer Registry.

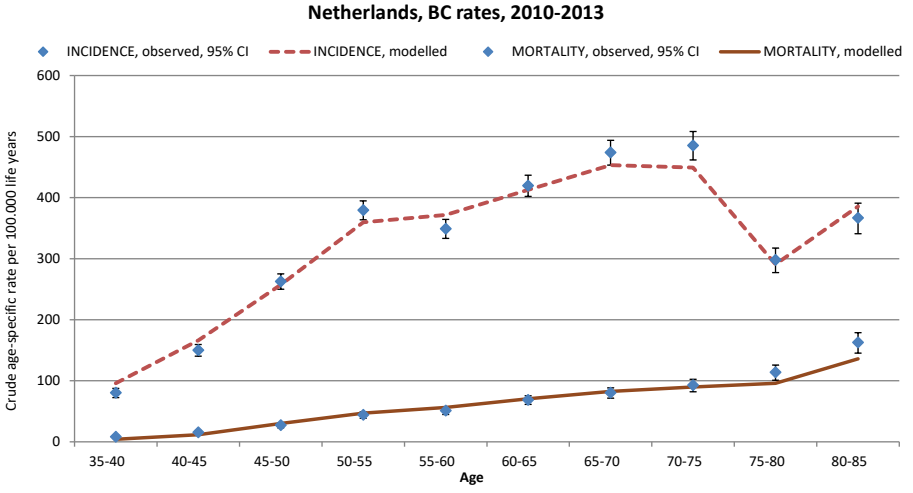


Figure S4. Fit of the model predictions with observed breast cancer incidence and mortality in the Netherlands, 2010–2013.

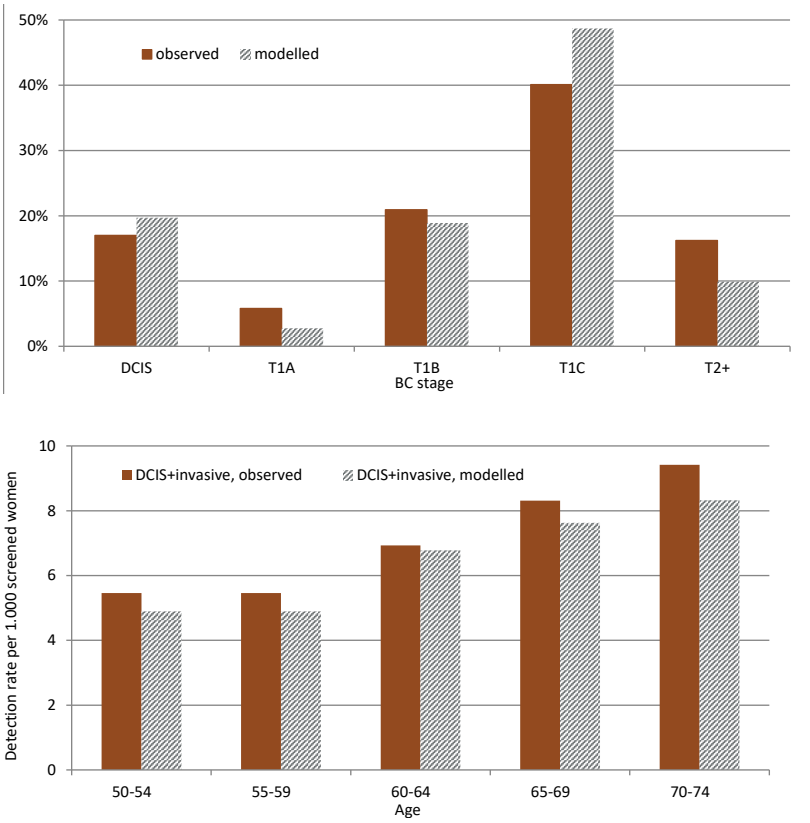


Figure S5. Fit of the model predictions with observed stage distribution (top, only screen detected cancers, 2010–2014) and detection rate (bottom, 2013) in the Netherlands.

6.6. Finland (North)

For the MISCAN model for breast cancer screening in Finland we modelled the female Finnish population born 1910–2016. As direct model inputs we included the screening histories between 1987 until 2016[7,8]. We calibrated age-specific and period-specific incidence hazards and levels, DCIS probabilities, durations and sensitivity. As the observed data was not conform the TNM stage classification, the BC stage distribution data was adjusted as follows: DCIS, localized (including T1A-T2+, node negative) and non—localized (including T1A-T2+, node positive). We fitted the Finnish model on age-specific breast cancer incidence in two periods simultaneously: 1975–1985 and 1995–2014 (data from the Finnish Cancer Registry).

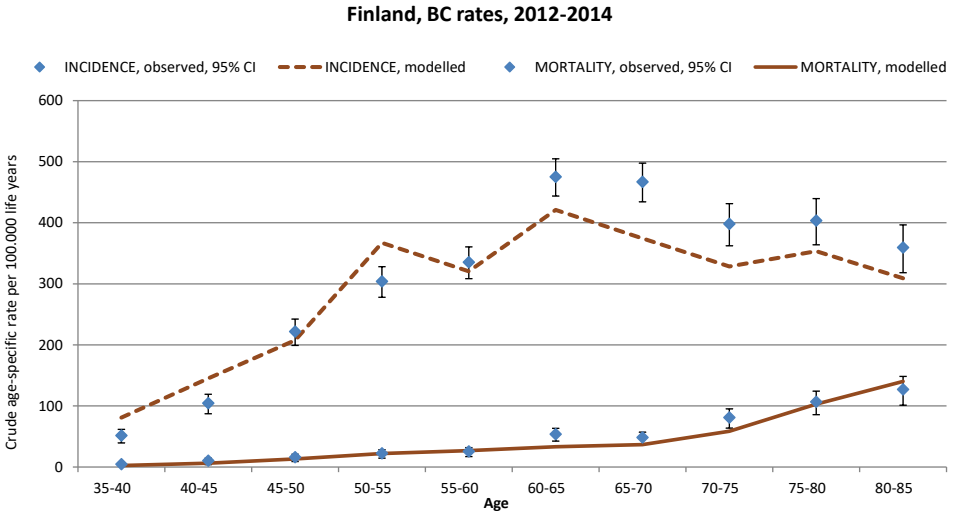


Figure S6. Fit of the model predictions with observed breast cancer incidence and mortality in Finland, 2012–2014.

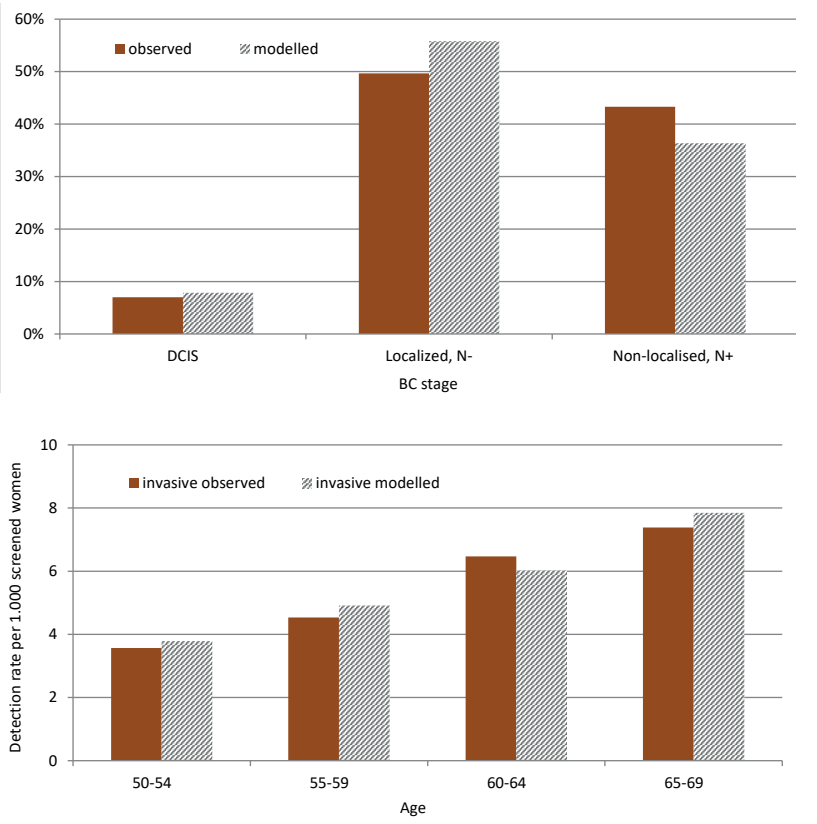


Figure S7. Fit of the model predictions with observed stage distribution (top, only screen detected cancers, 2006–2011) and detection rate (bottom, 2013) in Finland.

6.7. Italy (South)

The Italian MISCAN Breast model was calibrated on age-specific and stage-specific breast cancer incidence and mortality in Italy in 2006-2009 with data from the Italian Association of Cancer Registries (AIRTUM), which covers approximately one third of the total Italian female population. Stage distribution parameters were calibrated using data from the Cancer Screening National Monitoring reports. We modelled the age distribution of the female Italian population born between 1920 and 1980 using data from the Human Mortality Databases.

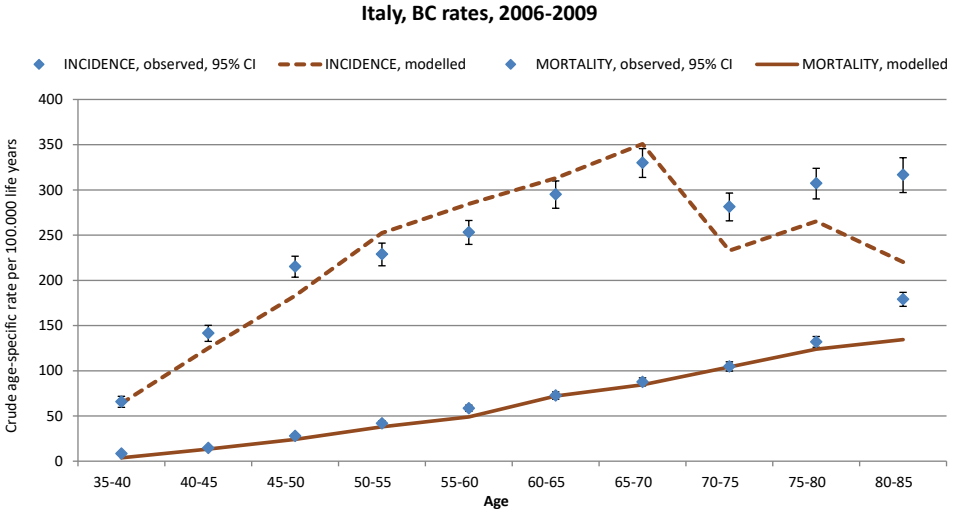


Figure S8. Fit of the model predictions with observed breast cancer incidence and mortality in Italy, 2006–2009**.**

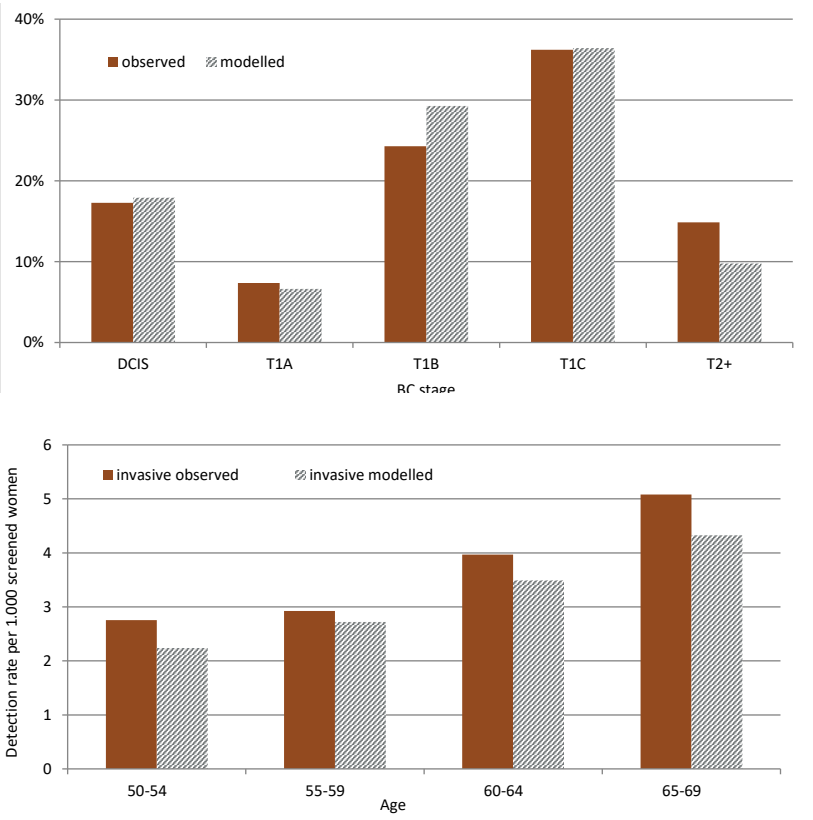


Figure S9. Fit of the model predictions with observed stage distribution (top, only screen detected cancers) and detection rate (bottom, 2013), in Italy.

6.8. Slovenia (East)

The Slovenian MISCAN Breast model was calibrated on age-specific breast cancer incidence in 1975–2014 (CR of Slovenia). As the implementation of the breast cancer screening program in Slovenia (DORA) started in 2008, the estimate of total coverage (opportunistic and organized) was mostly based on expert opinion for 1990–2008. We modelled the female Slovenian population born 1926–1982 using data from the Human Mortality Databases. We validated the Slovenian model by comparing the observed and modelled Interval cancer rates in 2014.

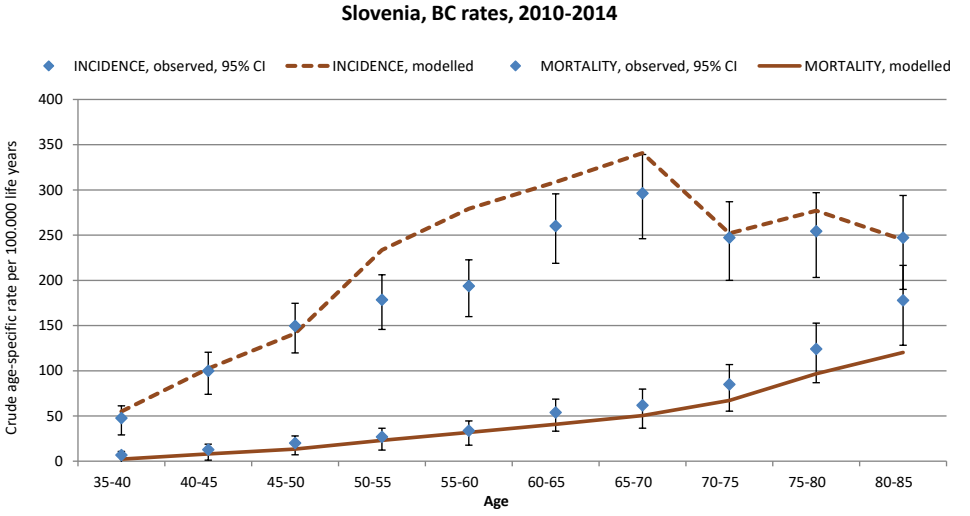


Figure S10. Fit of the model predictions with observed breast cancer incidence and mortality in Slovenia, 2010–2014.

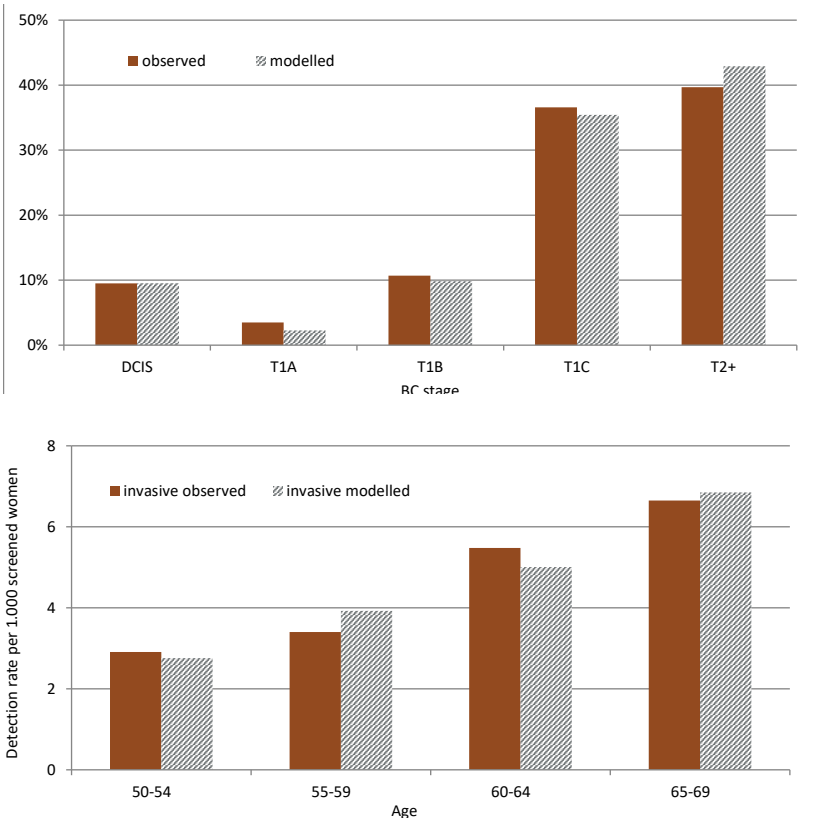


Figure S11. Fit of the model predictions with observed stage distribution (top, only screen detected cancers, 2011–2015) and detection rate (bottom, 2013) in Slovenia.

7. Model Validation

As the EU-TOPIA MISCAN-Breast model is also used to inform policy makers, it is important to give insight in the calibration (model fitting) methods and the accuracy of model outcomes[9,10]. The latter can be achieved by providing transparency, i.e. reporting on the structure of the model, the value of parameters and the rationale for the assumptions used (provided through the companion MISCAN-Breast model description), and by describing how the model has been validated.

To ensure the validity of our calibrated models and consequently the usefulness in predicting, monitoring, and improving the existing screening programs, it is very important to validate our model predictions against observed data. With this aim: i) a systematic review was performed to summarize the evidence for breast cancer specific mortality reduction due to screening in Europe[11]; ii) some studies ("best evidence") were identified and selected for validating the fully calibrated models for: Finland, Netherlands, Italy and Slovenia. These countries are exemplary for the Northern, Western, Southern and Eastern regions of Europe.

7.1. Best Evidence

7.1.1. Definition "Best Evidence"

As an extension of the systematic reviews, we identified those among all included studies in the review, that provides "best evidence " in observed data which the MISCAN model output can be compared to. We defined the best evidence study, judging the level of evidence of each study based on a group of factors. Those had a different hierarchic order and were reported in order of importance as follow:

1. Country was the strongest factor for this report as we looked for best data to validate models for exemplary countries.
2. Outcome prioritizes cancer specific mortality for all three cancer sites. Alternatively, incidence is a useful outcome to validate MISCAN-Colon models rated as Level II (Studies that assessed the effect of screening on CRC incidence reduction).
3. Study setting reflects the favoritism of data from actual cancer screening programs over other study settings, including randomized controlled trials and observational studies.
4. Study design/Risk of bias is a combination of study design and risk of bias and is based on the results of the prior quality assessment.

Table S3. List of factors forming the judgement of level of evidence of each study.

Criteria	Level	Type of Studies Retrieved
Country	I	Studies conducted in that specific country (national level).
	II	Studies conducted inside that specific country (regional level).
	III	Studies conducted in neighbouring countries within the same European region (national level).
	IV	Studies conducted in neighbouring countries within the same European region (regional level).
Outcome	I	Studies that assessed the effect of screening on cancer specific mortality reduction.
	II	Studies that assessed the effect of screening on cancer specific incidence reduction.
	III	Studies that assessed the effect of screening on overall mortality reduction.
Study setting	I	Screening program evaluation study.
	II	Research study.
Study design /Risk of bias	I	Randomized Control Trials with Low Risk.
	II	Randomized Control Trials with Moderate Risk / Observational studies with Low Risk (score of 8 or 9);
	III	Randomized Control Trials with High Risk / Observational studies with Moderate Risk (score 5 to 7).
	IV	Observational studies with High Risk (score from 0 to 4).

Thus, a study conducting a screening program evaluation within an exemplary country and investigating the impact of screening on cancer specific mortality will be considered as highest level of evidence. However, when no study will show all these factors at the same time the selection will be performed giving priority, respectively, to country, outcome, study setting, and combination of study design and risk of bias.

7.1.2. Best Evidence per European Region

No studies from Eastern Europe met the initial inclusion criteria and subsequently qualified as "best evidence".

Table S4. Best evidence for 3 of the 4 European countries and their respective point estimates on breast cancer mortality reduction due to mammography screening.

Study	Region	Country	Study Type	Target Age	Effect Sizes for Breast cCancer Mortality [1], (95%CI)
Heinävaara S, 2016[7]	North	Finland	CC	50–69	HR = 0.67 (0.49–0.90)
Puliti D, 2008[12]	South	Italy	CC	50–74	OR = 0.50 (0.42–0.60)
Paap E, 2014[13]	West	Netherlands	CC	50–75	OR = 0.42 (0.33–0.53)

[1] Attenders/non-attenders. CI = Confidence interval, HR = Hazard Ratio, OR = Odds ratio CC = Case Control study

7.2. Validation Results

7.2.1. The Netherlands

In a multi-region case-referent study, Paap[13] included breast cancer deaths in women aged 50–75 between 2004 and 2005 and estimated the benefit of the population-based screening program on breast cancer specific mortality to be as high as 58% (adjusted OR = 0.42; 95% CI 0.33-0.53) for screened compared to unscreened women.

The study included five of the nine regional screening organisations (which cover more than half of the target population for screening in the Netherlands): Stichting Bevolkingsonderzoek Noord-Nederland (BBNN), Stichting Kankerpreventie IKA, Stichting Kankerpreventie en screening Limburg (SKsL), Stichting Bevolkingsonderzoek Borstkanker Zuidwest Nederland (SBBZWN), and Stichting Vroege Opsporing Kanker Oost-Nederland (SVOKON).

Our estimate of 43.5% (odds ratio of 0.57) breast cancer mortality reduction due to screening with 100% attendance to screening lies outside of the confidence interval of the pooled odds ratio in the study. However, the model estimate is within the confidence interval of each individual region.

The estimates of reduction in breast cancer mortality of the Dutch best evidence could not be reproduced well, as the model estimates were considerably lower. The model estimates do therefore not overestimate the screening effect, as has been argued by critics, and may even be conservative.

Table S5. Estimates of breast cancer mortality reduction in Dutch best evidence[13] and modelled estimate for the same period of time and age group, screened vs. un-screened women.

Study	Age Group	Study Period	Study Estimates		Model Estimate [a]	
			Odds Ratio (95% CI)*	BC Mortality Reduction	Odds Ratio (95% CI) *	BC Mortality Reduction
Paap et al. 2014	50–75	2004–2005	Pooled 0.42 (0.33–0.53) BBNN 0.40 (0.22–0.74) IKA 0.38 (0.25–0.57) SKsL 0.24 (0.10-0.62) SBBZWN 0.49 (0.30–0.78) SVOKON 0.51 (0.30–0.87)	58%		43.5%

[a] 100% Attendance to screening assumed. * Estimates with correction for self-selection bias. Abbreviations: confidence interval (CI); breast cancer (BC).

7.2.2. Finland

Heinävaara[7] evaluated the long-term effect of organized mammography screening on incidence based mortality (IBM) in Finland in 1992–2011 among 50–84-year-old women using a case–control design with non-restrictive eligibility criteria of controls. Organised screening decreases mortality from breast cancer by 33% in women attending screening (HR = 0.67 [0.49–0.90], corrected for self-selection bias).

Table S6. Estimates of breast cancer mortality reduction in Finish best evidence[7] and modelled estimate for the same period of time and age group, screened vs. un-screened women.

Study	Age Group	Study Period	Study Estimates		Model Estimate [a]	
			Hazard Ratio (95% CI)	BC Mortality Reduction	Hazard Ratio (95% CI) *	BC Mortality Reduction
Heinävaara et al. 2016	50–84	1992–2011	0.67 (0.49–0.90)	33%	0.77	22%

[a] 100% Attendance to screening assumed. Abbreviations: confidence interval (CI); breast cancer (BC).

The simulation of the calibrated Finish MISCAN model could replicate the estimate for breast cancer specific mortality reduction due to screening. The model results underestimate the effect of BC screening on mortality reduction, but it is within the confidence interval of the estimates from Heinävaara et al.

7.2.3. Italy

The aim of the case–control study of Pulity[12] has been to evaluate the effectiveness of service screening programmes in reducing breast cancer mortality in the Italian areas participating in the IMPACT study (Piedmont, Tuscany, Umbria, Veneto, Emilia-Romagna). For women between 50 and 74 years who died between screening activation and 2002, the odds ratio comparing screened with unscreened women was 0.50 (95% CI: 0.42–0.59).

Table S7. Estimates of breast cancer mortality reduction in Italian best evidence [12] and modelled estimate for the same period of time and age group, screened vs. un-screened women.

Study	Age Group	Study Period	Study estimates		Model estimate [a]	
			Odds Ratio (95% CI)	BC Mortality Reduction	Odds Ratio (95% CI)*	BC Mortality Reduction
Puliti et al. 2008	50-74	1990-2002	0.50 (0.42–0.60)	50%	0.44	56%

[a] 100% Attendance to screening assumed. Abbreviations: confidence interval (CI); breast cancer (BC).

The estimate for breast cancer specific mortality reduction due to screening in Italy could be replicated with the simulation of the calibrated Italian MISCAN model. The model results overestimate the effect of BC screening on mortality reduction, but it is within the confidence interval of the estimates from Puliti et al.

7.2.4. Slovenia

Due to the lack of best evidence from Eastern Europe, we alternatively tried to validate the Slovenian model by comparing the observed and modelled Interval cancer rates–provided by the EU-TOPIA consortium members.

Per 5 year age-group, we modelled the clinically detected cancers in 2014, detected 12 and 24 months after the last negative screen (first/subsequent).

Table S8. Observed interval cancer rate in Slovenia and modelled estimate for the same period of time and age group.

		Interval Cancer Rate per 1,000 Visits			
		Diagnosed within the first year after test		Diagnosed within the second year after test	
		observed	modelled	observed	modelled
Initial[1]	50–54	0.32	0.76	1.29	0.89
	55–59	0	0.49	1.68	0.84
	60–64	1.08	0.69	0.36	1.18
	65–69	0	0.20	1.65	0.60
Subsequent[2]	50–54	0.73	0.31	0.24	0.52
	55–59	0.19	0.48	1.14	0.75
	60–64	0.38	0.57	0.95	0.79
	65–69	0.71	0.36	0.95	0.79

[1] Initial screening is the first screening examination of women within the screening programme, regardless of the organisational screening round in which the examination takes place. [2] Subsequent screening includes all screening examinations of women within the screening programme following an initial screening examination, regardless of the organisational screening round in which the examination takes place.

Keeping in mind that interval cancers are based on rather small numbers, we are satisfied with the results.

B) RESULTS OF THE SENSITIVITY ANALYSIS

Table S9. Harms-to-benefit-ratios in response to variation in input parameters, per country and screening strategy.

		Harm-to-Benefit-Ratios																			
		Overdiagnosed BC cases / BC Deaths Averted					False-Positives/ BC Deaths Averted					Overdiagnosed BC Cases/ LY Gained					False-Positives/ LY Gained				
Country	Screening strategy	Sens Highest	Sens Lowest	Referral Highest	Referral Lowest	Cover Age	Sens Highest	Sens Lowest	Referral Highest	Referral Lowest	Cover Age	Sens Highest	Sens Lowest	Referral Highest	Referral Lowest	Cover Age	Sens Highest	Sens Lowest	Referral Highest	Referral Lowest	Cover Age
Slovenia	45–74	0.51	0.52	0.52	0.52	0.55	43.36	49.97	85.01	36.38	38.34	0.04	0.04	0.04	0.04	0.04	3.30	3.82	6.50	2.78	2.96
	45–69	0.41	0.43	0.43	0.43	0.45	45.25	52.69	89.44	38.48	41.55	0.03	0.03	0.03	0.03	0.03	3.16	3.69	6.26	2.69	2.93
	50–74	0.54	0.55	0.55	0.55	0.58	31.49	36.32	66.14	23.01	27.65	0.04	0.04	0.04	0.04	0.05	2.57	2.98	5.43	1.89	2.28
	50–69*	0.44	0.45	0.45	0.45	0.48	32.26	37.65	68.44	23.97	29.52	0.03	0.03	0.03	0.03	0.04	2.43	2.83	5.16	1.81	2.23
Finland	45–74	0.40	0.41	0.38	0.38	0.38	31.19	36.95	75.70	31.92	29.20	0.03	0.03	0.03	0.03	0.03	2.36	2.79	5.72	2.41	2.20
	45–69	0.35	0.41	0.32	0.32	0.32	34.17	36.95	82.98	35.16	32.66	0.02	0.03	0.02	0.02	0.02	2.33	2.79	5.65	2.40	2.22
	50–74	0.54	0.53	0.51	0.51	0.50	25.57	29.16	59.40	20.59	23.67	0.04	0.04	0.04	0.04	0.04	2.07	2.36	4.80	1.67	1.90
	50–69*	0.36	0.33	0.34	0.34	0.34	26.88	30.51	63.39	21.65	25.47	0.03	0.02	0.02	0.02	0.02	1.96	2.23	4.63	1.58	1.85
Netherlands	45–74	0.30	0.28	0.30	0.30	0.30	16.79	19.17	45.86	18.78	16.22	0.02	0.02	0.02	0.02	0.02	1.18	1.35	3.21	1.32	1.14
	45–69	0.23	0.21	0.23	0.23	0.24	16.89	19.41	46.03	18.99	16.59	0.02	0.01	0.02	0.02	0.02	1.11	1.28	3.01	1.24	1.09
	50–74*	0.32	0.30	0.32	0.32	0.32	10.29	11.90	36.26	11.64	9.80	0.02	0.02	0.02	0.02	0.02	0.77	0.90	2.73	0.87	0.74
	50–69	0.25	0.23	0.25	0.25	0.26	10.16	11.84	35.72	11.57	9.91	0.02	0.02	0.02	0.02	0.02	0.71	0.83	2.50	0.81	0.70
Italy	45–74	0.28	0.26	0.28	0.28	0.30	54.97	61.45	58.23	24.59	47.29	0.02	0.02	0.02	0.02	0.02	3.87	4.35	4.11	1.74	3.40
	45–69	0.21	0.19	0.22	0.22	0.23	55.32	62.34	58.69	24.90	48.46	0.01	0.01	0.01	0.01	0.02	3.61	4.09	3.84	1.63	3.22
	50–74	0.29	0.28	0.30	0.30	0.32	43.68	48.70	46.29	15.74	37.46	0.02	0.02	0.02	0.02	0.02	3.30	3.69	3.51	1.19	2.89
	50–69*	0.23	0.21	0.23	0.23	0.24	43.08	48.46	45.73	15.63	37.63	0.02	0.01	0.02	0.02	0.02	3.02	3.40	3.21	1.10	2.69

Sens highest: application of the highest values for stage-specific sensitivity across all countries; Sens lowest: application of the lowest values for stage-specific sensitivity across all countries; Referral highest: application of the highest referral rates across all countries; Referral lowest: application of the lowest referral rates across all countries. Attendance: application of observed attendance; * Current screening strategy. BC: breast cancer; LY: Life years.

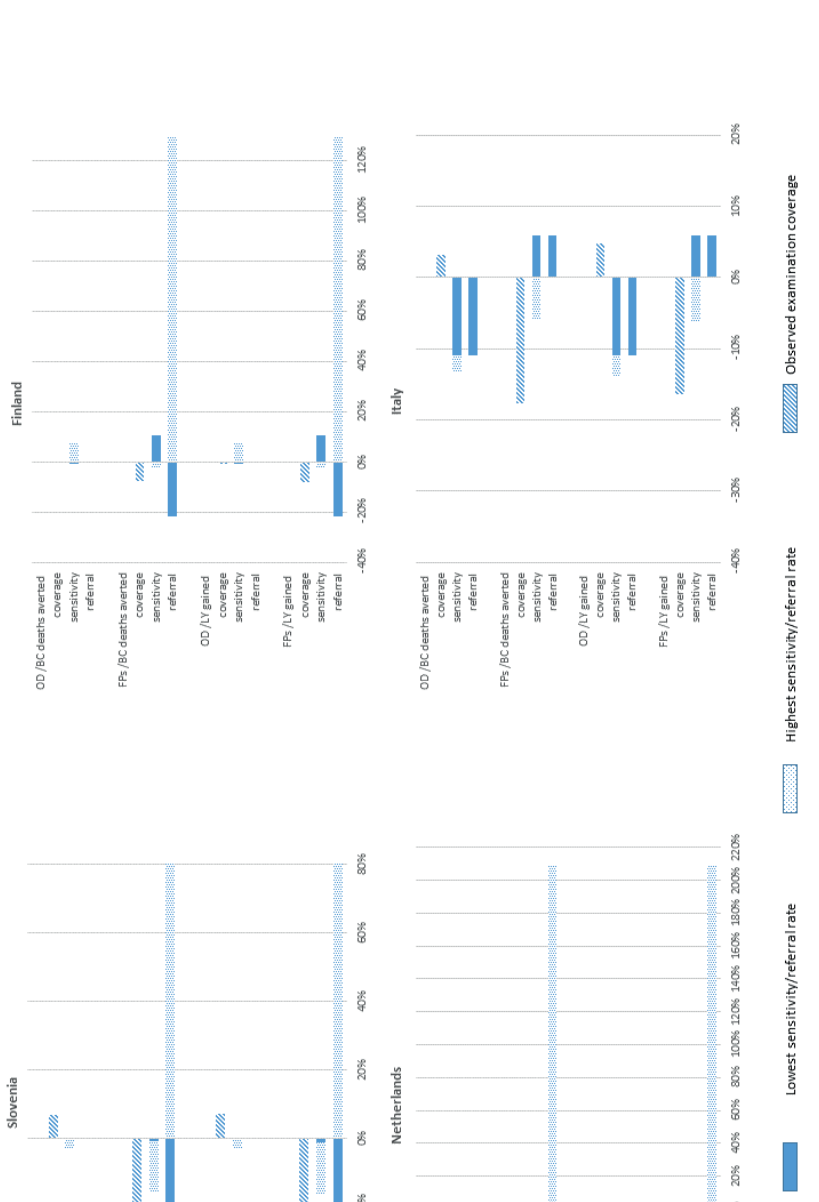
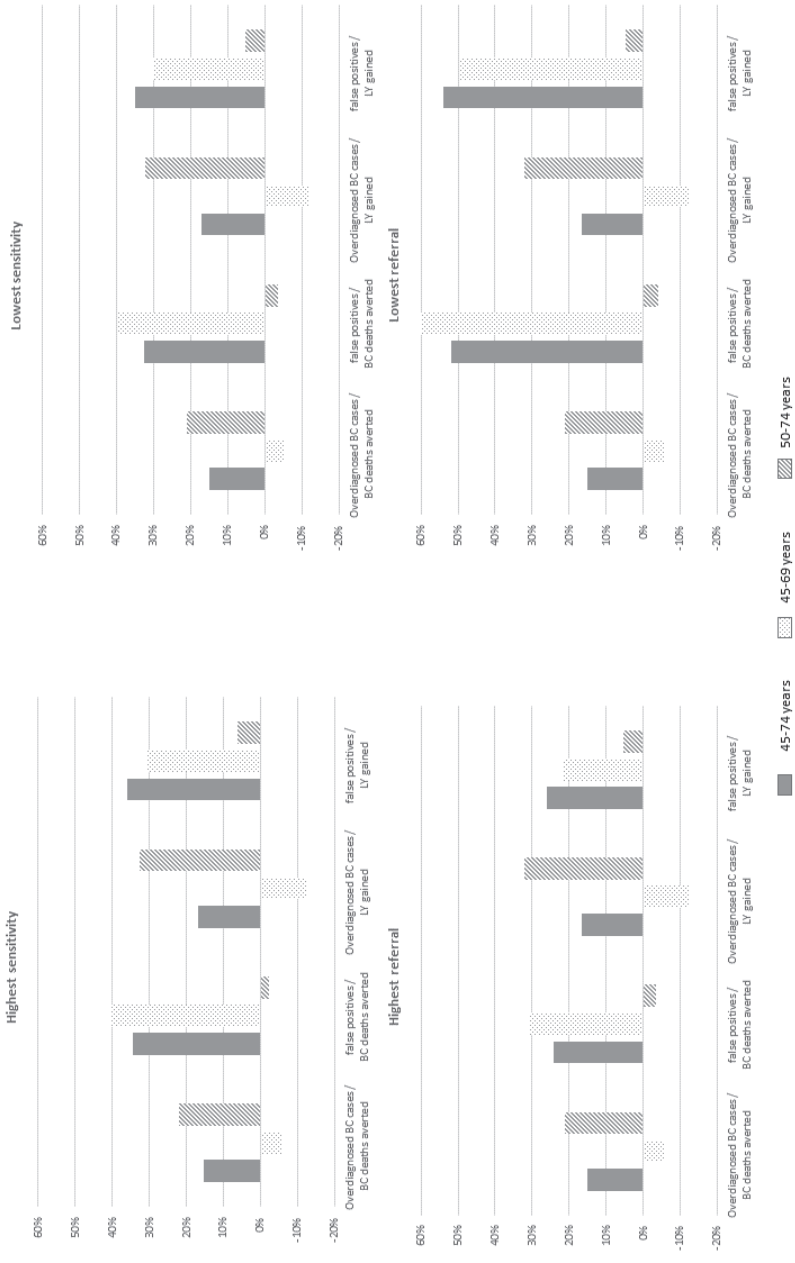


Figure S12. Percentage change in harms-to-benefit-ratios in response to variation in input parameters, per country, age-group 50–69

The figures contain—per country—the results of the sensitivity analysis for the reference age-group 50–69. For each of the four ratios, the bars present the percentage changes in response to variations in input parameters, compared to the base analysis. These variations include application of observed examination coverage from each country (striped bar), and application of the highest values for stage-specific sensitivity and referral rates (dotted bars) and the lowest values for stage-specific sensitivity and referral rates (blue bars). OD: Overdiagnosis; BC: Breast cancer; FP: False positive results; LY: Life years


Highest sensitivity
Lowest sensitivity
Highest referral
Lowest referral
60%
50%
40%
30%
20%
10%
0%
-10%
-20%
Overdiagnosed BC cases / BC deaths averted
false positives / BC deaths averted
Overdiagnosed BC cases / LY gained
false positives / LY gained
45-74 years
45-69 years
50-74 years

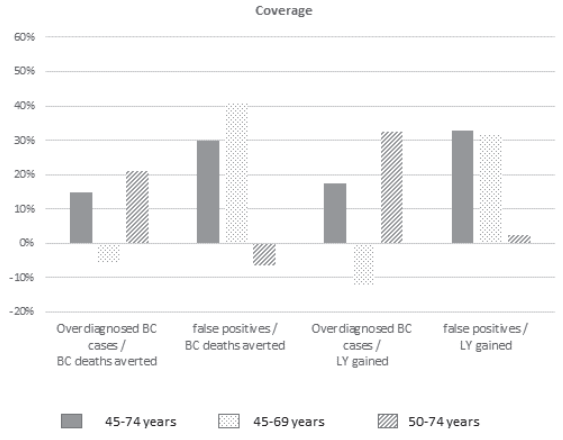
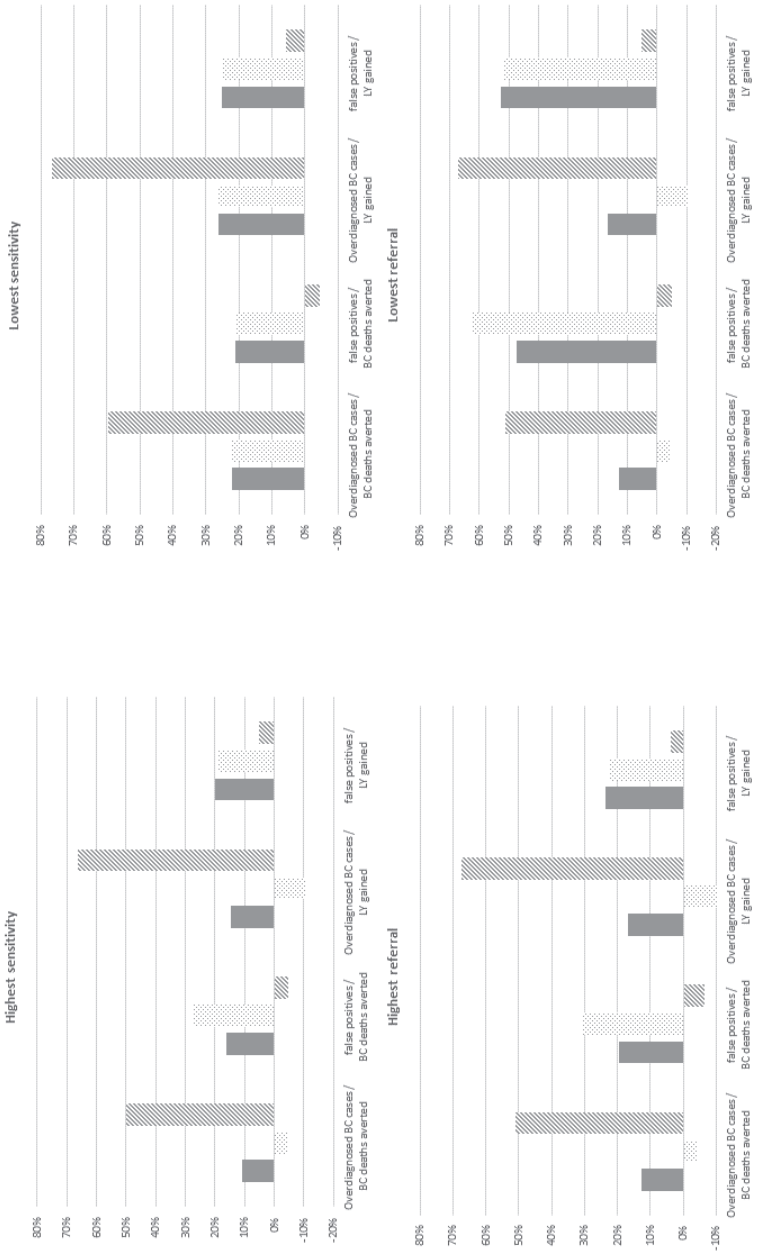


Figure S13. Percentage change in harms-to-benefit-ratios in comparison to the reference age-group 50–69, per screening scenario and varied parameter of the sensitivity analysis. SLOVENIA.


Highest sensitivity
Lowest sensitivity
Highest referral
Lowest referral
Overdiagnosed BC cases / BC deaths averted
False positives / BC deaths averted
Overdiagnosed BC cases / LY gained
False positives / LY gained
45-74 years
45-69 years
50-74 years

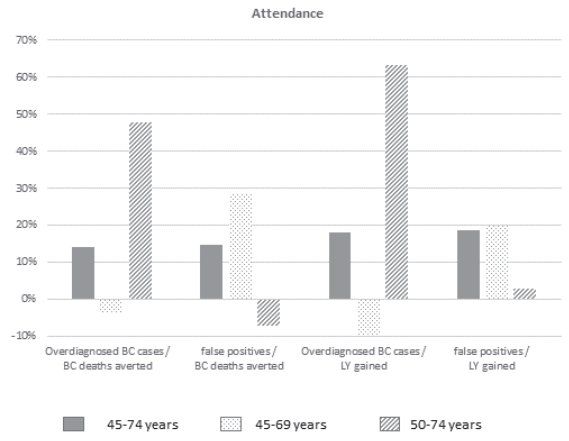
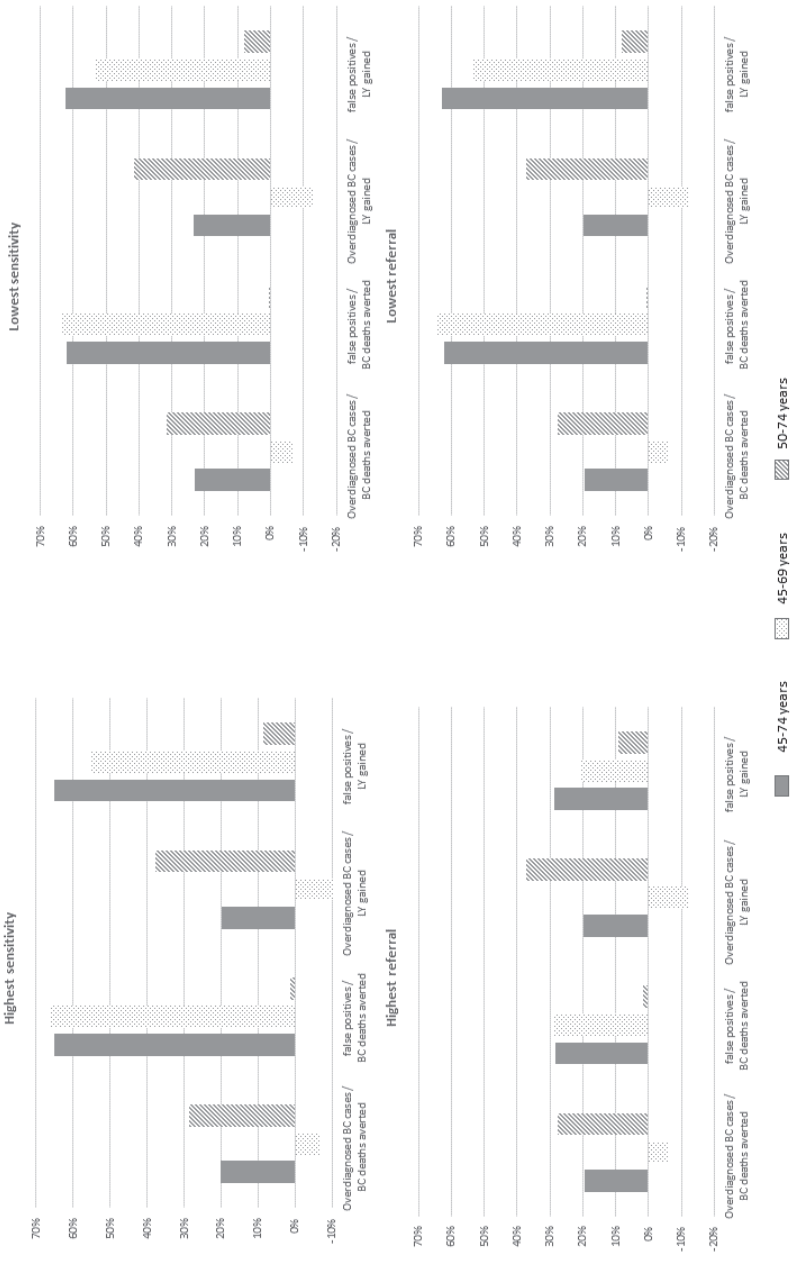


Figure S14. Percentage change in harms-to-benefit-ratios in comparison to the reference age-group 50–69, per screening scenario and varied parameter of the sensitivity analysis. FINLAND.


Highest sensitivity
Highest referral
Lowest sensitivity
Lowest referral
70%
60%
50%
40%
30%
20%
10%
0%
-10%
-20%
Overdiagnosed BC cases / BC deaths averted
false positives / BC deaths averted
Overdiagnosed BC cases / LY gained
false positives / LY gained
45-74 years
45-69 years
50-74 years

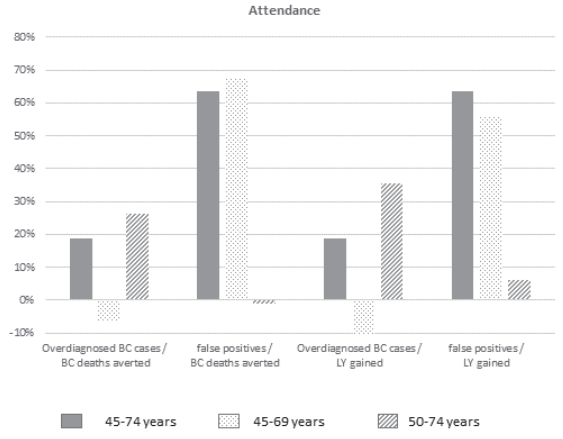
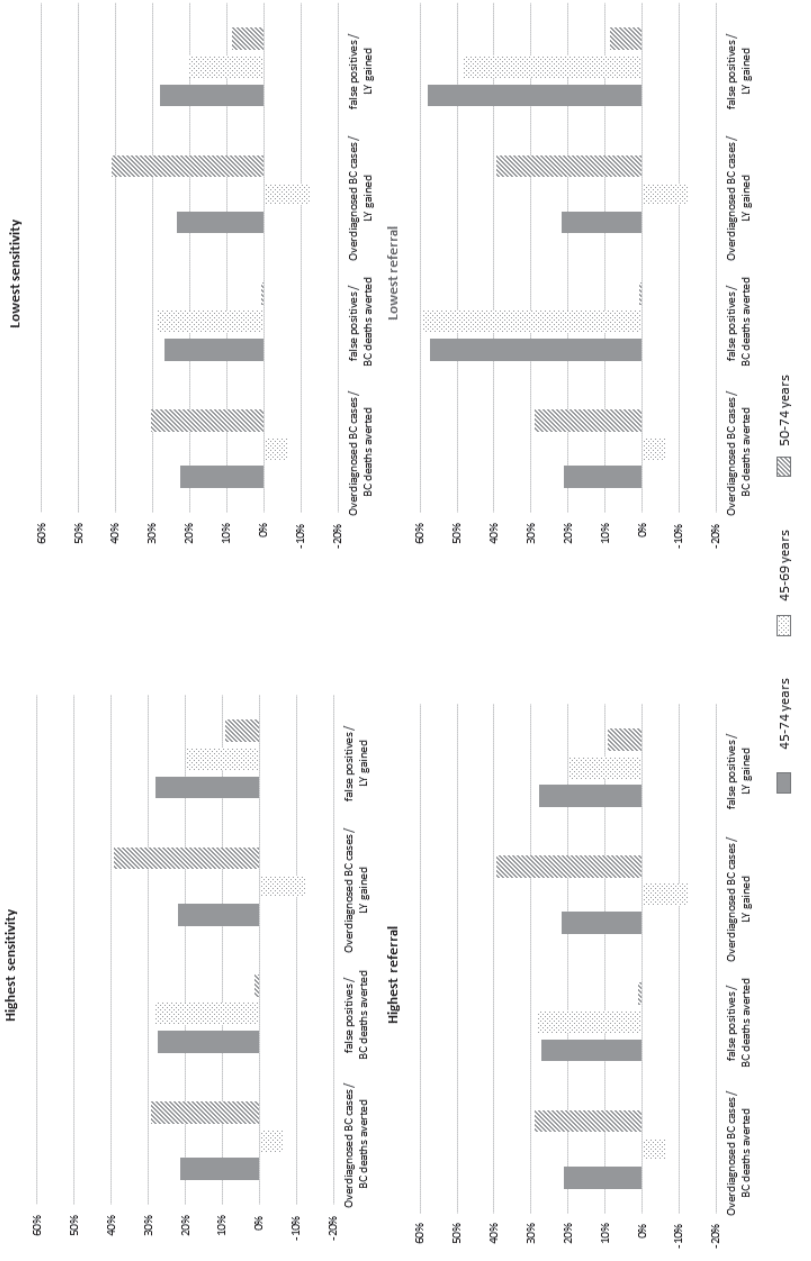


Figure S15. Percentage change in harms-to-benefit-ratios in comparison to the reference age-group 50–69, per screening scenario and varied parameter of the sensitivity analysis. THE NETHERLANDS.


Highest sensitivity
Lowest sensitivity
Highest referral
Lowest referral
60%
50%
40%
30%
20%
10%
0%
-10%
-20%
Overdiagnosed BC cases / BC deaths averted
false positives / BC deaths averted
Overdiagnosed BC cases / LY gained
false positives / LY gained
45-74 years
45-69 years
50-74 years

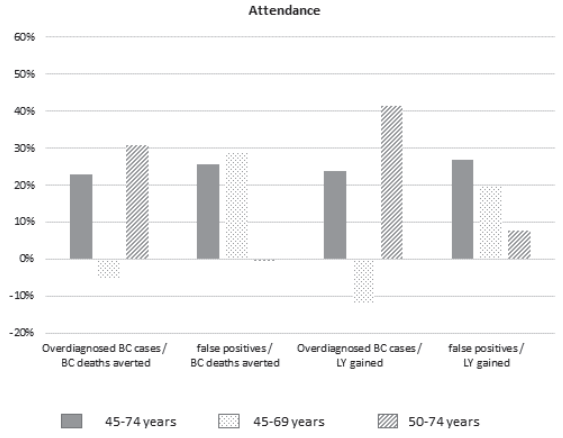


Figure S16. Percentage change in harms-to-benefit-ratios in comparison to the reference age-group 50–69, per screening scenario and varied parameter of the sensitivity analysis. ITALY

REFERENCES

1. Bjurstam, N.; Björneld, L.; Warwick, J.; Sala, E.; Duffy, S.W.; Nyström, L.; Walker, N.; Cahlin, E.; Eriksson, O.; Hafström, L.-O.; et al. The Gothenburg breast screening trial. Cancer 2003, 97, 2387–2396.
2. Tabar, L.; Vitak, B.; Chen, H.H.; Duffy, S.W.; Yen, M.F.; Chiang, C.F.; Krusemo, U.B.; Tot, T.; Smith, R.A. 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, 625-+.
3. Nyström, L.; Andersson, I.; Bjurstam, N.; Frisell, J.; Nordenskjöld, B.; Rutqvist, L.E. Long-term effects of mammography screening: Updated overview of the Swedish randomised trials. Lancet 2002, 359, 909–919.
4. de Koning, H.J.; Boer, R.; Warmerdam, P.G.; Beemsterboer, P.M.; van der Maas, P.J. Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. J. Natl Cancer Inst. 1995, 87, 1217–1223.
5. van Oortmarssen, G.J.; Habbema, J.D.; van der Maas, P.J.; de Koning, H.J.; Collette, H.J.; Verbeek, A.L.; Geerts, A.T.; Lubbe, K.T. A model for breast cancer screening. Cancer 1990, 66, 1601–1612.
6. Sankatsing, V.D.; Heijnsdijk, E.A.; van Luijt, P.A.; van Ravesteyn, N.T.; Fracheboud, J.; de Koning,H.J. Cost-effectiveness of digital mammography screening before the age of 50 in The Netherlands. Int. J. Cancer 2015, 137, 1990–1999.
7. Heinavaara, S.; Sarkeala, T.; Anttila, A. Impact of organised mammography screening on breast cancer mortality in a case-control and cohort study. Br. J. Cancer 2016, 114, 1038–1044.
8. Anttila, A.; Sarkeala, T.; Hakulinen, T.; Heinävaara, S. Impacts of the Finnish service screening programme on breast cancer rates. BMC Public Health 2008, 8.
9. Koleva-Kolarova, R.G.; Zhan, Z.; Greuter, M.J.W.; Feenstra, T.L.; De Bock, G.H. Simulation models in population breast cancer screening: A systematic review. Breast 2015, 24(4), 354–363.
10. Eddy, D.M.; Hollingworth, W.; Caro, J.J.; Tsevat, J.; McDonald, K.M.; Wong, J.B.; ISPOR-SMDM Modeling Good Research Practices Task Force. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Med. Decis. Making 2012, 32(5), 733–743.
11. Zielonke, N.; Gini, A.; Jansen, E.; Anttila, A.; Segnan, N.; Ponti, A.; Veerus, P.; de Koning, H.J.; van Ravesteyn, N.T.; Heijnsdijk, E.A.M.; et al. Evidence for reducing cancer specific mortality due to screening for breast cancer in Europe: a systematic review. Eur. J. Cancer 2020, 127, 191–206.
12. Puliti, D.; Miccinesi, G.; Collina, N.; De Lisi, V.; Federico, M.; Ferretti, S.; Finarelli, A.C.; Foca, F.; Mangode, L.; Naldoni, C.; et al. Effectiveness of service screening: A case-control study to assess breast cancer mortality reduction. Br. J. Cancer 2008, 99, 423–427.
13. Paap, E.; Verbeek, A.L.M.; Botterweck, A.A.M.; van Doorne-Nagtegaal, H.J.; Imhof-Tas, M.; de Koning, H.J.; Otto, S.J.; de Munck, L.; van der Steen, A.; Holland, R.; et al. Breast cancer screening halves the risk of breast cancer death: A case-referent study. Breast 2014, 23, 439–444.

Chapter 6

Overcoming barriers: modelling the effect of potential future changes of organized breast cancer screening in Italy

Nadine Zielonke, Carlo Senore, Antonio Ponti, Marcell Csanadi, Harry J. de Koning, Nicolien T. van Ravesteyn, Eveline A.M. Heijnsdijk, on behalf of the EU-TOPIA consortium.

Submitted

ABSTRACT

Objectives

Organized breast cancer screening in Italy started in the early 1990s. No nationally organized program exists, but regionally implemented initiatives with national coordination. In practice, screening does not achieve its full potential due to organisational and cultural barriers. The objective of this paper is to identify potential changes, their costs and effects, to overcome these barriers and to further optimize breast cancer screening in Italy.

Method

We simulated several scenarios using an online tool (EU-TOPIA evaluation tool) based on the MIcrosimulation SCreening ANalysis (MISCAN) model to assess the effects of potential measures to improve the breast cancer screening program in Italy. We performed an economic evaluation for increasing adherence for Southern Italy and harmonising screening intervals for the whole of Italy.

Results

Fully overcoming the barrier of low adherence in Southern Italy is predicted to almost triple the additional life-years between 2020 and 2050 compared to current screening while breast cancer mortality reduction increases from 4.5% to 14.5%. This future scenario has an acceptable ICER of €9,531 per QALY gained. Harmonising screening intervals and biennially screening all eligible women in Italy is predicted to gain 1% fewer life-years, while saving 18% of the total screening expenditures. Thus, this change would lead to a substantial decrease in costs, while maintaining almost all benefits.

Conclusion

Removing the most important barriers of breast cancer screening in Italy could result in substantial improvements at acceptable costs. The used online tools are valuable for stakeholders to quantify benefits, harms and costs of early cancer detection in Europe.

Abbreviations

AIRTUM – Associazione Italiana Registri Tumori (Italian Association of Cancer Registries)
DCIS – Ductal Carcinoma in Situ
GISMa – Italian group for mammography screening
GP – General Practitioner
LYG – Life years gained
ICERs – Incremental cost-effectiveness ratios
MISCAN – MIcrosimulation SCreening ANalysis
NHS – Italian National Health Service
QALYs – Quality adjusted life years
ONS – Osservatorio nazionale screening (National centre for screening monitoring)
WTP – Willingness to pay

INTRODUCTION

Early detection and diagnosis of breast cancer are effective ways to lower the burden of the disease[1-3]. In Italy, organized breast cancer screening began in the 1990 in the city of Florence, with inviting the 50–69-year-old female population. More regions followed and Italian guidelines have been initiated to recommend inviting women aged 50–69 to undergo mammography every two years, in accordance with EU Recommendations[4]. At present, there is no nationally organized breast cancer screening program in Italy, but regional health authorities have a mandate to implement breast cancer screening programs, following national guidelines. All resident women in the target age range must be invited, and all screening tests, ascertainment and treatments are free of charge[5]. The performance of all regional screening programs is regularly monitored by the National centre for screening monitoring (Osservatorio nazionale screening, ONS) on behalf of the Ministry of Health.

Despite the early detection efforts, in practice, the cancer screening programs fail to achieve their full potential. The European Union funded EU-TOPIA[6] (Towards improved screening for breast, cervical and colorectal cancer in all of Europe) project, which aims to improve health outcomes and equity of breast, cervical and colorectal cancer screening programs across Europe, developed a self-assessment tool for the purpose of identifying the barriers to the optimal operation of population-based breast cancer screening programs[7]. Details of this tool are described in the Supplementary materials Part 1. For Italy, barriers have been identified (Supplementary Table A1) that are common to those identified in other countries[8]. The introduction of screening programs in Italy has been slow and characterised by profound geographical differences[9 10]. Despite efforts to reduce and overcoming heterogeneity in screening between regions, discrepancies in adherence to breast cancer screening still exist, with southern regions still showing low attendance rates that do not reach the recommended minimum of the European Commission[5 9 11 12]. For this region, the invitation coverage was only 59% in 2018 and the adherence to the invitation was 38%[13]. Looking at Italy nationally, opportunistic and organized screening for breast cancer coexist, which complicates having comprehensive information about screening coverage[9 14]. Another effect of opportunistic screening is that it might lead to a screening interval that is not in line with EU Recommendations[4] because sub-groups of eligible women are using opportunistic screening options in addition to the program - thus leading to overscreening and inefficient utilization of resources. At the same time, another sub-group of eligible women has extended screening intervals due to a lack of human and financial resources within the screening organization and screening process – thus leading to underscreening.

The objective of this paper is to identify potential future changes (i.e. practical solutions) that can be initiated to the breast cancer screening programs - as well as their costs and benefits - to overcome these barriers and to improve screening in Italy. In this modelling study we want to

quantify the harms, benefits and costs of organized screening strategies for breast cancer in Italy.

METHODS

MISCAN breast model

We used the MIcrosimulation SCreening ANalysis (MISCAN) model, which has been described in detail elsewhere[15]. In brief, MISCAN simulates individual life histories and assesses the consequences of introducing a screening program on these life histories using the Monte Carlo method. Possible events in the life histories are birth and death of a person, onset of a pre-clinical ductal carcinoma in situ (DCIS), transitions between disease states, participation in screening and screen- or clinical detection of a cancer[15 17]. Originally, the model was calibrated to the Dutch situation[17].

The EU-TOPIA evaluation tool

Effects of chosen measures to further improve the breast cancer screening in Italy were assessed using the EU-TOPIA evaluation tool, which is based on the MISCAN model. We designed an online platform that allows stakeholders to use their country-specific data to quantify future harms and benefits of different cancer screening scenarios in their country (or region within a country)[18]. In this tool, stakeholders can download an Excel-template to provide country-specific data on their population demography, breast cancer epidemiology and breast cancer screening program. Once the stakeholders have filled out the data template, they can upload it and simulate the current cancer screening programs and impacts of potential changes in screening protocols (such as changing target ages or increasing screening attendance). The results are scaled to the country-specific population. A detailed user's guide can be found in the Supplementary Methods Part 2.

The evaluation tool was developed for three cancer sites (breast cancer, cervical cancer, and colorectal cancer) according to the same principles. In this paper, we use the breast cancer version of this tool. Detailed information on the cancer specific versions can be found on the tool's website[19]. The application and results of the colorectal version of the tool were reported in a previous publication[18]. The current version of the tool requires a registration (Supplementary Methods Part 2), and a direct approval is granted to European stakeholders involved in the field of cancer screening.

Model adjustments to the Italian situation

We adjusted and calibrated the MISCAN model to reflect the Italian demography (i.e. age distribution of the population and life expectancy), disease risk (i.e. breast cancer incidence and

stage distribution) and potential differences in the natural history of breast cancer. In developing the Italian model version, we used a specific calibration process described in Zielonke et al.[15]. Observed data on breast cancer incidence and stage distribution were provided by the (pooled data from 38) regional cancer registries (AIRTUM), while data about attendance and outcomes of the screening programs were provided by ONS and the Italian group for mammography screening GISMa. We calibrated the model by fitting the onset rate by age and the stage-specific dwell times against the observed incidence rate for the period 2006-2009. Opportunistic screening was included based on expert opinion[20]. We first validated the model replicating the data that was used in the calibration process (internal validation). Then, we externally validated the model against best evidence based on a recently published systematic review on breast cancer mortality reductions due to screening[2].

Simulated screening scenarios

For this study, we evaluated the effect of two barriers on screening outcomes and how to potentially overcome them in two separate analyses: the first one was focused on Southern Italy. The second one was focused on the whole of Italy, characterized by a combination of different sub-population screened at different intervals due to opportunistic screening and capacity barriers (Table 1). Screening performance indicators for ongoing programs were derived by the data provided in the EU-TOPIA monitoring tool, collecting quantitative data about screening activity.

First, we looked at the particular situation in Southern Italy, which comprises the regions of Abruzzo, Molise, Campania, Basilicata, Puglia, Calabria, Sicilia and Sardegna. We simulated the current screening (scenario 1) assuming biennial mammography screening from 50-69 with an adherence in 2013 (report year) of 14%, as presented in Table 1. In this study, adherence of screening was specified as the proportion (%) of the target population screened in the chosen report year after invitation. Then, we investigated the impact of investing in mobile screening units. Available evidence supports our hypothesis that short distance to facilities and generally increased screening capacities could increase adherence to screening[21]. Thus, we simulated four additional scenarios where we increase adherence by 20%, 30%, 40% and up to a maximum of 100% - equally added to the reported adherence of each of the 5-year age-groups. For the second analysis, we split the female Italian population into sub-groups. Based on expert opinion, we assumed that of all eligible women (ages 50-69), 20% are screened annually due to a combination of organized screening and opportunistic screening utilisation. Thirty percent of all eligible women are assumed to be screened with an extended screening interval of 28 months while half (50%) of all eligible women are screened biennially. We simulated the sub-groups separately and calculated their weighted average of the outcome parameters to obtain national results. In preparation for the simulation with the evaluation tool, we prepared the three separate Excel templates according to the sub-group characteristics presented in Table 1. In order to

fairly compare strategies consisting of different screening intervals, we adjusted the three Excel templates in a way, that invitational coverage, participation coverage and adherence were set to 100% while all other quality parameters (e.g. further assessment rate or detection rate) were adjusted to match the observed Italian data.

Table 1. Overview of assumptions of screening behaviour and simulated screening scenarios, per barrier and scenario

	Barrier 1: Southern Italy				
	Scenario 1[a]	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Screening interval (years)	2,0	2,0	2,0	2,0	2,0
Adherence by age[b]					
50-54	11,5%	31,5%	41,5%	51,5%	100,0%
55-59	14,0%	34,0%	44,0%	54,0%	100,0%
60-64	15,2%	35,2%	45,2%	55,2%	100,0%
65-69	14,8%	34,8%	44,8%	54,8%	100,0%
Average	*13,9%*	*33,9%*	*43,9%*	*53,9%*	*100,0%*

	Barrier 2: Italy (national)[c]				
	Sub-group 1: overscreened		Sub-group 2: underscreened		Sub-group 3: biennial
	Scenario 1[a]	Scenario 2	Scenario 1[a]	Scenario 2	Scenario 1[a]
Screening interval (years)	1,0	2,0	2,4	2,0	2,0
Adherence by age[b]					
50-54	100,0%	100,0%	100,0%	100,0%	100,0%
55-59	100,0%	100,0%	100,0%	100,0%	100,0%
60-64	100,0%	100,0%	100,0%	100,0%	100,0%
65-69	100,0%	100,0%	100,0%	100,0%	100,0%

[a] For each barrier/ sup-group, scenario 1 represents the current screening scenario.
[b] Adherence is specified as the proportion (%) of the target population screened in the index year after invitation.
[c] The proportions of the sub-groups of the female population eligible for screening are 20% for group 1, 30% for group 2 and 50% for group 3.

For each of the sub-groups, we simulated two scenarios: Scenario 1 represents the current screening (with existing barriers) and scenario 2 represents biennial screening from 2020 onwards (without this barrier). For the overscreened sub-group, we investigated the impact of investing in legislative measures, as well as on a training and information campaign for radiologists and general practitioners (GPs), to stop spontaneous activity in the Italian National Health Service (NHS) facilities. For the underscreened sub-group, we investigated the impact of investing in mobile screening units across the whole country to increase screening capacity and to subsequently decrease the screening interval.

Model outcomes and Cost-effectiveness analyses

Using the EU-TOPIA evaluation tool, the life histories of 10 million women were simulated. For each simulated scenario we estimated the breast cancer outcomes in (Southern) Italy for women aged 40-100 years in the period 2020-2050. These outcomes include the predicted number of breast cancer cases (DCIS and invasive), breast cancer deaths, breast cancer mortality reduction (% compared to no screening), number of false positives and overdiagnosis (as % of screen detected breast cancers), and resources required by the breast cancer screening program, such as number of screening tests and number of referrals. In addition, two ratios of harms and benefits are provided for each scenario: the number of screens needed to prevent one breast cancer death and the number of false positives per breast cancer death prevented.

Subsequently, we performed a cost-effectiveness analysis for each simulated scenario from a healthcare payer perspective and calculated direct medical costs including costs of screening, diagnostics and treatment. Costs of a screening mammogram were based on data from Foglia et al[22] and Mantellini et al[23]. We used costs of diagnosis and treatment costs by cancer stage reported by Fransisci et al[24]. In order to increase screening adherence in Southern Italy (i.e. to overcome barrier 1), we considered a maximum of 10 (à €180.000) mobile screening units a promising investment. In order to overcome barrier 2, we considered two different investments separately. First, we would invest €2.500.000 to stop opportunistic screening. Second, we would invest €18 million for a maximum of 99 mobile screening units for the underscreened sub-group. An overview of all costs and the disutility values is shown in Table 2.

We used the cost-effectiveness tool, which is integrated into the EU-TOPIA evaluation tool. When all simulations were completed, we counted several key outcomes, costs as well as quality adjusted life years (QALYs) gained for each scenario. Costs were discounted at 3% per year. Subsequently, in order to compare screening scenarios, incremental cost-effectiveness ratios (ICERs) were calculated for the strategies that try to overcome each barrier (Scenario 5 for barrier 1, scenario 2 for barrier 2), thus screening 100% of the eligible women in Southern Italy and screen all eligible women every two years. ICERs are the difference in costs divided by the difference in QALYs between a strategy and the current situation (Scenario 1 for each of the barriers). The ICER of a strategy therefore reflects the costs required to generate one additional QALY, compared to the current strategy. To compare the ICERs, we applied a willingness to pay (WTP) threshold of €23.000 per QALY gained. This threshold is frequently used in the international literature[17 25].

Table 2. Costs and disutilities used in the cost-effectiveness analysis, per barrier and scenario

	Costs (€)[a]								
	Barrier 1		Barrier 2						
Variable	Scenario 1	Scenario 5	Sub-group 1		Sub-group 2		Sub-group 3	Disutilities	Duration
Investment costs[b]	-	1,800,000	-	2,500,000	-	17,820,000	-		
Screening invitation letter	3	3	1.5[c]	3	3	3	3		
Mammography	45	45	60[d]	45	45	45	45	0.006	1 week
Diagnosis	190	190	252[d]	190	190	190	190	0.105	1 month
Treatment of DCIS	9.204	9.204	9.204	9.204	9.204	9.204	9.204	0.100	2 years
Treatment of T1A	11.816	11.816	11.816	11.816	11.816	11.816	11.816	0.100	2 years
Treatment of T1B	14.483	14.483	14.483	14.483	14.483	14.483	14.483	0.100	2 years
Treatment of T1C	14.272	14.272	14.272	14.272	14.272	14.272	14.272	0.250	2 years
Treatment of T2+	20.522	20.522	20.522	20.522	20.522	20.522	20.522	0.250	2 years
Death from breast cancer								0.400	6 months

[a] Costs were discounted at 3% per year.
[b] Investment costs are specified as an additional investment secondary to the resources routinely spent for the operation of the current program, in order to overcome the respective barrier.
[c] We assume half of the exams in this group to be performed outside the organised setting. Those women will not receive an invitation, thus these costs are lower for this sub-group.
[d] We adapted the costs of the screening process (mammography and diagnosis) for the sub-group of overscreened women based on results from Mantellini et al. They detected cost differences between opportunistic and organised screening in seven screening centres in Italy. Thus, the exams in this group that are performed outside the organised setting (50%) are 1,65 times more expensive.

Sensitivity analysis

To evaluate how assumptions and parameter values affected the screening outcomes and costs per QALY, we performed two sensitivity analyses for barrier 2. First, we used observed or estimated adherence of each sub-group instead of 100%: 75% adherence in the group of overscreened women, 45-50% adherence in the group of underscreened women and the observed 55-60% national average for biennial screening. Second, we varied the sub-group distribution for the simulations of barrier 2.

RESULTS

According to the EU-TOPIA evaluation tool, screening women in Southern Italy biennially from age 50 to 69 with an adherence to screening of 14% (current screening), would lead to 557,000 breast cancer diagnoses (DCIS and invasive, for women aged 40-100 years in the period 2020-2050), of which 7% would be screen-detected (Table 3). A total of 7,021 breast cancer deaths would be prevented (mortality reduction: 4.5%) compared to no screening and 6,000 life- years would be gained. Screening with the current adherence would to lead to 4.3% overdiagnosis and 336,000 false-positive results. Fully overcoming the barrier of low adherence, i.e. increasing adherence to 100% (scenario 5), is predicted to increase the share of screen detected cancers to 32% and would prevent 15,535 additional breast cancer deaths. Complete adherence would nearly triple the life-years compared to the current strategy (+10,500), while the percentage of overdiagnosis would remain the same (4.3%) and the number of false-positives would increase by 746% (derived from Table 3). Investing in up to 10 mobile screening units increased the total screening costs of the program (i.e. primary screening costs, costs for diagnostic follow-up and costs for cancer care) from €572million to €667 million (+17%, Table 3). Compared to the current screening strategy, the ICER was €9,531 per QALY gained for the adherence of 100%.

Table 4 presents the results of the simulations for barrier 2 for the whole of Italy. Continuing the current screening - characterized by a combination of screening intervals (i.e. the weighted average of the three sup-groups) – would lead to 1.3 million breast cancer diagnoses (DCIS and invasive) for women aged 40-100 years in the period 2020-2050, of which 32% would be screen detected. A total of 94,998 breast cancer deaths would be prevented (mortality reduction: 25%) and 27,560 life- years would be gained compared to no screening. If the described actions/ measures would be initiated to achieve biennial screening of all eligible women from 2020 onwards (i.e. the weighted average of scenario 2 of each sup-group), the EU-TOPIA evaluation tool predicts a similar number of breast cancer diagnoses and share of screen detected cancers. Moreover, overcoming this barrier would lead to an unchanged percentage of overdiagnosis (5%) while false positive results would decrease by 13% (derived from Table 4). Nearly the same

Table 3. Summary of screening outcomes and cost-effectiveness results for BARRIER 1 (low adherence in Southern Italy), by simulated scenario

	Screening outcomes[a]				
	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Number of screen tests (x 1,000)	5,717	+8,262	+12,386	+16,503	+35,487
Number of referrals (x 1,000)	373	+539	+808	+1,077	+2,315
Number of false positives (x 1,000)	336	+495	+745	+996	+2,170
Number of BC diagnoses[b] (% of screen detected)	557 (7%)	+6 (14%)	+9 (18%)	+11 (21%)	+18 (32%)
Number of BC deaths (x 1,000)	149	-4	-6	-8	-16
BC mortality reduction (%) compared to no screening	4.5	+2.8	+4.1	+5.3	+10
Prevented BC death	7,021	+4,398	+6,134	+8,298	+15,535
Overdiagnosis (%)[c]	4.3	-0.1	+/-	+0.1	-+/-
	Cost Effectiveness results[d]				
	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Primary screening costs	19,556,685	+19,040,730	+28,549,785	+38,046,525	+81,804,915
Costs for diagnostic follow-up	13,375,050	+4,611,702	+6,950,699	+9,315,608	+20,425,378
Costs for cancer care	539,731,388	-1,798,364	-2,656,144	-3,489,288	-7,521,792
Net costs compared to no screening	15,857,971	+21,908,068	+32,916,340	+43,962845	+94,888,501
Life-years gained vs no screening	6,000	+3,000	+4,300	+5,500	+10,500
QALY gained vs no screening	6,077	+2,866	+4,103	+5,240	+9,956
ICER[e]	-	7,645	8,023	8,391	9,531

[a] Results represent the difference in screening outcomes or costs between each scenario (2 to 5) and current screening (scenario 1) and are reported for women aged 40-100 in 2020-2050.

[b] Total of clinically and screen detected DCIS and invasive breast cancers (percentage of screen detected in brackets).

[c] Overdiagnosis is defined as women that would not have been diagnosed during their lives if they had never been screened as a percentage of all screen detected women.

[d] The results are presented for 1 million women followed from 2020 over their life-times. Both costs and effects are discounted with 3% per year.

[e] ICERs were computed as ratio between incremental net costs and benefits (QALY) compared to the current scenario (scenario 1).

BC: Breast cancer; QALY: Quality adjusted life-year; ICER: Incremental cost-effectiveness ratio

Table 4. Summary of screening outcomes and cost-effectiveness results for BARRIER 2 (screening intervals), by simulated scenario

	Screening outcomes[a]	
	Weighted average (scenario 1)	Weighted average (scenario 2)
Number of screen tests (x 1,000)	114,589	+13,153
Number of referrals (x 1,000)	8,131	-980
Number of false positives (x 1,000)	7,714	-971
Number of BC diagnoses[b] (% of screen detected)	1,294 (32%)	+2 (32%)
Number of BC deaths (x 1,000)	291	+2
BC mortality reduction (%) compared to no screening	24.6	+/-
Prevented BC death	94,998	-1,149
Overdiagnosis[c]	5.0	+/-
	Cost Effectiveness results[d]	
	Weighted average[d] (scenario 1)	Weighted average[d] (scenario 2)
Primary screening costs	129,107,349	-25,599,189
Costs for diagnostic follow-up	34,990,611	-6,376,965
Costs for cancer care	480,691,757	-89,562,403
Net costs compared to no screening	127,936,053	-30,440,225
Life-years gained vs no screening	27,560	-310
QALY gained vs no screening	27,402	-187
ICER[e]	162,799	3,578

[a] Results represent the difference in screening outcomes or costs between scenario 2 and current screening (scenario 1) and are reported for women aged 40-100 in 2020-2050.

[b] Total of clinically and screen detected DCIS and invasive breast cancers (percentage of screen detected in brackets).

[c] Overdiagnosis is defined as women that would not have been diagnosed during their lives if they had never been screened as a percentage of all screen detected women.

[d] The results are presented by 1 million women followed from 2020 over their life-times. Both costs and effects are discounted with 3% per year.

[e] ICERs were computed as ratio between incremental costs and benefits (QALY) compared to the previous scenario on the cost-effectiveness frontier.

BC: Breast cancer; QALY: Quality adjusted life-year; ICER: Incremental cost-effectiveness ratio; +/-: approximately the same results as scenario 1

number of breast cancer deaths would occur and 310 fewer life-years would be gained between 2020 and 2050 compared to current screening. Screening all eligible women in Italy every two years is predicted to lead to 11% more screen tests and less expensive cancer care and therefore to a saving of €122 million in total screening costs (derived from Table 4). While overcoming this barrier leads to a substantial decrease in referrals (-12%), false positive results (-13%) and costs (-18%), almost all benefits could be maintained. And therefore, if a mix of screening intervals would be maintained, the ICER would be 162,799 per QALY, compared to a situation where all eligible women are screened biennially (Table 4). Based on a WTP threshold of €23,000 per QALY gained, remaining in the current screening scenario is very cost-inefficient.

Sensitivity analysis

An overview of the results of the sensitivity analyses is provided in Table 5. Model-predicted base-case screening outcomes (weighted averages) were not particularly sensitive to the assumption of complete adherence for barrier 2. Overcoming the barrier, i.e. harmonising the screening interval throughout the country, would remain cost-saving (€83,072 per QALY gained) even when the observed adherence per sub-group would be applied. Varying the respective sub-group sizes showed only minor effects as well. Assuming the sub-groups would account for 30% (overscreened), 20% (underscreened) and 50% (biennial) led to savings of €41,569 per QALY gained compared to a situations where barrier 2 is present. Assuming the sub-groups would account for 20% (overscreened), 50% (underscreened) and 30% (biennial) led to savings of €31,586 per QALY gained compared to a situations where barrier 2 is present.

Table 5. Sensitivity analysis: Weighted average in comparison to the base case analysis[a], by adapted parameter

Screening outcomes[a]						
	Current screening (scenario 1)			overcoming the barrier (scenario 2)		
	Observed adherence	sub-groups 30/20/50	sub-groups 20/50/30	Observed adherence	sub-groups 30/20/50	sub-groups 20/50/30
Number of screen tests	-20%	+40%	+22%	-40%	+/-	-5%
Number of referrals	-37%	+10%	-3%	-40%	+/-	-5%
Number of false positives	-38%	+11%	-4%	-41%	+/-	-5%
Number of BC diagnoses[b] (% of screen detected)	+/- (32%)	+/- (32%)	+/-	+/-	-1%	-2%
Number of BC deaths	+9%	-1%	+1%	+9%	-1%	+1%
BC mortality reduction (%) compared to no screening	-26%	+4%	-2%	-29%	+2%	-4%
Prevented BC death	-26%	+4%	-2%	-29%	+2%	-4%
Overdiagnosis[c]	+8%	+3%	-1%	+8%	+1%	+2%
Cost Effectiveness results[d]						
	Current screening (scenario 1)			overcoming the barrier (scenario 2)		
	Observed adherence	sub-groups 30/20/50	sub-groups 20/50/30	Observed adherence	sub-groups 30/20/50	sub-groups 20/50/30
Primary screening costs	-34%	+14%	+/-	-37%	-5%	-5%
Costs for diagnostic follow-up	-28%	+12	+/-	-30%	-4%	-4%
Costs for cancer care	-17%	-19%	+/-	+2%	+/-	+/-
Net costs compared to no screening	-36%	+17%	+/-	-41%	-3%	-6%
Life-years gained vs no screening	-28%	+4%	+/-	-28%	-3%	-3%
QALY gained vs no screening	-27%	+4%	+/-	-28%	-3%	-3%
ICER[d]	83,072	41,569	31,586			

[a] Results are presented as percentage difference of the weighted average between of the respective sensitivity analysis and the base-case analysis.

[b] Total of clinically and screen detected DCIS and invasive breast cancers (percentage of screen detected in brackets).

[c] Overdiagnosis is defined as women that would not have been diagnosed during their lives if they had never been screened as a percentage of all screen detected women.

[d] ICERs were computed as ratio between incremental costs and benefits (QALY) compared to the previous scenario on the cost-effectiveness frontier.

BC: Breast cancer; QALY: Quality adjusted life-year; ICER: Incremental cost-effectiveness ratio; +/-: approximately the same results as base case analysis

DISCUSSION

Based on our results, two of the major barriers to the breast cancer screening programs in (Southern) Italy could be overcome by initiating feasible changes leading to better long-term outcomes. The employed online tools can be an effective and reliable resource for European policymakers that aim for informed decision-making on cancer screening in their country.

In Italy, like in most European countries, opportunistic and organized screening coexist. The extent of opportunistic screening in Italy is estimated to be 19%[20]. This is of concern, since organized screening programs are more likely to be attended by the socio-economically disadvantaged women[3 26-28] and low coverage in organized screening programs is associated with health and social inequalities[27], and opportunistic screening is associated with higher screening costs[23]. Moreover, it was shown in this as well as in other studies that the coverage of (organised) screening is of key importance in order to reach the full public health potential in terms of reduction in mortality from breast cancer[29 30 31]. A recent study conducted in Italy has shown that the reduction of social inequalities associated to organized breast screening programs is not only related to the increase of early diagnoses but also to the improved access to effective treatments[26]. This is of note considering that social disparities have been demonstrated to influence the entire cancer prevention and care pathway as well as survival and mortality and therefore organized screening programs may have the potential to eliminate at least some of the barriers encountered by disadvantaged women[32 33]. Based on our hypothesis that mobile mammography units help to increase participation, we showed that this strategy would be cost effective in regions with low screening adherence. Moreover, earlier studies[34] showed that decreasing geographical barriers is also an effective measure for decreasing social inequalities in participation in breast cancer screening even for the total population.

It can be debated whether we set the goal too high of overcoming barrier 1. Increasing screening adherence among eligible Southern Italian women from 14% to 100% might seem like an unrealistic and unfeasible attempt. But, investing resources in some areas for improvement is important. While the barrier assessment tool allows users to prioritise the most important barriers, it also stresses that this assessment should consider feasibility of ways to overcome barriers. Instead of aiming at fully overcoming the barrier, our results show that even an adherence of 54% (+40%, scenario 4) would lead to a doubling of screening benefits (e.g. breast cancer mortality reduction or LYG).

Public health decision makers need to get an insight into cancer screening's potentially beneficial and harmful short- and long-term effects. However, web-based cancer screening evaluation tools are scarce.

A specific strength of our analysis is the presented approach of simulating possible future changes, which can be adapted and followed by any registered user of the tools developed within the EU-TOPIA project. With the Barrier Assessment Tool, key problems and policy issues can be identified at every step of the screening process7. Solving those issues is vital in order to improve both the efficacy and equity of the current screening program. The Evaluation Tool helps to estimate the harms and benefits of the current screening program. But it also supports to define and evaluate the alternative screening scenarios reflecting on how key barriers can be overcome. Subsequently, the modelling quantifies the long-term impact of the proposed changes on screening outcomes. Continuous monitoring of screening activity would then permit to assess whether the implemented changes could actually achieve the expected impact. The EU-TOPIA evaluation tool is freely available and can be tailored to the specific situation in the user's country or region with a country, but also permits the use of EU-TOPIA benchmark data when stakeholders do not have all the required information. However, the MICAN breast model acts as the cornerstone of this analysis. Thus, an important strength is that a calibrated and validated model and its transparent methods was used. Many of the barriers analyzed in this study are common to different countries, as presented in two prior EU-TOPIA publications[7 8]. The existence of opportunistic screening beside organized programs as well as non-participation in either screening due to beliefs, values or practical issues are frequently barriers to effective breast cancer screening in Europe. Another strength of this paper is that it illustrates a pathway of evaluating the long-term impact of these barriers on screening outcomes and to also show the effect of overcoming them by initiating practical solutions. Previous publications have shown similar examples for colorectal screening and cervical screening highlighting that it is worth to spend efforts to remove barriers that affect the interval at which women participate in screening[35 36].

There are four noteworthy limitations of our analysis. Firstly, our analysis is based on the barrier assessment of one group of key informants who belong to the programme coordinator stakeholder group and who are experts in the national screening programs. Future analyses should also include the views of local providers, in particular in countries where screening is decentralized, like it is the case for Italy. Second, there are more potential measures to limit the use of opportunistic screening or to increase screening adherence than the ones in this paper. A recently published paper by Pelullo et al[28] looked at the very low attendance to breast and cervical cancer screening and the role of related determinants of the Southern Italian region Campania. Their findings underline the urgent need for an improved population education on cancer prevention and specifically on benefits and harms of cancer screening. Third, our simulations are based on a national model, which was calibrated to fit the observed Italian data for breast cancer. For a country with such heterogeneous screening practices, regional disparities in program implementation, program performances and population characteristics as Italy, using the webtool might have led to an under-or overestimation of certain outcome parameters.

Fourth, while our analysis focuses on the target ages of 50-69 years, some regions in Italy also invite younger (mainly Emilia-Romagna, Piedmont and Valle d'Aosta) or older (mainly Emilia-Romagna, Umbria Basilicata and Lombardy)[13] women to screening as well. However, we assume the effect of such regional disparities on our reported national results to be marginal.

Cancer screening programs in Italy are provided at regional level and implemented by Local Health Units. These programs consist of active invitation of all eligible women and participation – as well as subsequent treatment – is free of charge. In contrast, opportunistic screening in Italy is usually initiated by the recommendation of a woman's GP or by her own choice. In this context, the role of GPs should be carefully reconsidered. Although they are not directly involved in the organization and implementation of screening programs, they play a strategic role in promoting and motivating the participation to screening programs. GPs are usually supporting the programs by signing the screening invitation letter, but their more active involvement in all the screening phases might ensure a wider and more conscious access to screening for those women who usually refuse health services[28].

Our analysis shows that removing the most important barriers of the current Italian breast cancer screening programs could result in substantial improvements at acceptable costs. The used online tools are valuable for stakeholders to quantify benefits, harms and costs of early cancer detection in Europe. This Italian example illustrates a systematic approach and stepwise process that can be easily followed or adapted by other European countries or stakeholders.

REFERENCES

1. Lauby-Secretan B, Scoccianti C, Loomis D, et al. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med 2015;372(24):2353-8.
2. Zielonke N, Gini A, Jansen E, et al. Evidence for reducing cancer specific mortality due to screening for breast cancer in Europe: a systematic review. Eur J Cancer 2020;127:191-206.
3. IARC Working Group. IARC Handbooks of Cancer Prevention, Volume 15: Breast Cancer Screening. Lyon, 2016.
4. Council of the European Union. Council Recommendation of 2 December 2003 on Cancer Screening (2003/878/EC). OJ L 327: 34-38. 2003.
5. Giorgi Rossi P, Carrozzi G, Federici A, et al. Invitation coverage and participation in Italian cervical, breast and colorectal cancer screening programmes. J Med Screen 2018;25(1):17-23.
6. EU-TOPIA. EU-TOPIA – towards improved cancer screening for breast,cervical and colorectal cancer in all of Europe [www.eu-topia.org] 2016 [Accessed on 15.03.2021] [Available from: www.eu-topia.org.
7. Turnbull E, Priaulx J, van Ravesteyn NT, et al. A health systems approach to identifying barriers to breast cancer screening programmes. Methodology and application in six European countries. Health Policy 2018;122(11):1198-205.
8. Priaulx J, de Koning HJ, de Kok I, et al. Identifying the barriers to effective breast, cervical and colorectal cancer screening in thirty one European countries using the Barriers to Effective Screening Tool (BEST). Health Policy 2018;122(11):1190-97.
9. Petrelli A, Giorgi Rossi P, Francovich L, et al. Geographical and socioeconomic differences in uptake of Pap test and mammography in Italy: results from the National Health Interview Survey. BMJ Open 2018;8(9):e021653.
10. Damiani G, Federico B, Basso D, et al. Socioeconomic disparities in the uptake of breast and cervical cancer screening in Italy: a cross sectional study. BMC Public Health 2012;12:99.
11. Commission E. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth Edition. 2006.
12. Giorgi Rossi P, Djuric O, Navarra S, et al. Geographic Inequalities in Breast Cancer in Italy: Trend Analysis of Mortality and Risk Factors. Int J Environ Res Public Health 2020;17(11).
13. ONS. Osservatorio Nazionale Screening. Rapporto 2019, 2019 [Accessed on 21.03.2021].
14. Giordano L, Castagno R, Giorgi D, et al. Breast cancer screening in Italy: evaluating key performance indicators for time trends and activity volumes. Epidemiologia e Prevenzione 2015;39(3 Suppl 1):30-9.
15. Zielonke N, Geuzinge AH, Heijnsdijk EAM, et al. Extending Age Ranges in Breast Cancer Screening in Four European Countries: Model Estimations of Harm-to-Benefit Ratios. Cancers 2021(13(13), 3360).
16. Zielonke N, Geuzinge HA, Heijnsdijk EAM, et al. Extending age ranges in breast cancer screening in four European countries: Model estimations of harm-to-benefit-ratios European Journal of Cancer 2021 submitted.
17. Sankatsing VDV, Heijnsdijk EAM, van Luijt PA, et al. Cost-effectiveness of digital mammography screening before the age of 50 in The Netherlands. Int J Cancer 2015;137(8):1990-9.

18. Gini A, van Ravesteyn NT, Jansen EEL, et al. The EU-TOPIA evaluation tool: an online modelling-based tool for informing breast, cervical, and colorectal cancer screening decisions in Europe. Preventive Medicine Reports 2021;submitted.
19. EU-TOPIA. EU-TOPIA evaluation tool [https://miscan.eu-topia.org] 2019 [Accessed on 21.04.2021] [
20. Zielonke N, Kregting LM, Heijnsdijk EAM, et al. The potential of breast cancer screening in Europe. Int J Cancer 2020.
21. Mandrik O, Tolma E, Zielonke N, et al. Systematic reviews as a "lens of evidence": Determinants of participation in breast cancer screening. J Med Screen 2020:969141320930743.
22. Foglia E, Marinelli S, Garagiola E, et al. Budget impact analysis of breast cancer screening in Italy: The role of new technologies. Health Serv Manage Res 2020;33(2):66-75.
23. Mantellini P, Lippi G, . I costi dello screening. Un'analisi in logica activity based. Firenze Zadig Editore, 2011.
24. Francisci S, Guzzinati S, Capodaglio G, et al. Patterns of care and cost profiles of women with breast cancer in Italy: EPICOST study based on real world data. Eur J Health Econ 2020;21(7):1003-13.
25. Pashayan N, Morris S, Gilbert FJ, et al. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. JAMA Oncol 2018;4(11):1504-10.
26. Zengarini N, Ponti A, Tomatis M, et al. Absence of socioeconomic inequalities in access to good-quality breast cancer treatment within a population-wide screening programme in Turin (Italy). Eur J Cancer Prev 2016;25(6):538-46.
27. Menvielle G, Kulhánová I, Mackenbach JP. Assessing the impact of a public health intervention to reduce social inequalities in cancer. In: Vaccarella S, Lortet-Tieulent J, Saracci R, et al., eds. Reducing social inequalities in cancer: Evidence and priorities for research. Lyon: IARC, 2019.
28. Pelullo C, Cantore F, Lisciotto A, et al. Organized Breast and Cervical Cancer Screening: Attendance and Determinants in Southern Italy. Cancers 2021; 13(7):1578.
29. Chen TH, Yen AM, Fann JC, et al. Clarifying the debate on population-based screening for breast cancer with mammography: A systematic review of randomized controlled trials on mammography with Bayesian meta-analysis and causal model. Medicine (Baltimore) 2017;96(3):e5684.
30. Marmot MG, Altman DG, Cameron DA, et al. The benefits and harms of breast cancer screening: an independent review. Br J Cancer 2013;108(11):2205-40.
31. Gianino MM, Lenzi J, Bonaudo M, et al. Organized screening programmes for breast and cervical cancer in 17 EU countries: trajectories of attendance rates. BMC Public Health 2018;18(1):1236.
32. Bradley CJ, Given CW, Roberts C. Race, socioeconomic status, and breast cancer treatment and survival. J Natl Cancer Inst 2002;94(7):490-6.
33. Bigby J, Holmes MD. Disparities across the breast cancer continuum. Cancer Causes Control 2005;16(1):35-44.
34. Guillaume E, Launay L, Dejardin O, et al. Could mobile mammography reduce social and geographic inequalities in breast cancer screening participation? Preventive Medicine 2017;100:84-88.

35. Csanadi M, Gini A, Koning H, et al. Modeling costs and benefits of the organized colorectal cancer screening programme and its potential future improvements in Hungary. J Med Screen 2020:969141320968598.
36. Jansen EEL, Ivanus U, Jerman T, et al. The optimal HPV-screening protocol in Eastern-Europe: The example of Slovenia. Gynecol Oncol 2021;160(1):118-27.

SUPPLEMENTARY METHODS

EU-TOPIA Barrier to effective screening tool (BEST)

The European Union funded EU-TOPIA (Towards improved screening for breast, cervical and colorectal cancer in all of Europe) project aims to improve health outcomes and equity of breast, cervical and colorectal cancer screening programmes in ways that take full account of the different demographical, medical, political, economic and cultural contexts across Europe. The project aims to do so by providing national, regional, and local policymakers with tools to evaluate and quantify their cancer screening programmes. One of the tools created was a self-assessment process for the purpose of identifying the barriers to the optimal operation of population-based breast cancer screening programmes.

The online tool was used with workshop participants prior to and during an EU-TOPIA workshop in Turin, April 2019. This Excel version has been made available so that screening organisers, researchers and policy-makers in Europe can continue to use and apply the tool in future projects.

Background to the tool

The tool is based on a health systems approach to evaluating health services. This means that all parts of the screening process are important from start to finish. We identified six sub-systems in our model: generation of knowledge and effectiveness, identification of the population at risk, maximising informed participation, successful operation of the programme, adequate follow-up and ensuring effective treatment for those that need it[1 2].

Aim of the tool

The aim of the tool is to enable screening organisers, researchers and policy-makers to make a self-assessment of their organised breast, cervical and colorectal cancer screening programmes to identify the most important barriers to effectiveness and equity. The tool also allows users to prioritise the most important barriers, also considering feasibility, and identify ways to overcome barriers.

How to apply the tool

The tool should be applied separately for each cancer screening programme (breast, cervical and colorectal). The tool is the same for each cancer type but the barriers may differ between programmes. The tool has been developed primarily for use in countries, regions or municipalities where organised, population-based cancer screening programmes are already in place, or at least where a pilot programme is in operation. Where a programme has yet to be introduced, we suggest respondents consider the main barrier(s) to the introduction of such a programme and use this as the basis for discussions on how to overcome barriers.

Ideally, the tool should incorporate the views of all stakeholders, either by each stakeholder completing the tool separately (and the results amalgamated to determine an average) or by organising a stakeholder meeting to complete the tool using a consensus approach. The process of completing the tool (the thought processes involved and discussions) are as important as the results of the tool. Hence, it would be useful to make notes whilst completing the tool to capture any new insights raised from the discussion. Moreover, it should be emphasised that the barriers are the opinion of the respondents and that the impact of barriers varies, depending in part on measures taken to mitigate their effects.

Key sections of the tool

Firstly, we would like you to rate the importance of each barrier that exists in your country in terms of its impact on programme effectiveness and equity. Secondly, we would like you to select the three most important barriers to overcome. Thirdly, we would like you provide any examples you have from your country of ways that barriers have been overcome (or at least reduced). The latter section was primarily of interest to the EU-TOPIA workshop for sharing knowledge across countries. However, this exercise may still be useful for discussions about how to overcome remaining barriers.

What next?

The results of this self-assessment exercise will enable you to prioritise barriers to be reduced or overcome and whether the resource currently being allocated to overcoming barriers is being used effectively. In addition, this tool is part of several work packages from the EU-TOPIA project. Following on from this work package is another work package on 'road maps' to further develop and improve cancer screening programmes. The results from this activity inform the road maps. Please refer to the EU-TOPIA website for further information on the workshops and other tools available.

Table A1. Results of the barrier assessment from a group of Italian stakeholders*

Sub-system	Barrier	Effect[1]	Equity[1]	Rank[3]
Knowledge	Issues with establishing protocols, processes and legal frameworks (e.g. inadequate national governance structure, professionals with relevant knowledge)	3	3	
	Screening guidelines and protocols are not regularly updated or updates are delayed (e.g. by complex administration procedures)	2	1	
Identification	Population register is not accurate (e.g. not updated with changes of address)	2	2	
	Population register is not complete (e.g. some eligible people not included)	2	2	
Participation	Some people have beliefs and values that lead to non-participation in screening programme	4	4	1
	Some people experience practical issues that lead to non-participation in screening programme (e.g. inconvenient appointments, inadequate health insurance)	3	4	
	Inadequate public promotion of screening programme (e.g. primary care physicians are not sharing information or promoting screening)	3	3	
	Inadequate system for monitoring levels and patterns of screening participation (e.g. inequalities among some subgroups)	3	2	
	Inadequate response to low levels of uptake (informed participation) and patterns of screening participation (e.g. inequalities among some subgroups)	4	4	
Operation	Inadequate information technology (IT) systems (e.g. disjointed systems)	4	1	
	Insufficient human, physical and/or financial resources to operate screening programme (e.g. limited capacity, organisational or logistical issues)	4	3	2
	Inadequate adherence by providers to screening guidelines and protocols (e.g. opportunistic screening occurs outside the organised screening programme)	4	4	3
	Inadequate system for monitoring operational aspects of the screening programme (e.g. quality of screening experiences of those who participate)	3	2	
	Inadequate response to address quality issues relating to the operation of the screening programme	3	1	
Follow up	Insufficient human, physical and/or financial resources to conduct follow-up investigations for those that need it	4	3	
	Inadequate adherence by providers to follow-up guidelines and protocols (e.g. clinician's attitudes and established pattern of practice)	4	1	

	Inadequate system for monitoring people who require follow-up investigations but do not participate (e.g. due to personal beliefs or practical issues)	2	1
	Inadequate response to people who require follow-up investigations but do not participate (e.g. due to personal beliefs or practical issues)	2	1
	Inadequate sharing of follow-up information between national/regional screening organisations, providers of follow-up investigations and primary care	2	1
Treatment	Some people have beliefs and values that lead to them declining cancer treatment	2	4
	Insufficient human, physical and/or financial resources to provide treatment to those that need it	3	4
	Inadequate system for monitoring treatment information (e.g. treatment data not systematically linked to cancer screening data)	4	2
	Inadequate sharing of treatment information between national or regional screening organisations, providers of cancer treatment and primary care	4	2

*What is your job role? - Regional screening organisation representative, researcher
Does your country have an organised, population-based screening programme for breast cancer in operation? - Yes - regionally organised
Do you have expert knowledge of breast cancer screening (organised or opportunistic screening) in your country? - Yes
[1] Informants were asked to rate the importance of each barrier that exists in your country in terms of its impact on programme effectiveness and equity
[2] Informants were asked to select the three most important barriers to overcome

REFERENCES

1. Priaulx J, de Koning HJ, de Kok I, et al. Identifying the barriers to effective breast, cervical and colorectal cancer screening in thirty one European countries using the Barriers to Effective Screening Tool (BEST). Health Policy 2018;122(11):1190-97.
2. Turnbull E, Priaulx J, van Ravesteyn NT, et al. A health systems approach to identifying barriers to breast cancer screening programmes. Methodology and application in six European countries. Health Policy 2018;122(11):1198-205.

EU-TOPIA EVALUATION TOOL - USER GUIDE (VERSION JANUARY 2020)

1. Introduction

The overall aim of the EU-TOPIA project is to improve existing cancer screening programmes in Europe. To aid in the improvement of existing programmes, we have developed a web-based evaluation tool based on the MISCAN model to allow European policymakers and researchers to simulate outcomes of multiple cancer screening strategies for their own country. This document describes the structure and the standard process required for utilizing the EU-TOPIA evaluation tool from creating an account in section 3 to analyzing the results of your simulations in section 6.

2. MISCAN model

The EU-TOPIA evaluation tool uses a well-established microsimulation model for cancer (MISCAN-Breast). MISCAN was developed in the 1970s at the Department of Public Health of Erasmus MC, University Medical Center Rotterdam and was designed for evaluating the effect of cancer screening. MISCAN simulates individual life histories and assesses the consequences of introducing a screening programme on these life histories. The model estimates the effect of cancer screening in a dynamic population and can explain results of cancer screening trials and predict and compare the (cost-)effectiveness of different screening strategies. Our model has been calibrated on data from a number of countries exemplary for all European regions

3. Account management

3.1. Register

First things first: you need to register as a user of EU-TOPIA evaluation tool website.
You can do that by filling out the registration form (https://miscan.eu-topia.org/registration/register), click on the link in the confirmation e-mail to confirm your e-mail address and let the EU-TOPIA admins process your information for the final authorization.

3.2. Log in

Once the EU-TOPIA admins send you an e-mail to confirm your account, it is fully operative. Now you are able to log in into the EU-TOPIA evaluation tool (https://miscan.eu-topia.org/login) with your e-mail and password and perform simulations. When you log in the first time, you will be asked to download this user's guide and to declare to have read it.

3.3. Log out

You can log out from the application in anytime clicking the button "Logout" at the top right corner of the web site.

3.4. Account information / password

You can view your account information and/or change your password clicking on "My account" or on your user name (both at the top-right corner of the screen).

3.5. Help

If you need any help, please feel free to e-mail the EU-TOPIA research team at eu.topia@erasmusmc.nl or complete the contact form by clicking "Help" at the top-right corner of the screen.

4. Data collection

The EU-TOPIA evaluation tool was designed to allow users to simulate outcomes of several cancer screening strategies for their own country. Thus, this tool requires the users to upload specific demographic and screening data for their own country.

Section 4.1. describes how to download the excel data templates needed to collect country specific data.

Sections 4.2 to 4.4 describe the excel data templates needed to collect country specific data and give important general information for filling out the data tables. Instructions on how to fill in specific data templates are described in two parts: epidemiological data in section 4.5, and screening programme related monitoring data in section 4.6.

4.1. Download Excel Data Templates

1. Log on to the main dashboard page.
   Once you log in you will see the main evaluation tool dashboard (Figure 1).

Figure 1. EU-TOPIA evaluation tool, user's dashboard.

2. Click on “Download”.
3. You can download templates from each cancer site and save these on your computer (see Figure 2).
   For the breast specific cancer template, click the breast cancer icon.

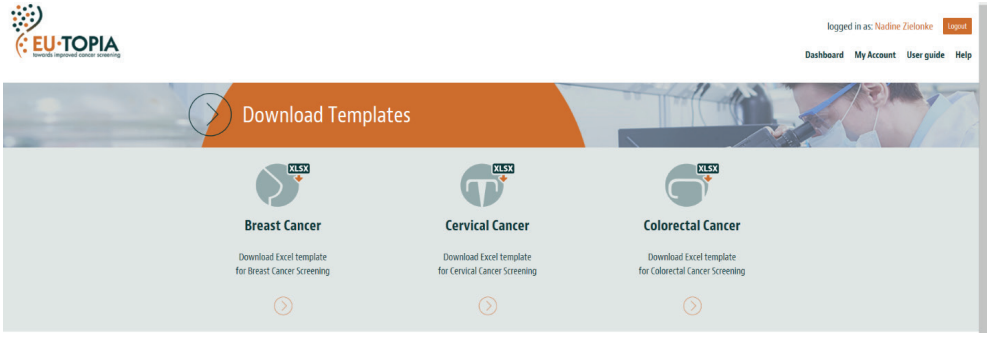


Figure 2. EU-TOPIA evaluation tool, download templates section.

4. Save the templates to your computer by clicking “Save as” in Save option list (see Figure 3)

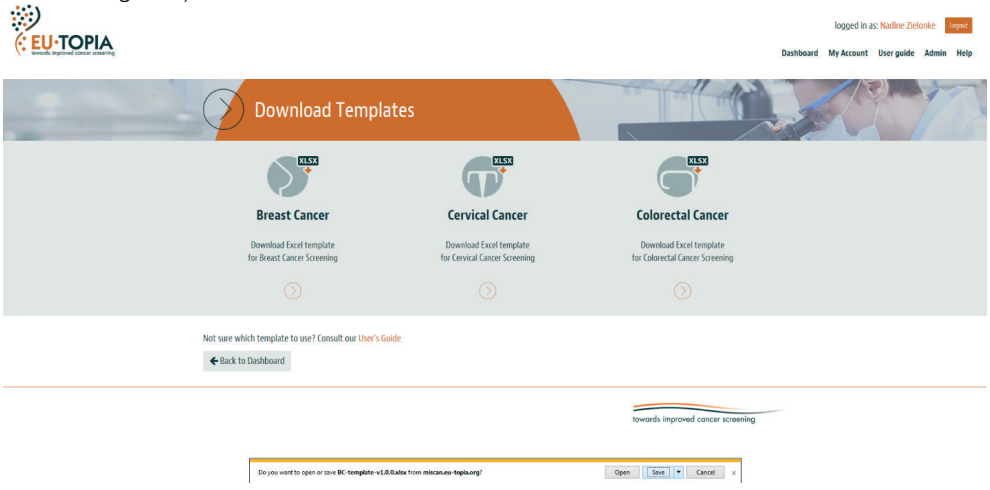


Figure 3. EU-TOPIA evaluation tool, saving excel data template.

4.2. Breast Cancer Tables

In this document you will find instructions on how to fill the Excel tables in order to simulate different BREAST cancer screening strategies for your country.

To tailor the model to your own country, you need to upload the following information:

- population and epidemiological data (6 tables)
- screening monitoring information (12 tables)

A core set of tables is required for the model to run your simulations. These are labelled as 'mandatory' in Table 1. Including information labelled as 'should have' or 'nice to have' will improve the quality of the information the model generates for your country. A more detailed description of each table will be provided from section 3.6.

Table 1. Overview of all data requirements for the EU-TOPIA web-based tool for breast cancer.

Table name	Brief description	Level of importance	More detailed instructions for filling this table.
Table0	Select your country	Mandatory	Section 4.4
eTable1	Population data	Mandatory	Section 4.5
eTable2	Cancer incidence	Mandatory	
eTable3	Cancer mortality	Mandatory	
eTable4	Relative survival by stage	Should have	
eTable5	Stage distribution	Should have	
eTable6	All-cause mortality by age	Should have	
sTable1	Screening strategy	Mandatory	Section 4.6
sTable2	Screening coverage	Mandatory	
sTable3	Screening history	Should have	
sTable4	Further assessment indication	Should have	
sTable5	Further assessment participation	Should have	
sTable6	Further assessment outcome	Should have	
sTable7	Outcome	Should have	
sTable8	Pathological sizes	Nice to have	
sTable9	Surgical treatment	Nice to have	
sTable10	Interval cancers	Should have	
sTable11	Opportunistic screening	Should have	
sTable12	Adjuvant treatment	Should have	

4.3. Essential information for filling out the data tables

Key instructions

- Fill out the tables in the order that they are presented in the excel template.
- 'Mandatory' data is necessary to have a good refinement of the model for the country-specific analysis
- The more accurate and complete the country-specific data that you are providing, the more accurate and country-specific will be the results of the simulations.
- Fill only the white cells in the data templates. All other cells are non-editable.
- Only use the data type specified by the information box when you click on an empty cell indicated.
- Be careful with copying and pasting data from other Excel files. If data is pasted into non-editable cells, the tool will not recognize the inputs.

Information about data quality

- After you submit the data, automatic data quality checks will be applied to make sure the inputs fall within a reasonable range.
- The more complete the input data, the better the model will simulate your scenarios.

Trouble shooting

- Make sure you always work with the most recent version of the user guide as we are constantly updating this document to answer as many questions as possible.
- If you encounter any problems when entering the data, please contact the EU-TOPIA research team at eu.topia@erasmusmc.nl

Organised vs. opportunistic screening

Based on the IARC Handbook of Cancer Prevention (IARC Working Group. IARC Handbooks of Cancer Prevention, Volume 15: Breast Cancer Screening. Lyon, 2016) we defined

- organised breast cancer screening as screening programmes organised at national or regional level, with an explicit policy, specifying age categories, method and interval for screening, a management team responsible for implementation and a health care team for decisions and care; a quality assurance structure and systematic monitoring of quality indicators; a method for monitoring of cancer occurrence in the target population. The programme policy includes active invitation of the entire target population and usually also active follow-up of screen-positive subjects (IARC 2005 -2016; and Miles et al. Cancer 2004).
- In contrast, opportunistic breast cancer screening refers to mammograms performed as a result of the initiative of women themselves or at time of routine health checks. The classification non-population based (opportunistic) screening applies to areas where individual invitations are not sent to women in the eligible population, or when

women undergo a mammography outside, or in addition to, the (existing) screening programme. Registration and monitoring of such screening activity is often unavailable.

4.4. Country

Table0: Country

Select your country from the drop-down menu if you submit national data (=preferred, Step 1 in Figure 1). In that case you can ignore the second drop-down menu (Region). If your country is not listed, please contact us at: eu.topia@erasmusmc.nl

If you want to submit regional instead of national data, please contact us as at eu.topia@erasmusmc.nl. After we assigned you to an index region, you need to select your country from the first drop-down menu (Step 1 in Figure 4) and the respective index region (A-Z) from the second drop-down menu (Step 2 in Figure 4).

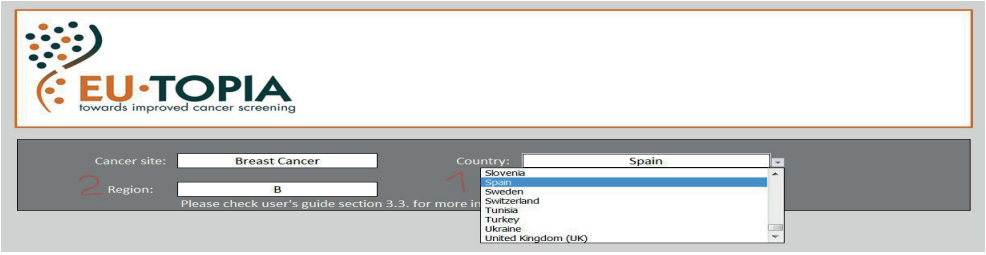


Figure 4. EU-TOPIA evaluation tool, Select your country (and region) in Table0.

Please be aware that we strongly advise you to use the same reference population in all tables! In case of a mixed situation - where most data is national and some is regional, or vice versa – leave a comment in the email accompanying the submitted data templates.

4.5. Epidemiological Data

All tables for demographic and epidemiological data are marked as “eTablex”.

Please fill out all of them for your country. In this section, we show you how to fill them out correctly.

eTable1: Population age composition

Level of importance	Mandatory
What does this table contain?	The 2018 female population and population projections (up to 2050) for your country
Format	Separated by calendar year and five-year age groups
Potential data sources	National statistical office
Potential data sources	We suggest you use the base case scenario (i.e. the scenario based on current population trends). If there are several sources for population data in your country, we suggest using the source which is also used for official government projections. A good resource is EUROSTAT population projections (http://ec.europa.eu/eurostat/data/database).

eTable2: Breast cancer incidence rates

Level of importance	Mandatory
What does this table contain?	The number of incident breast cancer cases (ICD-10: C50 and D05.1) and person years at risk in 1981-1985 and 2011-2015.
Format	Separated by five-year age groups, including Ductal Carcinoma In Situ (DCIS)
Potential data sources	National cancer registries
Trouble shooting	We prefer the data of 1981-1985 as this represents a period before any form of screening (organised or opportunistic) was introduced in Europe and hence can serve as the background incidence. However, if you know the cancer register was not complete at that time, please fill-out only the table for the period 2011-2015. For the respective 5-year period, you can either report the sum of cases (diagnoses) and the population, or the average over those five years. Both will be calculated into the incidence rate over that period. If your estimates are based on small numbers, we recommend using a longer time period in reporting incidence data (10 most recent years instead of 5). If you do not have direct access to national cancer registry data or detailed data are not available, we suggest checking the availability of cancer incidence data from the IARC cancer in five continents dataset (CI5, http://ci5.iarc.fr/CI5I-X/Pages/download.aspx).

eTable3: Breast cancer mortality rates

Level of importance	Mandatory
What does this table contain?	Mortality due to breast cancer (ICD-10: C50) in 1981-1985 and 2011-2015.
Format	Separated by five-year age groups
Potential data sources	National cause of death register
Trouble shooting	We prefer the data of 1981-1985 as this represents a period before any form of screening (organised or opportunistic) was introduced in Europe and hence can serve as the background mortality. However, if you know the death register was not complete at that time, please fill-out only the table for the period 2011-2015. For the respective 5-year period, you can either report the sum of cases (deaths) and the population, ,or the average over those 5 years. Both will be calculated into the incidence rate over that period.

eTable4: Relative survival

Level of importance	Should have
What does this table contain?	Probability of surviving 10 years after a diagnosis of Breast cancer (ICD-10: C50), observed in the most recent years
What does this table NOT contain?	Carcinoma in situ cases
Format	Separated by stage.
Potential data sources	National cancer registries
Trouble shooting	Tumour (T) of the breast cancer TNM staging system. The staging information should be based on the histopathological assessment of the tumour (pTNM), not the clinical (cTNM). If this information is not available, please leave a comment in the NOTES sheet of the data template. T describes the size of the tumour (area of cancer).tumour T1A status: between 0.1 and 0.5 cm across, independent of node status T1B status: between 0.5 cm and 1 cm across, independent of node status T1C status: between 1 cm and 2 cm across, independent of node status T2+ status: between 2 cm and 5 cm across, independent of node status Only numbers between 0 and 100 can be used as input (i.e. no words, notes, or symbols): typing 0.2, the web-based tool will read 0.2%; and typing 20, 20%. Survival probabilities that show greater survival in advanced stages compared to lower stages will be marked as “low-quality” data and you will be advised not to use this data when using the web-based tool. If your estimates are based on small numbers or detailed data are not available, we recommend to check published data from EUROCARE group (http://www.eurocare.it/)

eTable5: Breast cancer stage distribution

Level of importance	Should have
What does this table contain?	Stage distribution of breast cancer (ICD-10: C50 and D05.1) in 1981-1985 and 2011-2015
Format	Separated by stage Tumour, including Ductal Carcinoma In Situ (DCIS)
Potential data sources	National cancer registries
Trouble shooting	Tumour (T) of the breast cancer TNM staging system. The staging information should be based on the histopathological assessment of the tumour (pTNM), not the clinical (cTNM). If this information is not available, please leave a comment in the NOTES sheet of the data template. DCIS Ductal Carcinoma In Situ T describes the size of the tumour (area of cancer). T1A status between 0.1 and 0.5 cm across, independent of node T1B status between 0.5 cm and 1 cm across, independent of node T1C status between 1 cm and 2 cm across, independent of node T2+ status between 2 cm and 5 cm across, independent of node If your stage distribution adds up to more than 100%, the web-based tool will mark your data as “low-quality” data (additional function in the web-based tool). In that case please contact the EU-TOPIA research group at: eu.topia@erasmusmc.nl

eTable6: Population all-cause mortality

Level of importance	Should have
What does this table contain?	Current all-cause mortality rate for women
Format	By single ages (0-100)
Potential data sources	Life tables from national statistical offices
Trouble shooting	You are only allowed to input numbers (no words, notes, or symbols) between 0 and 1. Please make sure that values are not multiplied by 100,000 person-years (hence, values need to be reported considering 1 person-year). The inputs need to be age-specific mortality rates, so no age-specific probabilities of death. In the projections of the web-based tool, the all-cause mortality will be assumed to stable over time. As an alternative data source, we recommend your country’s life tables from the Human Mortality database (http://www.mortality.org/, use the ‘mx’ column in the data provided).

4.6. Screening Data

All tables for screening data are marked as "sTablex".

Please fill out all of them for your country. In this section, we show you how to fill them out correctly.

sTable1: Screening strategy

Level of importance	Mandatory
What does this table contain?	A summary of the current screening strategy of your country
What does this table NOT contain?	Data on opportunistic screening
Potential data sources	National breast cancer screening programme
Format	Separated by five-year age groups
Detailed content description	
Country (or Region):	Enter the country or area to which all tables refer.
Index year	All tables should report data from that index year
Starting year of the programme	When was the organised breast cancer screening programme introduced in your country/region?
End of the roll-out phase	Enter the respective year in case the roll-out was completed before the index year.
A1 Target population	Total number of age-eligible women obtained from official statistics (irrespective of the screening interval).
A2 Screening interval	Interval (in years) between routine screens decided upon in each screening programme dependent on the screening practice in your country.

Trouble shooting	As index year you should pick the most recent year you have complete (follow up) data on. Note that in sTable2 data will be required up to June of the following year. If screening is not implemented uniformly across the country or region on which you are reporting (i.e. there is regional variation in the rollout of screening or there is regional variation in the eligible age range or frequency) please report a screening policy which best represents the most common policy in your country. Consider for example a country where some regions invite women each year between age 45 and 49 and every two years between 50 and 69, whereas other regions only invite women between 50 and 74 biannually. Given the population sizes and the representativeness of the regions, you can either report only one of them as the screening strategy of your country. Alternatively, you can report a joined version of those two screening policies. We propose you enter a screening frequency that best reflects the screening practice in your country, i.e. if you invite women more often than every 24 months, you could enter e.g. 1.8 years. When you chose the screening scenarios (see section 5.4), you can simulate the effect of an optimal frequency of e.g. 2 years compared to the over- or underscreening currently present in your country. It is possible to indicate different screening intervals per age group. However, the model calculations are solely based on the frequency entered at age-group 50-54. If you are in doubt please contact us at: eu.topia@erasmusmc.nl If your data cannot be stratified by five-year age groups, put the total amount in the row marked as “unknown”. In a mixed situation, with data from some areas which can be stratified and other data that cannot be stratified, please fill separately the first rows for the former and the last row for the latter. Always check the total figures at the bottom of each table to be sure that the sum of the strata is the total number expected.

sTable2: Screening coverage

Level of importance		Mandatory
What does this table contain?		Screening invitations and screening tests
What does this table NOT contain?		Data on opportunistic screening
Potential data sources		National breast cancer screening programme
Format		Absolute number, separated by five-year age groups
Detailed content description		
B1	Individuals personally invited in index year	Enter the number of all personally invited women (not counting reminders or returned letters) in the period to which data refer (from January 1st to December 31st). Do not include invitations to intermediate mammograms (short term recalls) in this column.
B2	Individuals screened of invited in index year	This is a subset of B1. Enter the number of women who received a test – counting any test performed up to June of the following year (invitation cohort*). It is also acceptable, assuming steady state, to estimate this number using the number of attendees in the index year - regardless of their invitation date. Do not include tests referring to intermediate mammograms (short term recalls) in this column.
B3	Individuals screened in index year	Women who received a test in the index year – regardless of when invited (examination cohort**). Do not include tests referring to intermediate mammograms (short term recalls) in this column.
Trouble shooting		If your data cannot be stratified by five-year age groups, put the total amount in the row marked as "unknown". In a mixed situation, with data from some areas which can be stratified and other data that cannot be stratified, please fill separately the first rows for the former and the last row for the latter. Always check the total figures at the bottom of each table to be sure that the sum of the strata is the total number expected.

*Invitation cohort: It includes women attending screening until June 30th of the year following the reference one. It can be used to estimate the response rate (leaving at least 6-month interval available for responding, to the first invitation or to the reminder for women invited at the end of the reference year). Complete data about screening outcomes (including results of assessment and treatment, if indicated) might not be available for women in this cohort until the end of the year following the reference one. Therefore, to obtain accurate measures of screening performance it might be more efficient to refer to the examination cohort

**Examination cohort: It includes women screened during the reference year, independent of the invitation date. For women in this cohort the information about screening results is generally available by June 30th of the year following the reference one, which allows to get meaningful data about quality indicators within a reasonable interval.

You can decide, based on the availability of the data about outcomes, to use either the invitation or the examination cohort as the denominator for calculating screening performance indicators (sTables 4 to 7).

sTable3: Screening history

Level of importance		Should have
What does this table contain?		Screening invitations and screening tests by screening history
What does this table NOT contain?		Data on opportunistic screening
Potential data sources		National breast cancer screening programme
Format		Absolute numbers, separated by five-year age groups Data should be stratified per initial/subsequent tests: • Initial screening is the first screening examination of individual women within the screening programme, regardless of the organisational screening round in which the examination takes place. Include also screening tests performed in a population-based screening programme before receiving the first invitation (these examinations are often referred to as "spontaneous tests"). • Subsequent screening includes all screening examinations of individual women within the screening programme following an initial screening examination, regardless of the organisational screening round in which the examination takes place. • Unknown if initial or subsequent strata should be used for tests for which the above distinction is not available. The numbers collected in the three sub-tables should refer to strictly distinct sets of women. Always check the total figures at the bottom of the three tables to be sure that the sum of the strata is the total number expected.
Detailed content description		
C1-C4	Individuals personally invited in index year	It includes all personally invited women (not counting reminders or returned letters) in the period to which data refer. C1: women receiving their first invitation in the programme C2: women invited in the reference year who had already been invited in previous screening rounds C3 (sub-set of C2): women invited in the reference year who had been invited in previous screening rounds and had not attended the last invitation C4 (sub-set of C2): women invited in the reference year who had been invited in previous screening rounds and had attended the last invitation Please indicate the number of women invited from January 1st to December 31st of the index year. Do not include invitations to intermediate mammograms (short term recalls) in these columns.

C5-C8	Individuals screened of invited in index year	It is a subset of the women-invited-in-index-year (C1) who received a test – counting any test performed up to June of the following year (Invitation cohort). It is also acceptable, assuming steady state, to estimate this number using the number of attenders in the index year - regardless of their invitation date (examination cohort). Do not include tests referring to intermediate mammograms (short term recalls) in these columns. C5: women receiving their first test in the programme C6: women who had already received previous test in the programme C7 (sub-set of C6): women who had received previous tests in the programme, but who had not attended in the previous round C8 (sub-set of C6): women who had received previous tests in the programme and had attended in the previous round.
C9-C12	Individuals screened in index year	Women who received a test in index year – regardless of when invited. Do not include tests referring to intermediate mammograms (short term recalls) in these columns C9: women receiving their first test in the programme C10: women who had already received previous test in the programme C11 (Sub-set of C10): women who had received previous tests in the programme, but who had not attended in the previous round C12 (Sub-set of C10): women who had received previous tests in the programme and had attended in the previous round.
Trouble shooting		If your data cannot be stratified by five-year age groups, put the total amount in the row marked as “unknown”. In a mixed situation, with data from some areas which can be stratified and other data that cannot be stratified, please fill separately the first rows for the former and the last row for the latter. Always check the total figures at the bottom of each table to be sure that the sum of the strata is the total number expected.

sTable4: Further assessment indication

Level of importance		Should have
What does this table contain?		Further assessment as an additional diagnostic technique (either at screening or at recall) that are performed for medical reasons in order to clarify the nature of a perceived abnormality detected at the screening examination. It may include breast clinical examination, additional imaging and invasive investigations (cytology, core biopsy). Further assessment may have taken place on the same day as the screening examination or on recall. Please include among positive women also those undergoing assessment (as defined) on the same day as screening and not only those referred for assessment at a different date/ session
What does this table NOT contain?		Data on breast cancer detected by opportunistic screening
Potential data sources		National breast cancer screening programme
Format		Absolute numbers, separated by five-year age groups Data should be stratified per initial/subsequent tests: • Initial screening is the first screening examination of individual women within the screening programme, regardless of the organisational screening round in which the examination takes place. Include also screening tests performed in a population-based screening programme before receiving the first invitation (these examinations are often referred to as "spontaneous tests"). • Subsequent screening includes all screening examinations of individual women within the screening programme following an initial screening examination, regardless of the organisational screening round in which the examination takes place. • Unknown if initial or subsequent strata should be used for tests for which the above distinction is not available. The numbers collected in the three sub-tables should refer to strictly distinct sets of women. Always check the total figures at the bottom of the three tables to be sure that the sum of the strata is the total number expected.
Detailed content description		
D1	Individuals screened in index year	In this column you should report the number of women screened in the reference year : i.e. sTable2 column B3, if examination cohort, or column B2, if invitation cohort. This column refers to the denominator of the "Recall rate" indicator, i.e. if the numerator (number of further assessment recommended) has not been provided by all areas, then report the number of women screened in the areas where data on the number of further assessment recommendation are available.
D2	Positive	Women who have been recommended further assessment (it is a subset of D1).
D3	Negative	Women who have not been recommended further assessment (it is a subset of D1).

Trouble shooting	If your data cannot be stratified by five-year age groups, put the total amount in the row marked as "unknown". In a mixed situation, with data from some areas which can be stratified and other data that cannot be stratified, please fill separately the first rows for the former and the last row for the latter. Always check the total figures at the bottom of each table to be sure that the sum of the strata is the total number expected. Consider for example in a country where: - 20 regions provide relevant information for calculating compliance - 15 of these regions have data on recall In this case: - the number of women screened documented in table 2 will refer to the 20 regions - the number of women screened documented in table 3 will refer to the 15 regions.

sTable5: Further assessment participation

Level of importance		Should have
What does this table contain?		Further assessment participation among positive women.
What does this table NOT contain?		Data on opportunistic screening
Potential data sources		National breast cancer screening programme
Format		Absolute numbers, separated by five-year age groups. Each woman is counted only once. Data should be stratified per initial/subsequent tests: • Initial screening is the first screening examination of individual women within the screening programme, regardless of the organisational screening round in which the examination takes place. Include also screening tests performed in a population-based screening programme before receiving the first invitation (these examinations are often referred to as "spontaneous tests"). • Subsequent screening includes all screening examinations of individual women within the screening programme following an initial screening examination, regardless of the organisational screening round in which the examination takes place. • Unknown if initial or subsequent strata should be used for tests for which the above distinction is not available. The numbers collected in the three sub-tables should refer to strictly distinct sets of women. Always check the total figures at the bottom of the three tables to be sure that the sum of the strata is the total number expected.
Detailed content description		
E1	Positive	By default, the column will report the number of the women who have been recommended further assessment, column D2 in sTable4. Hence it refers to the denominator of the "Further assessment participation rate" indicator, so if the numerator (number of further assessment performed) has not been provided by all areas, then report the number of positive women in the areas where data on number of further assessment performance are available.
E2	Further assessment performed	Women who actually underwent further imaging and/or invasive assessment, irrespective of whether further assessment was complete or not (it is a subset of E1).
E3	Further assessment not performed	Women who did not undergo further assessment (it is a subset of E1).
Trouble shooting		If your data cannot be stratified by five-year age groups, put the total amount in the row marked as "unknown". In a mixed situation, with data from some areas which can be stratified and other data that cannot be stratified, please fill separately the first rows for the former and the last row for the latter. Always check the total figures at the bottom of each table to be sure that the sum of the strata is the total number expected.

sTable6: Further assessment outcome

Level of importance		Should have
What does this table contain?		Results of further assessment
What does this table NOT contain?		Data on opportunistic screening
Potential data sources		National breast cancer screening programme
Format		Absolute numbers, separated by five-year age groups. Each woman is counted only once. Data should be stratified per initial/subsequent tests: • Initial screening is the first screening examination of individual women within the screening programme, regardless of the organisational screening round in which the examination takes place. Include also screening tests performed in a population-based screening programme before receiving the first invitation (these examinations are often referred to as "spontaneous tests"). • Subsequent screening includes all screening examinations of individual women within the screening programme following an initial screening examination, regardless of the organisational screening round in which the examination takes place. • Unknown if initial or subsequent strata should be used for tests for which the above distinction is not available. The numbers collected in the three sub-tables should refer to strictly distinct sets of women. Always check the total figures at the bottom of the three tables to be sure that the sum of the strata is the total number expected.
Detailed content description		
F1	Individuals screened of invited in index year	By default, the column will report the number of women screened in the reference year, you reported in column D1, sTable4. Hence, it is also the denominator of the "Surgical referral rate" indicator. If the numerator (number of further assessment recommended) has not been provided by all areas, then report the number of women screened in the areas where data on surgical referral are available.
F2	Further assessment performed	By default, the column will report the number of women who underwent further imaging and/or invasive assessment, column E2 in sTable5. In the programme or areas where data is available on treatment referral, these are the women who actually underwent imaging and/or invasive further assessment, irrespective of whether further assessment was complete or not.

F3	Treatment/Surgery referral or inoperable ca	Women referred to open surgical biopsy or surgical intervention or neo-adjuvant therapy as a result of assessment, including also those with cancers that are not fit for surgery or other treatment (it is a subset of F2).
F4	Negative	This includes all other possible known results of assessment (it is a subset of F2). Please include also "Short Term recall", being a mammogram performed out of sequence with the screening interval (say at 6 or 12 months for programme with two-years screening interval), as a result of the screening test (not recommended by the European Guidelines) or as a result of further assessment.
Trouble shooting		If your data cannot be stratified by five-year age groups, put the total amount in the row marked as "unknown". In a mixed situation, with data from some areas which can be stratified and other data that cannot be stratified, please fill separately the first rows for the former and the last row for the latter. Always check the total figures at the bottom of each table to be sure that the sum of the strata is the total number expected.

sTable7: Outcome

Level of importance		Should have
What does this table contain?		Screening outcomes. Please indicate the most advanced lesion per woman.
What does this table NOT contain?		Data on opportunistic screening
Potential data sources		National breast cancer screening programme
Format		Absolute numbers, separated by five-year age groups. Each woman is counted only once. Data should be stratified per Initial/subsequent tests: • Initial screening is the first screening examination of individual women within the screening programme, regardless of the organisational screening round in which the examination takes place. Include also screening tests performed in a population-based screening programme before receiving the first invitation (these examinations are often referred to as "spontaneous tests"). • Subsequent screening includes all screening examinations of individual women within the screening programme following an initial screening examination, regardless of the organisational screening round in which the examination takes place. • Unknown if initial or subsequent strata should be used for tests for which the above distinction is not available. The numbers collected in the three sub-tables should refer to strictly distinct sets of women. Always check the total figures at the bottom of the three tables to be sure that the sum of the strata is the total number expected.
Detailed content description		
G1	Individuals screened of invited in index year	This column refers to column D1 in sTable4, which is also the denominator of the "Surgical referral rate" indicator, calculated in sTable6. If the numerator (number of further assessment recommended) has not been provided by all areas, then report the number of women screened in the areas where data on surgical referral are available.
G2	Further assessment performed	By default, the column will report the number of women screened in the reference year, column E2 in sTable5. In the programmes or areas where data is available on treatment referral, these are the women who actually underwent imaging and/or invasive further assessment, irrespective of whether further assessment was complete or not.
G3	Benign lesions, or no lesion	Negative screening test result
G4	CIS	Women with in situ carcinoma detected (ductal or lobular).
G5	Invasive breast cancer	Screen detected invasive cancers (any stage/pathological size)
G6	Other histology	Inflammatory node; non epithelial cancer

Trouble shooting	Screen detection means that the diagnostic assessment process following a positive primary screening examination has been completed. This process should usually be finished within six month. However, if organizational characteristics or constrains require a longer period, include all cancers diagnosed through screening in the period that suit your programme best . If your data cannot be stratified by five-year age groups, put the total amount in the row marked as “unknown”. In a mixed situation, with data from some areas which can be stratified and other data that cannot be stratified, please fill separately the first rows for the former and the last row for the latter. Always check the total figures at the bottom of each table to be sure that the sum of the strata is the total number expected.

sTable8: Pathological size

Level of importance		Nice to have
What does this table contain?		Pathological size of screen detected INVASIVE cancers. Please indicate the most advanced lesion per woman.
What does this table NOT contain?		Data on opportunistic screening
Potential data sources		National breast cancer screening programme
Format		Tumour (T) of the breast cancer TNM staging system. The staging information should be based on the histopathological assessment of the tumour (pTNM), not the clinical (cTNM). If this information is not available, please leave a comment in the NOTES sheet of the data template. T describes the size of the tumour (area of cancer) Absolute numbers, separated by five-year age groups. Each woman is counted only once. Data should be stratified per Initial/subsequent tests: • Initial screening is the first screening examination of individual women within the screening programme, regardless of the organisational screening round in which the examination takes place. Include also screening tests performed in a population-based screening programme before receiving the first invitation (these examinations are often referred to as "spontaneous tests"). • Subsequent screening includes all screening examinations of individual women within the screening programme following an initial screening examination, regardless of the organisational screening round in which the examination takes place. • Unknown if initial or subsequent strata should be used for tests for which the above distinction is not available. Screen detection means that the diagnostic assessment process following a positive primary screening examination has been completed. This process should usually be finished within six month. However, if organizational characteristics or constrains require a longer period, include all cancers diagnosed through screening in the period that suits your programme best . The numbers collected in the three sub-tables should refer to strictly distinct sets of women. Always check the total figures at the bottom of the three tables to be sure that the sum of the strata is the total number expected.
Detailed content description		
I1	T1A	Tumours between 0.1 and 0.5 cm across, independent of node status
I2	T1B	Tumours between 0.5 cm and 1 cm across, independent of node status
I3	T1C	Tumours between 1 cm and 2 cm across, independent of node status
I4	T2-T4	Tumours between 2 cm and 5 cm across, independent of node status

Trouble shooting	If your data is available in a different staging system than TNM, please consider the indication for the tumour size only and fill the columns based on that information. If your data cannot be stratified by five-year age groups, put the total amount in the row marked as "unknown". In a mixed situation, with data from some areas which can be stratified and other data that cannot be stratified, please fill separately the first rows for the former and the last row for the latter. Always check the total figures at the bottom of each table to be sure that the sum of the strata is the total number expected.

sTable9: Surgical treatment

Level of importance	Nice to have
What does this table contain?	Treatment of cancer
What does this table NOT contain?	Data on opportunistic screening
Potential data sources	National breast cancer screening programme
Format	Surgical treatment, separated by age groups. Each woman is counted only once. Data should be stratified as follows: • Screen detected in situ carcinoma (ductal or lobular). • Screen detected invasive carcinomas Data should be stratified according to the final surgery, for example in cases when breast conservation therapy is followed by mastectomy, the final surgery is mastectomy.
Detailed content description	
Trouble shooting	Only proportions larger than 0% can be entered. Screen detection means that the diagnostic assessment process following a positive primary screening examination has been completed. This process should usually be finished within six month. However, if organizational characteristics or constrains require a longer period, include all cancers diagnosed through screening in the period that suit your programme best . If your data cannot be stratified by a, put the total amount per treatment in the row marked as "unknown". In a mixed situation, with data from some areas which can be stratified and other data that cannot be stratified, please fill separately the first rows for the former and the last row for the latter.

sTable10: Interval cancers

Level of importance		Should have
What does this table contain?		Interval cancers
What does this table NOT contain?		Data on opportunistic screening
Potential data sources		National breast cancer screening programme
Format		Absolute numbers, separated by five-year age groups. Each woman is counted only once. Data should be stratified per Initial/subsequent tests: • Initial screening is the first screening examination of individual women within the screening programme, regardless of the organisational screening round in which the examination takes place. Include also screening tests performed in a population-based screening programme before receiving the first invitation (these examinations are often referred to as "spontaneous tests"). • Subsequent screening includes all screening examinations of individual women within the screening programme following an initial screening examination, regardless of the organisational screening round in which the examination takes place. • Unknown if initial or subsequent strata should be used for tests for which the above distinction is not available. The numbers collected in the three sub-tables should refer to strictly distinct sets of women. Always check the total figures at the bottom of the three tables to be sure that the sum of the strata is the total number expected.
Detailed content description		
M1	Interval cancers diagnosed within the first year after test	Number of women diagnosed with an interval cancer within 12 months since the last negative mammography
M2	Interval cancers diagnosed within the second year after test	Number of women diagnosed with an interval cancer between 12 and 24 months since the last negative mammography
M3	Interval cancers diagnosed within the third year after test	Number of women diagnosed with an interval cancer between 24 and 36 months since the last negative mammography. This only applies to countries with a three-year interval programme. In a two-year programme, column M3 will be empty.
M4	Last negative test in (reference year) - 2	Number of women with a negative primary screen and the number of negative further assessment performed 2 years before the reference year for data collection (i.e. previous screening round)

Trouble shooting	If the index year for your reported data is 2016, M1 should report on women screened in 2015 and diagnosed with an interval cancer in 2016, whereas M2 reports on women screened in 2014 and diagnosed with an interval cancer in 2016. M4 is the number of negative primary screens plus the number of negative further assessment in 2014 (two years prior to the index year) Interval cancers can occur in two scenarios: Women A: negative screen – breast cancer clinically detected after XX months interval cancer Women B: positive screen – negative further assessment – breast cancer clinically detected after XX months interval cancer Both interval cancers, A and B, should be counted here and would set the numerator of the interval cancer rate, whereas M3 is the denominator. If your data cannot be stratified by five-year age groups, put the total amount in the row marked as "unknown". In a mixed situation, with data from some areas which can be stratified and other data that cannot be stratified, please fill separately the first rows for the former and the last row for the latter. Always check the total figures at the bottom of each table to be sure that the sum of the strata is the total number expected.

sTable11: Opportunistic screening

Level of importance	Should have
What does this table contain?	An estimate of the proportion of women (of all women in the population) who underwent a mammography outside the organised programme in the past two years.
What does this table NOT contain?	Data on population-based, organised screening
Potential data sources	National surveys, National social insurance institution
Format	Separated by five-year age groups.
Trouble shooting	The classification non-population based (opportunistic) screening applies to areas where individual invitations are not sent to the women in the eligible population or when women undergo a mammography outside or additionally to the (existing) screening programme. Together with sTable2, this information enables the EU-TOPIA team to include the TOTAL screening reality into the simulation of your country or region. As an example of a mixed situation, think of women who have been invited to the organised screening programme three years ago, who did not follow the invitation but got screened opportunistically in the index year instead. Those women have to be counted in this table. If you have individual screening histories but you are unable to disentangle them into “organised” and “opportunistic” correctly, then the available data should be put into the respective tables for organised screening only (sTable2). If your data cannot be stratified by five-year age groups, put the total amount in the row marked as “unknown”.

sTable 12: Adjuvant treatment

Level of importance	Should have
What does this table contain?	Adjuvant treatment of all cancers (detected inside or outside a screening programme)
What does this table NOT contain?	Surgical treatment and radiology
Potential data sources	National breast cancer screening programme, hospital or insurance data
Format	Proportions of adjuvant treatment, separated by age groups. The data reported here should refer to all cancer cases detected in the index year plus the year prior to that (e.g. if your index year is 2016, the cancers you report should have been diagnosed in either 2015 or 2016), independent of when or whether the woman has ever been screened. Each woman is counted only once. Data should be stratified per cancer stage and by therapy • CIS: in situ carcinoma (ductal or lobular) • T1a/b: tumours between 0.1 and 1 cm across, independent of node status • T1c: tumours between 1 cm and 2 cm across, independent of node status • T2-T4: Node positive tumours between 2 cm and 5 cm across, independent of node status • No adjuvant therapy: no other treatment except for surgery or radiation • Chemotherapy: adjuvant to surgery and/or radiation • Hormonal therapy: adjuvant to surgery and/or radiation • Combined: a combination of chemotherapy and hormonal therapy

5. Simulation

In this section you will find instructions on how to use the EU-TOPIA evaluation tool to simulate different breast cancer screening strategies for your country.

After registering and filling out the Excel data templates, as described in section 3, you are ready to start a simulation.

5.1. Upload data

1. Log in into the EU-TOPIA evaluation tool (https://miscan.eu-topia.org/login) with your e-mail and password.
2. Go to the simulation section in the MISCAN web- tool dashboard.
3. Start by selecting a cancer type, breast cancer in this case, and giving your simulation a name. (Figure 5).
4. Then, upload the data for your country by clicking “Choose file” and select the data template that you completed as described in Section 4.
   Users from one of the exemplary countries (Finland, Italy, the Netherlands and Slovenia) can click on the “Choose” button and either directly select the data from their country in the drop-down menu or select “upload your own data” to upload the data template.

Press “Continue” (Figure 5).

Figure 5. EU-TOPIA evaluation tool, Screening type, name and data

5.2. Quality check of the data

After uploading your data, the web-based tool performs checks on data quality and completeness.

In the next step you are asked to verify your data (Figure 6). If you wish you can choose to use data already provided by the exemplary country of your region (North: Finland, South: Italy, West: the Netherlands, East: Slovenia) for non-mandatory data. If the data that you uploaded is missing (indicated by a red cross), incomplete or of insufficient quality (indicated by an orange cross) for specific non-mandatory tables, the tool will automatically use the respective data from the exemplary country.

If you did enter data in the template but the evaluation tool marks your data as of insufficient quality, please check if the format of your data is as specified in section 4, within the allowed limits and proportions add up to 100%, if applicable. Also, read the 'trouble shooting' part at the bottom of the respective instruction table in section 3 if present. If you still encounter problems with data quality, please contact us at eu.topia@erasmusmc.nl

Press "Continue" (Figure 6).

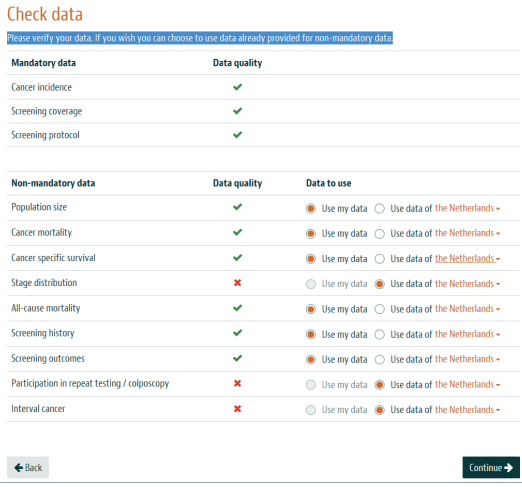


Figure 6. EU-TOPIA evaluation tool, verify your data

5.3. Exemplary countries

Based on the country that you registered from, the EU-TOPIA admins will assign you to one out of four European regions (see list below). Based on that region, model parameters from the exemplary country of that region are used in case non-mandatory data is missing (Figure 6). Also, the parameters that were calibrated as described in the model description (available

via the “information” icon at the dashboard), are based on the regional model of the exemplary country. Background information on all of the four exemplary countries is also available via the “information” icon at the dashboard in the fact sheets.

If you think that your country would be better represented by one of the other three exemplary countries, you can consult the EU-TOPIA research team at eu.topia@erasmusmc.nl

Northern Europe: Finland (exemplary country), Denmark, Estonia, Faroe Islands, Iceland, Latvia, Lithuania, Norway and Sweden.

Southern Europe: Italy (exemplary country), Cyprus, Gibraltar, Greece, Malta, Portugal and Spain.

Eastern Europe: Slovenia (exemplary country), Bulgaria, Czech Republic, Croatia, Hungary, Poland, Romania and Slovakia.

Western Europe: The Netherlands (exemplary country), Austria, Belgium, France, Germany, Ireland, Luxembourg, United Kingdom and Switzerland.

5.4. Selection of screening scenarios

Now you are asked to define the settings of each scenarios you want to simulate (Figure 7). You can add more scenarios (a maximum of 5 per simulation) by pressing the + sign.

Please note that the changes of the scenario parameters will be affective from 2020 onwards.
The default setting is the current status of your screening programme according to the mandatory data you provided. You are able to change the following parameters:

Sensitivity: Select whether the test sensitivity should be improved. If you pick this option, the sensitivity will be increased by 5% across all age-groups.

Target age: Select the starting age and the maximum age at which women should be screened in this simulation.
The actual ages at which screening is performed in the simulation are determined by the starting age and the screening interval. Therefore, the last screening age of a woman can be slightly lower than the maximum screening age if the selected starting age and screening interval do not result in a screening invitation at this maximum age. For example, if the user selects a starting age of 50, a maximum age of 65 and an interval of 2 years, the last screening performed will be at age 50 + 7*2years = 64 years old.

Screening interval: Select how many years there should be between each screening round.

Adherence: Select how much the adherence should be increased. As a default, the screening coverage entered in the data template (sTable2) is used. Please note that, for each age-group, the current total examination coverage is calculated from column J in sTable2 (Examination coverage) PLUS the information on opportunistic screening participation from sTable11. If sTable11 is left empty, only the information from sTable2 is used.
By choosing an increase in adherence, the respective value (2.5%, 5%, 10%, 20%, 30% and 40%) will be added to the examination coverage of each of the age-groups equally. If, for example, the current coverage is 60% in the age-group 50-54, 65% in the age-group 55-59, 70% in the age-group 60-54, and 75% in the age-group 65-69 and you choose an increase in adherence of 10%, the coverage will increase to 70%, 75%, 80% and 85%, respectively.

Stop screening: If you would like to simulate the effect of a full stop of (organised) screening activities for your country or region from 2020 onwards, you can do so by choosing the "Stop screening" option under "adherence". This option will overwrite all other parameter choices for this scenario.

No duplicate scenario can be selected.

Press "Start simulation" (Figure 7). Confirm that you are sure you want to start your simulation.

Figure 7. EU-TOPIA evaluation tool, choose scenarios and start simulation.

Your simulation will be submitted for processing. You will receive an e-mail as soon as your simulation is finished.

6. Results

In this section it will be described how to download and interpret the results of the EU-TOPIA evaluation tool.

6.1 Downloading the results

Once your simulation is finished and the results are ready, you will receive an e-mail with a link. Following the link, you will reach your simulation online and be able to download a PDF report with the results (Figure 8).

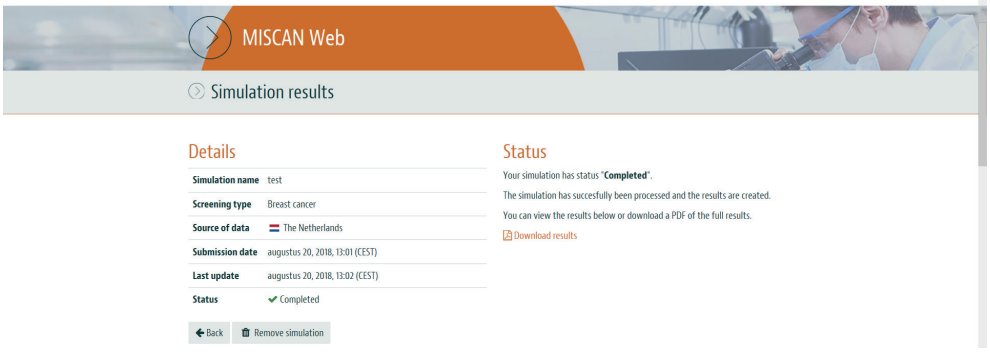


Figure 8. EU-TOPIA evaluation tool, view and retrieve your simulation results

Alternatively, you can also find the results by going to the RESULTS section available in the EU-TOPIA evaluation tool dashboard (Figure 9). You will reach the results of the simulation by clicking on the specific simulation you are interested in.

Figure 9. EU-TOPIA evaluation tool, access simulation results via dashboard

Here, an overview of all simulations you started is provided, including the status of the simulation, which can be:

Submitted: the server is currently running multiple other simulations, your simulation is in the queue.

Processing: your simulation is currently running on the server, your results will be ready soon.

Failed: something went wrong, the EU-TOPIA admins are informed and will contact you about a solution.

Completed: you will reach the results of the simulation by clicking on the simulation

On the page with the results of your simulation, you can find a summary of your simulated screening scenarios and first results: a graph with crude incidence rates per year (2020-2050), a graph with crude mortality rates per year (2020-2050), and a summary table with main outcomes for the population for the years 2020-2050, respectively.

You can retrieve the PDF of the simulation report by pressing
If you want, you can start a new simulation.

6.2. Description of the results.

In the simulation, the life histories of 10,000,000 women were simulated, including Ductal Carcinoma In Situ (DCIS). The (new) screening strategy, including the adherence to screening, was applied. The EU-TOPIA evaluation tool produces several types of outputs, of which table 1 and 2 as well as figure 1 and 2 are displayed in your web browser, while the downloadable PDF presents the remaining more detailed tables as well.

Box 1

The first overview you see is a summary of data sources for the model inputs. This is based on the data you verified for your simulation – which are either based on the country-specific data you provided or on the data from the exemplary country of your region for non-mandatory data (see Section 5.2.)

Table1

Here the settings of the selected scenario(s) are displayed. You can see the selected test sensitivity, target ages, screening interval and non-adherence reduction for each scenario. If current screening was selected, the settings as specified in the uploaded data template are used.

Current screening strategy

This table provides an overview of the current screening strategy and the adherence by age that all simulations are based on.

Figure 1

In this figure the crude breast cancer incidence rates per 100,000 women years between 2020 and 2050 are shown by age group for each of the scenarios.

Figure 2

In this figure the crude breast cancer mortality rates per 100,000 women years between 2020 and 2050 are shown by age group for each of the scenarios.

Table2

The results represents a summary of the screening outcomes in women aged 40 to 100 in 2020 to 2050 with the selected test sensitivity, target ages, screening interval and non-adherence reduction for each scenario.

The number of overdiagnosed breast cancers (DCIS and invasive) represent women that would not have been diagnosed during their lives if they had never been screened as a percentage of all diagnosed women. In addition, two ratios of harms and benefits are provided for each

scenario: the number of screens needed to prevent one breast cancer death and the number of false positives per breast cancer death prevented.

Tables 3-10

In the downloadable PDF, all numbers from Table2 are available separated by age-group in tables 3-10.

Chapter 7

Disability-Adjusted Life Years Averted Versus Quality-Adjusted Life Years Gained: A Model Analysis for Breast Cancer Screening

Maša Davidović, Nadine Zielonke, Iris Lansdorp-Vogelaar, Nereo Segnan, Harry J de Koning, Eveline A.M. Heijnsdijk

Value Health 2021 Mar;24(3):353-360

ABSTRACT

Objectives

To quantify the impact of mammography-based screening on the quality of life, disability-adjusted life years (DALYs) averted or quality-adjusted life years (QALYs) gained can be used. We aimed to assess whether the use of DALYs averted or QALYs gained will lead to different cost-effective screening strategies.

Methods

Using the microsimulation model MISCAN, we simulated different breast cancer screening strategies varying in starting age (starting at 45, 47, and 50 years), stopping age (stopping at 69, 72, and 74 years), and frequency (annual (A), biennial (B), combination of both (A+B) and triennial (T)). In total, we defined 24 different breast cancer screening strategies including no screening as a reference strategy. We calculated incremental cost-effectiveness ratios (ICERs) and compared which strategies were on the efficiency frontiers for DALYs and QALYs.

Results

Breast cancer screening averted between 46.00 and 105.58 DALYs and gained between 28.69 and 64.50 QALYs per 1,000 women. For DALYs there were five strategies on the efficiency frontier (T50-69, T50-74, T45-74, B45-74 and A45-74). The same strategies plus one (B45-72) were on the efficiency frontier for QALYs.

Conclusion

Using DALYs averted instead of QALYs gained to assess the effects on quality of life from breast cancer screening in the Dutch population yields differences in ICERs, but almost the same strategies were on the efficiency frontiers. Whether the choice in outcome measure leads to a difference in optimal policy depends on the cost-effectiveness threshold.

Abbreviations

QoL – Quality of life
DALYs – Disability-adjusted life years
QALYs – Quality-adjusted life years
CEA - Cost-effectiveness analyses
YLLs – The sum of years of life lost from premature death
YLDs – Years of life lived with disability
HRQoL – The health-related quality of life
ICERs – Incremental cost-effectiveness ratios
WHO – The World Health Organization

INTRODUCTION

Mammography screening is strongly recommended for all asymptomatic women aged 50 to 69 years with an average risk for breast cancer and it leads to early detection and early treatment of breast cancer[1,2]. Several randomized controlled trials have established its effectiveness demonstrating, on average, a 25% breast cancer mortality reduction from mammography screening among women 50 to 69 years old[3]. However, randomized controlled trials are limited in follow-up time and only a limited number of screening strategies can be evaluated.

Cost-effectiveness analysis evaluates the life-time harms and benefits of mammography screening and the resulting cost-effectiveness of different screening strategies to determine the optimal screening program[4]. Traditionally, cost-effectiveness analyses mostly considered life-years gained to calculate the ratio of costs per unit of effectiveness[4,5]. To capture not only the effects on mortality and morbidity, but also to compare different burden of diseases and the impact of screening itself, two alternative measures of life-years gained have been proposed: disability adjusted life-years (DALYs) and quality-adjusted life-years (QALYs). These are summary measures of population health that allow the combined impact of death and morbidity to be considered simultaneously[6]. DALYs are a measure of the burden of disease, and studies on cost-per-DALY mostly have been focusing on health interventions in lower income country contexts analyzing health effects in global health[7,8]. QALYs are more dominantly used in economic evaluations, but mostly in higher income countries[7,8]. The number of cost-effectiveness studies, which included costs per DALY averted or QALY gained has grown globally in recent years, with more than 600 and 7,000 reported studies by 2017 respectively, to measure health benefits[8–11]. Still, information on the impact of using DALYs versus QALYs on outcomes of cost-effectiveness analyses and optimal screening strategies is lacking.

The aim of this research was to determine whether using DALY averted or QALY gained in cost-effectiveness analyses of 24 screening strategies leads to differences in estimates of optimal strategies for breast cancer screening in the Netherlands.

METHODS

Using the microsimulation model MISCAN, we simulated different breast cancer screening strategies varying in starting age, stopping age, and frequency. To determine the optimal screening strategy, we evaluated the costs and effects of different screening strategies, using both DALY averted and QALY gained.

DALYs and QALYs

DALY is "a measure of the gap between the ideal health and the current health status"[6]. It focuses on the burden of a disease and health loss in the quality of life, due to combined effects of disability and premature mortality as a consequence of the disease. DALYs are based on the sum of years of life lost from premature death (YLLs) and years of life lived with disability (YLDs) (Figure 1)[6]. In contrast, QALY measures the quality of life in health gain. It assesses the health state of a person in which the benefits, in terms of length of life, are adjusted to reflect the quality of life[6]. If a person lives in perfect health, QALYs will be equal to life-years lived. However, due to a disease, quality of life can be affected, and a reduction of quality needs to be subtracted from the life-years depending on the severity of the disease. QALYs are calculated by multiplying the health-related quality of life (HRQoL) utilities of a specific health state, with the duration of that health state (Figure 1)[6]. In this study, the impact of a screening strategy on DALYs and QALYs (i.e. how many DALYs can be averted or QALYs gained by the intervention) is measured as the difference between a situation with and without screening (Appendix 1, Figure 1 in Supplemental Materials)[12].

Literature review

We performed a literature review to identify studies estimating disability weights or health-related quality of life (HRQoL) utilities related to screening, breast cancer and its consequences. Since 2010, the Global Burden of Disease Project (GBD)[13] has developed a new standardized approach and composed a list of disability weights to allow overall comparisons across regions and countries. In short, they defined cancer-specific health states, and calculated the values based on household surveys and a web-based survey. Salomon et al.[14] re-estimated disability weights from the GBD Study 2010, and have expanded and improved the description of health states with participants from four European countries, including the Netherlands. Therefore, we decided to use disability weights from this study[14,15]. The disability weights range from 0.036 for mastectomy to 0.54 for the terminal phase of breast cancer. To calculate QALYs, HRQoL utilities were extracted from three carefully selected papers that describe health states most comparable to the set defined by the GBD Study[16–18]. The utilities varied from 0.99 for mammography screening to 0.288 for the last month in the terminal phase of breast cancer (Table 1). Appendix 2 Table 1 presents the full list of HRQoL utilities and disability weights extracted from literature (in Supplemental Materials).

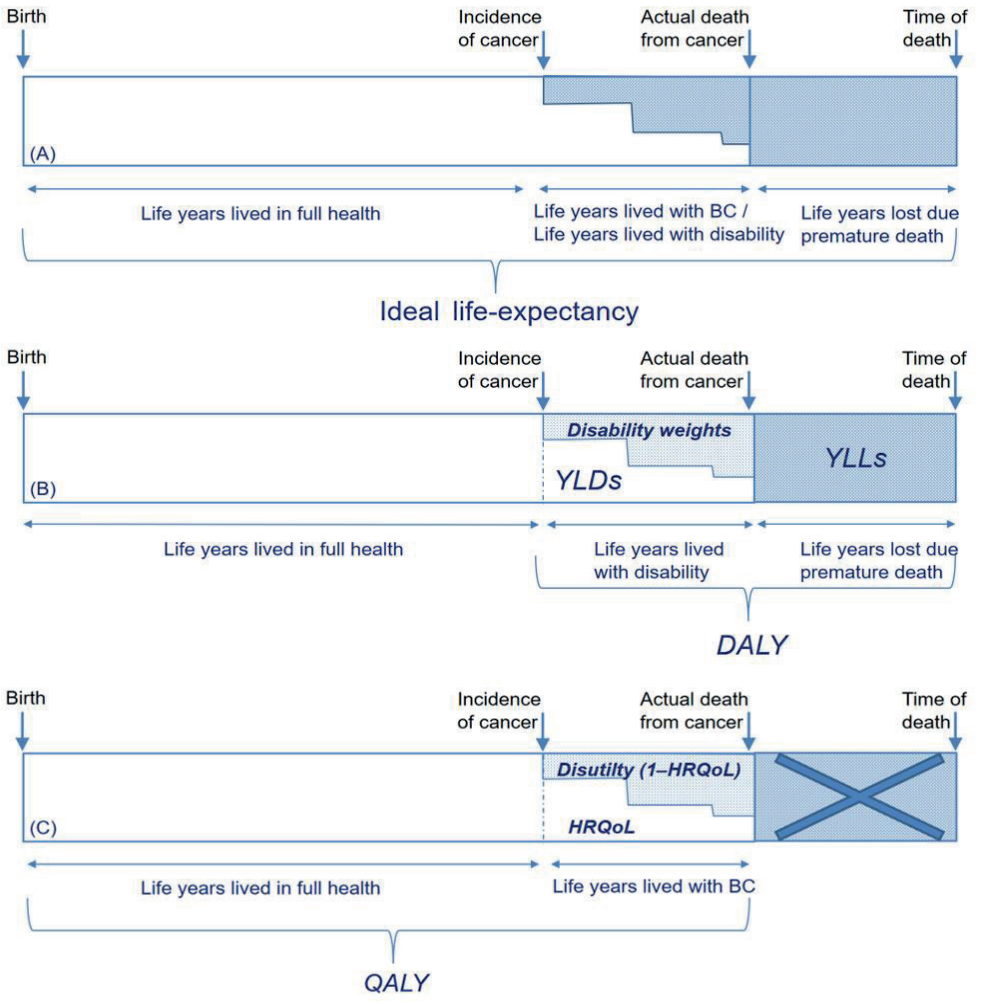


Figure 1. DALYs and QALYs.

Each box represents the lifespan from birth to death. A person can live the whole life in full health (ideal life-expectancy) (A). When the person is diagnosed with cancer, QoL is reduced. The effects on QoL can be measured on two ways. Firstly, we can estimate how long a person lived with disability due to disease (YLDs) and how many years the person lost due to disease (YLLs) (B). YLLs are calculated as a number of deaths and life expectancy at that age. To calculate YLDs, prevalence of a given health state is multiplied by a disability weight (DW). DW are anchored between 0 and 1, where 0 represents full health and 1 is equivalent to death. The sum of these two measures represents DALY for a specific person and specific disease. One DALY can be thought of as one year of "healthy" life lost. Secondly, we can estimate the loss of QoL during life years lived with disease (C). QALYs are calculated as a time spent in specific health state with HRQoL utility for that health state. One QALY is equal to 1 year of life in perfect health. As it is shown (B and C), disability weights and disutility are two parts of the same equation and could result in the same outcomes. However, disability weights and disutility are currently measured very differently and therefore are not necessarily complementary.

Table 1. Disability weights and health-related quality of life utilities and disutilities for breast cancer related health states

Health state descriptions	**Disability weight (95% CI)**		**Reference**
Diagnosis and primary therapy	0.288 (0.193-0.399)		
Controlled phase of breast cancer	0.049 (0.031-0.072)		
Metastatic phase of breast cancer	0.451 (0.307-0.600)		Global Burden Disease Study 2015 [15]
Terminal phase of breast cancer	0.540 (0.377-0.687)		
Mastectomy due to breast cancer	0.036 (0.020-0.057)		
Health state descriptions	HRQoL utility	Disutility[a]	Reference
Mammographic screening – 1 week	0.99	0.01	De Koning et al. (1991) [17]
Biopsy with benign results / due to false positive	0.89	0.11	De Haes et al. (1991) [16]
3 months – 1 year after mastectomy	0.84	0.16	De Haes et al. (1991) [16]
Breast cancer[b]	0.83 (0.40-0.90)[c]	0.17	Schleinitz et al. (2006) [31]
Base state – stable metastatic disease with no toxicity	0.71	0.29	Lloyd et al. (2006) [18]
Terminal illness – 1 month	0.29	0.71	De Haes et al. (1991) [16]

Abbreviations: CI – Confidence Interval; HRQoL utilities – Health Related Quality of Life Utilities;

[a] Disutility is calculated as 1 – HRQoL

[b] The value is the mean of utilities for stage 1 and stage 2 by TNM classification (for more, see Schleinitz et al. 2006) [31].

[c] Utility and standard error

MISCAN model

The MISCAN model is a well-established microsimulation model. It was designed to evaluate the effect of breast cancer screening. MISCAN simulates individual life histories and assesses the consequences of introducing a screening program on these life histories[19]. The model estimates the effect of cancer screening in a population and can explain results of cancer screening trials and predict and compare the cost-effectiveness of different screening strategies[19]. Our model has been calibrated on Dutch data, and the standard life expectancy of Dutch women was incorporated[20]. We used the MISCAN model to quantify the effects of mammography-based screening, comparing screened to non-screened life histories and to estimate DALYs, QALYs, DALYs averted and QALYs gained and costs for each screening strategy.

Screening strategies

We simulated different breast cancer screening strategies varying in starting age (starting at 45, 47, and 50 years), stopping age (stopping at 69, 72, and 74 years), and frequency (annual, biennial, triennial and combination of annual and biennial). In total, we defined 24 different breast cancer screening strategies including no screening as a reference strategy. Overall, we assumed an attendance of 80% to all screening invitations[21], which means that there is an independent probability of 80% to attend each screen. We assumed 100% diagnostic follow-up for women with a positive test.

Model parameters and assumptions

In our modeling analysis, one birth cohort of 10 million Dutch women, born in 1965, was simulated for every screening strategy and women were followed from age 40 to death. For each strategy, the total number of screens, the numbers of breast cancers diagnosed, breast cancers deaths, life-years from diagnosis to death, and age of death were predicted. We defined the possible life histories a woman might pass through from screening to death of breast cancer or other causes (Appendix 3, Figure 1 in Supplemental Materials). For women who had or have been diagnosed with breast cancer, two clinical stages were assumed: controlled breast cancer (Stage 0 (*ductal carcinoma in citu*), Stage 1-3) and metastatic breast cancer (Stage 4) based by TNM classification of malignant tumors staging system. Stage distribution is based on Dutch Cancer Registry data[22], and the proportion of controlled breast cancer cases, which progressed to metastatic stage was assumed by model. In the model, we also assumed that all women with breast cancer (controlled of metastatic) received lumpectomy or mastectomy with or without radiation. In order to investigate the impact of screening on a woman's quality of life for each life history we assigned the utility and disability-values for each health state obtained from the literature review. The model includes overdiagnosed cases; therefore, the impact of overdiagnosis in terms of disability and morbidity is included as well.

Additionally, to analyze the costs of each screening strategy, the number of events was calculated and multiplied with their respective costs. The costs were based on Sankatsing et al. (2015)[1], and were €65 for screening, €160 for biopsy, €9,000 for lumpectomy/mastectomy[23], €6,000 for controlled breast cancer, €16,000 for metastatic breast cancer and €18,000 for death from breast cancer (including palliative care). Women who went to several stages had the sum of the costs in those stages.

Outcomes

To estimate DALYs for each screening strategy, YLLs were calculated by multiplying the number of estimated breast cancer deaths in 5–year age groups with the respective remaining life-expectancy at that age[20]. The YLDs for each health state were calculated by multiplying the time spent in that health state with the number of women who went through that health state – provided by the MISCAN model – and with the specific disability weight. The sum of all health state-specific YLDs represents the total YLDs of one screening strategy. The sum of the total YLDs and YLLs produced the DALYs for that specific screening strategy

QALYs were calculated by following all life histories and multiplying the time spent in a specific health state with its score of 1 – HRQoL utility (disutility) and the number of women who went through that health state. Finally, these calculation methods were implemented for each screening strategy and the results were compared to the reference strategy (no screening) to calculate DALYs averted and QALYs gained (Appendix 4 Tables 1-7 presents an example of the calculation of DALYs and QALYs in Supplemental Materials).

Cost-effectiveness analysis

The net costs for each strategy were defined as the difference between the total costs of the strategy and the costs of the reference strategy. We conducted analyses with 3% annual discount for costs and effects as recommended by WHO[24], which means that costs and effects in the future will count less than current costs and effects. Undiscounted results are presented in the Supplement (Appendix 5 Tables 1 and 2 in Supplemental Materials). Finally, we identified the efficient frontier of non-dominated strategies and calculated incremental cost-effectiveness ratio (ICERs) among these strategies. We assumed the ICER threshold of €20,000 per QALY gained, recommended by the Health Council of the Netherlands[25]. For the purpose of our study and simplification of comparison between these two measures, we assumed an equal threshold for both QALY gained and DALY averted analysis.

RESULTS

The least intensive screening strategy considered (T50-69) led to additional 46.00 DALYs averted and 28.69 QALYs gained per 1,000 women (Table 2). The most expensive strategy is annual screening from age 45 to 74 years, which averted 105.58 DALYs and gained 64.50 QALYs per 1,000 women.

Table 2. The discounted net costs in Euros, LYG, DALYs averted and QALYs gained per 1,000 women per strategy

Screening strategy[a]	Net costs per strategy[b]	LYG	DALYs averted	QALYs gained
No screening	*Ref. (Total cost: 1,201,486)*	*Ref.*	*Ref.*	*Ref.*
T50-69	56,217	36.70	46.00	28.69
T50-74	64,326	40.10	49.68	30.48
T45-69	111,203	46.90	61.79	37.69
B50-69	118,112	46.90	58.30	36.60
T45-74	119,276	50.30	65.45	39.50
B50-72	129,089	49.80	61.40	38.19
B50-74[c]	140,301	51.70	63.44	39.00
B47-69	158,257	53.80	68.77	42.68
B47-72	169,153	56.70	71.84	44.24
B47-74	176,803	58.10	73.28	44.91
B45-69	190,112	57.80	75.32	46.20
B45-72	200,995	60.80	78.40	47.87
B45-74	208,447	62.20	79.88	48.52
A47-50_B50-69	212,050	57.30	74.23	45.79
A47-50_B50-72	226,701	61.20	78.34	47.91
A47-50_B50-74	234,261	62.40	79.69	48.39
A45-50_B50-69	262,407	62.00	81.53	49.84
A45-50_B50-72	277,127	65.90	85.63	51.98
A45-50_B50-74	284,729	67.10	86.97	52.46
A50-69	323,346	63.00	76.92	48.59
A50-74	349,540	67.50	81.58	50.92
A45-69	456,493	78.40	100.94	62.02
A45-74	462,471	83.00	105.58	64.50

Abbreviations: LYG – Life Years Gain, DALY – Disability Adjusted Life Years, QALY – Quality Adjusted Life Years

[a] Screening strategies consist of annual (A), biennial (B) and triennial (T) screening or combination of annual and biennial (A + B)

[b] Result is presented in Euros

[c] Current screening strategy in the Netherlands

Figure 2 depicts the relation between DALY averted or QALY gained and the net costs for each screening strategy. For DALYs there were five strategies on the efficiency frontier (T50-69, T50-74, T45-74, BB45-74 and A45-74). The same strategies plus one (B45-72) were on the efficiency frontier for QALYs. The ICERs ranged between €1,222 and €9,883 per each DALY averted, and between €1,960 and €15,899 per each QALY gained (Table 3). Therefore, assuming an ICER threshold of €20,000 per DALY averted or QALY gained, all evaluated screening strategies can be considered cost-effective and the most intensive scenario, screening annually from 45 to 74 years old, can be considered as optimal.

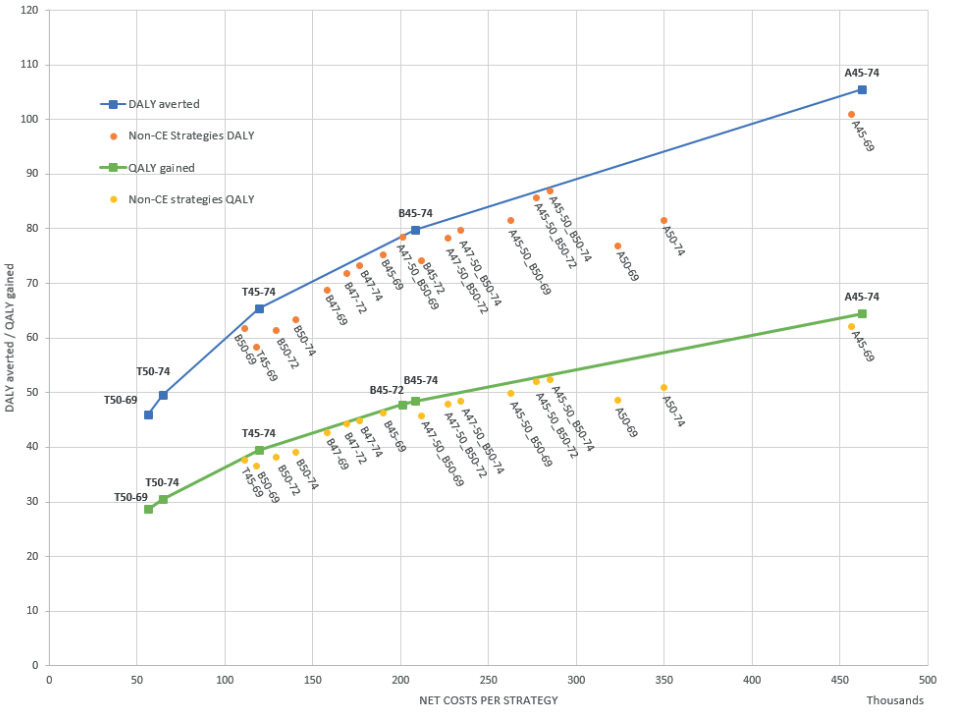


Figure 2. Efficiency frontier of screening strategies.

Each screening strategy is represented by a point. Both DALYs averted or QALYs gained and costs for each strategy are relative to a situation without screening and are presented per 1,000 women. Strategies consist of annual (A), biennial (B), triennial (T) screening and combination of annual and biennial (A+B). Strategies on the green (QALY gained) and blue line (DALY averted) represent cost-effective strategies. Dominated strategies are represented by dots only.

Table 3. Additional discounted costs; DALYs averted; and QALYs gained, and Incremental Cost-Effectiveness Ratio for cost-effective screening strategies

Cost-effective strategies – ICER							
DALY analyses				QALY analyses			
Screening strategy	Additional costs[a]	Additional DALY averted	ICER[b]	Screening strategy	Additional costs[a]	Additional QALY gained	ICER[b]
No screening	Ref.	Ref.	Ref.	No screening	Ref.	Ref.	Ref.
T50-69	56,217	46.00	1,222	T50-69	56,217	28.69	1,960
T50-74	8,109	3.67	2,207	T50-74	8,109	1.80	4,512
T45-74	54,950	15.77	3,484	T45-74	54,950	9.01	6,096
/	/	/	/	B45-72	81,720	8.37	9,776
B45-74	89,171	14.43	6,180	B45-74	7,452	0.66	11,364
A45-74	254,024	25.70	9,883	A45-74	254,024	15.98	15,899

Abbreviations: DALY – disability adjusted life years; ICER – incremental cost-effectiveness ratio; QALY – quality adjusted life years

[a] The results are presented in Euros, and reflect the additional costs relative to the less effective strategy in the ranking

[b] The results are presented in Euros per one DALY averted or QALY gained, and are calculated as the difference in costs divided by the difference in DALY averted or QALY gained between a strategy and the previous, less effective strategy in the ranking

DISCUSSION

Both DALYs averted and QALYs gained increased with expanding age limits for screening. The lowest results, in both DALYs averted and QALYs gained, were in the strategy with the shortest period of screening (T50 to 69). The screening strategy with the biggest age interval and the highest screening frequency (A45-74) had the highest impact on life-years gained, regardless of measure used. However, by increasing the age range or frequency of screening, the costs increased as well, consequently making this the most expensive strategy. All ICERs were below the cost-effectiveness threshold of €20,000 per additional QALY gained or DALY averted. In our analyses, the most cost-effective strategy in both DALY and QALY analyses, is annual screening from 45 to 74 (A45-74), with an ICER of €9,883 per DALY averted and €15,899 per QALY gained. Although almost the same strategies were on the efficiency frontier for DALYs and QALYs, the ICERs were different. Therefore, when choosing another threshold or simulating more intensive protocols, the cost-effective strategy might differ between DALYs and QALYs.

The costs per QALY are lower from results published elsewhere[4,26,27]. However, these discrepancies in cost-effectiveness estimates between our results and results reported by others can be explained by variations in screening and treatment costs (the Dutch costs are comparably low),

annual discount rate, geographical settings and simulated screening strategies. In comparison with a previous cost-effectiveness study performed with the same model[1], we used less detailed costs. The outcomes however, are comparable.

Interpreting and comparing our results using DALYs in cost-effectiveness analysis is challenging, because the DALY approach is used dominantly in evaluating the intervention costs in low- and middle-income countries, where population based mammography is not implemented. Additionally, our results show that the order of screening strategies remained almost the same regardless using DALY or QALY approach. However, there were differences between absolute estimates of DALYs averted and QALYs gained, because of differences in used utilities. DALY for example does not incorporate disabilities from the mammography and biopsy, whereas QALY do incorporate disutility from diagnosis and primary therapy. Also disability weights for having cancer were generally higher than the disutilities from having cancer. In order to test the effect of disutilities for each health state, we conducted a sensitivity analysis by replacing 1 – HRQoL disutilities with values used as disability weights for health states that overlapped (mastectomy; controlled breast cancer; metastatic breast cancer and terminal disease) or 0 (screening and biopsy). Utilities were replaced one at a time, in both the reference strategy (no screening) and an exemplary strategy (B50-69), while other utilities remained the same (respective 1 – HRQoL disutility). The QALYs gained were sensitive to varying disutilities for controlled and metastatic breast cancer, whereas changing the disutilities for screening, biopsy and mastectomy hardly had any effect.

Additionally, we tested the effect of varying all disutilities together by replacing the health disutilities (1 – HRQoL) for mastectomy; controlled breast cancer; metastatic breast cancer; and terminal disease with corresponding disability weights in all screening strategies, including the no screening strategy, at the same time. Thus, we re-calculated the ICERs and identified the non-dominated strategies. Compared with previous results, the same strategies were at the efficient frontier for both DALY averted and QALY gained used, except one that did not appear as non-dominant (B45-72) (Appendix 6 Figure 1 in Supplemental Material).

Therefore, in this breast cancer screening example, the difference between DALYs averted and QALYs gained was mainly caused by the difference in disability weights and disutilities for the controlled breast cancer state. The impact of utility used in such an analysis is previously reported as a major driver of differences between DALY and QALY measures[11].

When interpreting the results of our study, some limitations are noteworthy. Firstly, we only considered two stages of breast cancer: metastatic stage and controlled breast cancer (all stages except metastatic stage). To provide a more sophisticated analysis, additional research should investigate a more precise estimate for disability by cancer stage. Secondly, to estimate

DALYs and YLLs in our analysis, we followed the same method as used in the GBD study[15], except the remaining life expectancy. We used country-specific remaining life expectancy for each age provided by the World Health Organization (WHO)[20] which are slightly lower than the life expectancy used in the GBD study. Therefore, when we compared the DALYs and DALYs averted from our analysis with the GBD study results, there were slight differences. We believe the country-specific life expectancy provided by the WHO reflects our cohort more precisely. Thirdly, co-morbidity was not considered in our model analysis. A remaining challenge for disability and disutility weighting is adjusting for co-morbidities and co-disabilities[28]. Finally, since data sources including stage distribution and treatment approaches were not available, the simplifying assumptions of a constant duration of diagnosis and treatment, mammography screening, biopsy, metastatic breast cancer and terminal breast cancer was made. Although these limitations influence our ability to precisely estimate absolute values of DALYs averted and QALYs gained, both measures are expected to be influenced similarly. Thus, the impact of these limitations on the comparison between the outcomes is expected to be modest. However, these limitations can significantly affect a comparison between results of our study and results published in other CE analyses.

Comparing DALY and QALY outcomes in cost-effectiveness analyses is challenging. In theory, DALYs averted and QALYs gained could be equivalent if disutility of specific health state is equal to disability weight of the same health state. DALYs averted and QALYs gained are both used as outcome in cost-effectiveness analyses, providing a basis for extensive comparisons of the health effects of various policies and interventions[29]. Yet, they are not interchangeable, and the differences need to be considered[30–32].

The utilities are derived from different studies, using different methods, techniques and groups of subjects. The disability weights are more based on expert valuations and reflect on specific disease, while disutility is based on population or individual opinion and reflect on health states[30,32]. Additionally, age of disease onset and discounting procedures are important factors that can lead to variations in DALY averted and QALY gained results when the same intervention is evaluated[32].

The major advantage of using QALYs in cost-effectiveness analyses is that - besides accounting for an improving life expectancy, it also takes the health effects on the quality of life into account. In general, cost-effectiveness thresholds have been based on QALYs gained. However, using QALYs is associated with many challenges due to the complexity and diversity of health related utilities used in the calculation; and lack of capturing years lost due to the disease, which is relevant when evaluating fatal diseases as cancer[29–32]. DALYs provide a much more general overview of the burden of disease, for its universal set of disability weights used around the globe. Hence, DALYs averted can be used interchangeably across Europe, not only to evaluate

the quality of life among breast cancer patients, but also for other cancer types, because disability weights have the same value for different types of cancer, and there are only two cancer-specific health states: stoma and mastectomy[15]. Although differences in DALYs averted and QALYs gained were observed, conclusions of this cost-effectiveness analysis for breast cancer screening in the Netherlands is not likely to change based on the measure used (DALY averted versus QALY gained) when measuring health gain, as long as the recommended threshold for cost-effectiveness in the Netherlands is applied[11,25].

Nonetheless, depending on the researcher's preferences, the aim of the study and the country-specific context using DALYs averted or QALYs gained as a part of a cost-effectiveness analysis, should be carefully identified[11]. In general, results should be expected to be very similar when using DALYs or QALYs, except when the disutilities and disabilities are considerably different in important health states. When there are substantial differences, the reasons behind this should be evaluated. Future research should evaluate a wider range of life histories, including co-morbidities, health states duration and the effects of treatment.

In conclusion, our study demonstrates that using DALYs averted instead of QALYs gained to assess the effects on quality of life from breast cancer screening in the Netherlands yields differences in results. Nevertheless, the relation between these two tools remained constant through all analyses. To align with literature and accommodate comparison across studies, we therefore recommend using DALYs for burden of disease studies and QALYs in cost-effectiveness analyses.

REFERENCES

1. Sankatsing VD V, Heijnsdijk EAM, van Luijt PA, et al. Cost-effectiveness of digital mammography screening before the age of 50 in The Netherlands. Int J cancer. 2015;137(8):1990-1999. doi:10.1002/ijc.29572
2. Schunemann HJ, Lerda D, Quinn C, et al. Breast Cancer Screening and Diagnosis: A Synopsis of the European Breast Guidelines. Ann Intern Med. November 2019. doi:10.7326/M19-2125
3. Lauby-Secretan B, Scoccianti C, Loomis D, et al. Breast-cancer screening-viewpoint of the IARC Working Group. N Engl J Med. 2015;372(24):2353-2358. doi:10.1056/NEJMsr1504363
4. van Luijt PA, Heijnsdijk EAM, de Koning HJ. Cost-effectiveness of the Norwegian breast cancer screening program. Int J cancer. 2017;140(4):833-840. doi:10.1002/ijc.30513
5. van Luijt P. Overdiagnosis in the Dutch and Norwegian breast cancer screening program. 2016. http://hdl.handle.net/1765/94339. [Accessed August 28, 2020].
6. Gold MR, Stevenson D, Fryback DG. HALYs and QALYs and DALYs, Oh My: similarities and differences in summary measures of population Health. Annu Rev Public Health. 2002;23:115-134. doi:10.1146/annurev.publhealth.23.100901.140513
7. Leech AA, Kim DD, Cohen JT, Neumann PJ. Use and Misuse of Cost-Effectiveness Analysis Thresholds in Low- and Middle-Income Countries: Trends in Cost-per-DALY Studies. Value Heal. 2018;21(7):759-761. doi:10.1016/j.jval.2017.12.016
8. Neumann PJ, Thorat T, Zhong Y, et al. A Systematic Review of Cost-Effectiveness Studies Reporting Cost-per-DALY Averted. PLoS One. 2016;11(12):e0168512. doi:10.1371/journal.pone.0168512
9. Thorat T, Cangelosi M, Neumann P. Skills of the trade: The Tufts Cost-Effectiveness Analysis (CEA) Registry. J Benefit-Cost Anal. 2012;3:5. doi:10.1515/2152-2812.1114
10. Center for the Evaluation of Value and Risk in Health. The Cost-Effectiveness Analysis Registry. Institute for Clinical Research and Health Policy Studies, Tufts Medical Center. www.cearegistry.org. [Accessed August 28, 2020].
11. Feng X, Kim DD, Cohen JT, Neumann PJ, Ollendorf DA. Using QALYs versus DALYs to measure cost-effectiveness: How much does it matter? Int J Technol Assess Health Care. 2020;36(2):96–103. doi:10.1017/S0266462320000124
12. Fox-Rushby JA, Hanson K. Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis. Health Policy Plan. 2001;16(3):326-331. doi:10.1093/heapol/16.3.326
13. Salomon JA, Vos T, Hogan DR, et al. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. Lancet (London, England). 2012;380(9859):2129-2143. doi:10.1016/S0140-6736(12)61680-8
14. Salomon JA, Haagsma JA, Davis A, et al. Disability weights for the Global Burden of Disease 2013 study. Lancet Glob Heal. 2015;3(11):e712-23. doi:10.1016/S2214-109X(15)00069-8
15. Salomon JA, Vos T, Hogan DR, et al. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2129-2143. doi:10.1016/S0140-6736(12)61680-8.

16. 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):538-544. doi:10.1002/ijc.2910490411
17. de Koning HJ, van Ineveld BM, van Oortmarssen GJ, et al. Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. Int J cancer. 1991;49(4):531-537. doi:10.1002/ijc.2910490410
18. Lloyd A, Nafees B, Narewska J, Dewilde S, Watkins J. Health state utilities for metastatic breast cancer. Br J Cancer. 2006;95(6):683-690. doi:10.1038/sj.bjc.6603326
19. 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. doi:10.1016/0169-2607(85)90048-3
20. World Health Organization. Global Health Observatory (GHO): the data repository. World Health Organization. http://apps.who.int/gho/data/?theme=main&vid=61160. Published 2014. [Accessed August 19, 2020].
21. National Evaluation Team for Breast Cancer Screening (NETB). National Evaluation of Breast Cancer Screening in the Netherlands 1990-2011/2012. Rotterdam, The Netherlands; 2014.
22. IKCnet. Cijfersoverkanker.nl. Netherlands. www.cijfersoverkanker.nl. [Accessed August 10, 2020].
23. Saadatmand S, Tilanus-Linthorst MMA, Rutgers EJT, et al. Cost-effectiveness of screening women with familial risk for breast cancer with magnetic resonance imaging. J Natl Cancer Inst. 2013;105(17):1314-1321. doi:10.1093/jnci/djt203
24. Baltussen RMPM, Adam T, Tan-Torres Edejer T, et al. Making Choices in Health : WHO Guide to Cost-Effectiveness Analysis / Edited by T. Tan-Torres Edejer ... [et Al]. World Health Organization; 2003.
25. Zorginstituut Nederland. Cost-Effectiveness in Practice. https://english.zorginstituutnederland.nl/binaries/zinl-eng/documents/reports/2015/06/16/cost-effectiveness-in-practice/Cost-effectiveness+in+practice.pdf. [Accessed August 28, 2020].
26. Rashidian A, Barfar E, Hosseini H, Nosratnejad S, Barooti E. Cost effectiveness of breast cancer screening using mammography; a systematic review. Iran J Public Health. 2013;42(4):347-357.
27. Mittmann N, Stout NK, Tosteson ANA, Trentham-Dietz A, Alagoz O, Yaffe MJ. Cost-effectiveness of mammography from a publicly funded health care system perspective. C open. 2018;6(1):E77-E86. doi:10.9778/cmajo.20170106
28. Mittmann N, Stout NK, Lee P, et al. Total cost-effectiveness of mammography screening strategies. Heal reports. 2015;26(12):16-25.
29. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA. 2016;316(10):1093-1103. doi:10.1001/jama.2016.12195
30. Gold MR, Muennig P. Measure-dependent variation in burden of disease estimates: implications for policy. Med Care. 2002;40(3):260-266. doi:10.1097/00005650-200203000-00009
31. Schleinitz MD, DePalo D, Blume J, Stein M. Can differences in breast cancer utilities explain disparities in breast cancer care? J Gen Intern Med. 2006;21(12):1253-1260. doi:10.1111/j.1525-1497.2006.00609.x

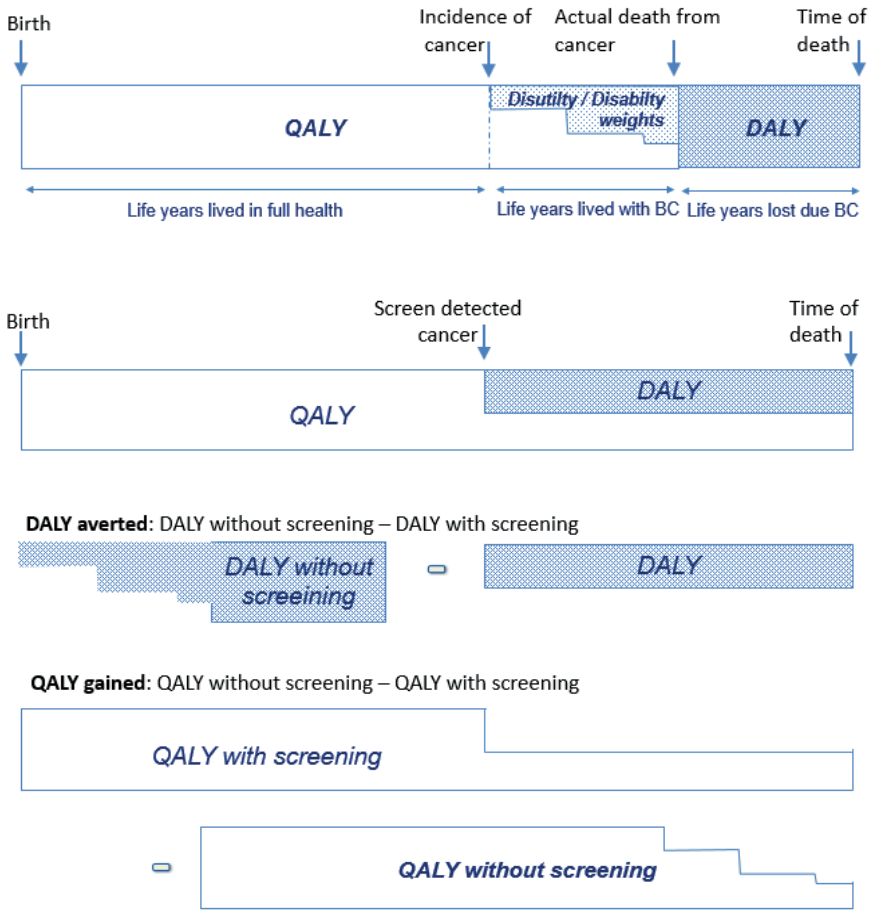
32. Sassi F. Calculating QALYs, comparing QALY and DALY calculations. Health Policy Plan. 2006;21(5):402-408. doi:10.1093/heapol/czl018

SUPPLEMENTAL MATERIAL

This supplement material gives a more details of theoretical background: on disability-adjusted life year (DALY) averted and quality-adjusted life year (QALY) gained (Appendix 1); The health-related quality of life (HRQoL) utilities and disability weights (DW) related to breast cancer (BC) health states (Appendix 2) used in our analysis; defined life histories which were followed in our DALY and QALY calculation (Appendix 3); the background of calculation on DALY averted and QALY gained (Appendix 4); ICER analysis without 3% discount rate for costs and effects (Appendix 5) and sensitivity analysis (Appendix 6).

Appendix 1: DALYs, QALYs, DALYs averted and QALYs gained

Appendix 1 Figure 1 summaries the DALYs and QALYs in one measure and depicts DALYs averted and QALYs gained. Lifespan is presented from birth to death combining DALYs and QALYs together. Until diagnosis of breast cancer, person can live in full health. From diagnosis to premature death due disease, quality of life is reduced, and disability rises. If breast cancer screening is conducted and breast cancer is detected, it affects the lifespan and quality of life. The difference in total DALYs without screening and DALYs with screening presents DALY averted (How many DALYs are averted due intervention). QALYs gained present difference of QALYs with screening and QALYs without it, indicating how many QALYs are gained due the intervention.



Appendix 1 Figure 1. Disability adjusted life years (DALYs) averted and quality adjusted life years (QALYs) gained

Appendix 2: HRQoL utilities and disability weights

Appendix 2 Table 1 shows disability weights extracted from the Global Burden Disease (GBD) Study 2015[1] for breast cancer and health states related to it.

Appendix 2 Table 1. Disability weights

Global Burden of Diseases Report (2015)[1]			
Cancer specific health states used in our research		**Health state description**	**Disability weight (95% CI)**
Diagnosis and primary therapy phase of breast cancer	Cancer, diagnosis and primary therapy	has pain, nausea, fatigue, weight loss and high anxiety	0.288 (0.193-0.399)
Controlled phase of breast cancer	Generic uncomplicated disease: worry and daily medication	has a chronic disease that requires medication every day and causes some worry but minimal interference with daily activities	0.049 (0.031-0.072)
Metastatic phase of breast cancer	Cancer, metastatic	has severe pain, extreme fatigue, weight loss and high anxiety	0.451 (0.307-0.6)
Terminal phase of breast cancer	Terminal phase, with medication (for cancers, end-stage kidney/liver disease)	has lost a lot of weight and regularly uses strong medication to avoid constant pain. The person has no appetite, feels nauseous, and needs to spend most of the day in bed	0.54 (0.377-0.687)
Mastectomy due to breast cancer	Mastectomy	had one of her breasts removed and sometimes has pain or swelling in the arms	0.036 (0.02-0.057)

In Table 2 (Appendix 2), the pre-selection list of HRQoL utilities that were extracted by specific criteria from the selected papers, are listed. For the final analysis, we selected HRQoL utilities based on similarity of health state descriptions to the disability weights.

Appendix 2 Table 2. Health-related quality of life (HRQoL) utilities

Health State descriptions	Lower interval	Utility [a]	Higher interval	Method of evaluation	Duration	Author(s)
Mammographic screening		0.99		VAS	1 week [b]	De Koning et al. (1991)[2]
Breast biopsy		0.89		VAS	5 weeks [b]	De Koning et al. (1991)[2]
State I disease [c]	0.50	0.91	1.00	SG (and TTO)	10 years survival	Schleinitz et al. (2006)[3]
State II disease [c]	0.26	0.75	0.99	SG (and TTO)	10 years survival	Schleinitz et al. (2006)[3]
State III disease [c]	0.25	0.51	0.94	SG (and TTO)	10 years survival	Schleinitz et al. (2006)[3]
State IV disease (estrogen receptor positive) [c]	0.00	0.36	0.75	SG (and TTO)	10 years survival	Schleinitz et al. (2006)[3]
State IV disease (estrogen receptor negative) [c]	0.00	0.40	0.79	SG (and TTO)	10 years survival	Schleinitz et al. (2006)[3]
Stable metastatic breast cancer, depending on toxicity from treatment		0.50-0.80		SG (and SG)		Earle et al. (2000)[4]
Base state – stable metastatic disease with no toxicity		0.715		VAS (and SG)	10 years survival	Lloyd et al. (2006)[5]
Terminal illness		0.288 [d]		VAS	1 month	De Haes et al. (1991)[6]
3 months – 1 year after mastectomy		0.844 [d]		VAS	10 months	De Haes et al. (1991)[6]
Terminal metastatic breast cancer		0.25		SG		Earle et al. (2000)[4]
Breast cancer, mastectomy		0.99		SG	Ten-years period	Carter et al. (1998)[7]
Disease-free after mastectomy, depending how long afterwards		0.98-1.00		SG/TTO		Earle et al. (2000)[4]

Abbreviations: VAS – Visual Analogue Scale; SG – Standard Gambling; TTO – Time Trade-Off

[a] Utilities are presented as mean / median (SD or adjustment value)

[b] Used assumptions on the durations

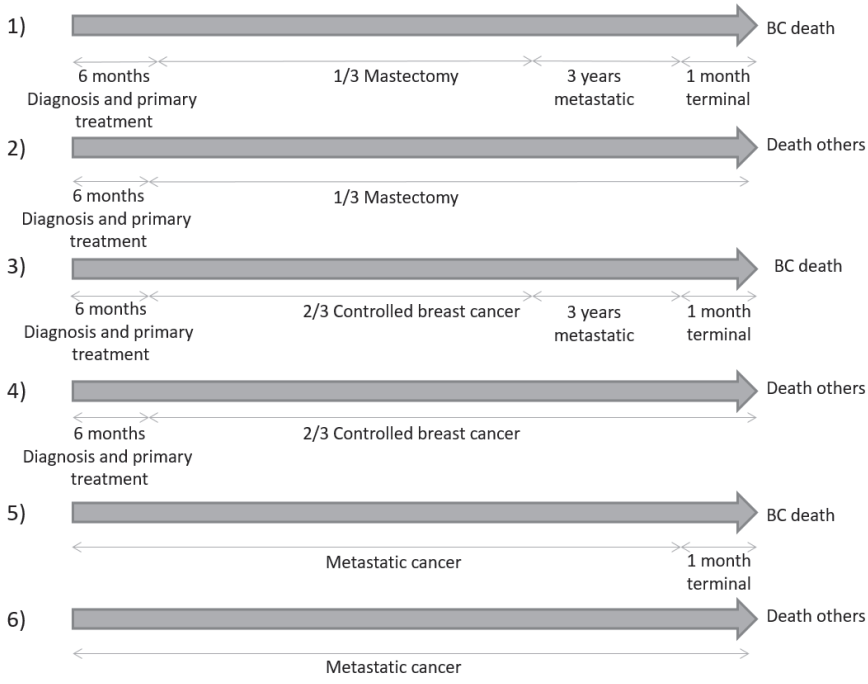
[c] Stages are classified by TNM classification system

[d] Transformed into a utility score using a power function of TTO=1-(1-VAS) ^ 1.82

Appendix 3: Life histories

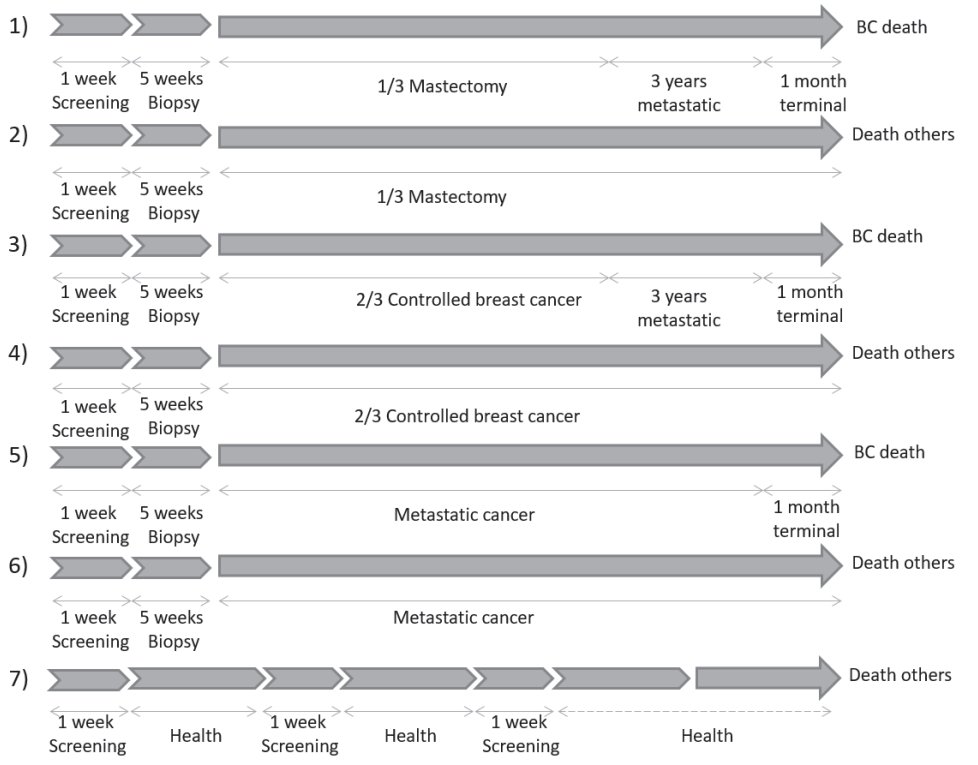
To obtain DALYs and QALYs, we defined possible life histories (LH), different in health phases which person go through, and causes of death (breast cancer or other cause) in order to investigate the impact of screening on a woman's quality of life. For both analyses, life histories started from screening to death. Each scenario can have diverse time duration, depending from time-duration of health state, and age of death. Following these different life histories was important for attaching different disability weights and utilities to different life periods and health states. In both DALY and QALY analysis, we assume that from all persons diagnosed with non-metastatic BC, 1/3 of them had mastectomy[8], and 2/3 lived with controlled BC (Stage 0 (*ductal carcinoma in citu*) – Stage 3, based on TNM staging system). Additionally, the time spent with mastectomy, controlled BC, and metastatic BC was calculated as the difference between life years lived from diagnosis to death and the sum of time spent in others health states.

For the DALYs analysis, specific health states time-durations were assumed which remained the same for all six scenarios (6 months for diagnosis and primary treatment, 3 years for metastatic and 1month for terminal phase of BC). Women diagnosed with BC spent the first six months in the diagnostic and treatment phase followed by either mastectomy (LH-1), (LH-2), or the controlled BC phase (LH-3, LH-4). Women who lived with controlled BC or with mastectomy could die from BC (LH-1, LH-3), or from other causes (LH-2, LH-4, and LH-6). Women who were diagnosed with metastatic BC did not go through the diagnostic and treatment phase for six months, but rather lived in the metastatic phase until they reach the terminal phase and died of BC (LH-5) or of other causes (LH-6). If persons died from BC (controlled or metastatic BC), it was assumed that 1 month prior death, disease progressed in terminal phase, and more 3 years before terminal phase disease metastasized, except for persons who were diagnosed with metastatic BC earlier. Figure 1 (Appendix 3) summarizes above described LH (1–6).



Appendix 3 Figure 1. Life histories (LH: 1–6) for disability-adjusted life years (DALYs) analysis

For the QALYs analysis, specific health states time-duration were extracted from literature which remained the same for all scenarios (1 week of screening, 5 weeks of biopsy, 3 years for metastatic and 1 month for terminal phase of BC). Every scenario started with mammography screening (Appendix 3 Figure 2). Women with a positive screen received a biopsy. We assumed one third of biopsies are leading to a cancer diagnosis[8]. Persons who had mastectomy, controlled BC, or metastatic BC could die from BC or from other causes. If person died from BC (LH-1, LH-3, LH-5), it was assumed that month prior death disease progressed in terminal phase, and before terminal phase disease metastasized for 3 years, except for persons who were diagnosed with metastatic BC immediately after biopsy where time-duration were calculated. Finally, screened women could be BC free and die from other causes and followed the screening schedule during her life (LH-7).



Appendix 3 Figure 2. Life histories (LH: 1–7) for quality-adjusted life years (QALYs) analysis

Appendix 4: Calculation method

MISCAN input

The calculation process is presented on the example of biennial screening strategy from 50 to 74 years that is current screening strategy in the Netherlands. DALYs and QALYs analyses were performed separately. The data from the MISCAN model for screened women is presented in Appendix 4 Table 1. Incidence of BC, number of BC deaths, and life years from diagnosis to death for women who died from BC or other causes were estimated by the model for both controlled BC (Stage 1 – model name for non-metastatic breast cancer) and metastatic BC (Stage 2 – model name for metastatic breast cancer). We assumed that one third of total BC (Stage 1) diagnosis had mastectomy, and two thirds lived with BC until death.

DALYs calculation

Following previously described life histories and assumptions on time duration of specific health states and on the clinical stages of disease, we calculate total years of life lived with disability (YLDs) for each life history separately.

In the life history of women who were diagnosed with BC, there were two possible scenarios: to die from BC or other causes. Women who were diagnosed with BC and died from it (Table 2 - Life history 3) (Appendix 4), during their life they spent 6 months in a phase of diagnosis and primary therapy, 3 years in metastatic state of BC and 1 month in terminal phase of BC. Women who were diagnosed with BC but died from other causes, spent 6 months in phase of diagnosis and primary therapy and the rest of time in state of controlled BC (Table 2 – Life history 4) (Appendix 4). These times are multiplied with corresponding disability weights and number of women estimated by the model. To calculate time lived with controlled BC, from total life years lived from diagnosis to death (from BC or other causes) we subtracted life years spent in other health states (the number of women multiplied with time coefficient (3 for 3 years' time-period; 0.5 for six months and 0.08 for 1 month) for each experienced health state). Obtained life years lived with controlled BC is multiplied with corresponding disability weights.

Appendix 4 Table 1. MISCAN output for biennial screening strategy from 50 to 74 years, per 1,000 women

BIENNIAL SCREENING STRATEGY FROM 50 TO 74 YEARS[A]								
		Incidence	**BC deaths**	**Other deaths**	**LY diag_BCdeath**	**LY diag_death**	**Visits**	**Total LY**
STAGE 1	Total	63	10	53	125	907	4,569	24,053
	Mastectomy	21	3	18	42	302		
	Controlled BC	42	7	35	83	605		
STAGE 2	Metastatic BC	14	8	6	80	95		

aResults are rounded to whole numbers
Abbreviations: BC – Breast Cancer; LY – Life years
BC deaths – number of persons died from breast cancer
LY diag_BCdeath - life years lived from diagnosis to death due breast cancer
LY diag_death - life years lived from diagnosis to death due other causes
Total LY - total life years lived in the cohort
Stage 1 – model name for non-metastatic breast cancer
Stage 2 – model name for metastatic breast cancer

Appendix 4 Table 2. Life History 3 and 4 – Had Controlled Breast Cancer and Died from Breast Cancer or Other Causes

Life History 3 - Had Controlled BC And Died From BC					
	DW	Number	Time	Time of Life	Total
Dg and Th	0.29	6.96	0.50	3.48	1.00
Metastatic BC	0.45	6.96	3.00	20.88	9.42
Terminal BC	0.54	6.96	0.08	0.58	0.31
	DW	Total LYL	LYL This State	LYL Other	
Controlled BC	0.05	83.24	58.30	24.94	2.86
			Total Life Years Lost in this Life History:		**13.59**
Life History 4 - Had Controlled BC And Died from Other Causes					
	DW	Number	Time	Time of Life	Total
Dg and Th	0.29	35.26	0.50	17.63	5.08
	DW	Total LYL	LYL This State	LYL Other	
Controlled BC	0.05	604.83	587.20	17.63	28.77
			Total Life Years Lost in This Life History:		**33.85**

Abbreviations: BC – Breast Cancer; Dg – Diagnosis; Th – Therapy; DW – Disability weight for specific health state; LYL – Life years lived
Total LYL – total life years lived from diagnosis to death
LYL this state – total life years lived with this state (total LYL minus LYL lived in other health states)
LYL other – life years lived in other health states

Women who were diagnoses with BC (Stage 1- model name for non-metastatic breast cancer) and had mastectomy during life, could die from BC or other causes. Time lived with mastectomy is calculated as a 1/3 of difference of total life years from diagnosis to death due BC and life years lived in other health states. That time is multiplied with the number of women estimated by model and DW for mastectomy. The number of women who had mastectomy and died from BC (Table 3 – Life history 1) (Appendix 4) is multiplied with 0.5 years' time-period and DWs for diagnosis and therapy, 3 years' time-period metastatic BC, 0,08 years' time-period terminal BC, respectably. The number of women who had mastectomy and died from other causes (Table 3 – Life history 2) (Appendix 4) is multiplied with 0.5 years' time-period and DW for diagnosis and therapy phase.

Appendix 4 Table 3. Life History 3 and 4 – Had Mastectomy and Died from Breast Cancer or Other Causes

Life History 1 - Had Mastectomy and Died from BC					
	DW	Number	Time	Time of Life	Total
Dg and Th	0.29	3.48	0.50	1.74	0.50
Metastatic BC	0.45	3.48	3.00	10.44	4.71
Terminal BC	0.54	3.48	0.08	0.29	0.16
	DW	Total LYL	LYL This State	LYL Other	
Mastectomy	0.04	41.62	29.15	12.47	1.05
			Total Life Years Lost in This Life History:		**6.41**
		Life History 2 - Had Mastectomy and Died from Other Causes			
	DW	Number	Time	Time of Life	Total
Dg and Th	0.29	17.63	0.50	8.82	2.54
	DW	Total LYL	LYL This State	LYL Other	
MastectomY	0.04	302.41	293.60	8.82	10.57
			Total Life Years Lost in This Life History:		**13.11**

Abbreviations: BC – Breast Cancer; Dg – Diagnosis; Th – Therapy; DW – Disability weight for specific health state; LYL – Life years lived
Total LYL – total life years lived from diagnosis to death
LYL this state – total life years lived with this state (total LYL minus LYL lived in other health states)
LYL other – life years lived in other health states

For women who were diagnosed with metastatic BC and died from BC (Table 4 – Life history 5) (Appendix 4), the number of metastatic BC is multiplied with 0.08 years (assumption on time duration women spent with terminal phase of BC) and DW for terminal phase of BC. The difference between life years spent from diagnosis to death from metastatic BC and the life years spent in terminal phase of BC is multiplied with DW for metastatic BC. The total life years from diagnosis to death of women who lived with metastatic BC and died from other causes (Table 4 – Life history 6) (Appendix 4) is multiplied with DW for metastatic BC.

Appendix 4 Table 4. Life History 5 and 6 – Had Metastatic Breast Cancer and Died from Breast Cancer or Other Causes

Life History 5 - Had Metastatic BC and Died from BC					
	DW	Number	Time	Time of Life	Total
Terminal BC	0.54	8.14	0.08	0.68	0.37
	DW	Total LYL	LYL This State	LYL Other	
Metastatic BC	0.45	80.13	79.45	0.68	35.83
			Total Life Years Lost in This Life History:		**36.20**
Life History 6 - Had Metastatic Cancer and Died from Other Causes					
	DW	Number	Time		Total
Metastatic BC	0.45	95.36	95.36		43.01
			Total Life Years Lost in This Life History:		**43.01**

Abbreviations: BC – Breast Cancer; DW – Disability weight for specific health state; LYL – Life years lived
Total LYL – total life years lived from diagnosis to death
LYL this state – total life years lived with this state (total LYL minus LYL lived in other health states)
LYL other – life years lived in other health states

Years of life lost due premature death (YLLs) is calculated as mortality rate per age groups (5 years interval, from 0 to 100 years) multiplied with life expectancy at that age interval (Appendix 4 Table 5). Mortality rate for five-years interval is calculated by MISCAN model. The life expectancy at each age interval is based on Dutch life expectancy[9]. The total YLLs in this screening strategy is calculated as a sum of all YLLs per each age group. Starting from age group 40-45 years onward, we used 3% discounted rate.

Appendix 4 Table 5. Years of life lost (YLLs) for biennial screening strategy from 50 to 74

Years of life lost - YLL			
Age of death	BC deaths	LE	**Total**
0-4	0.00	82.90	**0.00**
5-9	0.00	78.90	**0.40**
10-14	0.00	73.90	**1.16**
15-19	0.01	69.00	**1.89**
20-24	0.01	64.00	**2.08**
25-29	0.01	59.10	**2.41**
30-34	0.02	54.10	**4.13**
35-39	0.11	49.20	**17.78**
40-44*	0.22	44.30	**30.32**
45-49*	0.33	39.50	**35.25**
50-54*	0.45	34.70	**36.07**
55-59*	0.51	30.10	**30.98**
60-64*	0.54	25.70	**23.78**
65-69*	0.52	21.40	**16.33**
70-74*	0.48	17.20	**10.54**
75-79*	0.44	13.30	**6.40**
80-84*	0.38	7.90	**2.83**
85-89*	0.27	6.80	**1.49**
90-94*	0.14	4.60	**0.43**
95-99*	0.04	3.00	**0.08**
		Total:	**106.98**

Abbreviations: BC – Breast Cancer; LE – Life Expectancy
*Results are discounted using 3% discount rate

Finally, the sum of all YLDs from each life history and total YLLs represents total DALYs of this screening strategy (Appendix 4 Table 6). The total DALYs for this screening strategy is compared with reference screening strategy (no screening strategy). The difference represents how many years this intervention averts (DALY averted).

Appendix 4 Table 6. Final results of DALYs and DALYs averted for biennial screening strategy from 50 to 74

Total YLDs of this strategy:	**146.17**
Total YLLs of this strategy:	224.36
Total DALYs of this strategy:	370.52
DALY averted by this strategy:	**63.44**

Abbreviations: YLD – Years Lived with Disability; YLL – Years of Life Lost Due Premature Death; DALY – Disability Adjusted Life Years

QALYs calculation

For the QALY calculations, we used similar method as for the DALY analysis. We defined one additional life history for QALY analysis: Life history 7 – Followed Screening Program and Died from Other Causes (Appendix 4 Table 7), for women who did not diagnosed with BC during life and died from other causes. As well, we followed the assumption that each diagnosis of BC needs 3 biopsies in average[8] and quality of life lost due additional biopsies are calculated in Table 6. In our calculation we used disutility (1 – health related quality of life utility (HRQoL)) to estimate how many years of life is lost due lower quality of life. Disutility was described for screening, biopsy, mastectomy, controlled, metastatic and terminal BC. Specific health states time-duration (1 week of screening, 5 weeks of biopsy, 3 years for metastatic and 1 month for terminal BC) was multiplied with specific disutility and number of women who experienced that state. Time lived with mastectomy, controlled or metastatic BC is calculated as a difference of total life years from diagnosis to death (due BC or other causes) and life years lived in other health states. Finally, the sum of all years lost, due lower quality of life as a consequences of BC and related health states, is subtracted from total life lived in this cohort to estimate total QALYs for this screening strategy. This result is compared with no-screening, and the difference represents how much quality adjusted life years are gain due this screening program.

Appendix 4 Table 7. Followed Screening Program and Died from Other Causes

Life History 7 – Followed Screening Program and Died from Other Causes					
	Disutility	Number	Time	Time of Life	Total
Screening	0.01	4491.12	0.02	86.37	0.86
Biopsy Other	0.11	155.39	0.10	14.94	1.64
		Total Life Years Lost in This Life History:			**2.51**

Appendix 5: Cost-effectiveness analysis without 3% annual discount rate for costs and effects

The net costs for each strategy were defined as the difference between the total costs of the strategy and the costs of the reference strategy. In Appendix 5 Table 1 are presented the results of cost-effectiveness analysis without using 3% annual discount rate for costs and effects of each strategy in the same order as it is presented in CE analysis with discounted rates. The minimum DALY averted and QALY gained is by screening triennially from 50 – 69 (T50-69) with 190 averted and 100 gained years, respectively compared to no screening. The most beneficial regarding to DALY averted and QALY gained is screening annually from 45 to 74 (A45-74), with 402 DALY averted and 215 QALY gained.

Appendix 5 Table 1. The net costs in Euros, DALYs averted and QALYs gained per 1,000 women per strategy; without 3% annual discount rate for costs and effects

Screening strategy[a]	Net costs per strategy[b]	DALYs averted	QALYs gained
No screening	*Ref. (Total cost 2,348,122)*	*Ref.*	*Ref.*
T50-69	36,829	189.9	99.6
T50-74	44,1963	213.3	108.6
T45-69	98,156	229.8	123.9
B50-69	105,504	253.1	132.8
T45-74	130,120	237.3	126.0
B50-72	146,892	256.9	133.7
B50-74[c]	168,734	270.7	138.3
B47-69	176,806	265.0	142.8
B47-72	193,262	284.5	150.5
B47-74	207,869	294.1	153.7
B45-69	212,121	280.1	151.9
B45-72	228,549	299.7	159.8
B45-74	242,540	309.6	163.3
A47-50_B50-69	242,677	276.8	150.6
A47-50_B50-72	265,018	302.8	160.9
A47-50_B50-74	279,776	312.0	164.0
A45-50_B50-69	298,886	294.5	161.5
A45-50_B50-72	321,440	320.5	171.8
A45-50_B50-74	336,399	329.6	174.7
A50-69	466,600	311.5	166.9
A50-74	515,858	341.9	178.9
A45-69	595,010	371.5	203.2
A45-74	593,139	401.8	215.3

Abbreviations: DALY – Disability Adjusted Life Years, QALY – Quality Adjusted Life Years, Ref. - Reference
[a] Screening strategies consist of biennial (B), annual (A) screening or combination (A + B), and triennial (T)
[b] Result is presented in Euros
[c] Current screening strategy in the Netherlands

Cost-effectiveness (CE) analysis was performed by dividing the net costs for each strategy by its DALY averted or QALY gained. Annual discount rate for costs and effects were not used (for results with discounted rates, please see the manuscript). As cost-effective strategies we assumed the same strategies as presented in discounted analysis (see the manuscript) and calculated incremental cost-effectiveness ratios (ICERs) (Appendix 5 Table 2).

Appendix 5 Table 2. Additional costs, Additional DALYs averted, Additional QALYs gained and Incremental Cost-Effectiveness Ratio for cost-effective screening strategies; without 3% annual discount rate for costs and effects

Cost-effective strategies – undiscounted ICER							
DALY analysis				QALY analysis			
Screening strategy	Additional costs[a]	Additional DALY averted	**ICER[b]**	Screening strategy	Additional costs[a]	Additional QALY gained	**ICER[b]**
No screening	Ref.	Ref.	**Ref.**	No screening	Ref.	Ref.	**Ref.**
T50-69	36,829	190	**194**	T50-69	36,829	100	**370**
T50-74	7,368	23	**315**	T50-74	7,368	9	**823**
T45-74	61,308	40	**1,539**	T45-74	61,308	24	**2,526**
B45-74	137,036	56	**2,426**	B45-74	137,036	30	**4,503**
A45-74	350,599	92	**3,800**	A45-74	350,599	52	**6,742**

Abbreviations: DALY – disability adjusted life years; ICER – incremental cost-effectiveness ratio; QALY – quality adjusted life years, Ref. - Reference

a The results are presented in Euros, and reflect the additional costs relative to the less effective strategy in the ranking

b The results are presented in Euros per one DALY averted or QALY gained, and are calculated as the difference in costs divided by the difference in DALY averted or QALY gained between a strategy and the previous, less effective strategy in the ranking

Appendix 6: Sensitivity Analysis

We conducted a sensitivity analysis to measure and evaluate parameter uncertainty of the HRQoL utility values used in QALY analysis. We performed a sensitivity analysis by replacing health disutility (1 – HRQoL) with the same value used as disability weight for the health conditions that overlap: controlled breast cancer; metastatic breast cancer; terminal disease and mastectomy. Previous values of 0.170, 0.285, 0.712 and 0.116 were replaced with 0.049, 0.451, 0.540 and 0.036 respectively. Utilities for screening and biopsy, for which no corresponding disability weights exist in our model, we replaced with 0. Utilities were replaced one at a time, both in the reference strategy (no screening) and an exemplary strategy (B50-69), while other utilities remained the same (respective 1-HRQoL disutility). The differences in QALY gained comparing with results in the base case analysis were calculated and presented as percentage change (Appendix 6 Table 1). All other methodological procedures and assumptions remained the same as in the base case analysis.

An overview of the results of the sensitivity analysis is provided in Appendix 6 Table 1. Our model-predicted base-case cost-effectiveness results were sensitive to varying disutilities for controlled and metastatic breast cancer, whereas changing the disutilities for screening, biopsy and mastectomy hardly had any effect.

Appendix 6 Table 1. Results (QALY and QALYs gained) of the sensitivity analysis

Utilities			**No screening**	**Strategy B50-69**		**Percentage change**
Health state	1-HRQoL	DW	Total QALY	Total QALY	QALY gained	
Base case	/	/	23,800.51	23,837.10	36.60	Ref.
Screening	0.010	0	23,800.51	23,837.86	37.35	+2%
Biopsy	0.110	0	23,802.82	23,839.53	36.71	0%
Controlled BC	0.170	0.049	23,859.64	23,914.25	54.62	+49%
Metastatic BC	0.285	0.451	23,750.59	23,801.29	50.70	+39%
Terminal disease	0.712	0.540	23,800.86	23,837.38	36.52	0%
Mastectomy	0.116	0.036	23,820.05	23,862.61	42.55	+16%

Abbreviations: HRQoL – Health-Related Quality of Life utility, DW – disability weight, QALY – Quality Adjusted Life Years, Ref. - Reference

Additionally, we tested the robustness of our analysis by replacing the health disutilities (1 – HRQoL) for mastectomy; controlled breast cancer; metastatic breast cancer; and terminal disease with corresponding disability weights in all screening strategies, including the no screening strategy, at the same time. Thus, we re-calculated the ICERs and identified non-dominated strategies. In Appendix 6 Table 2 the results of each screening strategy, ordered by the net costs per strategy (same order as in the manuscript) are presented. The QALY gained ranged from 59 – 129, compared with 29 – 65 previously. The ICERs per QALY gained varied from €947 to €3,585 (Table 3). This is lower than in the previous analysis presented in the manuscript, where the range of ICERs per QALY gained was €1,960 to €15,899.

In Appendix 6 Figure 1 we present the efficiency frontier of the screening strategies. In our sensitivity analysis, one strategy did not appear as non-dominant (B45-72) compared with previous results, giving the results of the same strategies at the efficient frontier for both DALY averted and QALY gained used.

Appendix 6 Table 2. The discounted net costs in Euros, LYG, DALYs averted, QALYs gained with 1-HRQoL utilities and QALYs gained (with adjusted utilities) per 1,000 women per strategy

Screening strategy[a]	Net costs per strategy[b]	LYG	DALYs averted	QALYs gained with 1-HRQoL utilities	QALY gained with adjusted utilities
No screening	*Ref. (Total cost: 1,201,486)*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
T50-69	56,217	36.70	46.00	28.69	59.38
T50-74	64,326	40.10	49.68	30.48	65.18
T45-69	111,203	46.90	61.79	37.69	75.24
B50-69	118,112	46.90	58.30	36.60	74.61
T45-74	119,276	50.30	65.45	39.50	81.03
B50-72	129,089	49.80	61.40	38.19	79.47
B50-74[c]	140,301	51.70	63.44	39.00	82.75
B47-69	158,257	53.80	68.77	42.68	85.31
B47-72	169,153	56.70	71.84	44.24	90.15
B47-74	176,803	58.10	73.28	44.91	92.52
B45-69	190,112	57.80	75.32	46.20	91.53
B45-72	200,995	60.80	78.40	47.87	96.47
B45-74	208,447	62.20	79.88	48.52	98.86
A47-50_B50-69	212,050	57.30	74.23	45.79	90.45
A47-50_B50-72	226,701	61.20	78.34	47.91	96.94
A47-50_B50-74	234,261	62.40	79.69	48.39	99.06
A45-50_B50-69	262,407	62.00	81.53	49.84	97.48
A45-50_B50-72	277,127	65.90	85.63	51.98	103.96
A45-50_B50-74	284,729	67.10	86.97	52.46	106.08
A50-69	323,346	63.00	76.92	48.59	97.74
A50-74	349,540	67.50	81.58	50.92	105.13
A45-69	456,493	78.40	100.94	62.02	121.48
A45-74	462,471	83.00	105.58	64.50	128.99

Abbreviations: LYG – Life Years Gain, DALY – Disability Adjusted Life Years, QALY – Quality Adjusted Life Years, Ref. - Reference

[a] Screening strategies consist of annual (A), biennial (B) and triennial (T) screening or combination of annual and biennial (A + B)

[b] Result is presented in Euros

[c] Current screening strategy in the Netherlands

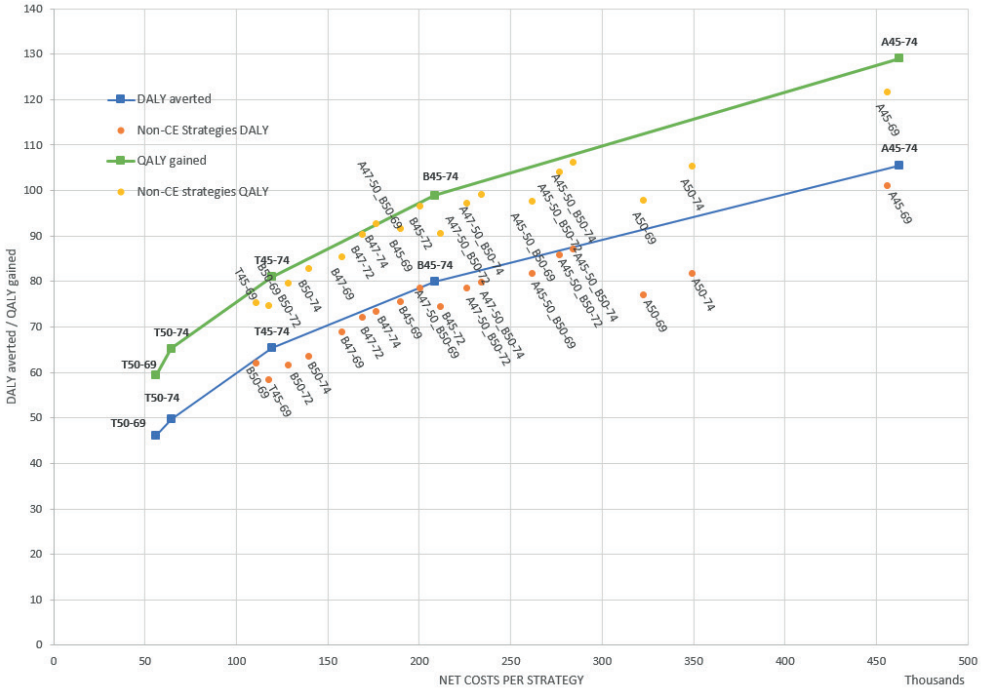
Appendix 6 Table 3. Additional discounted costs; DALYs averted; and QALYs gained (with adjusted utilities) and Incremental Cost-Effectiveness Ratio for cost-effective screening strategies

Cost-effective strategies – ICER							
DALY analysis				QALY analysis with adjusted utilities			
Screening strategy	Additional costs[a]	Additional DALY averted	**ICER[b]**	Screening strategy	Additional costs[a]	Additional QALY gained	**ICER[b]**
No screening	Ref.	Ref.	**Ref.**	No screening	Ref.	Ref.	**Ref.**
T50-69	56,217	46.00	**1,222**	**T50-69**	56,217	59.38	**947**
T50-74	8,109	3.67	**2,207**	**T50-74**	8,109	5.80	**987**
T45-74	54,950	15.77	**3,484**	**T45-74**	54,950	15.85	**1,472**
B45-74	89,171	14.43	**6,180**	**B45-74**	7,452	17.83	**2,108**
A45-74	254,024	25.70	**9,883**	**A45-74**	254,024	30.12	**3,585**

Abbreviations: DALY – disability adjusted life years; ICER – incremental cost-effectiveness ratio; QALY – quality adjusted life years, Ref. - Reference

[a] The results are presented in Euros, and reflect the additional costs relative to the less effective strategy in the ranking

[b] The results are presented in Euros per one DALY averted or QALY gained (with adjusted utilities), and are calculated as the difference in costs divided by the difference in DALY averted or QALY gained (with adjusted utilities) between a strategy and the previous, less effective strategy in the ranking



Appendix 6 Figure 2. Efficiency frontier of screening strategies.

Each screening strategy is represented by a point. Both DALYs averted or QALYs gained (with adjusted utilities) and costs for each strategy are relative to a situation without screening and are presented per 1,000 women. Strategies consist of annual (A), biennial (B), triennial (T) screening and combination of annual and biennial (A+B). Strategies on the green (QALY gained with adjusted utilities) and blue line (DALY averted) represent cost-effective strategies. Dominated strategies are represented by dots only.

REFERENCES

1. Global Burden of Disease Study 2015. Global Burden of Disease Study 2015 (GBD 2015) Disability Weights. Seattle, United States:Institute for Health Metrics and Evaluation (IHME); 2016. Available at: http://ghdx.healthdata.org/record/ihme-data/gbd-2015-disability-weights [Accessed July 23, 2020].
2. de Koning HJ, Martin van Ineveld B, van Oortmarssen GJ, et al. Breast cancer screening and cost effectiveness; Policy alternatives, quality of life considerations and the possible impact of uncertain factors. Int J Cancer. 1991;49(4):531–7. https://doi.org/10.1002/ijc.2910490410
3. Schleinitz MD, DePalo D, Blume J, Stein M. Can differences in breast cancer utilities explain disparities in breast cancer care? J Gen Intern Med. 2006;21(12):1253–60. https://doi.org/10.1111/j.1525-1497.2006.00609.x
4. Earle CC, Chapman RH, Baker CS, et al. Systematic overview of cost-utility assessments in oncology. J Clin Oncol. 2000;18(18):3302–17. https://doi.org/10.1200/JCO.2000.18.18.3302
5. Lloyd A, Nafees B, Narewska J, Dewilde S, Watkins J. Health state utilities for metastatic breast cancer. Br J Cancer. 2006;95(6):683–90. https://doi.org/10.1038/sj.bjc.6603326
6. de Haes JC, de Koning HJ, van Oortmarssen GJ, et al. The impact of a breast cancer screening programme on quality-adjusted life-years. Int J Cancer. 1991;49(4):538–44. https://doi.org/10.1002/ijc.2910490411
7. Carter KJ, Ritchey NP, Castro F, et al. Treatment of early-stage breast cancer in the elderly: a health-outcome-based approach. Med Decis Mak. 1998;18(2):213–9. https://doi.org/10.1177/0272989X9801800210
8. National Evaluation Team for Breast Cancer Screening (NTB). National evaluation of breast cancer screening in the Netherlands 1990-2011/2012. Thirteenth evaluation report. Rotterdam, The Netherlands:Erasmus MC University Medical Center Rotterdam, Radboud University Nijmegen Medical Centre; 2014. Available at: https://www.lrcb.nl/resources/uploads/2017/02/LETB-XIII-English.pdf [Accessed July 23, 2020].
9. WHO. Global Health Observatory data repository. Netherlands. Available at: http://apps.who.int/gho/data/?theme=main&vid=61160 [Accessed July 23, 2020].

Chapter 8

General Discussion

This thesis investigated the consequences of variations in breast cancer screening practices and potential ways to further optimize screening programs across Europe.

In order to inform screening recommendations, the effectiveness of breast cancer screening in Europe was reviewed and an established microsimulation model (MISCAN-Breast) was used to estimate the impact of current screening policies and the potential benefits of specific health policy changes. Furthermore, MISCAN-Breast was standardized into a user-friendly online application as part of the EU-TOPIA project.

In this chapter, I will first address the main findings of the research questions posed in the introduction of this thesis. Then, I will discuss directions for future research and highlight our general conclusions and recommendations based on the research described within.

MAIN FINDINGS

Part 1: The effectiveness of breast cancer screening

While the purpose of primary prevention is to prevent a disease from ever occurring, secondary prevention, which includes cancer screening, emphasizes early disease detection, and targets an asymptomatic population with an average risk of developing the disease. In Part 1 of this thesis, we reviewed the effectiveness of mammography screening in reducing breast cancer mortality and its impact on other benefits and harms.

Breast cancer mortality reduction due to mammography screening

In **Chapter 2**, we performed a systematic review to summarize the current evidence on reducing breast cancer mortality with mammography screening in Europe. Our results strengthen previous findings that mammography screening reduces mortality from breast cancer, but highlight that this occurs at varying magnitudes. This reflects differences in evaluation designs, target ages, ages of follow-up of breast cancer incidence or mortality, duration of follow-up since first invitation, control groups, and assessment methods of self-selection bias. There is an ongoing debate as to which study design is the gold standard for estimating the true effect of screening on cancer specific mortality[1-3]. Studies that included data from randomized controlled trials (RCT) as well as observational studies such as prospective and retrospective controlled cohort or case-control studies have been considered and underwent a methodologically sound quality appraisal in **Chapter 2**. In times of growing availability of high-quality data from screening programs, we think that estimates from contemporary observational studies can be considered more relevant today than those of the RCTs. The included RCTs were conducted more than 20 years ago when adherence to screening was lower and the quality of screening programmes and breast cancer care were less advanced

than today. Our investigation in **Chapter 2** highlighted that there are several observational studies employing methodologically appropriate approaches to provide strong evidence on screening effectiveness. For example, the included cohort studies with long follow-up periods (min. 5 years) and appropriate designs (regarding the representativeness of exposed women, and the ascertainment of either the exposure or outcome of interest) and studies that controlled for individual differences (such as age) related to the primary outcome. In addition, **Chapter 2** identified that the quantification of the actual effects of breast cancer screening is still lacking for Eastern Europe. One main explanation could be serious (financial) barriers to organizing screening services[4], while another key barrier in most of these countries is the presence of opportunistic screening leading to lack of data collection. This is an area of possible improvement and future research.

Benefits and harms of breast cancer screening

Systematic reviews of harms and benefits are important decision tools for stakeholders seeking to evaluate health policy interventions, such as breast cancer screening programs. In Chapter 3, we aimed to investigate the determinants of benefits and harms of different breast cancer screening approaches (mammography, ultrasonography, clinical breast examination and breast self-examination) and to summarize data from systematic reviews on those four screening approaches among the general population. Benefits included breast cancer mortality and all-cause mortality reduction as well as stage shift or detection of smaller tumours. Harms were expressed as overdiagnosis (i.e. diagnosis and treatment of breast cancer that normally would not have been clinically diagnosed), overtreatment, false-positive diagnosis, and radiation-induced deaths.

Overall, we found that systematic reviews of breast cancer screening focus on mammography more than on the other screening approaches, and evaluate benefits of screening more frequently than harms. All included systematic reviews are consistent in certifying a reduction in breast cancer mortality among women aged 50–69 years. However, important nuances exist across studies and regions at variable levels of credibility. Results for overdiagnosis ranged from 0% to 84% and varied by type of original evidence, the denominator (unscreened, screened detected, entire follow-up, etc.), the duration of follow up, the accounting for ductal carcinoma in situ (DCIS), adjustment for breast cancer risk and lead time. Therefore, the magnitudes in effects of benefits and harms of mammography remain heterogeneous. In general, the assessed reviews of RCTs have greater similarity in included studies but larger variability in quality assessment while reviews on observational studies show the opposite trend. Conducting a review of systematic reviews, as we did in Chapter 3, refers to reviewing the compiling evidence from multiple reviews into one accessible and usable document[5] and apply as a lens through which other types of studies should be appraised[6]. As discussed in Chapter 3, there are inherent difficulties in only relying on systematic reviews and meta-analysis without

understanding the strength of the primary evidence and the methodology of the analyses. Screening programs are heterogeneous and their harms and benefits ultimately depend on the design of the studies and their evaluation approach. As the strength of public health recommendations depend on the quality of evidence about harms and benefits[6], in our review of reviews we aimed to evaluate the certainty and quality of this evidence. Overall, we found that until new results from high-quality cohort or RCTs are published, additional systematic reviews on breast cancer screening with mammography would not be of great value.

Impact of optimized screening coverage on breast cancer mortality

Suboptimal (informed) participation in breast cancer screening is not only a barrier to achieving equity, but it is also a barrier to achieving the most important benefit in screening, reducing breast cancer specific mortality. In **Chapter 4**, we took the co-existence of organized and opportunistic breast cancer screening into account and illustrated, that breast cancer screening in Europe already has a substantial impact by preventing ten thousands of breast cancer deaths per year. Through introducing a hypothetical 100% examination coverage (i.e. 100% of the 50-69 year old women are screened), the number of breast cancer deaths in European women could be further reduced. The effect would be particularly notable (in relative terms) in Eastern Europe, where screening attendance is lower and programmes are not organized. Based on our results, countries that currently do not offer organized screening for the target age range of 50 to 69 years should strongly consider it. A secondary aim of this chapter was to provide an extensive overview of the amount of organized and opportunistic screening in Europe. This was determined by consulting national experts. The examination coverage of organized breast cancer screening (defined as the proportion of the target population screened in the chosen report year after invitation) showed marked regional disparities (59% in Northern Europe and 39% in Eastern Europe). The examination coverage of opportunistic screening was just as diverse (between 5% in Northern Europe and 32% in Southern Europe). Those add up to the total examination coverage, which ranged from 49% in Eastern Europe to 69% in Southern Europe, with even wider ranges across countries.

The information collected in **Chapters 2 to 4** thus provides the following answer to the research question:

How well is breast cancer screening really working?

Since the introduction of organized mammography screening, its role in the decline of breast cancer mortality is much debated. Several factors, aside from early detection, accounted for this decline, such as breast cancer awareness and more efficient treatment in multidisciplinary breast care centers. Of these factors, the contribution of breast cancer screening the most debated. This debate (at least partly) stems from disagreements over the validity and applicability of the available RCTs as well as the usefulness and interpretation of observational studies on breast cancer mortality. Already in 2013, Duffy et. al. concluded, that the controversy over conflicting

claims for mortality reduction in RCTs is largely artificial[7]. When the same screening and follow-up periods, the same denominator are maintained, all studies indicate a substantial reduction in breast cancer mortality with screening of around 20%. The European Commission Initiative on Breast Cancer's (ECIBC) guideline and the UK Panel on Breast Cancer agree that the best evidence for the relative benefit of screening on mortality reduction comes from RCTs of breast screening[8 9]. Nevertheless, we decided to include RCTs as well as observational studies in our systematic review from Chapter 2, as we think that estimates from contemporary observational studies may be considered more relevant today than those of the RCTs. Recent studies can take improvements in mammographic techniques and treatment into account that have occurred over the past 30 years. However, researchers need to account for the complexity in evaluating the long-term effect of breast cancer screening from observational data. The stringent use of widely acknowledged and transparent grading tools in order to appraise the quality of each included reference allowed us to highlight only those studies, which provide the most valid information. From this set of high quality studies we found that the impact of organized screening (target ages 50 to 69) on breast cancer specific mortality ranges from 12% to 58% in screening attenders versus non-attenders and from 4% to 31% in invited versus non-invited women. Based on this impact, we estimated that each year, breast cancer screening prevents nearly 21,700 breast cancer deaths in Europe. Introducing a hypothetical 100% coverage of screening in the advised target age groups, the number of breast cancer deaths of European women could be further reduced by almost 12,500 per year. The effect would be particularly notable in Eastern Europe. Even when programs to screen for breast cancer exist, much is still to be done. This includes increasing screening coverage through evidence-based interventions[10-12] and removing barriers to effective breast cancer screening[13 14]. There are no consistent conclusions about the magnitude of breast cancer mortality reduction among women younger than 50 years or older than 69 years. Thus, with moderate certainty of the evidence, the ECIBC continues to recommend screening at these ages only conditionally[15]. Similarly, the frequency at which overdiagnosis occurs remains a topic of strong debate[8 9 16].

Part 2: Modelling the impact of different interventions on the harms and benefits of breast cancer screening

In Part 2 of this thesis, we evaluated the impact of different screening interventions on the harms and benefits of breast cancer screening across different European countries, using the MISCAN-Breast model. Because the model parameters, which are fitted (calibrated) to match observed data, might differ across Europe, four different models were developed. From each European region, we selected an exemplary country with high quality observational data, including the screening behaviour of that population, to be representative for that region: the Netherlands for Western Europe, Italy for Southern Europe, Finland for Northern Europe and Slovenia for Eastern Europe (Figure 1, part 1). Then, to ensure the validity of these models and therefore their usefulness in predicting, evaluating, and improving the existing screening programs, we

validated our model predictions against observed data. As an extension of the systematic review noted in **Chapter 2**, we identified studies that provided "best evidence" in observed data, which the MISCAN model output can be validated against (Figure 1, part 2). We found that our country-specific models accurately estimated stage distribution, incidence and mortality rates as well as breast cancer mortality reductions due to screening in Europe (Figure 1, part 3). Finally, we used these models to design the EU-TOPIA evaluation tool. This powerful online tool allows users to simulate outcomes, benefits and harms of several cancer screening strategies for their own country (Figure 1, part 4).

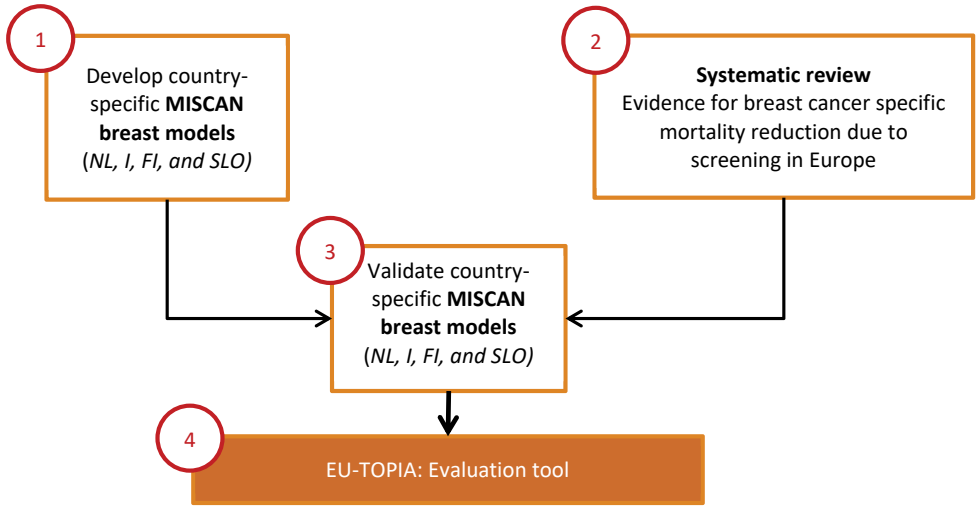


Figure 1: The consecutive steps of EU-TOPIA Evaluation tool development, based on the four exemplary countries (the Netherlands, Italy, Finland and Slovenia).

Effects of adjusting the screening age-range on harm-to benefit ratios

In 2003, the European Council published their first guidelines for organized mammography screening programs for early detection of breast cancer in asymptomatic women with a strong recommendation for inviting women ages 50-69, every two years[17]. Screening recommendations from some major medical agencies have strongly supported the recommendation for this age range[18,19]. However, with different demographic structures, background risks and health policies, there is no one-size-fits-all paradigm in organized breast cancer screening. While most European countries adopted the target age range for breast cancer screening suggested by the European Council, to increase the benefits for their female population, a few countries adopted a different age range, either inviting women younger than 50 years or inviting women beyond the age of 69 years (**Chapter 4**).

The effectiveness of screening below the age of 50 years is an important issue in breast cancer screening. While young women (<50 years) are known to be at lower risk of developing breast cancer compared to older women[1 20], tumors grow faster. However, younger women have higher breast density, which lowers the sensitivity of mammography and results in more false-positive results[21-25]. The most recent European screening guidelines suggest that screening could be beneficial for those women aged below 50 or over 69 years. For asymptomatic women aged 45 to 49 and 70 to 74 with an average risk of breast cancer, the ECIBC's Guidelines Development Group (GDG) suggests mammography screening in the context of an organized screening program[15], similar to the review performed for the IARC Handbook[1]. However, the evidence is not conclusive[9].

In **Chapter 5**, we therefore used the MISCAN-Breast models to assess how harm-to-benefit-ratios would vary if biennial breast cancer screening would be extended to younger and/or older age groups (45-69, 45-74, 50-74) when compared with the current strategy of screening women in the age group 50-69. We performed simulations for the four exemplary European countries noted previously (The Netherlands, Italy, Finland and Slovenia). Our research aimed to quantify the benefits and the harms of breast cancer screening in the varying age groups. The balance between benefits and harms is expressed in four harm-to-benefit-ratios: overdiagnosed breast cancer cases/averted breast cancer deaths, false-positive results/averted breast cancer deaths, overdiagnosed breast cancer cases/ life-years gained and false-positive results/ life-years gained.

Compared to the reference strategy 50–69, screening women at 45–74 or 50–74 years would be less beneficial in any of the four countries than screening women at 45–69, which would result in relatively fewer overdiagnoses per death averted or life-years gained. At the same time, false positive results per death averted would increase substantially. Our analysis in **Chapter 5** provides insight as to how harm-to-benefit-ratios of breast screening programs could be improved by adjusting the age range of screened women. Assuming different strategies, this modelling study represents meaningful information on the magnitude of harms and benefits and the results are likely to be relevant to other European countries as well. Our research focused on explaining how these four harm-to-benefit-ratios are affected by varying age-ranges in four countries. A valuable extension to this analysis would have been to include different screening frequencies and to include the cost-effectiveness of those strategies, as this is essential before any decision about potential screening policy changes can be made. Future research should combine the triad of benefits, harms and costs as key elements of health policy decision making. Such an analysis could consider additional screening effects, such as treatment related advantages or psychological harms that might occur throughout the phases of the screening pathway. These effects are all associated with women's health state utilities (quality-adjusted life years, QALYs) as well as disability-adjusted life years (DALYs). Adding such

units to the consideration of optimal age-range might even lead to results that contradict ours from **Chapter 5**. In a recently published systematic review, the authors determined whether new technology and therapies have improved population breast cancer outcomes at reasonable ranges of economic value. The results – all from modelling studies - were consistent in finding that biennial screening from ages 50-69 or 50-74 years was the most efficient (i.e., there were no other strategies that saved more QALYs for the same costs)[26].

Modelling feasible changes of breast cancer screening to overcome barriers

Despite the benefits of early detection, in practice, screening programs often fail to achieve their full potential. Therefore, in **Chapter 6** we assessed how breast cancer screening programs can be optimized. As part of the EU-TOPIA project, we developed a self-assessment tool to help identify the barriers to the optimal operation of population-based breast cancer screening programs[13]. Using this tool, barriers can be identified at every step of the screening process: knowledge generation, identification of the eligible population, maximising uptake (informed participation), successful operation of the programme, adequate follow-up and effective treatment for those who need it.

Using Italy as the example, we first identified barriers common to those identified in other countries, namely low adherence to screening (particularly notable in Southern Italy) and a combination of screening intervals due to opportunistic screening (leading to a screening interval < 2 years) and a lack of recourses (leading to a screening interval > 2 years). Subsequently, we modelled potential future changes (i.e. practical solutions) that can be initiated to the breast cancer screening programs to overcome these barriers and to improve screening in Italy. We also performed an economic evaluation, employing the cost-effectiveness tool that is integrated into the EU-TOPIA evaluation tool. We found that removing the most important barriers of breast cancer screening leading to increased adherence for Southern Italy and harmonized screening intervals for the whole of Italy, could result in substantial improvements at acceptable costs. For example, screening all eligible women in Italy every two years is predicted to lead to a substantial decrease in referrals (-12%), false positive results (-13%) and costs (-18%) while almost all benefits could be maintained. Overcoming this barrier would safe €162,799 per QALY gained. Our results demonstrated the effectiveness and reliability of EU-TOPIA tools as a resource for European policymakers aiming to improve cancer screening in their country.

The EU-TOPIA evaluation tool was the cornerstone of **Chapter 6**. Not only does it help to quantify the harms and benefits of existing screening programs, it also enables the definition and evaluation of alternative screening strategies that could be implemented to overcome key barriers. In this way, the EU-TOPIA evaluation tool gives users the opportunity to develop a road map for improving long-term cancer screening outcomes. By providing country-specific data (i.e. demographic, epidemiological, and cancer screening program related data) users can quantify

future harms and benefits of different cancer screening scenarios in their country. Users then define ways to overcome the barriers in the alternative scenario by 1) identifying input parameters which need to be changed and 2) defining the magnitude of change for each input parameters[27]. The Italian example demonstrates a systematic approach and stepwise process, which can be easily followed or adapted by other European countries or stakeholders. When alternative screening scenarios reflect more favorable long-term outcomes at feasible costs (as seen in **Chapter 6**), a next step is to identify (supportive) stakeholders to help to overcome these barriers in "real life". Subsequently, as the final part of the road map development, an action plan should be outlined for each of the simulated scenarios (Figure 2).

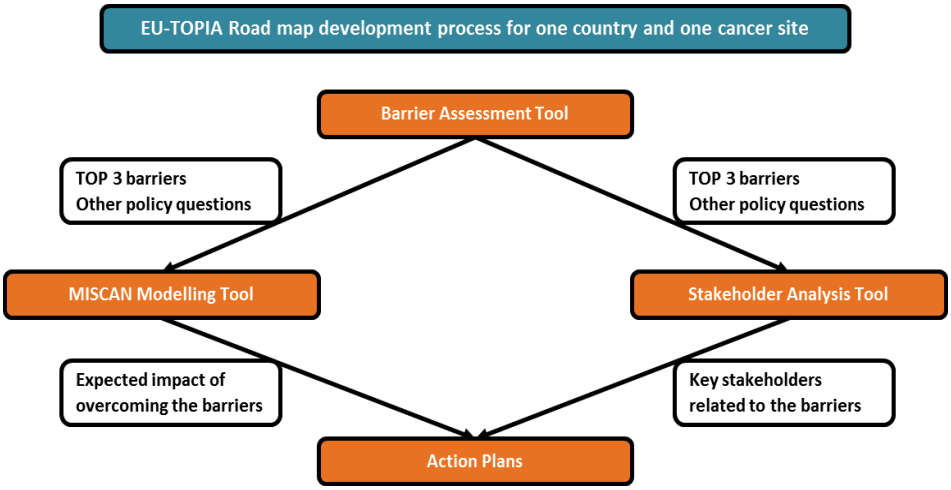


Figure 2: Road map development process in EU-TOPIA.

The intent of a road map is to inform national or regional policymakers about action plans which improve cancer screening programmes by overcoming barriers in a structured form. It also reports the quantification of the impact of the proposed changes on screening outcomes and identifies key stakeholders to implement the changes.

DALYs and QALYs: same but different

Cost-effectiveness analyses are valuable tools to determine an optimal screening intervention by comparing it with alternative strategies. Breast cancer screening affects the length of life (mortality) as well as quality of life (morbidity)[28]. To date, the best way to combine mortality and morbidity in one single unit is using health-adjusted life years, an umbrella term for quality adjusted life years (QALY) and disability adjusted life years (DALY). To use cost-effectiveness analyses more effectively to inform health policy, in **Chapter 7** we evaluated the impact of using DALYs versus QALYs on outcomes and optimal breast cancer screening strategies. We questioned if QALY and DALY were interchangeable summary measures of health. Using DALYs

averted instead of QALYs gained to assess the effects on quality of life from breast cancer screening in the Netherlands yields differences in cost-effectiveness. However, using DALYs averted or QALYs gained resulted in only subtle nuances, with nearly the same strategies on the efficient frontiers, and no effect on the conclusions of the cost-effectiveness analyses. The relationship between these two measures remained constant through all analyses and **Chapter 7** suggests that both can be used to determine the optimal breast cancer screening strategy. A number of other studies have discussed the theoretical differences between QALYs and DALYs[29 30]. They generally concluded that both measures have proven serviceable for resource allocation and priority setting in health policy. Nonetheless, to accommodate comparison across studies, results from **Chapter 7** recommend using DALYs for burden of disease studies and QALYs in cost-effectiveness analyses.

The information collected in **Chapters 5 to 7** thus provides the following answer to the research question:

How can harmful screening effects be reduced and positive effects enhanced? And are there better screening practices possible?

The scientific body of evidence on breast cancer screening that has been gathered in the past decades has led to the widespread use of mammography screening throughout Europe. However, there is no consensus about which screening strategy is optimal. The most recent systematic review on RCTs, which informed the European Breast Cancer Guidelines, estimated that for each breast cancer death avoided, approximately four overdiagnosed cases will be managed when women are invited to screening from age 50 to 69[9]. However, this figure remains tentative due to the potential bias in the overdiagnosis estimates. Such limited quantification abilities regarding the most important harm of breast cancer screening, emphasize the need for ways to complement RCTs. This is an area where simulation models are used to make projections about the effects of various screening strategies. After calibrating and validating MISCAN for four exemplary European countries, we concluded that - compared to the reference strategy 50–69 - screening women at 45–74 or 50–74 years would be less beneficial in any of the four countries than screening women at 45–69, which would result in relatively fewer overdiagnoses per death averted or life-years gained. At the same time, false positive results per death averted would increase substantially. Our results indicate that any conclusions on an improved balance between harms and benefits depend on the metric used to evaluate changes in screening strategy. That highlights the importance of taking preferences of different stakeholder groups (such as the target population, health professionals, and policymakers) into account[9 31]. Removing the most important barriers of the current breast cancer screening programs could result in substantial improvements and be cost effective or even cost saving. Based on our hypothesis that decreasing geographical barriers helps to increase participation, we showed that investing in mobile mammography units would be cost effective in regions with low screening adherence.

Both QALYs and DALYs can produce cost-effectiveness estimates that assist health policy makers with identifying changes that potentially improve their screening programs.

FUTURE CHALLENGES AND OPPORTUNITIES

Consideration on the comparability of harms and benefits

Overdiagnosis is inherent to breast cancer screening, which seeks to diagnose and mitigate the disease before it is clinically evident. The question we need to ask is then, how much overdiagnosis is acceptable for breast cancer screening before a program needs to change or to stop.

Despite research evidence on the effect of different screening strategies on harms and benefits and the ratios thereof, researchers are limited in the way they voice health policy recommendations. The reason for that is that health policy related decision-making depends on how the considered harms and benefits are prioritized by the people in charge of the program. More consideration should be given to the comparability of what are very different events. For example, the value of a life saved (QALYs gained) versus an overdiagnosed case (potentially resulting in DALYs averted): the consequences of the two different events are obviously of different magnitude. Or: Is overdiagnosis worse than false-positive results? The answer to that might differ depending on whom you ask: Women attending screening may have a different perspective and tolerance towards harms and benefits than policymakers. Making informed screening decisions requires transparency in values and preferences judgements. Through surveys and panel discussions, target population thresholds could be established and subsequently used to inform screening guidelines and recommendations. That would be a prerequisite for specifying the magnitude of benefit required for people to accept the burdens and harms associated with screening[32 33]. To avoid or clarify potential discrepancies, an acceptance threshold of harms and benefits of screening, translated into harms-to-benefit-ratios, should be included in the European Screening Guidelines.

Special considerations Eastern Europe

Breast cancer causes more than 51,000 deaths each year in Eastern Europe[34]. Chapter 2 presented a large amount of evidence on the effectiveness of screening programmes in Western, Northern and Southern Europe. Although many Eastern European countries have implemented some form of breast cancer screening, the key barrier in most of these countries is the presence of non-organized (opportunistic) screening programs, leading to lack of data collection and lack of good quality assurance systems. In addition, coverage of organized screening is commonly low and vulnerable minority groups are not reached[35 36]. One can expect that, when bringing the quality of the screening program to an acceptable level, breast cancer screening will be as

effective in Eastern European countries as it is in the rest of the continent. Thus, successfully implementing breast cancer screening in Eastern Europe can potentially impact the lives of thousands of women.

In order for screening programs to be efficient and effective (i.e. benefits are to be maximised and harms minimized), breast cancer screening programs should be adapted to their country-specific context to fit the culture, health system capacities and technologies available. Even then, cancer screening remains a complex intervention due to the diverse natural history of the disease, and the need to organize mass interventions in a healthy population. Moreover, organized screening puts further pressure on the available clinical and economic resources of a country. Given the more limited resources in some of the Eastern European countries, it is especially important to ensure that money is well spent and that citizens benefit optimally. The implementation of an efficient country-specific screening program is therefore difficult, but crucial.

During the EU-TOPIA project, we have built capacity throughout Europe, making individual organizations better equipped to evaluate and improve cancer screening programs and strengthening the European network. The tools developed during this project will be continued in the follow-up project EU-TOPIA EAST, so that researchers, program coordinators and policy makers can use them and can continue to develop road maps. Building on that, the road map development process can be taken to the next level by implementing feasible and country-specific interventions in Eastern Europe. In order to reach that goal:

- local health and social systems should be taken into account and detailed barrier and stakeholder analyses should be performed, leading to feasible changes to the current screening programs;
- the implemented programs should be monitored and evaluated using key indicators and sophisticated decision models (like MISCAN breast) to predict the long-term and country-wide benefits, harms and cost-effectiveness;
- the process of implementation should be aligned with European standards and best practices. The experiences of other EU Member States as well as the current country-specific situations should be analyzed and national guidelines and protocols for the implementation of the breast cancer screening programs should be developed; and
- detection of advanced and interval cancers should serve as an early indicator of the possible success of the pilot breast cancer screening program, as screening programs are long-term planned and costly.

Continuous monitoring of screening activities allows stakeholders to assess whether implemented changes are achieving the expected impact.

Future issues for Europe

Breast cancer screening programs commonly have an age-based approach, because age is the strongest risk factor for breast cancer for most women. However, women are not all the same and we know that at any given age there is variability in breast cancer risk due to individual risks of developing breast cancer, depending on many factors like genetic factors, lifestyle, or hormonal exposure. Recent scientific advances have largely improved our understanding of breast cancer genetics and other risk factors[1 37 38]. 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. Therefore, a tailored screening approach could both contribute to a reduction of possible harms and reduce the costs of service screening programs.

There is increasing interest in risk-stratified screening, with the number of publications in this area expanding rapidly over the last ten years. Two ongoing trials, My-PEBS (Europe) and the WISDOM trial (US) are presently testing age-based versus risk-based screening approaches that include genetic markers and family history information[39 40]. The Tailored Breast Screening Trial (TBST) in Italy uses a breast density classification to allocate women to a longer interval, decreasing the number of screening rounds in the 45-50-year age range[41]. The first two rounds of screening in the Dense Tissue and Early Breast Neoplasm Screening (DENSE) Trial in the Netherlands successfully demonstrated that offering a supplemental Magnetic resonance imaging (MRI) to women with extremely high breast density resulted in a reduction in interval cancers and a sharply reduced false-positive rate[42 43]. These first results led the Dutch government to assess the feasibility of supplemental MRI exams for women with extremely dense breasts to the national screening program[44].

The results of these trials might warrant significant organizational changes in current breast cancer screening programs[45 46]. The ultimate aim is to implement risk-stratified screening (i.e. high or low-risk pathways) that is justifiable from ethical, legal and societal viewpoints[47]. The feasibility of implementing personalized risk-based screening and prevention is dependent on how healthcare is funded and arranged, and can vary between countries[48]. Any future developments regarding tailored breast cancer screening will require more complex frameworks and shifts in the service organization of the screening program. This would include ethical principles and service infrastructure capability[49], the assessment and communication of individual risk, personalised invitation protocols, ensuring informed choice of the eligible women[50], knowledge management of health professionals [51] and the (annual) monitoring and evaluation of each program. All novel procedures and settings will need to adhere to the same high-quality standards that exist within current screening programmes

Recommendations for future research

- Elicit an acceptance threshold of harms-to-benefit-ratios, which can be included in the European Screening Guidelines. This would help to standardize the priorization of the considered harms and benefits.
- Extend the harm-to-benefit analysis of various age-based strategies of breast cancer screening to a cost-effectiveness analysis. Such an analysis would consider additional screening effects, such as treatment-related advantages or quality of life, as well as costs.
- Implement improved cancer screening programs in Eastern Europe taking into account all relevant country-specific factors, including ethical, cultural, environmental and socio-economic factors and resource differences.
- Develop country-specific road maps, taking barriers and resources into account, to ensure the implementation of feasible country-specific screening programs.
- Estimate the cost and effects of screening targeted to individual, age-specific, breast cancer risk based on a combination of risk factors including breast density, family history, and gene mutations.
- Evaluate the potential for risk-based breast screening in countries where a population-based screening programme is not yet fully implemented (e.g. Eastern Europe).

MAIN CONCLUSIONS

- Several methodologically appropriate approaches exist that are able to capture the true beneficial effect of mammographic screening.
- The reduction in breast cancer mortality due to breast cancer screening in attenders versus non-attenders ranged between 12%-58% in Europe.
- Yearly, 21,680 breast cancer deaths have already been prevented due to mammography screening. The number of breast cancer deaths could be further reduced substantially (by more than 12,000 annually) in all European regions, if 100% of the eligible women aged 50-69 in Europe would be screened every two years.
- Starting screening women at 45 years, five years prior to the currently recommended age-range of 50-69 years, could improve the ratio of overdiagnosed breast cancer to breast cancer deaths averted.
- The major barriers to the breast cancer screening programs in (Southern) Italy could be overcome by initiating feasible changes leading to better long-term outcomes. The employed online tools can be an effective and reliable resource for European policymakers aiming for informed decision-making on cancer screening in their country.
- DALY and QALY can both be used to determine the optimal breast cancer screening strategy.

REFERENCES

1. IARC Working Group. IARC Handbooks of Cancer Prevention, Volume 15: Breast Cancer Screening. Lyon, 2016.
2. Autier P, Boniol M. Breast cancer screening: evidence of benefit depends on the method used. BMC Med 2012;10:163.
3. Lauby-Secretan B, Loomis D, Straif K. Breast-Cancer Screening--Viewpoint of the IARC Working Group. N Engl J Med 2015;373(15):1479.
4. IARC. Cancer Screening in the European Union. Report on the implementation of the Council Recommendation on cancer screening. IARC, Lyon 2017;Available at: 2017_ cancerscreening_2ndreportimplementation_en.pdf.
5. Hasanpoor E, Hallajzadeh J, Siraneh Y, et al. Using the Methodology of Systematic Review of Reviews for Evidence-Based Medicine. Ethiop J Health Sci 2019;29(6):775-78.
6. Murad MH, Asi N, Alsawas M, et al. New evidence pyramid. Evid Based Med 2016;21(4):125-7.
7. Duffy SW, Chen THH, Smith RA, et al. Real and artificial controversies in breast cancer screening. Breast Cancer Management 2013;2(6), 519–528.
8. Marmot MG, Altman DG, Cameron DA, et al. The benefits and harms of breast cancer screening: an independent review. Br J Cancer 2013;108(11):2205-40.
9. Canelo-Aybar C, Ferreira DS, Ballesteros M, et al. Benefits and harms of breast cancer mammography screening for women at average risk of breast cancer: A systematic review for the European Commission Initiative on Breast Cancer. J Med Screen 2021:969141321993866.
10. Jepson R, Clegg A, Forbes C, et al. The determinants of screening uptake and interventions for increasing uptake: a systematic review. Health Technol Assess 2000;4(14):i-vii, 1-133.
11. Camilloni L, Ferroni E, Cendales BJ, et al. Methods to increase participation in organised screening programs: a systematic review. BMC Public Health 2013;13:464.
12. Luckmann R, Costanza ME, White MJ, et al. A 4-year randomized trial comparing three outreach interventions to promote screening mammograms. Transl Behav Med 2019;9(2):328-35.
13. Priaulx J, de Koning HJ, de Kok I, et al. Identifying the barriers to effective breast, cervical and colorectal cancer screening in thirty one European countries using the Barriers to Effective Screening Tool (BEST). Health Policy 2018;122(11):1190-97.
14. Turnbull E, Priaulx J, van Ravesteyn NT, et al. A health systems approach to identifying barriers to breast cancer screening programmes. Methodology and application in six European countries. Health Policy 2018;122(11):1198-205.
15. European Commission Initiative on Breast Cancer. Recommendations from the European Breast Guidlines. 2021;Available at: https://ecibc.jrc.ec.europa.eu/recommendations/.
16. de Gelder R, Heijnsdijk EAM, van Ravesteyn NT, et al. Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiol Rev 2011;33:111-21.
17. Council of the European Union. Council Recommendation of 2 December 2003 on Cancer Screening (2003/878/EC). OJ L 327: 34-38. 2003.

18. Lauby-Secretan B, Scoccianti C, Loomis D, et al. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med 2015;372(24):2353-8.
19. Siu AL, Force USPST. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med 2016;164(4):279-96.
20. Bray F, Colombet M, Mery L, et al. Cancer Incidence in Five Continents, Vol. XI (electronic version). Lyon: International Agency for Research on Cancer Available from: http://ci5iarcfr 2017.
21. van den Ende C, Oordt-Speets AM, Vroling H, et al. Benefits and harms of breast cancer screening with mammography in women aged 40–49 years: A systematic review. Int J Cancer 2017;141(7):1295-306.
22. Mandrik O, Zielonke N, Meheus F, et al. Systematic reviews as a 'lens of evidence': Determinants of benefits and harms of breast cancer screening. Int J Cancer 2019;145(4):994-1006.
23. Wanders JO, Holland K, Veldhuis WB, et al. Volumetric breast density affects performance of digital screening mammography. Breast Cancer Res Treat 2017;162(1):95-103.
24. Engmann NJ, Golmakani MK, Miglioretti DL, et al. Population-Attributable Risk Proportion of Clinical Risk Factors for Breast Cancer. JAMA Oncol 2017;3(9):1228-36.
25. Carney PA, Miglioretti DL, Yankaskas BC, 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.
26. Jayasekera J, Mandelblatt JS. Systematic Review of the Cost Effectiveness of Breast Cancer Prevention, Screening, and Treatment Interventions. Journal of Clinical Oncology 2020;38(4):332-50.
27. Gini A, van Ravesteyn NT, Jansen EEL, et al. The EU-TOPIA evaluation tool: An online modelling-based tool for informing breast, cervical, and colorectal cancer screening decisions in Europe. Prev Med Rep 2021;22:101392.
28. Molassiotis A, Tyrovolas S, Gine-Vazquez I, et al. Organized breast cancer screening not only reduces mortality from breast cancer but also significantly decreases disability-adjusted life years: analysis of the Global Burden of Disease Study and screening programme availability in 130 countries. ESMO Open 2021;6(3):100111.
29. Airoldi M, Morton A. Adjusting life for quality or disability: stylistic difference or substantial dispute? Health Econ 2009;18(11):1237-47.
30. Feng X, Kim DD, Cohen JT, et al. Using QALYs versus DALYs to measure cost-effectiveness: How much does it matter? Int J Technol Assess 2020;36(2):96-103.
31. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. Ann Intern Med 2009;151(10):738-47.
32. Helsingen LM, Zeng L, Siemieniuk RA, et al. Establishing thresholds for important benefits considering the harms of screening interventions. BMJ Open 2020;10(12):e037854.
33. Zeng L, Helsingen LM, Kenji Nampo F, et al. How do cancer screening guidelines trade off benefits versus harms and burdens of screening? A systematic survey. BMJ Open 2020;10(12):e038322.
34. Ferlay J, Ervik M, Lam F, et al. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.fr/today, accessed [14/05/2021].

35. World Cancer Report: Cancer Research for Cancer Prevention. Lyon: International Agency for Research on Cancer (IARC), 2020.
36. Porta M. A Dictionary of Epidemiology (5 ed.).Oxford University Press, 2014.
37. Saadatmand S, Geuzinge HA, Rutgers EJT, et al. MRI versus mammography for breast cancer screening in women with familial risk (FaMRIsc): a multicentre, randomised, controlled trial. Lancet Oncol 2019;20(8):1136-47.
38. Geuzinge HA, Obdeijn IM, Rutgers EJT, et al. Cost-effectiveness of Breast Cancer Screening With Magnetic Resonance Imaging for Women at Familial Risk. JAMA Oncol 2020;6(9):1381-89.
39. Esserman LJ, Study W, Athena I. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. NPJ Breast Cancer 2017;3:34.
40. MyPeBS H-. International Randomized Study Comparing personalized, Risk-Stratified to Standard Breast Cancer Screening In Women Aged 40-70. 2020;Available at: https://cordis.europa.eu/project/id/755394.
41. Paci E, Mantellini P, Giorgi Rossi P, et al. [Tailored Breast Screening Trial (TBST)].Tailored Breast Screening Trial (TBST). Epidemiologia e Prevenzione 2013;37(4-5):317-27.
42. Bakker MF, de Lange SV, Pijnappel RM, et al. Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. N Engl J Med 2019;381(22):2091-102.
43. Veenhuizen SGA, de Lange SV, Bakker MF, et al. Supplemental Breast MRI for Women with Extremely Dense Breasts: Results of the Second Screening Round of the DENSE Trial. Radiology 2021;299(2):278-86.
44. van Sonderen JF, van Kerkhof LWM, van Klink-De Kruijff IE, et al. Verkorte uitvoeringstoets aanvulling bevolkingsonderzoek borstkanker met MRI voor vrouwen met zeer dicht borstweefsel. RIVM-briefrapport. 2020. 2020-011.
45. Allweis TM, Hermann N, Berenstein-Molho R, et al. Personalized Screening for Breast Cancer: Rationale, Present Practices, and Future Directions. Ann Surg Oncol 2021.
46. French DP, Astley S, Brentnall AR, et al. What are the benefits and harms of risk stratified screening as part of the NHS breast screening Programme? Study protocol for a multi-site non-randomised comparison of BC-predict versus usual screening (NCT04359420). BMC Cancer 2020;20(1):570.
47. Pashayan N, Antoniou AC, Ivanus U, et al. Personalized early detection and prevention of breast cancer: ENVISION consensus statement. Nat Rev Clin Oncol 2020;17(11):687-705.
48. Rainey L, van der Waal D, Jervaeus A, et al. Are we ready for the challenge of implementing risk-based breast cancer screening and primary prevention? Breast 2018;39:24-32.
49. McWilliams L, Woof VG, Donnelly LS, et al. Risk stratified breast cancer screening: UK healthcare policy decision-making stakeholders' views on a low-risk breast screening pathway. BMC Cancer 2020;20(1):680.
50. McWilliams L, Woof VG, Donnelly LS, et al. Extending the screening interval for women at low risk of breast cancer as part of a risk-stratified Breast Screening Program: Views from two Professional stakeholder groups. Int J Behav Med 2021;28(Suppl 1):S103-S03.
51. Blouin-Bougie J, Amara N, Simard J. Toward a Population-Based Breast Cancer Risk Stratification Approach? The Needs and Concerns of Healthcare Providers. J Pers Med 2021;11(6).

Summary

Breast cancer is a major public health problem in Europe. It is by far the most frequently diagnosed neoplasm in European women as it accounts for nearly one third of all new cancers in women. At present, women in Europe have a 1:7 chance of developing breast cancer during their lifetime. Breast cancer remains the leading cause of death in European women. But encouragingly, breast cancer mortality has been declining in most of Europe since the 1980's. This favourable trend in breast cancer mortality is essentially due to advancements in early diagnosis and improved treatment. However, great inequity persists in cancer incidence and mortality rates across Europe.

Breast cancer screening means checking a woman's breasts for cancer before there are signs or symptoms of the disease. It aims to reduce morbidity associated with advanced stages of the disease and subsequently breast cancer mortality. At present, breast cancer screening programmes are well established in most European countries. Most of them adopted biennial screening for breast cancer in the minimum target age range (50–69 years, as recommended by the European Union). But disparities exist in terms of the status of implementation, the extent to which screening programmes are organized or coexist with opportunistic screening activity, as well as the invitation coverage and the attendance to screening.

With this thesis, we aimed to investigate the consequences of variations in breast cancer screening practices and potential ways to further optimize screening programs across Europe.

Part 1: The effectiveness of breast cancer screening

In Chapter 2, we performed a systematic review to summarize the current evidence on reducing breast cancer mortality with mammography screening in Europe. Our results strengthen previous findings that mammography screening reduces mortality from breast cancer, but highlight that this occurs at varying magnitudes. For this systematic review, we included randomized controlled trials (RCTs) as well as observational studies such as prospective and retrospective controlled cohort or case-control studies. The stringent use of widely acknowledged and transparent grading tools in order to appraise the quality of each included reference allowed us to highlight only those studies, which provide the most valid information. From this set of high quality studies we found that the impact of organized screening (target ages 50 to 69) on breast cancer specific mortality ranges from 12% to 58% in screening attenders versus non-attenders and from 4% to 31% in invited versus non-invited women.

In Chapter 3, we aimed to investigate the determinants of benefits and harms of different breast cancer screening approaches (mammography, ultrasonography, clinical breast examination and breast self-examination) and to summarize data from systematic reviews on those four screening approaches among the general population. Overall, we found that systematic reviews of breast cancer screening focus on mammography more than on the other screening approaches, and

evaluate benefits of screening more frequently than harms. All included systematic reviews are consistent in certifying a reduction in breast cancer mortality among women aged 50–69 years. However, important nuances exist across studies and regions at variable levels of credibility. Results for overdiagnosis varied widely by type of original evidence and several methodological differences, e.g. the denominator, the duration of follow up or the accounting for ductal carcinoma in situ.Therefore, the magnitudes in effects of benefits and harms of mammography remain heterogeneous.

We estimated in Chapter 4, that breast cancer screening prevents nearly 21,700 breast cancer deaths in Europe each year. Cancer screening programs can only be effective in reduction in mortality from breast cancer if a high proportion of people within the target population make an informed choice to participate. Our aim in Chapter 4 was to investigate what the effect would be of an increased or even complete breast cancer screening coverage on breast cancer mortality for each European country. We found that introducing a hypothetical 100% coverage of screening in the advised target age group (meaning, all eligible women aged 50-69 are invited and follow this invitation), the number of breast cancer deaths of European women could be further reduced by almost 12,500 per year. The effect would be particularly notable in Eastern Europe.

Part 2: Modelling the impact of different interventions on the harms and benefits of breast cancer screening

In Part 2 of this thesis, we evaluated the impact of different screening interventions on the harms and benefits of breast cancer screening across different European countries, using the microsimulation model (MISCAN-Breast).

Breast cancer screening causes harms and benefits. The balance between the two varies by age. In Chapter 5, we used the MISCAN-Breast models of four European countries (the Netherlands, Finland, Italy and Slovenia) to assess how harm-to-benefit-ratios would vary if biennial breast cancer screening would be extended to younger and/or older age groups (45-69, 45-74, 50-74) when compared with the current strategy of screening women in the age group 50-69. We found that in all countries, adding screening between the ages 45 and 49 or 70 and 74 resulted in more life-years gained and more breast cancer deaths averted, but at the expense of increases in harms (i.e. the number of overdiagnoses and false-positive diagnoses,). Adapting the age range of breast cancer screening is an option to improve harm-to-benefit ratios, but the prioritization of considered harms and benefits affects the interpretation of the results.

Despite the benefits of early detection, in practice, screening programs often fail to achieve their full potential. Therefore, in Chapter 6 we assessed how breast cancer screening programs can be further optimized. Using Italy as the example, we first identified barriers common to those identified in other countries, namely low adherence to screening and a combination of screening

intervals due to opportunistic screening (leading to a screening interval < 2 years) and a lack of recourses (leading to a screening interval > 2 years). Subsequently, we modelled potential future changes (i.e. practical solutions) that can be initiated to the breast cancer screening programs to overcome these barriers and to improve screening. Our analysis shows that removing the most important barriers of the current Italian breast cancer screening programs could result in substantial improvements and could be cost effective or even cost saving. This Italian example illustrates a systematic approach and stepwise process that can be easily followed or adapted by other European countries or stakeholders.

Cost-effectiveness analyses are valuable tools to determine an optimal screening intervention by comparing it with alternative strategies. Breast cancer screening affects the length of life (mortality) as well as quality of life (morbidity). To date, the best way to combine mortality and morbidity in one single unit is using health-adjusted life years, an umbrella term for quality adjusted life years (QALY) and disability adjusted life years (DALY). To use cost-effectiveness analyses more effectively to inform health policy, in Chapter 7 we evaluated the impact of using DALYs versus QALYs on outcomes and optimal breast cancer screening strategies. We questioned if QALY and DALY were interchangeable summary measures of health. We found that using DALYs averted instead of QALYs gained to assess the effects on quality of life from breast cancer screening in the Netherlands yields differences in cost-effectiveness. However, these differences were only subtle, with nearly the same strategies on the efficient frontiers, and no effect on the conclusions of the cost-effectiveness analyses. Since the relationship between these two measures remained constant through all analyses, we conclude that both can be used to determine the optimal breast cancer screening strategy.

CONCLUSIONS

Based on the results of the studies described in this thesis, we derived the following conclusions:

- Several methodologically appropriate approaches exist that are able to capture the true beneficial effect of mammographic screening.
- The reduction in breast cancer mortality due to breast cancer screening in attenders versus non-attenders ranged between 12%-58% in Europe.
- Yearly, 21,680 breast cancer deaths have already been prevented due to mammography screening. The number of breast cancer deaths could be further reduced substantially (by more than 12,000 annually) in all European regions, if 100% of the eligible women aged 50-69 in Europe would be screened every two years.
- Starting screening women at 45 years, five years prior to the currently recommended age-range of 50-69 years, could improve the ratio of overdiagnosed breast cancer to breast cancer deaths averted.

- The major barriers to the breast cancer screening programs in (Southern) Italy could be overcome by initiating feasible changes leading to better long-term outcomes. The employed online tools can be an effective and reliable resource for European policymakers aiming for informed decision-making on cancer screening in their country.
- DALY and QALY can both be used to determine the optimal breast cancer screening strategy.

Samenvatting

Borstkanker is een belangrijk volksgezondheidsprobleem in Europa. Het is veruit de meest voorkomende vorm van kanker bij Europese vrouwen: bijna een derde van alle nieuwe kankers bij vrouwen is borstkanker. Op dit ogenblik hebben vrouwen in Europa een kans van 1 op 7 om tijdens hun leven borstkanker te ontwikkelen. Borstkanker blijft de voornaamste doodsoorzaak bij Europese vrouwen. Maar bemoedigend is dat het sterftecijfer voor borstkanker in het grootste deel van Europa sinds de jaren tachtig is gedaald. Deze gunstige trend in het sterftecijfer voor borstkanker is vooral te danken aan de vooruitgang op het gebied van vroegtijdige diagnose en verbeterde behandeling. Er blijven echter grote ongelijkheden bestaan tussen de Europese landen in het aantal kankergevallen en de sterftecijfers.

Borstkankerscreening betekent dat de borsten van een vrouw op kanker worden gecontroleerd, voordat er tekenen of symptomen van de ziekte zijn. Doel is het voorkomen van gevorderde stadia van deze ziekte en daarmee ook de sterfte als gevolg van borstkanker te verminderen. Screeningprogramma's voor borstkanker zijn momenteel in de meeste Europese landen goed ingeburgerd. De meeste van deze landen voeren een tweejaarlijkse screening op borstkanker in de minimumleeftijdsgroep (50-69 jaar, zoals aanbevolen door de Europese Unie). Er bestaan echter verschillen in de stand van uitvoering, de mate waarin screeningprogramma's georganiseerd zijn of naast opportunistische screeningactiviteiten bestaan, en in de dekking van de uitnodigingen en de deelname aan screening.

Het doel van dit proefschrift was om de gevolgen van variaties in borstkankerscreeningpraktijken te onderzoeken (Deel 1) en mogelijke manieren om screeningprogramma's in heel Europa verder te optimaliseren (Deel 2).

Deel 1: De effectiviteit van borstkankerscreening

In hoofdstuk 2 hebben we een systematisch onderzoek uitgevoerd om een overzicht te krijgen van de huidige gegevens over het terugdringen van borstkankersterfte door mammografisch screenen in Europa. Onze resultaten versterken eerdere bevindingen dat mammografie-screening de sterfte aan borstkanker vermindert, maar benadrukken dat dit in verschillende mate gebeurt. Voor deze systematische review hebben we zowel gerandomiseerde gecontroleerde trials (RCT's) als observationele studies, zoals prospectieve en retrospectieve gecontroleerde cohort- of case-controlstudies, meegenomen. Het strikte gebruik van algemeen erkende en transparante instrumenten om de kwaliteit van elke geïncludeerde referentie te beoordelen,-liet ons toe alleen studies-met de meest valide informatie te selecteren.-Uit deze reeks studies van hoge kwaliteit bleek dat het effect van georganiseerde screening (richtleeftijd 50 -69 jaar) op de borstkankerspecifieke sterfte varieert van 12% tot 58% bij deelnemers aan de screening versus niet-deelnemers en van 4% tot 31% bij uitgenodigde versus niet-genodigde vrouwen.

In hoofdstuk 3 hebben we ons gericht op het onderzoeken van de determinanten van voor- en nadelen van verschillende screeningsmethoden voor borstkanker (mammografie, echografie, klinisch borstonderzoek en borstzelfonderzoek) en op het samenvatten van gegevens uit systematische reviews over deze vier screeningsmethoden onder de algemene bevolking. In het algemeen vonden we dat systematische onderzoeken naar borstkankerscreening zich meer richten op mammografie dan op de andere screeningsmethoden, en dat de voordelen van screening vaker worden geëvalueerd dan de nadelen. Alle systematische onderzoeken die we hebben meegenomen, bevestigen consistent een vermindering van borstkankersterfte onder vrouwen van 50-69 jaar. Er bestaan echter belangrijke nuances tussen de studies en regio's op verschillende niveaus van geloofwaardigheid. De resultaten voor overdiagnose liepen sterk uiteen naar gelang van het type oorspronkelijk bewijs en diverse methodologische verschillen, bv. de noemer (het berekenen van overdiagnose voor vrouwen van alle leeftijden of alleen screeningsleeftijden), de duur van de follow-up of het in aanmerking nemen van ductaal carcinoma in situ (een mogelijke voorloper van borstkanker). Daarom is er nog altijd geen eenduidige conclusie te trekken over de voordelen en nadelen van mammografie.

In hoofdstuk 4 hebben we geschat dat borstkankerscreening elk jaar bijna 21.700 sterfgevallen aan borstkanker in Europa voorkomt. Kankerscreeningprogramma's kunnen alleen effectief zijn in het terugdringen van de sterfte aan borstkanker, als een groot deel van de mensen binnen de doelpopulatie een geïnformeerde keuze maakt om deel te nemen. Ons doel in hoofdstuk 4 was te onderzoeken wat het effect zou zijn van een verhoogde of zelfs volledige dekking van de borstkankerscreening op de borstkankersterfte voor elk Europees land. We ontdekten dat bij een hypothetische dekking van 100% van de screening in de geadviseerde leeftijdsgroep (d.w.z. alle in aanmerking komende vrouwen tussen 50-69 jaar worden uitgenodigd en volgen deze uitnodiging), het aantal sterfgevallen door borstkanker bij Europese vrouwen met bijna 12.500 per jaar verder zou kunnen worden teruggebracht. Het effect zou vooral in Oost-Europa opmerkelijk zijn.

Deel 2: Modelleren van het effect van verschillende ingrepen op de schade en baten van borstkankerscreening

In deel 2 van dit proefschrift evalueerden we het effect van verschillende screeninginterventies op de voor- en nadelen van borstkankerscreening in verschillende Europese landen, met behulp van het microsimulatiemodel (MISCAN-Breast).

De balans tussen de voor- en nadelen van borstkanker screening verschilt met de leeftijd. In hoofdstuk 5 hebben we de MISCAN-modellen van vier Europese landen (Nederland, Finland, Italië en Slovenië) gebruikt om na te gaan hoe de verhouding tussen voor- en nadelen zou variëren als de screening op borstkanker zou worden uitgebreid naar jongere en/of oudere leeftijdsgroepen: 45-69 jaar, 45-74 jaar en 50-74 jaar. Wij hebben deze resultaten vergeleken

met de huidige strategie waarbij vrouwen in de leeftijdsgroep 50-69 worden gescreend. We ontdekten dat in alle landen het toevoegen van screening tussen de leeftijden 45 en 49 jaar of 70 en 74 jaar resulteerde in meer gewonnen levensjaren en meer vermeden sterfgevallen door borstkanker (voordelen). Maar dat gebeurde ten koste van een toename in het aantal overdiagnoses en vals-positieve diagnoses (nadelen) De verhouding tussen kosten en baten van extra borstkankerscreening kan verbeterd worden door aanpassing van de te testen leeftijdsgroepen, maar de keuze voor deze aanpassing is afhankelijk van de prioritering van de genoemde voor-en nadelen.

Ondanks de voordelen van vroege opsporing bereiken screeningprogramma's in de praktijk vaak niet hun volledige potentieel. Daarom hebben we in hoofdstuk 6 onderzocht hoe de screeningprogramma's voor borstkanker verder geoptimaliseerd kunnen worden. Met Italië als voorbeeld hebben we eerst barrières geïdentificeerd die ook in andere landen zijn vastgesteld, namelijk een lage deelname aan screening en variaties van screeningintervallen als gevolg van opportunistische screening (wat leidt tot een screeningsinterval kleiner dan 2 jaar) en een gebrek aan middelen (wat leidt tot een screeningsinterval groter dan 2 jaar). Vervolgens zijn met behulp van het MISCAN-model mogelijke toekomstige veranderingen (d.w.z. praktische oplossingen) gesimuleerd die in de borstkankerscreeningprogramma's kunnen worden doorgevoerd om deze belemmeringen weg te nemen en de screening te verbeteren. Onze analyse toont aan dat het wegnemen van de belangrijkste barrières van de huidige Italiaanse borstkankerscreeningprogramma's tot aanzienlijke verbeteringen zou kunnen leiden en kosteneffectief of zelfs kostenbesparend zou kunnen zijn. Dit Italiaanse voorbeeld illustreert een systematische aanpak en een stapsgewijs proces dat gemakkelijk kan worden gevolgd of aangepast door andere Europese landen of belanghebbenden.

Kosteneffectiviteitsanalyses zijn waardevolle instrumenten om een optimale screeninginterventie te bepalen door deze te vergelijken met alternatieve strategieën. Screening op borstkanker beïnvloedt zowel de lengte van het leven (mortaliteit) als de kwaliteit van het leven (morbiditeit). Tot op heden is het gebruik van voor gezondheid gecorrigeerde levensjaren de beste manier om mortaliteit en morbiditeit in één enkele eenheid te combineren. Dit is een overkoepelende term voor enerzijds voor kwaliteit gecorrigeerde levensjaren (QALYs) en anderzijds voor invaliditeit gecorrigeerde levensjaren (DALYs). Om kosteneffectiviteitsanalyses effectiever te gebruiken voor het informeren van gezondheidsbeleid, evalueerden we in hoofdstuk 7 de invloed van het gebruik van QALY's versus DALY's op uitkomsten en optimale strategieën voor borstkankerscreening. We vroegen ons af of QALY's en DALY's uitwisselbare samenvattende maten van gezondheid zijn. We ontdekten dat het gebruik van gewonnen QALY's in plaats van voorkomen DALYs om de effecten van borstkankerscreening op de kwaliteit van leven in Nederland te beoordelen, verschillen in kosteneffectiviteit oplevert. Deze verschillen waren echter slechts subtiel, met geen effect op de conclusies van de kosten-effectiviteitsanalyses. Omdat de relatie tussen deze twee

maatstaven in alle analyses constant bleef, concluderen we dat beide gebruikt kunnen worden om de optimale borstkankerscreeningstrategie te bepalen.

CONCLUSIES

- Op basis van de resultaten van de in dit proefschrift beschreven studies kunnen de volgende conclusies worden getrokken:
- Er bestaan verschillende methodologisch geschikte benaderingen die in staat zijn het werkelijke gunstige effect van mammografische screening in beeld te brengen.
- De vermindering van borstkankersterfte als gevolg van borstkankerscreening bij deelnemers versus niet-deelnemers varieert in Europa tussen 12%-58%.
- Jaarlijks zijn reeds 21.680 sterfgevallen door borstkanker voorkomen als gevolg van mammografie-screening. Het aantal sterfgevallen zou in alle Europese regio's nog aanzienlijk verder kunnen worden teruggebracht (met meer dan 12.000 per jaar), in de ideale situatie dat 100% van de in aanmerking komende vrouwen in de leeftijdsgroep 50-69 jaar in Europa om de twee jaar zou worden gescreend.
- Door vrouwen vanaf 45 jaar te screenen, dus vijf jaar vóór de momenteel aanbevolen leeftijdsgrens van 50-69 jaar, zou de verhouding tussen het aantal overgediagnosticeerde borstkankergevallen en het aantal vermeden borstkankerdoden kunnen worden verbeterd.
- De belangrijkste belemmeringen voor de borstkankerscreeningprogramma's in Italië kunnen worden weggenomen door haalbare veranderingen in gang te zetten die leiden tot betere resultaten op de lange termijn. De gebruikte online instrumenten kunnen een effectieve en betrouwbare bron zijn voor Europese beleidsmakers die zich richten op geïnformeerde besluitvorming over kankerscreening in hun land.
- Voor kwaliteit gecorrigeerde levensjaren (QALY) en invaliditeit gecorrigeerde levensjaren (DALY) kunnen beide worden gebruikt om de optimale strategie voor borstkankerscreening te bepalen.

About the author

Nadine Zielonke was born on June 6th 1979 in Wismar, the former German Democratic Republic. She completed her secondary education at the Geschwister Scholl Gymnasium in Wismar in 1997. The same year, she started studying Sociology and Political Science at the University of Rostock. After her intermediate examination, she transferred to study Demography. In 2011, she studied Political Science at Umeå University, Sweden. She obtained her Diploma in Demography as the first graduate ever in Germany in 2003. For her Diploma thesis, Nadine investigated causes and motives of teenage pregnancies in Germany. Between 2003 and 2005, she worked for several research institutions and cooperate companies throughoutGermany in order to gain work experience and a sense for her professional future.

In 2005, Nadine took the leap and moved to Vienna, Austria, where she worked as a research assistant at the Vienna Institute of Demography of the Austrian Academy of Sciences. During that time, she did research on the gender gap in life expectancy, and successfully wrote a research proposal that secured funds by the Austrian Science Fund. In the same year, Nadine attended the International Max Planck Research School for Demography (IMPRSD) in Rostock, Germany. IMPRSD was Europe's largest collaborative international program of research training in population studies.

In 2007, Nadine was offered the position of Manager of the Austrian National Cancer Registry at Statistics Austria. For seven years, her responsibilities entailed leading a team of registrars and analysts, analysing and publishing data, collaborating with other cancer registers and stakeholders as well as international representation. During that time, Nadine attended the Summer School in Cancer Epidemiology at the International Agency for Research on Cancer (IARC), which is the specialized cancer agency of the World Health Organization, located in Lyon, France. After the birth of her daughter Jasmine Dunja in 2014, she moved to the Netherlands.

In 2016, Nadine began working as a scientific researcher at Erasmus MC, Department of Public Health (Rotterdam, the Netherlands). During this time, she was consortium member of EU-TOPIA (Towards improved screening for breast, cervical and colorectal cancer in all of Europe), a five year project funded by the European Commission's Horizon 2020 programme. EU-TOPIA offered Nadine the scope for her PhD research on the evaluation and optimization of breast cancer screening in Europe. She developed and validated microsimulation models for several countries to determine the long-term health outcomes, benefits, harms and costs of breast cancer screening.

The results of this research are described in this thesis.

Since April 2021, Nadine is working as a researcher at the NFK (Nederlandse Federatie van Kankerpatiëntenorganisaties), which is the Dutch Federation of Cancer Patients Organisations. NFK strives for a better quality of life, better quality of care, and better access to care for (former) cancer patients and their families.

PHD PORTFOLIO

Summary of PhD training and teaching
PhD student: Nadine Zielonke
PhD period: 2016 - 2020

Erasmus MC department: Public Health
Promotors: Prof. dr. H.J. de Koning
Co-promotor: Dr. N. T. van Ravesteyn

PhD Training	Year	Workload (ECTS)
Courses at the Netherlands Institute for Health Sciences (NIHES)		
Cancer Epidemiology	2016	1.4
Planning and Evaluation of Screening	2016	1.4
Health Economics	2016	0.7
International Comparison of Health Care Systems	2017	1.4
Advanced Topics in Decision Making in Medicine	2018	2.4
From Problem to Solution in Public Health	2018	0.9
Other Courses		
Scientific Integrity	2016	0.3
Biomedical English Writing and Communication	2017	3.0
Logframe: the basis for good project writing	2018	0.1
Employability outside academia (PCDI)	2019	1.4
Presentations		
Poster presentation ICSN	2017	2.0
Oral presentation DGD Jahrestagung Cologne	2018	0.5
Oral presentation MEMTAB	2018	0.5
Poster Presentation ICSN	2019	0.5
Oral presentations VO meetings, Department of Public Health	2016-2020	2.0
Presentations at EU-TOPIA consortium meetings	2016-2020	2.0
Oral presentation Finnish Cancer Registry	2019	0.5
Other Activities		
Seminars at the Department of Public Health (attended)	2016-2020	3.0
Workshop at ESHPM: When is it too expensive? (attended)	2017	0.3
Junior representative at the Department of Public Health (2019)	2019	5.0
Junior MISCAN user group	2019	1.1
Supervision of Master Student	2017	3.0
Peer review activities (e.g. Cancers, The Breast, Health Economics)	2020	1.5
Total		**34.9***

*1 ECTS=28Hrs

PUBLICATION LIST

In this thesis

Zielonke N, Gini A, Jansen EEL, Anttila A, Segnan N, Ponti A, Veerus P, de Koning HJ, van Ravesteyn NT, Heijnsdijk EAM; EU-TOPIA consortium. Evidence for reducing cancer-specific mortality due to screening for breast cancer in Europe: A systematic review. Eur J Cancer. 2020 Mar;127:191-206.

Mandrik O, Zielonke N, Meheus F, Severens JLH, Guha N, Herrero Acosta R, Murillo R., Systematic reviews as a 'lens of evidence': Determinants of benefits and harms of breast cancer screening. Int J Cancer. 2019 Aug 15;145(4):994-1006.

Zielonke N, Kregting LM, Heijnsdijk EAM, Veerus P, Heinävaara S, McKee M, de Kok IMCM, de Koning HJ, van Ravesteyn NT; EU-TOPIA collaborators. The potential of breast cancer screening in Europe. Int J Cancer. 2021 Jan 15;148(2):406-418.

Zielonke N, Geuzinge HA, Heijnsdijk EAM, Heinävaara S, Senore C, Jarm K, de Koning HJ, van Ravesteyn NT, on behalf of the EU-TOPIA consortium. Extending age ranges in breast cancer screening in four European countries: Model estimations of harm-to-benefit-ratios. Cancers. 2021, 13(13), 3360;

Zielonke N, Senore C, Ponti A, Csanádi M, de Koning HJ, van Ravesteyn NT, Heijnsdijk EAM, on behalf of the EU-TOPIA consortium The effect of potential future changes of organized breast cancer screening in (southern) Italy. Submitted

Davidovi M, Zielonke N, Lansdorp-Vogelaar I, Segnan N, de Koning HJ, Heijnsdijk EAM. Disability-Adjusted Life Years Averted Versus Quality-Adjusted Life Years Gained: A Model Analysis for Breast Cancer Screening. Value Health. 2021 Mar;24(3):353-360.

EU-TOPIA publications

Gini A, Jansen EEL, Zielonke N, Meester RGS, Senore C, Anttila A, Segnan N, Mlakar DN, de Koning HJ, Lansdorp-Vogelaar I, EU-TOPIA consortium. Impact of colorectal cancer screening on cancer-specific mortality in Europe: A systematic review. Eur J Cancer. 2020 Mar;127:224-235

Jansen EEL, Zielonke N, Gini A, Anttila A, Segnan N, Vokó Z, Ivanus U, McKee M, de Koning HJ, de Kok IMCM, EU-TOPIA consortium. Effect of organised cervical cancer screening on cervical cancer mortality in Europe: a systematic review. Eur J Cancer. 2020 Mar;127:207-223.

Gini A, van Ravesteyn NT, Jansen EEL, Heijnsdijk EAM, Senore C, Anttila A, Novak Mlakar D, Veerus P, Csanádi M, Zielonke N, Heinävaara S, Széles G, Segnan N, de Koning HJ, Lansdorp-Vogelaar I; EU-TOPIA consortium. The EU-TOPIA evaluation tool: An online modelling-based tool for informing breast, cervical, and colorectal cancer screening decisions in Europe. Prev Med Rep. 2021 Apr 30;22:101392.

Gini A, Buskermolen M, Senore C, Anttila A, Novak Mlakar D, Veerus P, Csanádi M, Jansen EEL, Zielonke N, Heinävaara S, Széles G, Segnan N, de Koning HJ, Lansdorp-Vogelaar I. Development and Validation of Three Regional Microsimulation Models for Predicting Colorectal Cancer Screening Benefits in Europe. MDM Policy Pract. 2021 Jan 29;6(1):2381468320984974.

Other publications

Mandrik O, Tolma E, Zielonke N, Meheus F, Ordóñez-Reyes C, Severens JL, Murillo R. Systematic reviews as a "lens of evidence": Determinants of participation in breast cancer screening. J Med Screen. 2021 Jun;28(2):70-79.

Mohamad Al-Ali B, Madersbacher S, Zielonke N, Schauer I, Waldhoer T, Haidinger G.Impact of gender on tumor stage and survival of upper urinary tract urothelial cancer : A population-based study. Wien Klin Wochenschr. 2017 Jun;129(11-12):385-390

Waldhoer T, Berger I, Haidinger G, Zielonke N, Madersbacher S. Sex Differences of ≥ pT1 Bladder Cancer Survival in Austria: A Descriptive, Long-Term, Nation-Wide Analysis Based on 27,773 Patients. Urol Int. 2015;94(4):383-9

Pinter M, Hucke F, Zielonke N, Trauner M, Sieghart W, Peck-Radosavljevic M. Epidemiological trends of hepatocellular carcinoma in Austria. Dig Dis. 2014;32(6):664-9.

Pinter M, Hucke F, Zielonke N, Waldhör T, Trauner M, Peck-Radosavljevic M, Sieghart W. Incidence and mortality trends for biliary tract cancers in Austria. Liver Int. 2014 Aug;34(7):1102-8

Woehrer A, Slavc I, Waldhoer T, Heinzl H, Zielonke N, Czech T, Benesch M, Hainfellner JA, Haberler C; Incidence of atypical teratoid/rhabdoid tumors in children: a population-based study by the Austrian Brain Tumor Registry, 1996-2006. Cancer. 2010 Dec 15;116(24):5725-32.

Wibmer C, Leithner A, Zielonke N, Sperl M, Windhager R. Increasing incidence rates of soft tissue sarcomas? A population-based epidemiologic study and literature review. Ann Oncol. 2010 May;21(5):1106-11.

Words of Appreciation

The last six years of my life have been full of learning and teachers: not only did I learn about breast cancer screening and microsimulation, but also about life and people. The path towards this thesis has been circuitous. Its completion is thanks in large parts to the special people who challenged, supported and stuck with me along the way.

To start with, I want to thank my promotor, **Harry**, for the opportunity to be part of this fantastic project. I greatly admire your knowledge and insight. I really appreciated all your critical revisions with great eye for detail as well as your quirky presentations. And thanks to you I will always remember dancing among other researches going crazy, on a boat in Rotterdam…

Nicolien, thanks for keeping me on task, for always offering me actionable steps instead of fluff advice. You have a very supportive way to convey your knowledge, which was helpful at every step of the way. I value your wittiness, humour and honesty. Thank you for seeing this through with me!

I thank you, **Eveline,** for your responsiveness, immediate feedback and exemplary pragmatism. Thanks for your patience when teaching me Chinese (aka MISCAN breast) and for always being available for corrections, discussions or motivations at any time. I appreciate you a lot!

I wish to thank all the people of the **EU-TOPIA consortium** whose assistance was a milestone in the completion of this thesis. At the CPO of Turin, Carlo Senore, Antonio Ponti and Nereo Segnan. At the Finnish cancer registry, Ahti Anttila, Sirpa Heinävaara, and Tytti Sarkeala. At the National Institute for Public Health of Slovenia, Katja Jarm, Urška Ivanuš, Dominika Novak-Mlakar and Tatjana Kofol. At the Syreon Research Institute, Marcell Csanádi, Zoltán Vokó, János Pitter and György Széles. At the London School of Health and Tropical Medicine, Martin McKee and Jennifer Priaulx. At the National Institute for Health Development of Tallinn, Piret Veerus. I looked forward to all of our consortium meetings! It has been an enriching experience for me to work, write, city walk and escape with you.

I would also like to thank the **members of the committee**. I consider myself very lucky that you were willing to invest your precious time in reading and evaluating this dissertation and being part of the defence ceremony.

I want to express my appreciation for **Valery Lemmens'** as well as **Sabine Siesling's** support to continue working on this thesis at the Integraal Kankercentrum Nederland (IKNL) in 2020.

A PhD trajectory can be an exciting and fascinating endeavour, however, quite often it is also marked by periods of concern and frustration. I was in the lucky position to have amazing **colleagues and friends at the MGZ**, colleagues from the screening section, and

especially the **breast cancer team** who shared their knowledge, doubts or happy moments with me, offered advice and encouragement in times of difficulties, and never became tired of discussing scientific and non-scientific issues. Special thanks to Valérie, Timor, Farsia, Mirjam, Arvind, Lucie, Famke, Amarens, Elleke, Lindy, Lisanne, Sarocha, Jeroen and Erhan for all the lovely lunch breaks, borrels, good conversations, outings and drinks. **Maša**, it was a pleasure to co-supervise you through your master thesis. I am very proud of you! I am grateful for my time as junior representative at the **JVO,** as I love working with people who help other people.

Dear **Na-2314 peeps**, my former roomies Dayna, Clare, Erik, Amir and Andrea. We saw each other struggle, lose motivation, and feel stressed. But we had each other's back, had good laughs, celebrated every accepted paper, practiced resilience and became good friends. Thank you for walking this piece of the way together! People like you make a difference. **Dayna** and **Clare**, we seal the deal and form the triad of 2021 defences. How very special to be each other's paranymphs!

Petra en **Arry**, I would like to recognize the invaluable assistance that you both provided during my PhD trajectory. Thank you, Petra, for many honest conversations, doses of hand cream and photo shoots.

Ik wil ook mijn nieuwe collega's van de **NFK** bedanken. Bedankt voor de gezelligheid vanaf dag één. Leuk dat ik in zo'n warm nest terecht ben gekomen. Ik ben blij bij jullie een nieuwe uitdaging gevonden te hebben.

Getting through my dissertation required more than academic support, and I have many, many people to thank for listening to and, at times, having to tolerate me over the past six years.

Here I would like to thank my friends from Germany, Austria and the Netherlands, who supported me in many different ways. Liebe **Katja**, **Wencke**, **Inga**, **Elke** und **Ani**, wir wohnen nicht mehr um die Ecke voneinander. Aber wenn wir uns sehen, ist es sofort gut und wir machen da weiter, wo wir aufgehört haben. Danke für eure Freundschaft. Ich vermisse euch sehr! Dear **Mieke, Annika**, **Eva** and **Marieke**, lieve Snoopjes, I am so IMPRSD that our friendship lasts so long. We don't see each other often, but we enjoy sharing and celebrating all of our demographic events with each other. Bedankt! **Ondine, Isabelle**, **Beatrice** and **Ana**, let's never not have the time for another ladies night. We all know how much we need it. Thank you all for the mental support, reflections and fun moments.

Liebe Familie, es war ungeheuer wichtig für mich zu hören, wie stolz ihr auf mich seid. Auch wenn ihr gar nicht immer genau wusstest, womit ich mich da so ausführlich beschäftige... Das hat mich angetrieben und bestärkt, nicht aufzugeben und die erste Frau Doktor der Familie zu werden (so müsst ihr mich ab jetzt aber nicht – unbedingt – nennen!).

Mein liebes Schneckchen, liebe **Jasmine**. Ein gutes Vorbild für dich zu sein ist mein grösster Antrieb für alles, was ich tue. Die Herausforderungen anzunehmen und diese Doktorarbeit zu Ende zu bringen hat viel damit zu tun, dass ich dich stolz machen möchte. Auf mich. Und auf dich selbst. Denn wer kann schon von sich behaupten, mit 7 Jahren das Cover eines Buches gestaltet zu haben? Ich danke dir dafür.

And to conclude, **Vinodh**, Liebling, thank you for being by my side through all of this endeavor. I thank you for the many brainstorming sessions where I fully milked your qualities as editor. And as a generally super smart man. You were there when I was at my lowest, having doubts and a tough time. And you were also there when I was having the best of time. So thank you for always having my back. I thank you for providing happy distractions to rest my mind outside of my research, for the endless jokes, and all the delicious meals. Keep them coming xxx

Thank you !
Danke!
Bedankt!

Nadine, December 2021