KCE REPORT 423
HEALTH TECHNOLOGY ASSESSMENT

Federaal Kenniscentrum voor de Gezondheidszorg
Centre Fédéral d'Expertise des Soins de Santé
Belgian Health Care Knowledge Centre

EVALUATION OF BREAST CANCER SCREENING IN BELGIUM

JANIS LUYTEN, MATTIAS NEYT, CÉCILE CAMBERLIN, CÉLIA PRIMUS-DE JONG, HELENA VAN DEYNSE, ROOS LEROY

2026 www.kce.fgov.be

COLOPHON

Title:	Evaluation of breast cancer screening in Belgium
Authors:	Janis Luyten (KCE), Mattias Neyt (KCE), Cécile Camberlin (KCE), Célia Primus-de Jong (KCE), Helena Van Deynse (KCE), Roos Leroy (KCE)
Programme manager:	Els Van Bruystegem (KCE), Imgard Vinck (KCE)
Scientific director:	Irina Cleemput (KCE)
Health information research specialist:	Nicolas Fairon (KCE)
Data extraction, transformation and upload:	Stephan Devriese (KCE)
External experts:	Martine Berlière (UCLouvain), Michel Candeur (Centre de Coordination et de Référence pour le dépistage des cancers (CCRef)), Pino Cusumano (Groupe santé CHC, Liège), Denis Danthine (CHU de Liège), Pieter De Visschere (UZ Gent), Jan Decloedt (AZ Sint-Blasius, Dendermonde), Joëlle Desreux (CHU de Liège), François Duhoux (Cliniques universitaires Saint-Luc, Bruxelles), Julie Francart (Belgian Cancer Registry (BCR)), Mathieu Goossens (Centrum voor Kankeropsporing (CvKO)), Veronica Mendez (Hôpital Universitaire de Bruxelles (H.U.B.)), Institut Jules Bordet, Bruxelles), Mirela Roman (CHU Saint-Pierre, Bruxelles), Julie Soens (Clinique Saint-Jean, Bruxelles), Chantal Van Ongeval (UZ Leuven), Ilse Vanhoutte (AZ Sint-Lucas, Gent), Liv Veldeman (UZ Gent), Hans Wildiers (UZ Leuven)
Stakeholders:	Evandro de Azambuja (Belgian Society of Medical Oncology (BSMO)), Juliette Berguet (Baob Brussels), Wissam Bou Sleiman (Association Belge des Hôpitaux), Jean-Benoît Burrion (BruPrev), Patricia Calloud de Faudeur (EnVie), Ellen Coussement (Kom op tegen Kanker), Freya Couvin (FOD Volksgezondheid, Veiligheid van de Voedselketen, Leefmilieu – SPF Santé publique, Sécurité de la Chaîne alimentaire, Environnement (FOD VVVL – SPF SPSCAE)), Sophie Denoël (Sciensano), Joke Dierckx (Departement Zorg), Emilie Dubaere (Belgian SocieTy for Radiotherapy and Oncology (BeSTRO)), Rita Heeb (NAzorg bij BORstsparende ingreep of AMputatie (Naboram)), David Hercot (Vivalis), Claire Huyghebaert (Mutualités Libres (ML)), Brigitte Marlière (Vivre comme avant), Patrick Martens (CvKO), Marta Moreira Carrico (Rijksinstituut voor ziekte- en invaliditeitsverzekering – Institut national d'assurance maladie-invalidité (RIZIV – INAMI)), Patrick Neven (Belgische Vereniging Senologie (BVS)), Mirela Roman (Europa Donna), Dirk Van Gestel (Belgian Board of Oncology (BeBO)), Martine Van Hecke (Testaankoop), Koen Van Herck (BCR), Andries Van Schuerbeek (RIZIV – INAMI), Anne Vandenbroucke (Europa Donna), Mathilde Verougstraete (Institut pour l'égalité des femmes et des hommes – Instituut voor de gelijkheid van vrouwen en mannen), Geert Villeirs (Belgian Society of Radiology (BSR)), Christine Vrancken (EnVie)
External validators:	Philippe Autier (International Prevention Research Institute (i-PRI), France), Ian Cromwell (Canada's Drug Agency – L'Agence des médicaments du Canada (CDA – AMC), Canada), Geertruida de Bock (Rijksuniversiteit Groningen & Universitair Medisch Centrum Groningen (UMCG), the Netherlands)
Acknowledgements:	Carlos Canelo (Institut de Recerca Sant Pau, Barcelona, Spain), Laurens Carp (Universitair Ziekenhuis Antwerpen (UZA)), Cecile Colpaert (AZ Turnhout), Sabine Corachan (formerly KCE), Cindy De Gendt (BCR), Jonathan De Ro (BCR), Robin Delvaux (BCR), Jens Detollenaere (KCE), Sophie Gerkens (KCE), Peter C. Gøtzsche (Institute for Scientific Freedom, Copenhagen, Denmark), Jillian T. Henderson (Kaiser Permanente, Center for Health Research, Portland, USA), Max Lelie (Vrije Universiteit Brussel (VUB)), Gian Paolo Morgano (European Commission, Joint Research Centre (JRC), Ispra, Italy), Sien Ombelet (formerly KCE), Koen Putman (VUB), Lenneke Sicking (Departement Zorg), Inge Truyen (BCR), Nancy Van Damme (BCR), Liselot Van De Walle (formerly

	Departement Zorg), Walli Van Doren (RIZIV – INAMI), Bart Van Gool (BCR), Isabelle Vanden Bempt (UZ Leuven), Guillaume Vanhulle (Instituut voor de gelijkheid van vrouwen en mannen – Institut pour l'égalité des femmes et des hommes), Julie Verbeeck (BCR), Leen Verleye (KCE), Raf Verscuren (Jessa Ziekenhuis, Hasselt; UZ Leuven; RZ Heilig Hart Tienen)
Reported interests KCE members:	KCE is a federal institution funded by INAMI/RIZIV, by the Federal Public Service of Health, Food Chain Safety and Environment, and by the Federal Public Service of Social Security. KCE's mission is to advise policymakers on decisions relating to health care and health insurance on the basis of scientific and objective research. It is expected to identify and shed light on the best possible solutions, in the context of an accessible, high-quality health care system with due regard for growing demand and budgetary constraints. KCE has no interest in companies (commercial or non-commercial, i.e. hospitals and universities), associations (e.g. professional associations, unions), individuals or organisations (e.g. lobby groups) that could be positively or negatively affected (financially or in any other way) by the implementation of the recommendations. All experts involved in the writing of the report or the peer-review process completed a declaration of interest form. Information on potential conflicts of interest is published in the colophon of this report. All members of the KCE Expert Team make yearly declarations of interest and further details of these are available upon request.
Reported interests:	All experts and stakeholders consulted within this report were selected because of their involvement in the topic of breast cancer screening. Therefore, by definition, each of them might have a certain degree of conflict of interest to the main topic of this report.
AI-disclaimer:	In the development of this work, the authors utilised Microsoft Copilot 365 and DeepL to assist with organising the content of the text, occasionally draft sentences starting from the author's original ideas and modify text for clarity. Following its use, the content was carefully reviewed and revised by the authors, who accept full responsibility for the final version of the manuscript.
Secretarial support:	Andrée Mangin (KCE)
Layout:	Ine Verhulst
Disclaimer:	• **Several external experts (cf. supra) were consulted about a (preliminary) version of the scientific report. They were selected based on their expertise in the topic of this report; potential conflicts of interest were considered in this process. Their comments were discussed during meetings and/or via email. They did not co-author the scientific report and do not necessarily agree with its content.** • **Subsequently, the KCE research team invited three independent experts ('validators', cf. supra), with no prior involvement in the study, to conduct a peer review of the final version of the scientific report. They were selected based on their expertise in the topic of this report; potential conflicts of interest were considered in this process. The validators decided whether or not the report was scientifically valid. They did not co-author the scientific report and did not necessarily all three agree with its content. Their name and affiliation are included in the colophon.** • **Finally, this report has been approved by a majority of votes by the Executive Board.** • **KCE is exclusively responsible for the policy recommendations. Any (potential) persisting errors or omissions are the sole responsibility of KCE.**
Publication date:	09 July 2026
Project number:	2024-11-HTA

Domain:	Health Technology Assessment (HTA)
MeSH:	Breast Neoplasms; Mammography; Mass Screening; Early Detection of Cancer; Cost-Effectiveness Analysis; Overdiagnosis
NLM Classification:	WP 870 - Breast Neoplasms
Language:	English
Format:	Adobe® PDF™ (A4)
Legal depot:	D/2026/10.273/46
ISSN:	2466-6459

How to refer to this document?	Luyten J, Neyt M, Camberlin C, Primus-de Jong C, Van Deynse H, Leroy R. Evaluation of breast cancer screening in Belgium. Health Technology Assessment (HTA) Brussels: Belgian Health Care Knowledge Centre (KCE). 2026. KCE Reports 423. https://doi.org/10.57598/R423C.

This document is available on the website of the Belgian Health Care Knowledge Centre.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABBREVIATION	DEFINITION
AAPC	Average annual percent changes
ABUS	Automated breast ultrasound
ACR	American College of Radiology
ACS	American Cancer Society
AMCP	Academy of Managed Care Pharmacy
ASR	Age-standardized rate
ADH	Atypical ductal hyperplasia
AHRQ	Agency for Healthcare Research and Quality
ALH	Atypical lobular hyperplasia
AMSTAR	A MeaSurement Tool to Assess systematic Reviews
ATC	Anatomical therapeutic chemical
aOR	Adjusted odds ratio
BC	Breast cancer
BCR	Belgian Cancer Registry
BCSC	Breast Cancer Surveillance Consortium
BFM	Budget of financial means
BI-RADS	Breast Imaging Reporting and Data System
BMI	Body mass index
BRCA	Breast cancer gene
BSS	BreastScreen Singapore
BruPrev	Brussels screening organisation
CA$	Canadian Dollar
CBPL	Commissie voor de bescherming van de persoonlijke levenssfeer
CBA	Cost-benefit analysis
CBE	Clinical breast examination
CCA	Cost-consequence analysis
CCRef	Walloon screening organisation (‘Centre de coordination et de Référence’)
CCS	Charlson Comorbidity Score
CE	Cost-effectiveness
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CEM	Contrast-enhanced mammography
CESM	Contrast-Enhanced Spectral Mammography

CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CHP	Cyto-histopathology
CI	Confidence interval
CIR	Cumulative indidence rate
CMA	Cost-minimisation analysis
CMS	Centers for Medicare & Medicaid Services
CNBSS	Canadian National Breast Screening Studies
COI	Conflict of interest
CPAP	Canadian Patnership against Cancer
CPVP	Commission de la protection de la vie privée
cStage	Clinical stage
CT	Computerized tomography scan
cTNM	Clinical tumour node metastasis staging
CUA	Cost-utility analysis
CvKO	Flemish screening organisation (‘Centrum voor Kankeropsporing’)
DBT	Digital breast tomosynthesis
DD	Screning follow-up database
DCIS	Ductal carcinoma in situ
DENSE	Dense Tissue and Early Breast Neoplasm Screening trial
DKK	Danish Krone
DM	Digital mammo
EBSP	Edinburgh Randomised Breast Screening Project
ECIBC	European Commission Initiative on Breast Cancer
EQ-5D	EuroQOL five dimensions questionnaire
ER	Oestrogen receptor
ESP	European standard population
EU	European Union
EU-27	27 member states of the European Union (EU)
EUnetHTA	European network for Health Technology Assessment
EUNICE	European Network for Information on Cancer
EUSOMA	European Society of Breast Cancer Specialists
FaDia	Fatal diameter
FDA	Food and Drug Administration
FFDM	Full-field digital mammography
FN	False negative

FOD VVVL – SPF SPSCAE	Federal Public Service Health, Food Chain Safety and Environment (‘Federale overheidsdienst Volksgezondheid, Veiligheid van de Voedselketen en Leefmilieu’/’Service public fédéral Santé publique, Sécurité de la Chaîne Alimentaire et Environnement’)
FP	False positive
FU	Follow-up
FUD	First follow-up date
GDG	Guidelines Development Group
GP	General practitioner
GRADE	Grading of Recommendations, Assessment, Development, and Evaluations
HER2	Human epidermal growth factor receptor 2
HIP	Health Insurance Plan
HQO	Health Quality Ontario
HR	Hazard ratio
HRQoL	Health-related quality of life
HTA	Health technology assessment
IBC	Invasive breast cancer
ICD(-0)	International Classification of Diseases (for Oncology)
ICER	Incremental Cost-Effectiveness Ratio
ICMD	International Consortium on Mammographic Density
ICTRP	International Clinical Trials Registry Platform
ID	Incidence data breast cancer
IDR	Incidence density ratio
IE	Incremental effect
IKNL	Integraal Kankercentrum Nederland
IMA – AIM	Intermutualistic Agency (‘Intermutualistisch Agentschap’/‘Agence Intermutualiste’)
INAHTA	International Network of Agencies for Health Technology Assessment
INSZ - NISS	Social security number (‘Identificatienummer van de Sociale Zekerheid’ – ‘Numéro d'Identification de la Sécurité Sociale’)
IRR	Incidence rate ratio
IQR	Interquartile range
IQWiG	Institute for Quality and Efficiency in Health Care
ISPOR	The Professional Society for Health Economics and Outcomes Research
JBI	Joanna Briggs Institute
J-START	Japan Strategic Anti-cancer Randomised Trial

KCE	Belgian Health Care Knowledge Centre
KSZ-BCSS	Crossroads Bank for Social Security ('Kruispuntbank van de Sociale Zekerheid' – 'Banque Carrefour de la Sécurité Sociale')
LCIS	Lobular carcinoma in situ
LYG	Life year(s) gained
LYS	Life year(s) saved
MEA	Managed Entry Agreement
MeSH	Medical Subjects Headings
MISCAN	MIcro-simulation SCreening ANalysis
MRI	Magnetic resonance imaging
MyPeBS	My Personal Breast Cancer Screening trial
NA	Not available/not applicable
NBCF	National Breast Cancer Foundation
NCR	Netherlands Cancer Registry
ND	Not described
NHMRC	National Health and Medical Research Council
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHDI	National Institute for Health and Disability Insurance (RIZIV – INAMI)
NOK	Norwegian Krone
NPC	National Pharmaceutical Council
NR	Not reported
NRDO	National Registry of Diseases Office
NRSI	Nonrandomized study of intervention
NUH PASS	National University Hospital Patient Affordability Simulation System
OECD	Organisation for Economic Co-operation and Development
OR	Odds ratio
PET	Positron emission tomography
PICO	Population, interventions, comparator, outcomes
PPV	Positive predictive value
PR	Prevalence ratio
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
pStage	Pathological stage
pTNM	Pathological tumour node metastasis staging
QA	Quality assurance
QALY	Quality-adjusted life year

QoL	Quality of Life
QUADAS	Quality of primary diagnostic accuracy studies
RCT	Randomised controlled trial
RE Tomo	Reggio Emilia Tomosynthesis trial
RESR	Revised European Standardized Rate
RIZIV – INAMI	National Institute for Health and Disability Insurance ('Rijksinstituut voor Ziekte- en Invaliditeitsverzekering'/'Institut National d'Assurance Maladie-Invalidité') (NIHDI)
RoB	Risk of bias
ROBINS	Risk of bias in non-randomised studies of interventions
RQ	Research question
RR	Relative risk, risk ratio
RT	Radiotherapy
SA	Sensitivity analysis
SCL-90	Symptom Checklist-90
SD	Screening database
SEER	Surveillance, Epidemiology, and End Results Program
SGP	Singaporean Dollar
SiMRiSc	Simulation Model on Radiation Risk and Breast Cancer Screening
SM	Synthesis mammography
SMR	Standardised mortality ratio
SPECT	Single-photon emission computed tomography
SR	Systematic review
STAI-6	Spielberger State-Trait Anxiety Inventory state scale
STATbel	Directorate General Statistics – Statistics Belgium
TNM	Tumour node metastasis staging
TOSYMA	Tomosynthesis plus synthesised mammography trial
TP	True positive
TTO	Time trade-off
UICC	Union for International Cancer Control
UKCCCR	United Kingdom Co-ordinating Committee on Cancer Research
US$	United States Dollar
USPSTF	U.S. Preventive services Task Force
VABB	Vacuum-assisted breast biopsy
VAS	Visual analogue scale
VIKZ	Flemish Institute for Care Quality ('Vlaams Instituut voor Kwaliteit van Zorg')

WHO	World Health Organization
WISDOM	Women Informed to Screen Depending on Measures of Risk
WTP	Willingness to pay
YOLS	Years of life saved

SCIENTIFIC REPORT

1 INTRODUCTION

1.1 Breast cancer screening

Breast cancer is the most frequently diagnosed cancer among women in Belgium and other high-income countries, as well as one of the leading causes of cancer-related mortality.[1] Therefore, population breast cancer screening programmes have been initiated with the aim of detecting breast cancer in an earlier stage and conceivably reducing breast cancer mortality.

In Belgium, the breast cancer screening programme was installed in 2001 in the Flemish region (northern region) and in 2002 in Brussels (capital region) and the Walloon region (southern region). From the calendar year a woman turns 50 years old until age 69, she is biennially invited to have a screening mammography. The procedure is free of charge for participants.

Recently updated international guidelines recommend to widen the age eligibility criteria. To name one, the Guidelines Development Group of the European Commission Initiative on Breast Cancer (ECIBC) suggests extending mammography screening to women aged 45 to 49 (Conditional recommendation, Moderate certainty of the evidence) and to women aged 70 to 74 (Conditional recommendation, Moderate certainty of the evidence).[2]

In parallel to the regional screening programmes, women within and beyond the targeted age group, are also offered 'opportunistic' breast cancer screening, which may not only comprise a mammography but also ultrasound of the breasts. While mammograms performed within the frame of the screening programmes can easily be identified in the administrative databases using the dedicated invoicing codes, opportunistic screening mammograms (and ultrasounds) in asymptomatic women cannot be distinguished from diagnostic mammograms (and ultrasounds) in symptomatic women. As a consequence the magnitude of 'opportunistic' breast cancer screening in Belgium cannot be estimated.

This report evaluates the balance between benefits and harms as well as the cost-effectiveness of the currently implemented breast cancer screening programmes, a potential extension to younger and/or older women, various screening intervals, and the combination of mammography and ultrasound as a screening tool.

1.2 Research questions and objectives

Given the above mentioned challenges related to breast cancer screening, this report addresses the following research questions:

1. What is the clinical effectiveness and cost-effectiveness of invitation to breast cancer screening with mammography in a population of adult women
   a. aged 50-69 years (i.e. current screening programme in all Belgian regions)
   b. aged 40-49 years
   c. aged ≥70 years

   without suspected breast cancer and without a specifically increased risk of breast cancer?
2. What is the clinical effectiveness and cost-effectiveness of invitation to
   a. biennial vs. annual, and
   b. biennial vs. triennial

breast cancer screening with mammography in a population of adult women without suspected breast cancer and without a specifically increased risk of breast cancer?

3. What is the clinical effectiveness and cost-effectiveness of breast cancer screening using mammography combined with ultrasound compared to mammography alone in a population of adult women without suspected breast cancer and without a specifically increased risk of breast cancer?

Ultimately, the study was undertaken to inform Belgian policy makers on whether or not to adapt the current breast cancer screening programme, based on the best available evidence.

1.3 Target audience

This report is primarily intended for Belgian policymakers responsible for public health and health care, both at the federal and regional level. In the second place, the report will be of interest to healthcare professionals, who are involved in breast cancer screening and/or who provide care for patients affected by breast cancer.

We acknowledge that the information needs of some other stakeholders, more precisely patient associations and individual women, may not be fully addressed in this report. Yet, based on the most important results, decision aids will be drafted with the aim of informing eligible women on the benefits and harms resulting from participating to breast cancer screening.

1.4 Disclaimer

Throughout this report we use the terms ‘patients with (suspected) breast cancer’ alongside ‘women with (suspected) breast cancer’. The KCE authors acknowledge there are adults assigned female at birth who are transgender, non-binary, or who do not identify as a woman or with the sex they were assigned at birth. By using these terms we do not intend to exclude anybody or discriminate against any group.

1.5 Methodology

To provide an answer to the research questions listed in Section 1.2, the following methods were used:

- A systematic review of the clinical effectiveness and safety of breast cancer screening in each of the three age categories, the various screening intervals and the two imaging modalities (Chapter 3);
- A systematic review of the cost-effectiveness and cost-utility of breast cancer screening in each of the three age categories, the various screening intervals and the two imaging modalities (Chapter 4);
- The analysis of a dedicated linked database that was developed by the Belgian Cancer Registry (BCR), and which comprised breast cancer tumour data (source: BCR), healthcare insurance data (source: the Intermutualistic Agency (IMA-AIM)), Vital status data (source: Crossroads Bank for Social Security) and breast cancer screening data (source: the regional screening organizations, i.e. CvKO, BruPrev, CCRef; Chapter 5);
- A primary Belgian cost-effectiveness and cost-utility analysis of breast cancer screening (Chapter 6), based on the information identified in Chapters 3 and 4, and Belgian real-world data as described in Chapter 5;
- A targeted literature review on ethical considerations and principles related to breast cancer screening (Chapter 7);
- The consultation of stakeholders, through:

- Two face-to-face meetings and, where indicated, email communication in the course of the project with medical professionals and experts in breast cancer screening (hereinafter called 'the (clinical) experts'; see the colophon). The clinical experts all have profound expertise and experience in the diagnosis and treatment of patients with breast cancer. They work in academic and non-academic centres, geographically spread over the country, and know the Belgian context (e.g. fees, invoicing rules) very well.
- Face-to-face and online consultations with stakeholders, including patient representatives, regional and federal health authorities, screening organizations, representatives of health care providers, etc. (hereinafter called 'the stakeholders'; see the colophon). During the stakeholder meeting, preliminary results were presented and feedback was collected from all involved parties before finalizing the policy recommendations.

- A review of this report by three independent scientific experts (hereinafter called 'the validators'; see the colophon).

This health technology assessment (HTA) is performed in line with the procedures available in the KCE Process Book as well as the third edition of the KCE guidelines for economic evaluations.[3]

1.6 Structure of the report

In **chapter 2** we provide the reader with concise information on breast cancer in Belgium and discuss at greater length (breast) cancer screening.

Chapter 3 provides the systematic literature review on the clinical harms and benefits of breast cancer screening with mammography, with mammography and ultrasound and on various screening intervals.

Chapter 4 is dedicated to the systematic literature review on the cost-effectiveness of breast cancer screening with mammography, mammography and ultrasound and various screening intervals.

In **chapter 5** we describe the cohort that was selected for this study: the patient and tumour characteristics, the observed survival and the various cost groups (based on the available administrative databases).

In **chapter 6** we conduct a hypothetical primary economic evaluation of the currently organised breast cancer screening programmes in Belgium using the evidence and information gathered in the previous chapters.

In **chapter 7** the ethical aspects that should be considered in breast cancer screening are discussed.

In **chapter 8** we conclude with a discussion of the most important findings of this report.

1.7 Out of scope

The following topics were considered out of scope for the present study: budget impact and impact on the health system of a potential extension of the current screening programme to younger and/or older age groups, the potential increase or decrease of the screening frequency, and/or the potential adding of ultrasound to the screening programme. In the present study we do not elaborate on the barriers and facilitators to (breast cancer) screening, nor on the initiatives to increase the uptake of (breast cancer) screening in, among others, difficult to reach groups.

While in chapter 2 various breast imaging techniques and modalities are briefly discussed, the comparative evaluation of their effectiveness was not assessed in this study. In chapter 3 the clinical harms and benefits of breast cancer screening with mammography and ultrasound are discussed. Where available, results are separately presented for women with dense and non-dense breasts. However, we opted to not elaborate on (supplemental) imaging modalities for breast cancer screening in women with dense breast tissue in this report, as we are awaiting the results of the DENSE2 trial (see Appendix 1, Box 2).

The current approach for breast cancer screening, which is largely based on age and fixed screening intervals, dates back to an era with limited risk stratification.[4] Given the heterogeneity of breast cancer and the imbalanced distribution of risk factors across the population, several RCTs have been set up to assess more tailored, risk-based screening and prevention strategies (e.g. the MyPeBS study in Europe, the WISDOM study in the USA, the Tailored Breast Screening Trial (TBST) trial in Italy; see also Appendix 1, Box 3).[5, 6] The underlying rationale is intensifying screening of those at high risk, while reducing screening among low-risk individuals who are unlikely to benefit from screening, but who are nevertheless subject to the harms of screening (e.g. overdiagnosis, false positives, unnecessary biopsies and overtreatment; see chapter 2 section 2.2.3). The topic was considered out of scope for this report, as the full results of these trials were not available yet at the time this report was concluded.

Finally, while artificial intelligence (AI) is increasingly being integrated into healthcare systems, including (breast) cancer screening, it was considered out of scope for the present report. Despite promising improvements in several outcomes, issues with overdiagnosis, false positives and demographic variations in accuracy have highlighted the need for further research in this domain.[7, 8]

1.8 Position of this report relative to prior KCE reports

This report is an update of and hence replaces KCE report 11 (breast cancer screening in general; 2005),[9] KCE report 129 (breast cancer screening in women younger than fifty years; 2010),[10] KCE report 176 (breast cancer screening in women over seventy years; 2012),[11] and KCE report 216 (decision aids in breast cancer screening; 2014).[12] The present report differs from the previous in several ways, e.g. updated mortality data of two low risk of bias RCTs have been included, additional outcomes (e.g. overdiagnosis) have been assessed, GRADE assessments and Summary of Findings tables have been included, cost-effectiveness was assessed and the screening frequency as well as addition of ultrasound have been evaluated.

2 BACKGROUND

2.1 Breast cancer in Belgium

2.1.1 Incidence

Breast cancer is not a single disease, but rather a collection of distinct histological and molecular entities,[13] with some cases growing slowly or not at all while other cases grow rapidly and aggressively.[14] As a consequence, screening and treatment have different efficacies.[15]

Breast cancer is the **most commonly diagnosed cancer in women**, while it is a rare cancer in men.[1] In 2023, a total of 11 533 invasive breast cancers were diagnosed in women in Belgium.[16] The lifetime risk (up to 84 years) to be diagnosed with breast cancer was 15.4% in women, while it was 'only' 0.2% in men.[16]

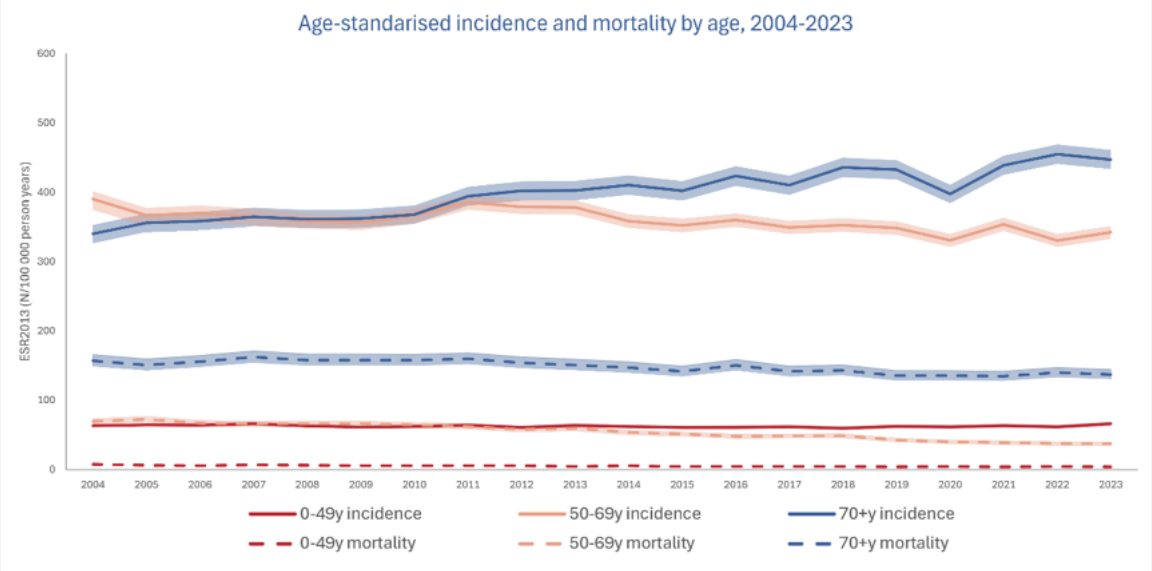
Data collected by cancer registries from 21 European countries, including Belgium, indicated that **breast cancer incidence** increased from 1978 to 2019, with the largest increases observed in 1978–1987 and 1988–1997 (with average annual percent changes (AAPC[a]) of up to 4.55 (95% confidence interval (CI): 2.56–6.57).[17] Following screening implementation, substantial increases in incidence of

[a] The Average Annual Percentage Change (AAPC) is the mean relative change in incidence risk between year x and year x+1. An AAPC equal to 1.07 (which indicates a 7% increase), applied to a cancer risk of 100 per 100 000 women in year x, results in a risk of 107 per 100 000 one year later.

in situ[b] and stage I cancer were observed (with AAPCs up to +10 in the first ten years of screening), while the incidence of stage IV cancer declined or remained stable in most countries (with AAPCs down to −6.16, 95% CI: −8.28 to −4.00). These trends were particularly pronounced among age groups targeted by screening (mostly 50–69 years).[17] However, the results of this study have been questioned, because – among others - trend analyses were based on data restricted to six to ten years after screening introduction and missing data may in part explain the drops in incidence rates before or after screening introduction.[18] Moreover, staging procedures have changed over time, making interpretation of stage shifts across different time periods hard. As regards **breast cancer mortality**, the data from the European cancer registries indicated that rates decreased most pronounced in 1998–2007 and 2008–2018/19 (with AAPCs down to −5.40, 95% CI: −9.70 to −0.89). During the first ten years after initiation of screening programmes, age-standardised mortality rates declined significantly in most countries (with AAPCs from −2.98 (95% CI: −4.63, −1.29) to −0.81 (95% CI: −1.29, −0.32)), yet, in several countries, statistically significant decreases in age-standardised breast cancer mortality were already recorded before screening implementation (AAPCs from −2.54 (95% CI: −3.15, −1.93) to −0.85 (95% CI: −1.66, −0.03)). Of note, in countries with later initiation of screening programmes, no marked shifts in mortality trends between the pre- and post-screening periods were evident.[17]

In Belgium, incidence rates for female invasive breast cancer remained stable between 2004 and 2023 (AAPC: 0.1, 95% CI: -0.1-0.3). As is illustrated in Figure 1, the incidence remained stable in the same period for the youngest age group (up to 49 years), while it decreased in the age group 50-69 years, i.e. the group invited for population screening (AAPC: -0.6, 95% CI: -0.8- -0.3) and increased in the oldest age group (AAPC: 1.4, 95% CI: 1.1-1.6).[19]

Figure 1 – Age-standardised incidence and mortality by age, 2004-2023



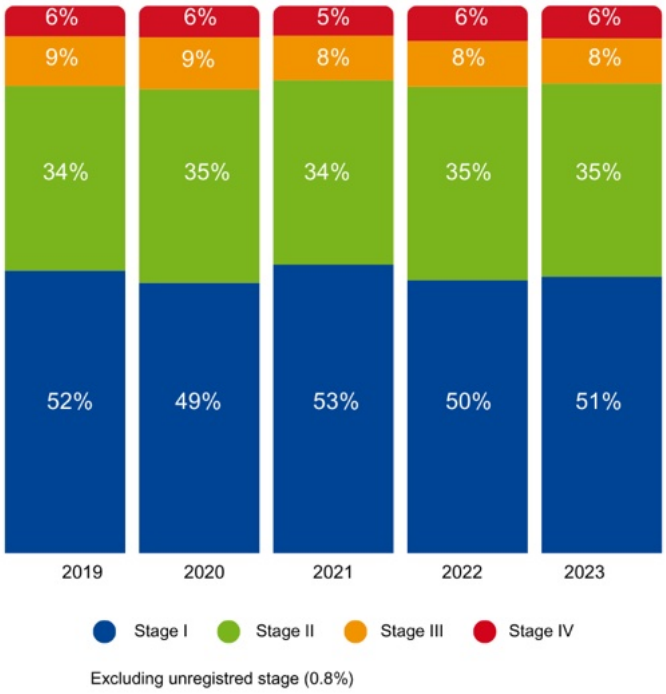
***Notes**: ESR2013: Incidence rate standardised to the 2013 revised European Standard Population (ESP).*
***Source**: Belgian Cancer Registry, 2025,[16] and 2026 (mortality data for 2022 and 2023; personal communication)*

b The authors note that stage-specific incidence trends should be interpreted with caution, especially in countries with a high proportion of unknown stage, but also as a stage migration (from stage II to III) might have occurred over time, especially around 2003, due to reclassification of tumours with more than three positive lymph nodes as stage III instead of stage II.[17] The latter was noticed for the Netherlands, Belgium (Flanders), and Germany among 0–49-year-olds. In addition, the increased incidence of in situ cancers may also partly reflect increased efforts to collect these data.[17.]

In 2023, 44.9% of all female breast cancers were diagnosed in the age group 50-69 years.[16] The median age of women diagnosed with breast cancer was 64 years.

Overall, 84.7% of all female breast cancers were diagnosed at clinical stage I or II, after exclusion of 'unregistered cases'[c].[19] In the age group 50-69 years, there were relatively more diagnoses at stage I, which has been attributed to breast cancer screening.[16] Over the last years, there was little variability in the 'combined stage'[d] distribution at diagnosis in the age group invited for screening (Figure 2).

Figure 2 – Invasive breast cancer – Stage* distribution over time in women aged 50-69 years



***Notes**: *: The stage refers to a 'combined stage', where known pathological stage prevailed over known clinical stage, except when there was clinical proof of distant metastasis (cM1). Note that these data are restricted to invasive breast cancer, i.e. excluding stage 0 (i.e. non-invasive breast cancer).*
***Source**: Belgian Cancer Registry, 2025[16]*

2.1.2 Mortality and net survival probability

In Europe, breast cancer is the most common cancer cause of death among women. In the EU-27 countries, 16.7% of all cancer deaths in women were due to breast cancer.[1] In Belgium, the most frequent cancer related cause of death is lung cancer, followed by breast cancer. In addition, breast cancer is the most important cause of premature death[e] in women.[21] In 2021, 2 048 women and 21 men died due to breast cancer.[16]

---

[c] Stage is reported according to the TNM Classification of Malignant Tumours, 8th edition.[20] If stage is unknown, not applicable or not submitted to the Belgian Cancer Registry, the stage is reported as 'unregistered stage'.[19]

[d] Combined stage: pTNM prevails over cTNM, unless cM1 or in case of neo-adjuvant therapy.

[e] Defined as the conditions with the highest burden in term of years of life lost before 75 years.

Over the past 20 years, the 5-year net survival probability[f] for women diagnosed with invasive breast cancer improved.[16] For women diagnosed in 2014-2023, the 5- and 10-year net survival was 91.7% and 86.4%, respectively (Table 1). In males diagnosed with breast cancer in the same period, the net survival probability was 88.2% (95% CI: 83.6 – 93.1) after 5 years and 81.4% (95% CI: 68.0-97.4[g]) after 10 years.[19]

Stage at diagnosis is inversely related to net survival. For women diagnosed at stage I, the net survival exceeded 100%, which indicates that the observed survival probability for this subgroup is higher than the one for the matched general population. Women diagnosed with breast cancer after age 70 years have a significantly lower 1-, 3-, 5- and 10-year net survival probability compared to younger women (data not presented).[19]

Table 1 – Net survival probability for women diagnosed with invasive breast cancer, 2014-2023

Stage*	N at risk	5 year (95%CI)	10 year (95%CI)
Total	108 604	91.7% (91.4%-92.1%)	86.4% (85.4%-87.5%)
I	46 448	100.9% (100.6%-101.3%)	100.6% (99.1%-102.1%)
II	38 188	93.6% (93.0%-94.2%)	86.8% (84.7%-89.0%)
III	7 348	76.9% (75.2%-78.6%)	64.0% (60.5%-67.8%)
IV	7 435	42.3% (40.8%-43.8%)	20.4% (17.5%-23.6%)
Unregistered cases	9 204	88.7% (87.5%-89.8%)	82.9% (80.4%-85.4%)

***Notes**: *: The stage refers to a 'combined stage', where known pathological stage prevailed over known clinical stage, except when there was clinical proof of distant metastasis (cM1). Note that these data are restricted to invasive breast cancer, i.e. excluding stage 0 (i.e. non-invasive breast cancer).*
***Source**: Belgian Cancer Registry, 2025[19]*

2.2 Cancer screening

2.2.1 Principles of screening

Screening should **not** be considered a type of **primary prevention**: it does not prevent the disease from occurring, but aims to interrupt the progression from early to mid-stage disease, mainly by early detection and prompt treatment.[23]

Cancer screening programs have as main objective to **decrease cancer-specific mortality**. This goal is achieved through the detection of cancers at an early stage of their progression, before they become symptomatic. The direct consequence of effective cancer screening is the decrease in the incidence of advanced stage cancers which will lead to reductions in cancer-specific mortality. The implicit assumption is that early detection, i.e. before the development of clinical symptoms, will lead to a more favourable prognosis because treatment initiated at that stage will be more effective and resulting in less morbidity than provided at a later stage of the disease.[24-28] With respect to breast cancer, this may include for instance the possibility for a breast conserving surgical approach rather than a

---

[f] In Belgium the data on the cause of death are not reliable, therefore breast cancer specific survival is not computed. Alternatively, the relative survival framework can be applied to estimate the net survival, i.e. the survival probability when other causes of death beside the cancer type under study are excluded. The net survival was estimated with the Pohar-Perme method[22] using the regional lifetables of 2023 (stratified on gender, age, calendar year and region) obtained from Statbel. Net survival may exceed 100%, when the observed survival probability for patients with the studied cancer type is higher than the one for the matched general population (i.e. no excess mortality due to cancer).[19]

[g] The net survival estimates are based on 'only' 1017 men with breast cancer, which explains the much wider confidence intervals, compared to those presented in Table 1.

mastectomy,[29] the possibility to avoid chemotherapy,[30, 31] or radiation therapy,[32] and less need to perform axillary lymph node dissection.[33]

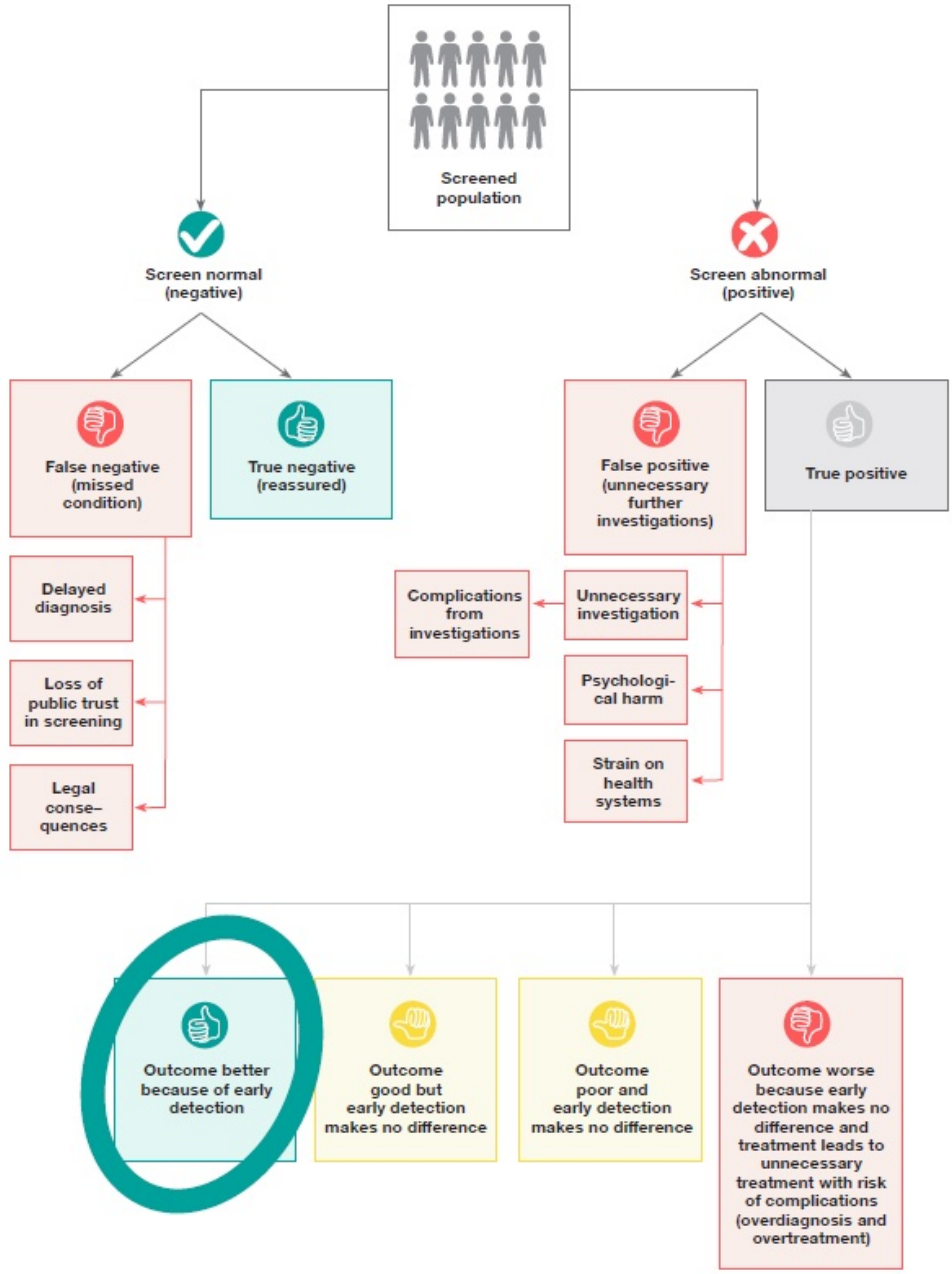
Importantly, screening is not the same as early diagnosis. While screening invites a-symptomatic people to undergo testing, early diagnosis is intended to detect conditions as early as possible among individuals with symptoms.[34]

The World Health Organization (WHO) developed in 1968 criteria for assessing the appropriateness of implementing a screening programme (Appendix 2).[35] These criteria have since then been adapted and amended, especially with the advent of genetic screening.[36] Yet, the **essential criteria for screening** remain:[23, 35, 36]

- The availability of an effective and reliable screening method;
- The availability of an intervention to reduce or improve the outcome;
- The safety and acceptability of the test to the individual;
- The overall benefits of the screening should outweigh its potential harms to the individual or the population;
- An economic evaluation of the screening program, using a health system or societal perspective, should be conducted to assess the full costs and effects of implementing, operating and sustaining the screening program.

Ideally, a screening method (e.g. genetic test, ultrasound scan) should be **easily administered**, **inexpensive** and should impose **minimal discomfort** on participants, given that a large fraction of those screened may not be at risk of the adverse outcome.[23] In addition, the screening method must correctly identify those at risk, and this should be evidence-based.

Figure 3 – Possible outcomes from a screening programme



***Source**: World Health Organization (WHO) - Screening programmes: a short guide, 2020*[34]

The **validity** of a screening method (or diagnostic test) is assessed through the calculation of its

- **Sensitivity**, the proportion of those who have the adverse outcome (disease) who are correctly identified, i.e. the 'true positives' (Figure 3);
- **Specificity**, the proportion of those who do not have the adverse outcome who are correctly identified, i.e. the 'true negatives'.

Ideally, a screening method would have both a high sensitivity and a high specificity. A low specificity, for instance, implies that the screening programme would identify many people without the outcome as having the outcome ('false positives', see section 2.2.3). This may result in unnecessary investigations (which may lead to complications), psychological harm and strain on the health system

(Figure 3).[34] A low sensitivity of the screening method will result in many 'false negatives': individuals who have the outcome would be informed that they do not have the outcome.

Yet, even a true positive result does not always mean that early detection will improve the outcome. For some screened individuals, early detection will not make any difference in the outcome from treatment. Even more, some may be harmed because the true positive screening results is in fact overdiagnosis (see section 2.2.3).[34]

Not only the validity of a screening method should be assessed, but also the ethical consequence, at both the individual and population level.[23] This is further explored in chapter 7.

2.2.2 Measuring the effectiveness of a (cancer) screening programme

Measuring the performance of a screening programmes is complex.[34] It requires the selection of **appropriate outcome measures**, which should be linked to the objectives of the programme. Some of these outcomes merit some additional thoughts:

- **Incidence**: logically, at the start of a screening programme a rise in the incidence of (early-stage) cancer is expected, which is followed by a decrease in advanced-stage cancer over time, which will result in a decline of cancer-related death.
- **Cancer stage at diagnosis**: changes in the proportion of early-stage cancer should be interpreted with caution, as these may (in part) be explained by overdiagnosis (cf. infra). Yet, incidence trends of advanced-stage cancer have been proposed as an appropriate measure to evaluate the effect of a cancer screening programme, because this outcome is independent of treatment.[37, 38]
- **Cancer-specific and all-cause mortality**: in randomised controlled trials (RCTs) on screening effectiveness, cancer-specific mortality is often selected as the primary outcome as the aim of screening is the prevention of premature death caused by cancer.[39] However, the validity of this end-point is based on the assumption that the cause of death can be assessed accurately, an assumption which has been challenged by several studies on the accuracy of death certificates (e.g. due to lack of blinded outcome assessment, sticky diagnosis bias[h], slippery linkage bias[i]).[14, 40] Moreover, cancer-specific mortality does not take deaths resulting from the screening programme (e.g. radiation-induced breast cancers, deaths due to complications after surgery, overdiagnosis leading to overtreatment and possibly death) into consideration, nor does it take potential adverse consequences from a cancer diagnosis (e.g. increased cardiovascular mortality and risk of suicide) into account.[14, 41-43] Therefore, the focus on disease-specific survival as primary outcome measure has been put into question. Since all-cause mortality does not require judgements about the cause of death and since it captures unexpected lethal side effects of screening and medical care, this measure has been proposed as a more important outcome to assess the effectiveness of a screening programme.[40, 43, 44] In this report, we opted to present results for both outcome measures.
- **Quality of life and psychological distress**: one of the objectives of many cancer screening programmes is to improve quality of life by avoiding burdensome treatment for advanced stage cancer, which is often associated with many side effects that may affect patients in the long term. Yet, cancer screening may lead to anxiety and stress when waiting for the test result or more

---

h Sticky diagnosis bias: too many deaths attributed to the target cancer in the screened group and/or too few deaths attributed to target cancer in the control group, leading to disease-specific mortality being biased against screening.[40]

i Slippery linkage bias: too few deaths due to invasive testing after a suspicious screening result or due to the (unnecessary) treatment of early disease (e.g. in case of overdiagnosis) are attributed to screening, leading to disease-specific mortality being biased in favour of screening.[40]

importantly, the distress related to false-positive (or intermediate) results. Moreover, not all individuals with screen-detected cancer benefit in terms of quality of life, as screening may also lead to overdiagnosis. These patients are unnecessarily labelled as cancer patients and undergo unnecessary treatment, with potential adverse consequences (e.g. discomfort related to the test, or more severe morbidity such as bowel perforation after colonoscopy), which may have an important negative impact on quality of life. These harmful effects on the quality of life should hence not be overlooked.[45] Questionnaires that assess quality of life should be sufficiently responsive also to more subtle reactions in order to properly evaluate the psychological distress that can be provoked by screening.[46]

Last, specifically with regard to breast cancer screening: (ductal) carcinoma in situ ((D)CIS, a non-invasive or pre-invasive breast cancer, or stage 0 breast cancer) is much more likely to be detected with screening mammography.[14] Despite the fact that it has been suggested that less than half of the in situ cases become invasive[j], these women are often treated with surgery, anti-oestrogens and radiotherapy.[14]

2.2.3 Biases and caveats

Measuring the effectiveness of a (cancer) screening programme, is subject to various biases. In the following paragraphs, we discuss the most important ones with respect to (breast) cancer screening.

- **Selection bias** (also known as 'healthy user bias', or 'healthy screening bias'),[48] is difficult to avoid in cohort studies, since individuals who participate in screening programmes often differ from those who do not take up their invitation.[23] For example, women with a family history of breast cancer may be more likely to attend mammography screening. In more general terms, individuals who actively engage in preventive health measures, such as cancer screenings and vaccinations, generally exhibit better overall health and maintain healthier lifestyles compared to those who do not. Therefore, when cohorts that participated in screening programmes are compared to cohorts that were invited, but decided not to participate, these underlying differences between cohorts must be taken into account[k].[49]
- **Length-time bias** results from screening programmes being more likely to detect tumours with slow progression, which are, on average, less aggressive than those diagnosed based on clinical signs or symptoms (Figure 4). Slower growing breast cancers, for instance, are more likely to be detected by screening prior to the development of symptoms than rapidly growing tumours.[23] As a result, screen-detected tumours have in general a more favourable prognosis, not necessarily because of detection at an earlier stage of the disease, but because screening is more likely to identify slow-growing, well-differentiated tumours. Length time bias is an overestimation of survival duration due to the relative excess of asymptomatic, slow progressing screening detected cancers, while fast progressing cancers are detected when they cause symptoms. Length-time bias can lead to an overestimation of the effectiveness of screening in improving patient outcomes.[50]

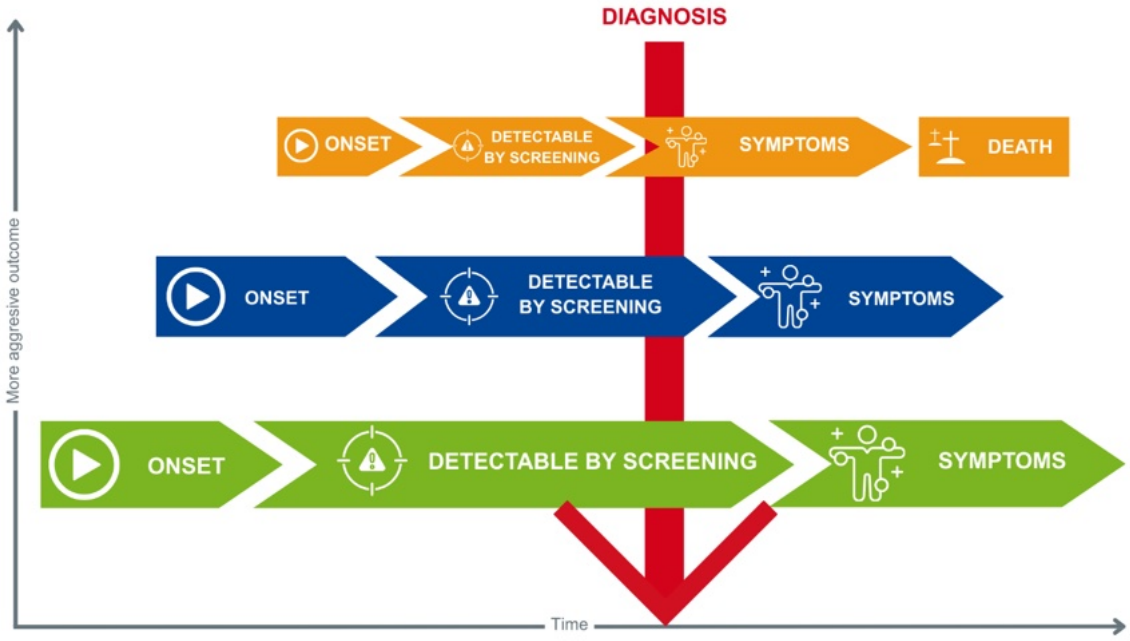
---

j For instance, in the Canadian National Breast Screening Study (CNBSS; see also Appendix 7.5.5) more than 80% of women who participated were free of invasive breast cancer at twenty years after CIS diagnosis.[47]

k Methods have been proposed to correct for this bias, e.g. by including a correction factor (D_r), which is the ratio of disease or mortality rates in participants invited but not attending screening to those in a comparable population not yet invited to screening.[49] This factor is used to adjust observed mortality or disease incidence. It can be estimated from comparable groups in randomized trials or other observational studies, however the choice of studies by the authors will heavily influence their resulting adjusted estimates.[49]

- Cancers detected between screening rounds are called **interval cancers**. They are thus on average more fast growing and more aggressive tumours than screen-detected cancers,[51] and do not benefit from screening programmes.

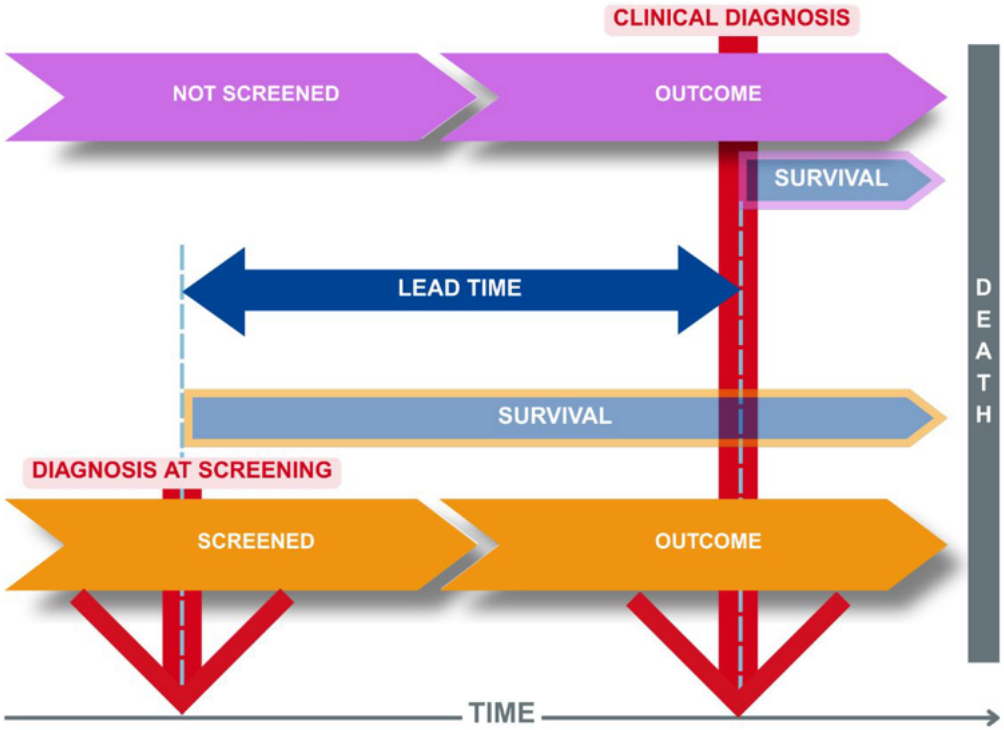
Figure 4 – Representation of length-time bias



***Source**: Figure based on Figure 9.5 in Carneiro, 2017*[23]

- **Lead-time** is the time interval between the screen-detection of a cancer and the occurrence of symptoms. Lead time has no impact on cancer outcome (Figure 5).[23] In the context of evaluating survival benefits for tumours detected through screening versus those identified clinically, it is essential to account for lead time. A screen-detected tumour is diagnosed earlier than it would be through clinical detection, potentially resulting in an apparent extension of survival. However, (a portion of) this apparent survival benefit may be attributed to earlier detection rather than an actual improvement in prognosis. Consequently, when reporting survival based on time since diagnosis, tumours detected through screening may appear to exhibit superior survival outcomes.[50, 52]

Figure 5 – Representation of lead-time bias



***Source**: Figure based on Figure 9.4 in Carneiro, 2017*[23]

- Another important caveat of screening is **overdiagnosis**. Overdiagnosis in screening is the detection of a disease or abnormal condition that would not have caused the person any harm during lifetime if left undiscovered.[45, 53, 54]

  In the context of cancer, overdiagnosis refers to:

  - Detection of a tumour that would have regressed spontaneously;
  - Detection of a slow or non-growing tumour which would not have posed a problem for the patient during his or her lifetime;
  - Detection of a tumour at a time when the patient is more likely to die from other diseases or old age.

As a result, the diagnosis of one of these conditions through screening does not lead to any tangible improvement in the patient's health, wellbeing, or life expectancy. On the contrary, overdiagnosis and overtreatment comes with adverse consequences: a previously healthy individual is now labelled as a sick person, causing unnecessary fear and anxiety, sometimes stigma, and it may have (financial) implications for obtaining life or health insurance.[55] In addition, overdiagnosis does not only lead to unnecessary exposure to potentially burdensome treatments, complications, etc., but also to unnecessary costs for the 'patient' and healthcare system.

Overdiagnosis is a major contributor to length time bias. Consequently, higher rates of overdiagnosis lead to exaggerated and overly optimistic survival rates for the screened population.[38, 56, 57]

Recent research pointed to the important impact of the screening setting on overdiagnosis estimates of ductal carcinoma in situ (DCIS).[58] The proportion of overdiagnosed DCIS increased with shorter follow-up, older age when first attended screening, shorter screening interval, and increased compliance. From an individual perspective, the problem of overdiagnosis is counterintuitive. Patients

diagnosed with cancer through a screening programme, and subsequently treated for that cancer, are likely to perceive the screening and treatment as life-saving. However, for those patients with an overdiagnosed cancer, the disease would never have caused any harm or clinical symptoms had it not been detected through screening.

It is important to understand that overdiagnosis can only be measured at the population level. An indication of overdiagnosis after the introduction of a screening programme, would be an increase in early-stage cancers, while there is no impact on the incidence of late-stage cancers.[23]

Overdiagnosis cannot be measured at the level of the individual screened person.[14] It is thus impossible to determine which individual benefitted and which person was harmed by the screening. Also at the point of early diagnosis, it is impossible to set overdiagnosed cases apart from disease that would eventually become symptomatic later in life. Hence, all screen-detected tumours are managed in the same way, causing harm to an individual with overdiagnosed disease (e.g. through side effects of surgery, radiation and systemic therapy, risks of anaesthesia, pain, discomfort),[59-61] while not providing any benefit.[62]

At the societal level, subsequent diagnostic procedures and overtreatment, as a consequence of overdiagnosis, are costly and contribute to waste in healthcare spending.[63]

Several approaches have been proposed for overdiagnosis. Two of them, the excess incidence approach and the lead time approach are further elaborated in Appendix 3.[64]

- Last, overdiagnosis should not be confused with **false-positive results**, in which case the suspicion of cancer obtained through screening is later dismissed through follow-up tests.[65] False-positive results are also an undesirable outcome of screening, as the associated anxiety over a potential (cancer) diagnosis, along with the subsequent investigations, can have harmful effects on individuals (Figure 3).

In contrast to overdiagnosis, false-positive results are easier to quantify, as patients with false-positive results can be clearly distinguished from those with true positive results. For this outcome, observational studies can also provide informative quantitative data about screening effects.[14]

2.2.4 Study designs to estimate the effect of screening programmes

The various biases and caveats described above make it difficult to evaluate screening programmes.[23] Obviously, cohort studies which compare the length of survival in those detected through screening with those not detected through screening are liable to lead-time and length-time bias. Case-control studies comparing the screening history in cases compared to controls are susceptible to selection and recall bias[l]. Non-randomized trials in which outcome measures are compared with historical or neighbourhood controls are prone to selection bias.[23]

As most of these biases are removed by random allocation, the best method to evaluate the harms and benefits of screening programmes, is the randomized controlled trial (RCT).[23] By taking the moment of randomisation as starting point for survival time rather than the time of diagnosis, lead-time bias is avoided. By including all patients in the analysis, length-time bias and bias due to overdiagnosis are averted.[66]

---

[l] Recall bias is a type of bias that occurs when participants in a research study or clinical trial do not accurately remember a past event or experience or leave out details when reporting about them. Recall bias is more likely to occur when the event happened a long time ago or when the study participant has a poor memory.

2.3 Breast cancer screening

2.3.1 Guidelines on breast cancer screening

The Guidelines Development Group (GDG) of the European Commission Initiative on Breast Cancer (ECIBC) developed guidelines on breast cancer screening for asymptomatic women with an average risk of breast cancer (Table 2).[2] The strength of their recommendation in favour of mammography screening is strong for the age group 50-69 years, but conditional for the other age groups (see the legend to Table 2 for explanation on conditional and strong recommendations).

Table 2 – European guidelines on breast cancer screening (ECIBC)

Age group (year)	Recommendation	Recommendation strength	Certainty of the evidence
	For asymptomatic women with an average risk of breast cancer, the ECIBC's guideline development group…		
40-44	suggests not implementing mammography screening. (issued in June 2016)	Conditional*	Moderate
45-49	suggests mammography screening over no mammography screening, in the context of an organised population-based screening programme. (issued in September 2021)	Conditional*	Moderate
50-69	recommends mammography screening over no mammography screening, in the context of an organised population-based screening programme. (issued in June 2016)	Strong§	Moderate
70-74	suggests mammography screening over no mammography screening, in the context of an organised population-based screening programme. (issued in June 2016)	Conditional*	Moderate

***Notes**: *: When a recommendation is conditional, this implies for (1) patients/women: the majority of individuals in this situation would want the suggested course of action, but many would not. Decision aids may be useful in helping patients to make decisions consistent with their individual risks, values, and preferences; (2) clinicians: different choices will be appropriate for individual patients, and clinicians must help each patient arrive at a management decision consistent with the patient's values and preferences. Decision aids may be useful in helping individuals to make decisions consistent with their individual risks, values, and preferences; (3) policy makers: policy making will require substantial debate and involvement of various stakeholders. Performance measures about the suggested course of action should focus on whether an appropriate decision-making process is duly documented.*

§: When a recommendation is strong, this implies for (1) patients/women: most individuals in this situation would want the recommended course of action, and only a small proportion would not; (2) clinicians: most individuals should follow the recommended course of action. Formal decision aids are not likely to be needed to help individual patients make decisions consistent with their values and preferences; (3) policy makers: the recommendation can be adopted as policy in most situations. Adherence to this recommendation according to the guideline could be used as a quality criterion or performance indicator.

***Source**: European Commission Initiative on Breast Cancer (ECIBC) - https://cancer-screening-and-care.jrc.ec.europa.eu/en/ecibc/*

As shown in Appendix 4, which summarizes the screening policies in Belgium's neighbouring countries as well as a selection of other European and non-European countries, these recommendations have not been implemented in several European countries. Of note, the predominant screening interval in the presented countries is biennial screening, with as an exception the UK pursuing triennial screening. Countries offering annual screening (i.e. Italy in pilot studies and a selection of Canadian provinces) limit this screening interval to women below fifty years, after which biennial screening is

offered until the upper age limit is reached. In all selected countries, the primary screening modality is mammography screening, which is supplemented with clinical breast examination in France.

2.3.2 *Imaging techniques and modalities*

Several imaging techniques have been used for breast cancer screening; the most common one is currently digital mammography. They are briefly described in the following section. It was considered beyond the scope of this study to assess and compare these modalities' performances.

In Belgium, like in many countries, **screen-film mammography**, where the images are stored directly on film, has been replaced by **digital mammography** (also referred to as full-field digital mammography (FFDM)) with the electronic images of the breast being stored as a computer file. They both use x-rays to produce images of the breast. Screening mammograms are usually performed in two standard views, the craniocaudal (i.e. from the top) and the mediolateral oblique (i.e. from the side) views. In case of diagnostic mammography, more x-ray images are taken to obtain views of the breast from several angles, which leads to a higher total dose of radiation.[67]

Digital mammography has the advantage that digital information can be enhanced, magnified, or manipulated for further evaluation more easily than information stored on film. In addition, digital images can be shared electronically, enabling virtual (remote) consultations between radiologists and breast surgeons.

Three-dimensional mammography, which is also known as **digital breast tomosynthesis** (DBT), produces multiple mammographic projections serially along an arc and reconstructs these data via post-processing techniques into thin slice images to create a quasi-three-dimensional assessment of breast tissue.[68] However, for comparison with prior mammograms and accurate interpretation of calcifications two-dimensional images are required.[69] **Synthetic mammography** (SM) refers to the two-dimensional reconstruction of the tomosynthesis slice datasets. Synthetic mammography can be used as an alternative to additional mammography projection (i.e. additional to the DBT), resulting in a reduced radiation exposure as well as image acquisition time.[70]

Breast ultrasound (a.k.a. **sonography**) uses high-frequency sound waves, hence no radiation, to produce detailed images of the breast. As a screening tool, ultrasound may be used as a supplement to mammography in women with dense breast tissue by aiding in the detection of mammographically occult or "masked" cancers.[71] Limitations of handheld ultrasound include operator dependence, long acquisition time when performed by a physician, and limited reproducibility.[70] **Automated Breast Ultrasound (ABUS)** is a supplemental screening tool (approved by FDA), designed for women with heterogeneously and extremely dense breast.[72] ABUS provides reproducible images for breast lesion location, size measurement, and characterization, and is often used alongside mammography.

Breast **magnetic resonance imaging (MRI)**, a non-invasive imaging technology that produces three dimensional detailed anatomical images, is the most sensitive imaging method for breast cancer detection, mostly because of its ability to detect neovascularity.[70] Breast MRI may be used for (supplemental) screening in women with extremely dense breasts and in high-risk women,[70, 73] such as women with genetic-based increased risk and their untested first-degree relatives, women who received chest radiation therapy before age 30, and women with a calculated risk of 20% or more based on strong clinical evidence.[74]

Contrast-enhanced mammography (CEM) is a recently introduced breast imaging technique which uses contrast medium to depict vascularity potentially associated with breast cancer. It thus provides anatomic and functional information at the same time.[73] The evalution of CEM in the screening setting is limited so far.[70]

The American College of Radiology (ACR) has developed the Breast Imaging Reporting and Data System (BI-RADS), a tool that provides a lexicon and reporting scheme for imaging of the breast and which may simplify outcome monitoring.[75] It applies to mammography, ultrasound, and MRI. The various assessment categories can be found in Appendix 5.

2.3.3 Quality assurance in breast cancer screening

The European Commission Initiative on Breast Cancer published a person-centred quality assurance (QA) scheme (2025), which defines a common set of quality and safety requirements for breast cancer services.[76] Breast cancer services that comply with these requirements can apply for the European QA scheme certification.

With regard to screening, the QA scheme does not only list requirements (e.g. the screening programme must invite and reach over time the entire target population that includes at least women aged 50-69; the screening programme has a systematic call/recall system in place), it also provides performance indicators (e.g. screening coverage[m], participation rate, time between screening mammogram and issuing of results).[77] In addition, the QA scheme stipulates that the screening programme centres providing breast cancer screening must have a written policy for informing women about a healthy lifestyle (including nutrition and physical activity).[76]

2.3.4 Breast cancer screening in Belgium

Over the last decades, Belgium evolved through six state reforms into a federal state. The federated entities (the regions) have become competent for a broad range of matters, including public health, e.g. health promotion and disease prevention, vaccination and the screening programmes for breast, colorectal and cervical cancer.[78] The screening programmes are coordinated at the regional level by dedicated screening organizations (i.e. Centrum voor Kankeropsporing (CvKO) in Flanders, BruPrev in Brussels and Le Centre de Coordination et de Référence pour le dépistage des cancers (CCRef) in the Walloon region), who have developed their own procedures and criteria for invitations and results communication (for details see Appendix 6).

While breast cancer screening is thus a regional competence, the screening mammographies, the second reading as well as supplemental testing and treatment in case of suspicious screening results are covered by NIHDI (National Institute for Health and Disability Insurance ('Rijksinstituut voor Ziekte- en Invaliditeitsverzekering' (RIZIV)/'Institut National d'Assurance Maladie-Invalidité' (INAMI))).[n]

In addition to the biennial invitation provided by the screening organizations, physicians can also prescribe a screening mammography for patients within the eligible age range (50-69 years).[79] This alternative pathway does not affect the remainder of the screening process. Last, women of any age may also have a mammography (whether or not supplemented by an US or other modality) beyond the official screening programme, which is also known as opportunistic screening. In these cases the dedicated invoicing codes for mammographies within the context of screening programmes cannot be used, but invoicing codes for diagnostic mammographies must be used. As a consequence, opportunistic mammography and diagnostic mammography cannot be distinguished in administrative databases, hence the proportion of opportunistic mammographies cannot reliably be estimated. Of note, unlike in the official screening programmes, opportunistic mammography does not include a systematic second reading,[79] and there is a patient cost-sharing. Yet, while the fee schedule stipulates that an US cannot be invoiced on the same day as a screening mammography, this is not the case for opportuntistic/diagnostic mammographies.

---

[m] According to the final list of performance indicators for breast cancer screening, developed by the ECIBC: Screening coverage: the numerator is the number of women screened (which can also include tests performed in the opportunistic setting) and the denominator is the total number of eligible (or target) women within a given period; participation rate: the numerator is the number of women screened and the denominator is the total number of women invited.

[n] https://www.riziv.fgov.be/nl/thema-s/verzorging-kosten-en-terugbetaling/wat-het-ziekenfonds-terugbetaalt/onze-tegemoetkoming-bij-vroegtijdige-opsporing-van-bepaalde-kwaadaardige-aandoeningen

2.3.4.1 *Indicators of the population breast cancer screening programmes, as reported by the screening organizations and BCR*

Periodically the three screening organizations publish together with the BCR, results on key performance indicators (Table 3). Although the calculation of certain indicators is performed differently between the regions, a closer look at the data leads to some interesting insights. While the response rate[o] is highest in Flanders by a substantial margin, the difference in coverage[p] between the three regions is still present, but lower in magnitude. The reason is that in Brussels and Wallonia, in contrast to Flanders, most women received a mammography (with or without another modality) outside the regional screening programmes (for opportunistic screening or diagnostic purposes) and only a minority a screening mammography via the screening programmes organised by the regions.[q] However, since both response rate and coverage are calculated differently in Flanders than in Brussels and Wallonia, the results should be interpreted with caution. The indicators also show that the recall rates - for both initial and repeat screenings - are substantially higher in Brussels and Wallonia than in Flanders.

In addition to participation and recall rates, also differences in specificity[r] of the screening programmes can be observed, with the screening programme in Flanders having a higher specificity (98%) than the programmes in Brussels and Wallonia (90% and 89%, respectively). In the years for which results are presented, the sensitivity (around 70%)[s] as well as the stage distributions of screen-detected cancers are fairly similar between the regions.

The data in Table 3 is presented purely for descriptive purposes. They were not used in the primary health economic evaluation performed in this study. All data used in this evaluation are described elaborately in Chapters 5 and 6.

---

o For Brussels and Wallonia, response rate was calculated by BCR as the number of women with a screening mammography of the regional screening programme divided by the eligible target population. In Flanders, response rate was calculated by restricting the numerator to women participating in the regional screening programme for whom a participation date was recorded. The denominator is the invited population.

p Coverage in Brussels and Wallonia was calculated by BCR as the total number of women with a screening mammography invoiced within the frame of the regional screening programme and of women with a bilateral mammography outside the regional screening programme, divided by the eligible target population. Coverage in Flanders was calculated as the total number of 1) women participating in the regional screening programme, 2) women with a bilateral mammography outside the regional screening programme, 3) women who are not eligible for participation that year, and 4) women with a mammography within or outside the regional screening programme, divided by the target population with health insurance.

q As the Belgian invoicing codes do not distinguish between an opportunistic mammography and a diagnostic mammography, an unknown proportion of the mammographies outside the regional screening programmes are for diagnostic instead of screening purposes. Coverage is therefore an overestimation of a screened population (either via opportunistic screening or the regional screening programme). Coverage and response rate were calculated differently in Flanders compared to Brussels and Wallonia (see above).

r Specificity was calculated using as numerator the number of negative screening mammograhies in year x for which no cancer was diagnosed 24 months later and as denominator the sum of number of negative screening mammograhies in year x for which no cancer was diagnosed 24 months later and number of positive screening mammographies in year x for which no cancer was diagnosed 24 months later.

s While sensitivity between the regions is highly similar in the presented years, larger differences can be observed for 2021 (Brussels: 79.4%, Wallonia: 78%, Flanders: 70%). Sensitivity was calculated using as numerator the number of screen-detected cancers in year x and as denominator the sum of number of screen-detected cancers in year x and interval cancers in year x. While sensitivity for Flanders was available for 2022 (i.e. 68%), more detailed information on the invited population was missing and therefore 2021 was preferred over 2022.[80, 81, 82]

Table 3 – Key indicators of breast cancer screening as reported by BCR and screening organizations

	Brussels		Flanders		Wallonia	
	2022	2024	2021	2024	2022	2024
Target population with health insurance	124 775	125 484*	889 876	896 199	497 361	497 129*
Not eligible due to medical reasons	3581	3431*	94 981	*NA*	14 700	14 845*
Eligible target population (2 years)	121 194	122 053*	794 895	*NA*	482 661	482 284*
Mammography 20XX-1	5507	5408*	187 162	*NA*	10 173	10 088*
Mammography 20XX-1 outside programme**	/	/	66 781	*NA*	86 848	88 779*
Eligible for invitation (2 years)	115 687	116 645*	607 733	*NA*	385 640	383 417*
Invited	*NA*	59 683*	438 957	388 448	*NA*	*NA*
Participants	10 442 (for 2y)	10 942 (for 2y)	241 726	218 631	19 417 (for 2y)	21 187* (for 2y)
Response rate[o]	8.60%	9.00%*	53.20%	54.00%	4.00%	4.40%*
Coverage[p]	44.30%	46%*	63.50%	66.90%	47.60%	50.30%*
Recall rate (1st participation)	17.5%	19.2%	6.4%	6.5%	17.8%	14.5%
Recall rate (subsequent participation)	5.1%	5.3%	2.4%	2.4%	7.1%	5.2%
Sensitivity[r]	72.50%*	*NA*	70.30%	*NA*	71.40%*	*NA*
Specificity[s]	90.10%*	*NA*	97.60%	*NA*	89.40%*	*NA*
	2020-2021		2018		2020-2021	
Screen-detected cancers	84 (+0 unknown stage)		1200 (+25 unknown stage)		123 (+0 unknown stage)	
Stage§ 0	15.50%		15%		9.80%	
Stage I	46.40%		56%		53.70%	
Stage II	29.80%		24%		32.50%	
Stage III	7.10%		4%		4.10%	
Stage IV	1.20%		1%		0%	
Interval cancers	24 (+0 unknown stage)		615 (+53 unknown stage)		46 (+0 unknown stage)	
Stage 0	8.30%		6%		2.20%	
Stage I	41.70%		46%		50.00%	
Stage II	29.20%		35%		32.60%	
Stage III	8.30%		9%		13.00%	
Stage IV	12.50%		4%		2.20%	

***Notes**: For each region, the left column represents the information from the last available report in which (quasi-) complete information on the eligible target population, sensitivity and specificity is available and the right column represents information from the last available report.§: The stage refers to a 'combined stage', where known pathological stage prevailed over known clinical stage, except when there was clinical proof of distant metastasis (cM1). Note that these data are restricted to invasive breast cancer, i.e. excluding stage 0 (i.e. non-invasive breast cancer). In case of neo-adjuvant systemic treatment cTNM data are used to define combined stage, but in case no cTNM data have been transferred to the BCR, upTNM data are used to define combined stage.*

[o]For Brussels and Wallonia, response rate was calculated by BCR as the number of women with a screening mammography of the regional screening programme divided by the eligible target population. In Flanders, response rate was calculated by restricting the numerator to women participating in the regional screening programme for whom a participation date was recorded. The denominator is the invited population.

[p]Coverage in Brussels and Wallonia was calculated by BCR as the total number of women with a screening mammography invoiced within the frame of the regional screening programme and of women with a bilateral mammography outside the regional screening programme, divided by the eligible target population. Coverage in

Flanders was calculated as the total number of 1) women participating in the regional screening programme, 2) women with a bilateral mammography outside the regional screening programme, 3) women who are not eligible for participation that year, and 4) women with a mammography within or outside the regional screening programme, divided by the target population with health insurance.
[r]Sensitivity was calculated using as numerator the number of screen-detected cancers in year x and as denominator the sum of number of screen-detected cancers in year x and interval cancers in year x. While sensitivity for Flanders was available for 2022 (i.e. 67.5%), more detailed information on the invited population was missing and therefore 2021 was preferred over 2022.
[s]Specificity was calculated using as numerator the number of negative screening mammograhies in year x for which no cancer was diagnosed 24 months later and as denominator the sum of number of negative screening mammograhies in year x for which no cancer was diagnosed 24 months later and number of positive screening mammographies in year x for which no cancer was diagnosed 24 months later.
**: Data not complete at moment of publication. **: In Brussels, having had a mammography outside the official screening programme is not an exclusion criterium for invitations. BCR = Belgian cancer registry. NA = Not available.*
***Sources**: Brussels,[81, 83] Flanders,[82, 84, 85] & Wallonia.[80]*

3 SYSTEMATIC REVIEW OF CLINICAL EFFECTIVENESS AND SAFETY OF BREAST CANCER SCREENING

Key points

1) Invited for screening with mammography vs. not invited for screening in women aged 50-69 years, 40-49 years, and 70+ years

The currently available RCT evidence is limited to trials performed in the period 1960 to 1990, and may not represent current screening, diagnostic and/or treatment approaches. Yet, while recent observational studies (e.g. cohort, case-control studies) comparing attendees to non-attendees represent current practices better, they are affected by selection bias ('healthy user bias'), recall bias and unmeasured confounders, resulting in overestimations of screening benefits.

All age groups[t]

After 13 years of FU, inviting adult women to mammography screening results in a statistically significant reduction of death ascribed to breast cancer, while it does not have an effect on all-cause mortality.

Invitation to screening with mammography results in a statistically significant reduction of the probability of being diagnosed with advanced disease (both for stage II or higher and stage III or higher).

Invitation to screening with mammography results in a statistically significant increase of overdiagnosis 7-9 years after randomisation. Among adult women at average risk of breast cancer, and invited to screening, 27% of all cancers (invasive and non-invasive) detected were overdiagnosed.

<50 years

The results for women younger than fifty years old, are comparable to the all age group for death ascribed to breast cancer, as well as for all-cause mortality.

[t] "All age groups" refers to results obtained in women of whatever age. In the other sections of the key points the results are restricted to data obtained for the age groups as defined in the research questions (<50 years, 50-69 years and ≥70 years).

Yet, for stage shift the effect of screening is clearly less pronounced and/or no longer statistically significant (only stage II or higher significant).

In this young age group, invitation to screening with mammography results in a statistically significant risk of overdiagnosis. Among women younger than 50 years old and at average risk of breast cancer, and invited to screening, 23% of all cancers (invasive and non-invasive) detected were overdiagnosed.

50-69 years

The results for death ascribed to breast cancer for women between fifty and sixty-nine years old, are more pronounced compared to the all age group. For all-cause mortality no relevant data could be retrieved from the RCTs.

In this age group, invitation to screening results in a statistically significantly reduced risk of being diagnosed with advanced disease (both for stage II or higher and stage III or higher).

Invitation to screening with mammography results in a statistically significant risk of overdiagnosis. Among women between 50 and 69 years old and at average risk of breast cancer, and invited to screening, 21% of all cancers detected (invasive and non-invasive) were overdiagnosed.

≥70 years

For women of seventy years and older, mammography screening does not result in a significantly decreased breast cancer related mortality risk nor in a reduced all-cause mortality risk.

In addition, for this age group, no relevant data were retrieved from RCTs with regard to stage shift and on overdiagnosis.

2) Various breast cancer screening intervals

Currently, there is no relevant RCT data available to answer the second research question (i.e. what is in women at average risk of breast cancer the clinical effectiveness of a) biennial screening with mammography vs. annual screening with mammography, and b) biennial screening with mammography vs. triennial screening with mammography).

3) Invited for screening with mammography and ultrasound vs. invited for screening mammography alone

The currently available RCT evidence is limited. So far only one large and methodologically very well performed RCT in Japanese women younger than fifty years old has been published.

After eleven years of follow-up, invitation to screening with the combination of mammography and ultrasonography does not have a statistically significant effect on all-cause mortality, nor on breast cancer mortality compared to being invited to mammography alone.

The proportion of false-positive results was statistically significantly higher than with mammography screening alone, which leads to unnecessary harm caused by subsequent investigations while there is no gain in health benefits.

The aim of this chapter[u] is to provide an overview of the available evidence on the clinical effectiveness and safety of **invitation to breast cancer screening with mammography** (and ultrasound, cf. section 3.2.4) in a population of **adult women without suspected breast cancer and without a specifically increased risk of breast cancer**[v]. We performed – when data were available - intention-to-treat analyses, by including all randomised women. However, for the sake of readability, in the following chapter "invitation to breast cancer screening with mammography" is abbreviated to "breast cancer screening with mammography", and "adult women without suspected breast cancer and without a specifically increased risk of breast cancer" is abbreviated to "women at average risk of breast cancer".

3.1 Methods

3.1.1 Research questions and outcome measures

A systematic literature search was performed to answer the following research questions (RQ):

RQ1: What is the clinical effectiveness of breast cancer screening with mammography in women at average risk of breast cancer and

a. aged 50-69 years (i.e. current breast cancer screening programme)?
b. aged 40-49 years?
c. aged 70 years and older?

RQ2: What is in women at average risk of breast cancer the clinical effectiveness of

- biennial screening with mammography vs. annual screening with mammography?
- biennial screening with mammography vs. triennial screening with mammography?

RQ3: What is the clinical effectiveness of breast cancer screening using mammography combined with ultrasound vs. mammography alone?

For all RQs, the following outcome measures of effect were selected:

- All-cause mortality
- Mortality ascribed to breast cancer
- Mortality ascribed to any cancer
- Other cause mortality
- Breast cancer stage at diagnosis
- Overdiagnosis
- Interval cancers

---

[u] The very first draft of this chapter was co-authored by Sien Ombelet.

[v] In the recent systematic reviews (SR) on breast cancer screening, various definitions have been used to describe the breast cancer risk in the target population: "asymptomatic women with an average risk of breast cancer",[86] "women at average breast cancer risk",[87] "women not suspected of breast cancer",[14] moderately increased risk for breast cancer, [with] average risk [referring] to those without factors placing them at higher-than-average risk of cancer (i.e. about 12.5% lifetime risk), whereas women with moderately increased risk (i.e.,12.5 to 20% lifetime risk) will include individuals with an elevated risk of breast cancer (e.g., dense breasts, one first degree relative with history of breast cancer)".[13] We opted for the phrasing adopted by the Institute for Quality and Efficiency in Health Care (IQWiG).[88]

- 'Treatment-related' morbidity
- Consequences of false-positive screening results
- Consequences of false-negative results
- Health-related quality of life (HRQoL)
- Radiation exposure

All outcome measures were binary and differences between intervention groups are presented as

- Relative risks (RR) and 95% confidence intervals (CI)
- Absolute effect estimates (/100 000) and 95% CI.

3.1.2 Systematic literature review

3.1.2.1 Systematic literature search

The systematic literature search was performed in two stages. In a first step, systematic reviews and Health Technology Assessments (HTAs) summarising benefits and harms of breast cancer screening were searched. The searches for systematic reviews were performed in the Cochrane Database of systematic reviews, Ovid-Medline and Embase. The search was limited to the period 2020 – 24 May 2024, to ensure the inclusion of longer follow-up data in the meta-analyses. The search strategies for each database, using a combination of Medical Subject Headings (MeSH) terms and free text words can be found in Appendix 8. A full update (i.e. all databases) was performed in December 2024; a final update in Medline for all searches was performed in March 2026.

The websites of the members of the International Network of Agencies for Health Technology Assessment (INAHTA) were searched for dedicated HTA reports in April 2024. An overview of the included HTA institutes can be consulted in Appendix 10).

In a second step, these data were amended through a systematic literature search for randomised controlled trials (RCTs), which were published after the search dates of the SRs retrieved and selected in the first step (period covered: 2020 – 12 December 2024; for the search strategies, see Appendix 8). In February 2026, an additional publication of one of the included RCTs (i.e. the J-START trial) was identified through an automatic journal alert and included in the report as it provided longer follow-up data.[89]

Although observational studies may better reflect current practices (since most trials on breast cancer screening are old and may not represent current screening and/or treatment approaches), observational studies carry important limitations, such as self-selection bias and recall bias (see section 2.2.3). For that reason, the update of the search was limited to RCTs. Yet, when observational studies were included in the identified SRs, their results are also included in the present report, to illustrate the differences in results between randomised and non-randomised studies. For clarity reasons, they are discussed in separate sections and not included in the Key Points.

When interpreting the results, it should be kept in mind that the differences in results between RCTs and observational studies on breast cancer screening may also (in part) be attributed to differences in analytic approach. While in observational studies analyses are by definition based on the actual intervention received (also known as 'per protocol' analysis, this is: screened versus not screened), this was not the case in the included RCTs. In these RCTs an 'intention to treat' approach was applied, i.e. participants were analysed according to their original group assignment, or in other words: invited for screening versus not invited). As is illustrated in Appendix 9.4, in the included RCTs average

attendance rates[w] ranged between 60% and 90% (and depended on the age group and screening round). It's important to note that compliers and non-compliers in screening programmes may differ *a priori* in aspects which are not related to screening but which nevertheless affect the risk of death (from the disease).[91]

3.1.2.2 Study selection

The selection of systematic review (SR) records and HTAs was performed by two researchers (RL and SO), based on pre-defined inclusion criteria (i.e. population, interventions, comparator, outcomes (PICO), study design, period and language) provided in Appendix Table 69. The screening of titles and abstracts of the retrieved studies was followed by a full text screening. At that stage, the following exclusion criteria were applied:

- Search of the review limited to one database
- Search strategy restricted to MeSH terms
- In- and exclusion criteria not clearly formulated
- Narrative instead of systematic review
- Lack of quality assessment of included studies
- RQ under study not covered

All conflicts in study selection were resolved by discussion between both reviewers. The number of excluded full texts and the reasons for exclusion can be found in Appendix Table 70 and 71. The selection of RCTs was performed by one researcher (RL).

3.1.3 Quality assessment of full texts

Two reviewers independently assessed the methodological quality of all selected HTAs and SRs using the second version of A MeaSurement Tool to Assess systematic Reviews (AMSTAR-2) criteria (Appendix Table 72).[92]

The authors selected the Cochrane Collaboration's tool for assessing risk of bias for RCTs,[93] and the risk of bias in non-randomised studies of interventions (ROBINS-I) tool for observational studies.[94]

When applicable, disagreements were resolved through discussion.

More details on the reported methods to assess the quality and provide the synthesis of primary studies included in the SRs that were selected for this report, can be found in Appendix 9.2.

3.1.4 Forest plots, summary of findings and assessment of the certainty of the evidence

The forest plots presented in this chapter, which were developed by the authors of this report, are based on:

- the numerators (i.e. number of events) and denominators presented in the Cochrane SR (2013) which were only to a limited extend updated in the unpublished SR (2024),[14, 95] the SR by Canelo-Aybar et al. (2021),[86] the SR by Bennett et al. (2024) and the manuscript by Tarone et al. (1995);[96]
- the risk of bias (RoB) assessment conducted by Canelo-Aybar (2021),[86] using the ROB-1 tool.[93]

---

[w] Several approaches have been proposed to correct for non-compliance in breast cancer screening studies,[90,91] yet their application was considered beyond the scope of the present study.

The forest plots comprise two sub-analyses: a) RCTs with a low risk of selection bias, and b) RCTs with a high risk of selection bias. The pooled results of both subgroups were considered the primary analysis. For the sake of completeness, the results restricted to RCTs with a low risk of selection bias (i.e. UK Age and Malmö trials) are also presented (and briefly discussed).

For the meta-analyses the fixed effects method was used. This method makes the assumption that the observed variation in treatment effects in the different studies is entirely due to sampling variation, and that the underlying treatment effect is the same in all study populations.[97]

We applied the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to determine the certainty of evidence across the body of evidence for each outcome,[98] which was also performed by other review authors (i.e. Bennett et al. (2024),[13] Glechner et al. (2023),[99] Canelo-Aybar et al. (2022),[100], Canelo-Aybar et al. (2021)[86], and Gøtzsche & Jørgensen[14]). GRADE criteria for assessing the certainty of the evidence include study limitations, inconsistency of results, indirectness of evidence, imprecision, and publication bias.[101] The GRADE assessment comprises four categories: high, moderate, low, and very low certainty evidence (Box 1)

Box 1 – GRADE Working Group grades of evidence

- High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
- Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
- Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

***Source**: Balshem et al., 2011*[101]

The GRADE summary of findings tables (developed in GRADEpro) are provided in Appendix 9.6; they do not only provide the relative risk (RR) and its 95% confidence interval (CI) and the certainty assessment, but also the absolute effect estimate and 95% CI for each outcome.

The GRADE guidelines to communicate the findings of systematic reviews of interventions,[102] were not applied since the (clinical) experts involved in this report (see colophon) expressed their concerns about the applicability in the field of breast cancer screening, of the decision thresholds for judgments on health benefits and harms as provided in the paper by Morgano et al. (2025).[103] At the moment of development of this report no breast-cancer-screening specific decision thresholds were publicly available, and the development of *de-novo* thresholds was considered beyond the scope of this research, hence this approach was not implemented.

Whenever possible results at 13 years of follow-up were presented as main outcome. This time span was chosen as a trade-off had to be made between achieving sufficiently long follow-up on the one hand and minimizing data limitations (e.g. loss to follow-up and women not invited to screening being screened (also referred to as 'contamination')) on the other hand.

3.2 Results

3.2.1 Results of the search

3.2.1.1 Included SRs and HTA reports

The search for SRs yielded 344 electronic records (following de-duplication), published between 2020 and 2024, which were screened based on title and abstract. From these, seventeen were selected as potentially relevant, of which eleven were excluded (see Appendix Table 70 for reasons for exclusion).

In addition, eight HTA reports were identified which were published in 2020 or later and which included a medical harms-benefit analysis. Of these reports, four full texts were screened for eligibility, of which two were included.[13, 88]

The following SRs and HTA reports were selected for this report:

- Bennett et al., 2024,[13] to inform the Canadian Task Force on Preventive Health Care
- Henderson et al., 2024,[87] to advise the US Preventive Services Task Force
- Mathieu et al., 2024[104]
- Glechner et al., 2023[99]
- Canelo-Aybar et al., 2022,[100] to inform the European Commission Initiative on Breast Cancer (ECIBC)
- HTA report IQWiG, 2022[88]
- Canelo-Aybar et al., 2021,[86] to inform the European Commission Initiative on Breast Cancer (ECIBC)
- Zielonke et al., 2020[105]

The update of the search for SRs (December 2024) revealed two more records. After full text assessment, it was decided not to include either in this review. Nicholson et al. (2024) was excluded as the publication was confined to the recommendations following the SR performed by Henderson et al. (2024).[87, 106] Autier et al. (2024) was excluded as it is a SR and meta-analysis of cohort studies.[49]

In the search for RCTs, no more recent, completed trials were identified. Yet, this search yielded an unpublished SR,[14] which is an update of the previous Cochrane SRs on breast cancer screening with mammography.[95, 107-110] We carefully compared the unpublished manuscript with the preceding (published) version from 2013: the reported results and risk of bias assessments remained unchanged, while for two trials (Canadian National Breast Screening Study (CNBSS) and UK Age) longer follow-up results have been added. In addition, the authors amended the published Cochrane SR with a GRADE assessment of the certainty of the evidence and Summary of Findings Tables.[14] This unpublished SR was also included in the present report (see Appendix 9.3.1.8 for more details).[14]

Mathieu et al. (2024) included studies which contained observational or modelled/simulated data; they evaluate a wide range of breast cancer screening modalities. In the present review, data extraction and reporting is restricted to observational studies.[104]

Zielonke et al. (2020) published a SR of high quality (according to AMSTAR-2 criteria, Appendix 9.2.2), but their focus was on effectiveness of screening stratified by European region and they did not perform a meta-analysis.[105] Moreover, the RCTs included in their SR were also included in Bennett et al. (2024),[13] and Canelo-Aybar et al. (2021).[86] Therefore there was no apparent added value of including the results of Zielonke et al. in this summary.

The data extraction tables for the SRs can be found in Appendix 9.3.1; no data extraction table was developed for the Zielonke SR.

3.2.1.2 Breast cancer screening trials included in the SRs and HTA reports

Table 4 provides a summary of the breast cancer screening trials included in the SRs and HTA reports, highlighting the recruitment period and age range of the targeted populations. Of note, none of the trials have a target population which corresponds to the current target population for breast cancer screening in Belgium (i.e. women between 50 and 69 years[x]), while several trials focus on a younger age group (CNBSS-1, Gothenburg trial, UK Age trial and J-START). It is notable that none of these trials assess the harms and benefits of breast cancer screening in minority groups, subgroups with poor access to care or women with particular risk profiles.[14]

For more detailed information about each trial (e.g. design and intervention), we refer the reader to Appendix 9.4.

Table 4 – Summary of breast cancer screening trials

RCTs	Recruitment period and targeted population
Health Insurance Plan (HIP) of Greater New York	December 1963 – June 1966 Women aged 40 to 64 years
Malmö Mammographic Screening Trial (MMST)	**MMST I** October 1976 – September 1978 Women aged 45 to 70 years
	MMST II September 1978 – November 1990 Women aged 45 to 49 years
Swedish Two-County Trial	**Kopparberg** October 1977 – January 1980 Women aged 40 to 74 years
	Östergötland May 1978 – March 1981 Women aged 40 to 74 years
(Edinburgh Randomised Breast Screening Project (EBSP)[y])	1979 – 1981 Women aged 45 to 64 years
Canadian National Breast Screening Study (CNBSS)	**CNBSS-1** January 1980 – March 1985 Women aged 40 to 49 years
	CNBSS-2 January 1980 – March 1985 Women aged 50 to 59 years

x More precisely, women living in Belgium are invited for a biannual mammography from the year they turn fifty years old up and including the year they turn sixty-nine years old.

y The Edinburgh trial had significant methodological flaws, including inadequate randomisation and major baseline differences between the intervention and control groups, particularly in socioeconomic status and entry date definitions. Eligibility and exclusion procedures varied between groups, with broader criteria and earlier entry dates for the control group, raising concerns about comparability. Recruitment practices were inconsistent, with the control group enrolled later, and exclusion due to prior breast cancer also differed between groups. Therefore, results of this trial will not be discussed further throughout this chapter.

RCTs	Recruitment period and targeted population
Stockholm breast cancer trial	First cohort: March 1981 – April 1982 Second cohort: April 1982 – May 1983 Women aged 40 to 64 years
Gothenburg breast screening trial	December 1982 – April 1984 Women aged 39 to 49 years
United Kingdom Co-ordinating Committee on Cancer Research (UKCCCR) trial of Breast Screening Frequency	1989 – 1996 Women aged 50 to 62 years
UK Age trial	1991 – 1997 Women aged 39 to 41 years
J-START	2007 – 2009 Women aged 40 to 49 years
AgeX Pilot study	2009 Women aged 47 to 73 years

3.2.1.3 Quality appraisal

The SR by Bennett et al. (2024) met most AMSTAR-2 criteria,[13] however, they did not provide full explanation of the included study designs and all sources of funding for included studies, and the data extraction was not performed independently in duplicate, although the data were verified by a second reader. Risk of bias assessment for non-randomised trials was done using the Joanna Briggs Institute (JBI) checklists,[111] which do not include an assessment of reporting bias (as required by AMSTAR-2, Appendix 9.2.2).[92] Another shortcoming were the forest plots: the forest plots on breast cancer mortality and all-cause mortality did not include events and denominators for intervention and control groups. None of the forest plots comprise the RoB analysis of included studies.

The SR by Canelo-Aybar et al. (2022) met most AMSTAR-2 criteria,[100] yet, the items study design explanation, comprehensive search strategy and duplicate study selection and data extraction were scored as “probably yes”. The funding sources of the included studies were not included.

The SR by Glechner et al. (2024) met almost all AMSTAR-2 criteria.[99] Three items were scored as “probably yes”: study design explanation, comprehensive search strategy and duplicate data extraction.

The unpublished update of the previous Cochrane SR on breast cancer screening with mammography, met all AMSTAR-2 criteria.[14] Yet, the authors assigned the CNBSS trial to the group of studies with a low risk of selection bias, while several other (review) authors expressed concerns about the inclusion of symptomatic patients in CNBSS, leading to potentially biased randomisation.[13, 112-114] Based on these publications, we considered the CNBSS as having a potential risk of selection bias, while Gøtzsche and Jørgensen did not.

The other selected SRs had more shortcomings, as is illustrated in Table 74 in Appendix 9.2.2. The risk of bias assessments from other systematic reviews can be found in Appendix 9.2.3.

3.2.2 Findings – Invited for screening with mammography vs. not invited for screening in women aged 50-69 years, 40-49 years and 70+ years

For the first research question (i.e. What is the clinical effectiveness of breast cancer screening, in women aged 50-69 years, 40-49 years and 70+ years?), the SR by Bennett et al. (2024) and the unpublished update of the Cochrane SR (Gøtzsche & Jørgensen, 2024) were used as a starting point,[13, 14] given their high quality (Appendix Table 74) and recent search dates (i.e. July 2023 and 28 February 2023, respectively).

This information was amended with data provided in the SR by Henderson et al. (2024, which was published to advise the US Preventive Services Task Force),[87] the SR by Mathieu et al. (2024),[104] the HTA report published by the German Institute for Quality and Efficiency in Health Care (IQWiG, 2022),[88] as well as the SR by Canelo-Aybar et al. (2021; SR for the European Commission Initiative on Breast Cancer).[86]

Bennett et al. (2024), Henderson et al. (2024), Gøtzsche & Jørgensen (2024) and Canelo-Aybar et al. (2021), reviewed screening effectiveness in all age groups,[13, 14, 86, 87] while the IQWiG report focused on the younger (45-49 years[z]) and older (70+ years) age groups,[88] and Mathieu et al. (2024) looked specifically at older women (75+ years).[104]

3.2.2.1 All-cause mortality

Key points

All age groups[aa]

In a general population of adult women at average risk of breast cancer, invitation to screening with mammography does not have a statistically significant effect on all-cause mortality after 13 years of follow-up (FU) (RR: 0.99, 95% CI: 0.97-1.01; absolute effect: 61 fewer per 100 000, 95% CI: from 182 fewer to 61 more; low certainty of evidence). The certainty of the evidence was downgraded one level due to serious risk of selection bias in six out of eight included RCTs and another level due to indirectness because of the age of the trials.

<50 years

In a general population of adult women younger than 50 years old and at average risk of breast cancer, invitation to screening with mammography does not have a statistically significant effect on all-cause mortality after 13 years of FU (RR: 0.99, 95% CI: 0.94-1.04; absolute effect: 21 fewer per 100 000, 95% CI: from 126 fewer to 84 more; low certainty of evidence). The certainty of the evidence was downgraded one level due to serious risk of selection bias in four out of six included RCTs and another level due to indirectness because of the age of the trials.

50-69 years

No relevant RCT data were identified.

≥70 years

In a general population of adult women of 70 years or older and at average risk of breast cancer, invitation to screening with mammography does not have a statistically significant effect on all-

---

z Although the authors of the IQWiG report focused for the first research question on mammography screening in women aged 45 to 49 years, they also report on data obtained in women younger than 45 years old (depending on availability of data).[88]

aa "All age groups" refers to results obtained in women of whatever age. In the other sections of the key points the results are restricted to data obtained for the age groups as defined in the research questions (<50 years, 50-69 years and ≥70 years).

cause mortality after 7.9 years of FU (RR: 0.98, 95% CI: 0.93-1.03; absolute effect: 567 fewer per 100 000, 95% CI: from 1986 fewer to 851 more; very low certainty of evidence). The certainty of the evidence was downgraded one level due to serious risk of selection bias in the included RCT, one level due to indirectness because of the age of the trials and another level because the outcome was reported in only one (inadequately) randomised trial.

ALL AGE GROUPS

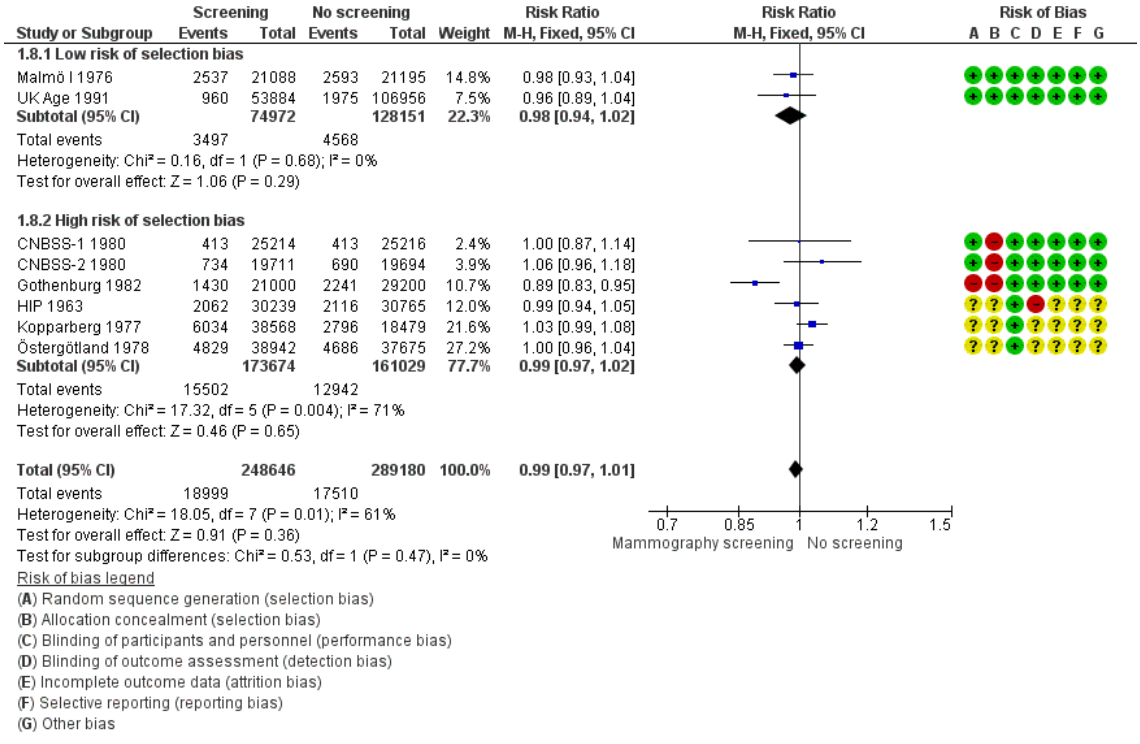
A. Randomised controlled trials

Screening with mammography **does not result in a statistically significantly reduced risk of all-cause mortality** in a general population after thirteen years (relative risk (RR): 0.99, 95% confidence interval (CI): 0.97-1.01, p=0.36; Figure 6). Gøtzsche & Jørgensen (2013; 2024) and Bennett et al. (2024) arrived at the same conclusion.[13, 14, 95] The outcome all-cause mortality was not assessed by Canelo-Aybar et al. (2021).[86]

Comparable results were obtained after seven (RR: 0.98, 95% CI: 0.96-1.01, p=0.20; Figure 63 in Appendix 9.5.1.1) and after twenty-two years follow-up (RR: 1.01, 95% CI: 0.99-1.04, p=0.35; Figure 64 in 9.5.1.1).

The pooled mortality estimates of adequately randomised trials point to the same conclusions Figure 6, Figure 63, Figure 64).

Figure 6 – Forest plot for mammography screening vs. no screening, outcome: all-cause mortality (13 years FU), all ages



***Sources**: Gøtzsche & Jørgensen, 2013; Gøtzsche & Jørgensen, 2024,[14, 95] and Canelo-Aybar et al., 2021.[86]*

VARIOUS AGE SUBGROUPS[bb]

A. Women aged 40-49 years

Randomised controlled trials

The pooled results of six RCTs providing data for women under age fifty years, indicate that screening with mammography **does not result in a reduction of all-cause mortality** in this age group (RR: 0.99, 95% CI: 0.94-1.04, p=0.65; Appendix Figure 65). After seven years of FU, the same conclusion is drawn (RR: 1.01, 95% CI: 0.96-1.07, p=0.73; Appendix Figure 66). These results are in line with IQWiG's and Cochrane's SRs.[14, 88, 95]

The meta-analysis restricted to RCTs with a low risk of selection bias and providing data up to thirteen years after randomisation, did not identify an effect of screening on all-cause mortality either (RR: 0.98, 95% CI: 0.91-1.05, p=0.49).

B. Women aged ≥50 years

Randomised controlled trials

After thirteen years of FU, screening with mammography **does not result in a reduction of all-cause mortality** in women of fifty years or older (RR: 0.99, 95% CI: 0.97-1.02, p=0.61; Appendix Figure 67). For this age group, the seven-year FU results are restricted to studies with a high risk of selection bias. Based on these pooled data, it can be concluded that after seven years of FU screening with mammography results in a 3% lower risk of all-cause mortality (RR: 0.97, 95% CI: 0.94-1.00, p=0.03; Appendix Figure 64). Other review authors come to the same conclusion.[13, 14, 95]

The meta-analysis restricted to RCTs with a low risk of selection bias, points to no effect of screening on all-cause mortality either after thirteen years of FU (RR: 0.98, 95% CI: 0.93-1.03, p=0.36).

C. Women aged 50-69 years

No data were identified.

D. Women aged ≥ 70 years

Randomised controlled trials

The IQWiG authors identified for the older age groups usable data from one RCT (Swedish Two-County; at 7.9 years after randomisation), for which they considered the potential for bias as high. They also conclude that screening with mammography **does not result in a reduction of all-cause mortality** in women of seventy years or older (RR: 0.98, 95% CI: 0.93-1.03).[88]

Mathieu et al. (2024) provided no pooled data for all-cause mortality.[104]

[bb] In these sections, the results are presented according to the reported age groups (which may deviate from the age groups defined in the research questions).

3.2.2.2 Mortality ascribed to breast cancer

Key points

All age groups

In a general population of adult women at average risk of breast cancer, invitation to screening with mammography results in a statistically significant reduction of death ascribed to breast cancer after 13 years of FU (RR: 0.81, 95% CI: 0.74-0.87; absolute effect: 81 fewer per 100 000, 95% CI: from 111 fewer to 55 fewer; low certainty of evidence). The certainty of the evidence was downgraded one level because the risk of selection bias was high in seven out of nine RCTs and another level due to indirectness because of the age of the trials.

<50 years

In a general population of adult women younger than 50 years old and at average risk of breast cancer, invitation to screening with mammography results in a statistically significant reduction of death ascribed to breast cancer after 13 years of FU (RR: 0.84, 95% CI: 0.73-0.96; absolute effect: 49 fewer per 100 000, 95% CI: from 82 fewer to 12 fewer; low certainty of evidence). The certainty of the evidence was downgraded one level because the risk of selection bias was high in six out of eight RCTs and another level due to indirectness because of the age of the trials.

50-69 years

In a general population of adult women of 50-69 years old and at average risk of breast cancer, invitation to screening with mammography results in a statistically significant reduction of death ascribed to breast cancer after 13 years of FU (RR: 0.74, 95% CI: 0.66-0.83; absolute effect: 169 fewer per 100 000, 95% CI: from 221 fewer to 111 fewer; low certainty of evidence). The certainty of the evidence was downgraded one level because the risk of selection bias was high in five out of six RCTs and another level due to indirectness because of the age of the trials.

≥70 years

In a general population of adult women of 70 years old or older and at average risk of breast cancer, invitation to screening with mammography has no effect on breast cancer mortality (RR: 0.92, 95% CI: 0.62-1.37; absolute effect: 45 fewer per 100 000, 95% CI: from 215 fewer to 109 more; very low level of evidence). The certainty of the evidence was downgraded one level due to indirectness because of the age of the trial and two levels due to imprecision one level because the risk of selection and other biases was high for the RCT that had the highest weight in the meta-analysis, one level due to indirectness because of the age of the trials and another level because the 95% CI includes a positive as well as a harmful effect.

In all trials, this outcome was the main focus. Yet, review authors have warned that **breast cancer mortality is not a reliable outcome** and (in some studies) **biased in favour of screening**, in essence due to differential misclassification of cause of death.[14, 95] In the Health Insurance Plan (HIP, New York, 1963) trial, for instance, assignment of cause of death was unblinded in 72% of women affected by breast cancer, and post-randomisation exclusion of women who previously had breast cancer was 'most completely ascertained for screened women', hence the reliability of the reported breast cancer mortality rates can be questioned.[14, 95, 115]

ALL AGE GROUPS

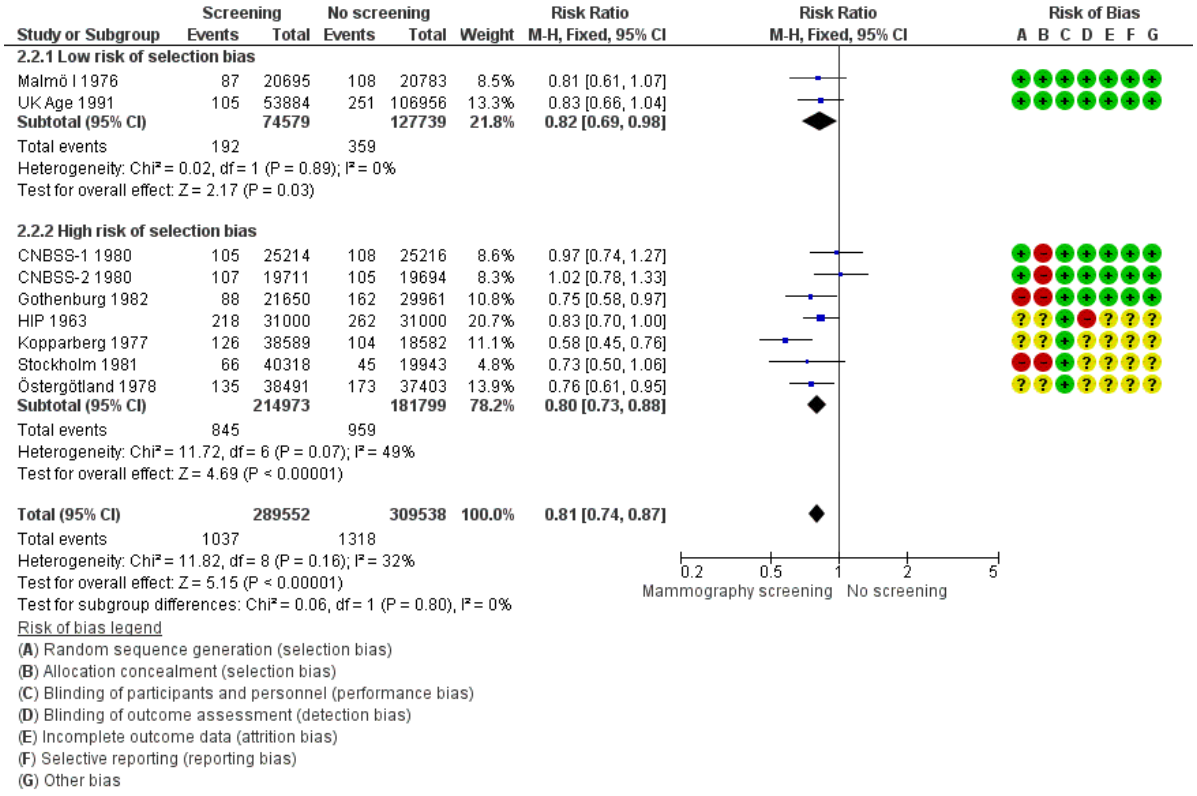
A. Randomised controlled trials

After thirteen years of FU, **screening with mammography does result in a 19% reduction of breast cancer mortality in** a general population of adult women at average risk of breast cancer (RR: 0.81, 95% CI: 0.74-0.87, p<0.001; Figure 7). Also, when the results are restricted to the two RCTs with a

low risk of selection bias, the point estimate hints to an 18% reduced risk of breast cancer mortality with screening mammography after thirteen years (RR: 0.82, 95% CI: 0.69-0.98, p=0.03).

Yet, while the effect of screening on breast cancer mortality was also statistically significant after seven years of FU (RR: 0.80, 95% CI: 0.72-0.90, p<0.001; Appendix Figure 69), this was no longer the case after twenty-four years of FU (RR: 0.95, 95% CI: 0.86-1.04, p=0.26; Appendix Figure 70).

Figure 7 – Forest plot for mammography screening vs. no screening, outcome: breast cancer mortality (13 years FU), all ages



***Sources**: Gøtzsche & Jørgensen, 2013; Gøtzsche & Jørgensen, 2024,[14, 95] and Canelo-Aybar et al., 2021.[86]*

The Canadian SR update by Bennett et al. (2024), notified a 15% reduction in the risk of breast cancer mortality with screening mammography for a general population (RR: 0.85, 95% CI: 0.78-0.93).[13] This result is reported for each age subgroup and is based on 'short-case accrual' data, i.e. reported deaths among cases of breast cancer diagnosed during the screening intervention period. The review authors assume that these data are less prone to bias from contamination because women in the control group would not have been screened until the trial was over[cc]. Bennett's sensitivity analysis, which is restricted to three RCTs with a moderate risk of bias (UK Age, Malmo I & II), is in line with our analyses (RR: 0.88, 95% CI: 0.77-1.00).[13] Gøtzsche and Jørgensen pooled three RCTs and conclude that screening with mammography does not have an effect on breast cancer mortality (RR: 0.90, 95% CI: 0.79-1.02).[14] The other review authors do not provide results on this outcome for all ages together.

[cc] Bennett et al. (2024) argue that data which include deaths occurring in all cases diagnosed to the end of the follow-up period (called 'long-case accrual') may underestimate the benefits of screening as women in the control group are more likely to have been screened after the trial.[13]

B. Observational studies

Bennett et al. (2024) pooled three cohort studies and one ecological study[dd], with the latter performed for a screening adherence analysis, which provided a relative risk reduction of 52% (RR: 0.48, 95% CI: 0.41-0.57).[13] Likewise, the pooling of seven case-control studies[ee] also pointed to a statistically significant reduction in breast cancer mortality (OR: 0.56, 95% CI: 0.49-0.64).

VARIOUS AGE SUBGROUPS

Both the meta-analyses by Canelo-Aybar et al. (2021) and the IQWiG report (2022) provided outcome estimates stratified by age group.[86, 88] A subgroup analysis by age group was also performed by Bennett et al. (2024), but these results were not incorporated in this report as they concluded that there was no significant heterogeneity between different age groups. They therefore used the same relative risks for all age groups in their calculations.[13] Gøtzsche & Jørgensen (2013; 2024) provided data separately for women under and above 50 years of age.[14, 95]

Of note, the IQWiG authors as well as Gøtzsche and Jørgensen point to an **ongoing randomised trial** investigating an extension of the existing mammography screening programme in the UK to younger age groups (47-49 years) and older age groups (> 70 years), the AgeX trial (registered at clinicaltrials.gov under NCT01081288).[14, 88, 116] The primary outcome of this trial is breast cancer mortality. Study completion is foreseen end of 2026. The feasibility and acceptability of randomisation was assessed in a pilot study (see also Table 85 in Appendix 9.4).[117]

A. Women aged 40-49 years

Randomised controlled trials

In women younger than age fifty years, the pooled result points to a **statistically significant reduction of 16% in breast cancer mortality** after thirteen years of FU (RR: 0.84, 95% CI: 0.73-0.96, p= 0.009; Appendix Figure 71). This is not the case when the meta-analysis is restricted to the two trials with a low risk of selection bias (RR: 0.80, 95% CI: 0.65-1.00, p=0.05). After seven years of FU, the effect of screening on breast cancer mortality is not statistically significant either (RR: 0.89, 95% CI: 0.77-1.04, p=0.16; Appendix Figure 72).

For short-case accrual, Canelo-Aybar et al. (2021) report in women aged 40-49 years an effect of mammography screening on breast cancer mortality that is not statistically significant (RR: 0.88, 95% CI: 0.76-1.02).[86] The exclusion of six trials at high risk of bias[ff], led to a similar result (RR: 0.86, 95% CI: 0.73-1.02). For long-case accrual, comparable results were obtained (RR: 0.92, 95% CI: 0.83-1.02; based on eight RCTs[gg]), which was not affected by exclusion of studies at high risk of bias.[86]

The IQWiG report calculated incidence density rates for women aged 45-49 years for different follow-up times; a follow-up time of ten years was chosen for the main analysis, as this follow-up time ensured according to the review authors that the largest number of studies could be included (i.e. 8 RCTs[hh]).[88] Of note, in contrast with the other SRs and our work, the IQWiG meta-analyses included the Edinburgh trial (Table 78 in Appendix 9.4).[88, 118] This trial was excluded from the other SRs as it was considered unreliable due to substantial baseline imbalances.[14, 95] A meta-analysis of these data showed a reduction in the risk of breast cancer mortality of 15%, which was borderline non-significant (incidence density ratio (IDR): 0.85, 95% CI: 0.73-1.002). A sensitivity analysis including the results of the UK

---

dd Three cohort studies and one ecological study: Coldman et al. (2014), Morrel et al. (2017), Duffy et al. (2021), Choi et al. (2021)

ee Seven case-control studies: Paap et al. (2014), Pocobelli et al. (2015), Massat et al. (2016), Ripping et al. (2017), van der Waal et al. (2017), Maroni et al. (2021), De Troeyer et al. (2023)

ff Six trials at high risk of bias: HIP, CNBSS 1, Gothenburg, Malmo II, Swedish Two County, Stockholm

gg Eight RCTs: HIP, CNBSS 1, UK Age, Gothenburg, Malmo I & II, Swedish Two County, Stockholm

hh Eight RCTs: HIP, CNBSS 1, UK Age, Gothenburg, Malmo, Swedish Two County, Stockholm, Edinburgh

Age study at 9 years follow-up instead of 10.7 years, yielded comparable results (IDR: 0.83; 95% CI: 0.70-0.99).

B. Women aged 50-69 years

Randomised controlled trials

For the subgroup of women aged 50-69 years, the pooling of the six RCTs, of which five[ii] were considered having a high risk of selection bias, resulted in a **statistically significant risk decrease of 26%** (RR: 0.74, 95% CI: 0.66-0.83; Appendix Figure 73). However, when the results are restricted to the Malmö I trial, which had a low risk of selection bias, mammography screening led to a non-statistically significant reduction in mortality ascribed to breast cancer (RR: 0.83, 95% CI: 0.66-1.04).

Canelo-Aybar et al. (2021) report, in line with our main finding, a statistically significant risk decrease of 23% (RR: 0.77, 95% CI: 0.66-0.90).[86]

C. Women aged ≥50 years

Randomised controlled trials

For women aged fifty years and older at randomisation, screening with mammography resulted in a **statistically significant reduction of 23% in breast cancer mortality** after thirteen years (RR: 0.77, 95% CI: 0.69-0.86, p<0.001; Appendix Figure 74) and a reduction of 28% after seven years of FU (RR: 0.72, 95% CI: 0.62-0.85, p=0.31; Appendix Figure 75).

When the results are restricted to the Malmö I trial, i.e. the only trial assessed as having a low risk of selection bias, there was no effect of screening on the risk of death due to breast cancer after seven (RR: 0.80, 95% CI: 0.51-1.24, p=0.31) nor after thirteen years of FU (RR: 0.86, 95% CI: 0.64-1.16, p=0.32).

Gøtzsche and Jørgensen pooled for this age group two RCTs and conclude that screening with mammography does not have an effect on breast cancer mortality (RR: 0.94, 95% CI: 0.77-1.15).[14]

D. Women aged ≥ 70 years

Randomised controlled trials

Data from RCTs are limited for the subgroup of women of seventy years or older, as only the Swedish Two-County trial and the Malmö I trial provided data that could be further analysed. In this age group, **screening with mammography does not have an effect on breast cancer mortality** (RR: 0.92, 95% CI: 0.62-1.37; p=0.69; Appendix Figure 76). It is important to note that the pooled estimator is heavily determined by the Swedish Two County study due to the more than fifteen times higher weighting, which can be observed in Appendix Figure 76.

When the results are restricted to a small subsample of the Malmö I trial, the conclusion does not change (RR: 0.98, 95% CI: 0.20-4.83, p=0.98).

Both the IQWiG authors (2022; IDR 0.91, 95 % CI: 0.61-1.36, p= 0.655) as well as Canelo-Aybar et al. (2021; RR 0.77, 0.54-1.09) conclude that screening with mammography does not have an effect on breast cancer mortality in the seventy+ age group.[86, 88] Gøtzsche and Jørgensen consider the relative effect for this subgroup as not estimable, since the only available RCT was suboptimally randomised.[14]

[ii] Five RCTs: HIP, Gothenburg, Stockholm, Kopparberg, Östergötland

Observational studies

Both the systematic review by Bennett et al. (2024) and the systematic review by Henderson et al. (2024) report the results of a large observational study[jj] (n= 1 058 013 women aged 70-84 years) which applied an emulated trial[kk] methodology.[13, 87, 120] This study, assessed by Henderson et al. (2024) as 'fair quality', reported a reduced breast-cancer mortality risk for screening between age 70 and 74 (hazard ratio (HR): 0.78, 95% CI: 0.63-0.95), but the absolute risk difference was small and not statistically significant (1.0 fewer deaths per 1000 screened, 95% CI: -2.3-0.1). For continued screening above the age of 74 years, no reduced mortality risk was found (HR: 1.00, 95% CI: 0.83-1.19).[13, 87, 120]

Mathieu et al. (2024) evaluated health benefits and harms of screening in older women (> 75 years).[104] When only looking at the two included cohort studies[ll] with a moderate risk of bias (all other observational studies had a serious or critical risk of bias), the effectiveness of continued screening varied between a non-statistically significant breast-cancer death reduction of 13% (HR: 0.87, 95% CI: 0.55-1.37) for screening in women > 75 years of age, to a 7.7% cumulative survival rate improvement for patients aged 75 years or older (p < 0.001; 95% CI not provided in the paper).[104, 121, 122]

3.2.2.3 Mortality ascribed to any cancer

Key points

All age groups

In a general population of adult women at average risk of breast cancer, invitation to screening with mammography does not have an effect on mortality due to any cancer after 9-23 years of FU (RR: 1.00, 95% CI: 0.96-1.03; absolute effect: 0 fewer per 100 000, 95% CI: from 102 fewer to 76 more; low certainty of evidence). The certainty of the evidence was downgraded one level because the risk of selection bias was high in five out of seven RCTs and another level due to indirectness because of the age of the trials.

Various age groups

No relevant RCT data were identified.

ALL AGE GROUPS

Randomised controlled trials

Screening with mammography does not have an effect on mortality ascribed to any type of cancer (RR: 1.00, 95% CI: 0.96-1.03, p=0.82; Figure 8). When the results are restricted to the two trials with a low risk of selection bias (i.e. Malmö, with 9 years FU and UK Age, with 23 years FU), a similar conclusion can be drawn (RR: 0.98, 95% CI: 0.93-1.03, p=0.42).

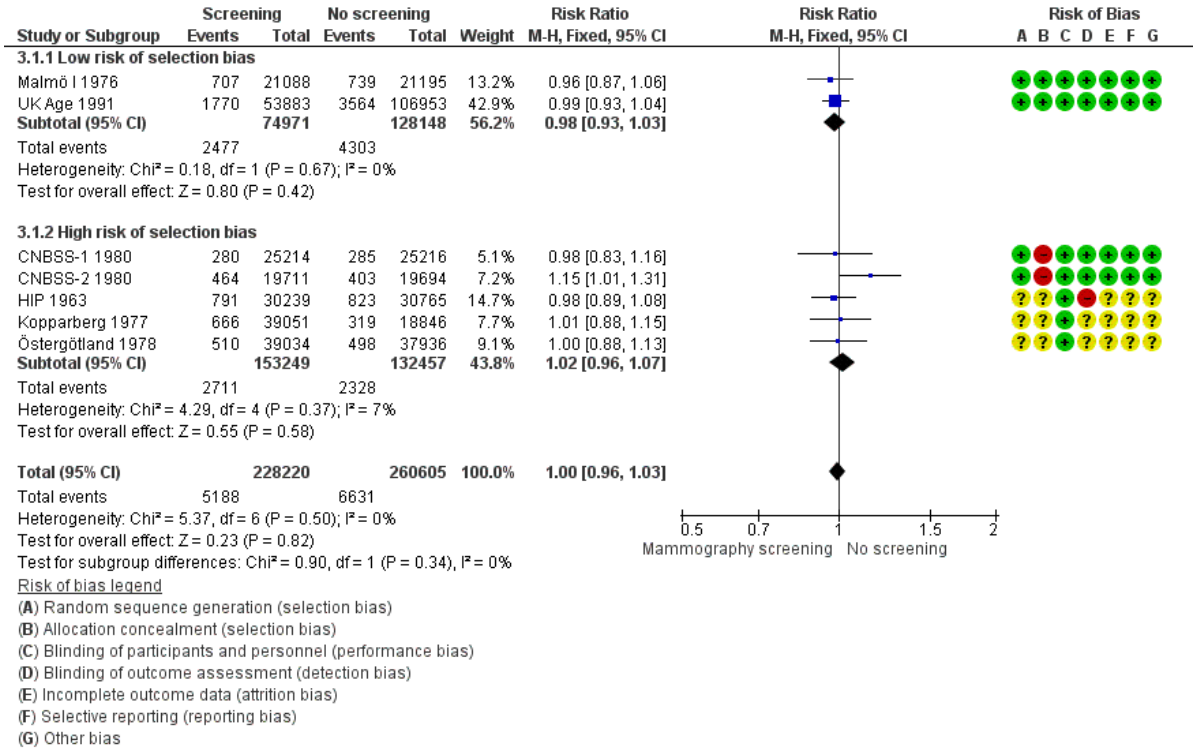
Gøtzsche and Jørgensen come to comparable results.[14] No other review authors included this outcome.

[jj] Large observational study with an emulated trial methodology: García-Albéniz et al. (2020)

[kk] Target trial emulation is a novel method using large observational studies (big data) to emulate a target trial when a RCT is not available. The method can improve the strength of observational studies by emulating key principles of randomized trials to provide a better measure of causal inference than mere association.[119]

[ll] Two cohort studies: Yang et al. (2020), Richman et al. (2023)

Figure 8 – Forest plot for mammography screening vs. no screening, outcome: any cancer mortality (9 to 23 years FU), all ages



***Sources**: Gøtzsche & Jørgensen, 2013; Gøtzsche & Jørgensen, 2024.*[14, 95]

VARIOUS AGE SUBGROUPS

No relevant data were identified.

3.2.2.4 Other cause mortality

Key points

All age groups

No data available

<50 years

In a general population of adult women younger than 50 years old and at average risk of breast cancer, invitation to screening with mammography does not have an effect on other-cause mortality after 17 years of FU (RR: 1.01, 95% CI: 0.97-1.05; absolute effect: 30 more per 100 000, 95% CI: from 91 fewer to 151 more; low certainty of evidence). The certainty of the evidence was downgraded one level due to serious risk of selection bias in four out of five included RCTs and another level due to indirectness because of the age of the trials.

50-69 years

In a general population of adult women of 50-69 years old and at average risk of breast cancer, invitation to screening with mammography does not have an effect on other-cause mortality after 17 years of FU (RR: 0.99, 95% CI: 0.95-1.04; absolute effect: 64 more per 100 000, 95% CI: from 318 fewer to 254 more; low certainty of evidence). The certainty of the evidence was downgraded

one level because the risk of selection bias was high in all included RCTs and another level due to indirectness because of the age of the trials.

≥70 years

In a general population of adult women of 70 years or older and at average risk of breast cancer, invitation to screening with mammography does not have an effect on other-cause mortality (RR: 0.99, 95% CI: 0.94-1.04; absolute effect: 279 fewer 100 000, 95% CI: from 1677 fewer to 1118 more; very low certainty of evidence). The certainty of the evidence was downgraded one level due to serious risk of selection bias, one level due to high unexplained heterogeneity of results and another level due to indirectness because of the age of the trials.

ALL AGE GROUPS

No relevant data were identified.

VARIOUS AGE GROUPS

Only Canelo-Aybar et al. (2021) provided data with regard to 'other cause mortality' (i.e. other than breast cancer).[86] Their results are very comparable to ours.

A. Women aged 40-49 years

Randomised controlled trials

For the age group of 40-49 years, **screening with mammography does not have an effect on other cause mortality** (RR: 1.01, 95% CI: 0.97-1.05, p=0.66; Appendix Figure 77). When the data are restricted to the trial with a low risk of selection bias (i.e. UK age), the same conclusion could be drawn (RR: 0.98, 95% CI: 0.93-1.04, p=0.54).

B. Women aged 50-69 years

Randomised controlled trials

Similarly, in women aged 50-69 years, **screening with mammography does not impact other cause mortality** (RR: 0.99, 95% CI: 0.95-1.04, p=0.69; Appendix Figure 78). Of note, all included RCTs were considered having a high risk of selection bias.

C. Women aged 70-74 years

Randomised controlled trials

Also for women aged 70-74 years old, the data were limited to trials with a high risk of selection bias. They pointed to no effect of **screening with mammography on other cause mortality** (70-74 years, RR: 0.99, 95% CI: 0.94-1.04, p=0.76, Appendix Figure 79).

3.2.2.5 Breast cancer stage at diagnosis

Key points

All age groups

In a general population of adult women at average risk of breast cancer, invitation to screening with mammography results in a statistically significant reduction of the probability of being diagnosed with advanced disease (stage II or higher: RR: 0.66, 95% CI: 0.61-0.72; absolute effect: 309 fewer per 100 000, 95% CI: from 354 fewer to 254 fewer; low certainty of evidence - stage III or higher: RR: 0.66, 95% CI: 0.52-0.83; absolute effect: 76 fewer per 100 000, 95% CI: from 107 fewer to 38 fewer; low certainty of evidence). The certainty of the evidence was downgraded one level because

the risk of selection bias was high in four out of five c.q. two out of three RCTs and another level due to indirectness because of the age of the trials.

<50 years

In a general population of adult women younger than 50 years old and at average risk of breast cancer, invitation to screening with mammography

- results in a statistically significant reduction of the probability of being diagnosed with stage II or higher (RR: 0.88, 95% CI: 0.78-0.99; absolute effect: 49 fewer per 100 000, 95% CI: from 90 fewer to 4 fewer; low certainty of evidence). The certainty of the evidence was downgraded one level because the risk of selection bias was high in four out of five RCTs and another level due to indirectness because of the age of the trials.
- does not affect the probability of being diagnosed with stage III or higher (RR: 0.98, 95% CI: 0.74-1.29; absolute effect: 2 fewer per 100 000, 95% CI: from 21 fewer to 23 more; very low certainty of evidence). The certainty of the evidence was downgraded one level because the risk of selection bias was high in three out of four RCTs, one level due to indirectness because of the age of the trials and another level due to imprecision because the 95% CI includes a positive as well as a harmful effect.

50-69 years

In a general population of adult women of 50-69 years old and at average risk of breast cancer, invitation to screening with mammography

- results in a statistically significant reduction of the probability of being diagnosed with stage II or higher (RR: 0.73, 95% CI: 0.66-0.81; absolute effect: 221 fewer per 100 000, 95% CI: from 278 fewer to 155 fewer; low certainty of evidence). The certainty of the evidence was downgraded one level because the risk of selection bias was high in all included RCTs and another level due to indirectness because of the age of the trials.
- results in a statistically significant reduction of the probability of being diagnosed with stage III or higher (RR: 0.62, 95% CI: 0.48-0.80; absolute effect: 65 fewer per 100 000, 95% CI: from 89 fewer to 34 fewer; low certainty of evidence). The certainty of the evidence was downgraded one level because the risk of selection bias was high in all included RCTs and another level due to indirectness because of the age of the trials.

≥70 years

No relevant RCT data were identified.

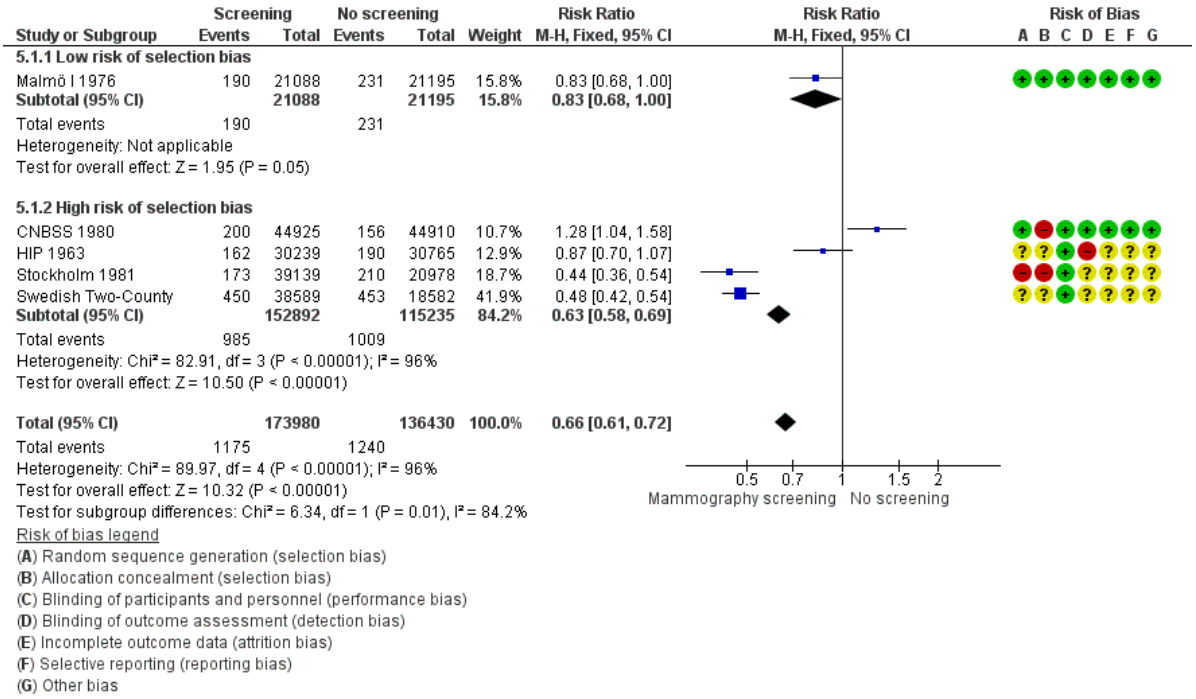
ALL AGE GROUPS

Randomised controlled trials

For this outcome, the forest plots are based on data provided in the SR by Bennett et al. (2024), which are actually based on data reported by Tarone et al. (1995).[13, 96] The meta-analysis of Canelo-Aybar et al. (2021) does not provide pooled data for all age groups together.[86] Gøtzsche & Jørgensen (2013; 2024) did not select this outcome. [14, 95]

The meta-analysis indicates that mammography screening results in a **statistically significant reduction of 34% of breast cancers diagnosed at stage II or higher** compared to usual care (RR: 0.66, 95% CI: 0.61-0.72, p<0.001; Figure 9). When the data are restricted to the single trial with a low risk of selection bias, the Malmö I study, no statistically significant result is observed (RR: 0.83, 95% CI: 0.68-1.00, p=0.05).

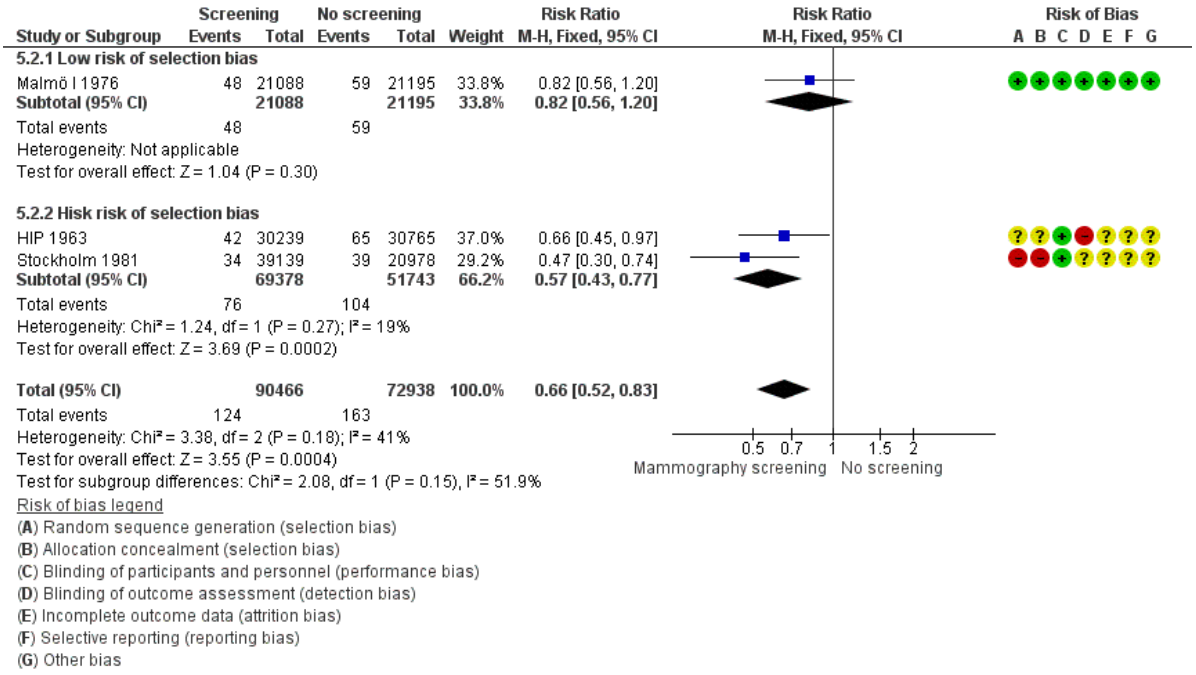
Figure 9 – Forest plot for mammography screening vs. no screening, outcome: stage II or higher at diagnosis, all ages



***Sources**: Bennett et al., 2024,[13] Tarone et al., 1995,[96] and Canelo-Aybar et al., 2021.[86]*

Similar conclusions can be drawn for the effect of screening on the risk of being diagnosed with breast cancer stage III or higher or tumour size ≥40 mm. The pooled data of three RCTS suggest that mammography screening results in a **statistically significant 34% lower risk of being diagnosed at stage III or higher** (RR: 0.66, 95% CI: 0.52-0.83, p<0.001; Figure 10).[13] But, when the data are restricted to the Malmö I trial, the single study with a low risk of selection bias, there is no evidence for a statistically significant effect of screening with mammography on this outcome (RR: 0.82, 95% CI: 0.56-1.20, p=0.30).

Figure 10 – Forest plot for mammography screening vs. no screening, outcome: stage III or higher at diagnosis, all ages



***Sources**: Bennett et al., 2024,[13] Tarone et al., 1995,[96] and Canelo-Aybar et al., 2021.[86]*

Observational studies

The SR by Bennett et al. (2024) also included observational studies.[13] One cohort study[mm], which compared screening with mammography vs. no screening, and which provided data for all age groups together, concluded that there was a reduction in stage II+ at diagnosis in those screened with mammography (IRR: 0.72, 95% CI: 0.68-0.76).[13]

VARIOUS AGE GROUPS

A. Women aged 40-49

Randomised controlled trials

In women younger than fifty years old, mammography screening **results in a statistically significantly reduced risk of being diagnosed with breast cancer at stage IIA or higher** (RR: 0.88, 95% CI: 0.73-0.99, p=0.04; Appendix Figure 80). Canelo-Aybar et al. (2021) come to the same conclusion.[86]

When the evidence is restricted to the single trial with low risk for selection bias, UK Age (1991), the point estimate is the same, but the 95% CI is wider and includes the null value, hence the result is not statistically significant (RR: 0.88, 95% CI: 0.73-1.05, p=0.16). The pooling of four studies with high risk of bias indicates the same result.

> *Of note, the SR by Bennett et al. (2024) included for age-group 40-49 (and 50-59) years only one RCT, the CNBSS study, which was assessed as very low level of evidence.[13, 96] They reported for women aged 40-49 years, a 55% higher risk for being diagnosed at Stage II or*

[mm] One cohort study: Puliti et al. (2017)[123]

higher in those screened with mammography compared to usual care (RR: 1.55, 95% CI: 1.23-2.11).[96]

It is important to note that the data collected in the CNBSS 1 (women aged 40-49 years) and 2 (women aged 50-59 years) studies are differently reported by Bennett et al. (2024) (who based their data on Tarone (1995)) and by Canelo-Aybar et al. (2021).[13, 86] *Most probably this is due to the fact that Tarone (and thus Bennett et al.) did not include patients for whom the number of nodes affected (i.e. the criterion for being assigned stage II or higher) was unknown (as reported in Miller et al. (1992a, 1992b; Table 7)),*[13, 96, 124, 125] *while Canelo-Aybar et al. (2021) did include those cases in the numerators of his analyses.*[86] *Moreover, while Canelo-Aybar et al. (2021) used as denominator all women randomised,*[86] *Tarone and Bennett et al. present as denominator all women diagnosed with invasive cancer.*[13, 96]

With regard to stage III+ breast cancer or tumour sizes ≥ 40 mm, result in women < 50 years old screening does not result in an increased risk of being diagnosed with advanced disease (RR: 0.98, 95% CI: 0.74-1.29, p=0.88; Appendix Figure 82). This result is also reported by Canelo-Aybar et al. (2021) used as denominator all women randomised,[86]

When the data are restricted to the single trial with a low risk of selection bias (i.e. UK age trial), the conclusion is the same: screening has no statistically significant impact on the risk of being diagnosed with **stage III+ breast cancer or tumour sizes ≥ 40 mm** (RR: 0.85, 95% CI: 0.57-1.28, p=0.44).

B. Women aged 50-69 years

Randomised controlled trials

No trials with a low risk of selection bias provide data for this outcome in this age group.

The pooled results of four trials with a high risk of selection bias point to a **27% lower risk of being diagnosed with a stage II+ breast cancer as a result of screening** (RR:0.73, 95% CI: 0.66-0.81), p<0.001; Appendix Figure 81). The result obtained by Canelo-Aybar et al. (2021) and based on the same trials is somewhat different (RR: 0.80, 95% CI: 0.64-1.00), as their estimate is based on a random effects model, while we opted for a fixed effects model.[86]

Also for this age group, Bennett et al. (2024) included only the CNBSS study, which was assessed as providing very low level of evidence.[13, 96] They concluded that in women aged 50-59 years screening with mammography does not have a statistically significant effect on the risk of being diagnosed at Stage II or higher (RR: 1.55, 95% CI: 1.23-2.11).[96]

Using the **stage III or higher** definition, in women aged 50-69 years, **screening reduces the risk of being diagnosed with advanced disease with 38%** (RR: 0.62, 95% CI: 0.48-0.80, p<0.001; Appendix Figure 83); yet, this result is again solely based on three trials with a high risk of selection bias. This result is also reported by Canelo-Aybar et al. (2021).[86]

C. Women aged 50-74 years

Randomised controlled trials

Canelo-Aybar et al. (2021) identified one trial (the Swedish Two County trial, assessed as having a high risk of selection bias) which provided data for this outcome obtained in women aged 50-74 years.[86] Their results indicate that the risk of being diagnosed with an advanced stage breast cancer is 36% (stage II+ breast cancer, RR: 0.64; 95% CI: 0.55-0.73) to 37% (stage III or higher, RR: 0.63, 95% CI: 0.45-0.89) lower for those who are screened compared to the control group[nn].[86]

[nn] Because Canelo-Aybar et al. (2021) did not provide raw data for numerators and denominators, no forest plot could be developed.[86]

D. Women aged ≥ 70 years

Observational studies

The meta-analysis by Bennett et al. (2024) included also observational studies.[13] One quasi-experimental study[oo] compared the diagnosis of advanced stage breast cancer defined as stages III and IV as per the TNM classification before and after breast cancer screening programme implementation in older age groups.[126] In both screening periods (i.e. 1998-2002 and 2003-2011), a reduction in the diagnosis of advanced stage breast cancer was observed in the subgroup of 70-75 years old (IRR: 0.79, 95% CI: 0.71-0.87, and 0.88, 95% CI: 0.81-0.97, respectively), whereas in the older sub-group (76-80 years old) this was not the case (IRR: 1.04, 95% CI: 0.94-1.17, and 1.02, 95% CI: 0.92-1.13, respectively).[13]

Mathieu et al. (2024) identified one observational study that investigated whether prior mammography use was associated with cancer stage at diagnosis in older women (> 65 years), based on a retrospective cohort study using a linked Medicare-Tumour Registry database.[127] They report an increased risk of late-stage disease among non-users of mammography in women aged 75-84 years (adjusted odds ratio (aOR): 3.64, 95% CI: 2.96-4.48) and ≥85 years (aOR: 6.87, 95% CI: 3.97-11.9). A cohort study which compared primary breast cancers detected by mammography vs. detected by a physician or by the patient, indicated that mean tumour size was higher in the latter group (3.02 cm vs 1.53 cm, p<0.001).[128] Of note, both studies were assessed by Mathieu et al. as having a serious risk of bias.[104]

3.2.2.6 Overdiagnosis

Key points

All age groups

In a general population of adult women at average risk of breast cancer, invitation to screening with mammography results in a statistically significant increase of overdiagnosis 7-9 years after randomisation (RR: 1.29, 95% CI: 1.23-1.35; absolute effect: 332 more per 100 000, 95% CI: from 263 more to 400 more; low certainty of evidence). Among adult women at average risk of breast cancer, and invited to screening, 27% of all cancers (invasive and non-invasive) detected were overdiagnosed.

The certainty of the evidence was downgraded one level because the risk of selection bias was high in five out of seven RCTs and another level due to indirectness because of the age of the trials.

After 10-15 years of FU, statistically significantly more women were diagnosed with breast cancer in the screened group compared to the control group (RR: 1.10, 95% CI: 1.05-1.16, p<0.001). Overall, 9.2% of all cancers detected in the screening group were overdiagnosed.

<50 years

In a general population of adult women younger than 50 years old and at average risk of breast cancer, invitation to screening with mammography results in a statistically significant increase of overdiagnosis 7-9 years after randomisation (RR: 1.20, 95% CI: 1.11-1.31; absolute effect: 178 more per 100 000, 95% CI: from 98 more to 276 more; low certainty of evidence). Among women younger than 50 years old and at average risk of breast cancer, and invited to screening, 23% of all cancers (invasive and non-invasive) detected were overdiagnosed.

The certainty of the evidence was downgraded one level because the risk of selection bias was high in two out of three RCTs and another level due to indirectness because of the age of the trials.

[oo] De Glas et al. (2014)[126]

50-69 years

In a general population of adult women of 50-69 years old and at average risk of breast cancer, invitation to screening with mammography results in a statistically significant increase of overdiagnosis (RR: 1.26, 95% CI: 1.10-1.44; absolute effect: 482 more per 100 000, 95% CI: from 185 more to 815 more; very low certainty of evidence). Among women of 50-69 years old and at average risk of breast cancer, and invited to screening, 21% of all cancers (invasive and non-invasive) detected were overdiagnosed.

The certainty of the evidence was downgraded one level because the risk of selection bias was high in the RCT, another level due to indirectness because of the age of the trials and another level due to imprecision because the outcome was reported in only one inadequately randomised trial.

≥70 years

No relevant RCT data were identified.

For a correct interpretation of the results presented in this section, the reader is referred to chapter 2, section 2.2.2, which elaborates on the concept of overdiagnosis and the methods used to calculate this outcome.

Various approaches to calculate overdiagnosis were used in the SRs. Bennett et al. (2024) estimated overdiagnosis using a cumulative-incidence approach[pp].[13] Canelo-Aybar et al. (2021) calculated overdiagnosis as the difference in the cumulative number of breast cancers in the groups invited and not invited to screening, expressed: (a) as a percentage of the number of cancers in the screening group (population perspective) or (b) as a percentage of the cancers diagnosed during the screening phase of the trial in the invited to screening group (individual perspective).[86] In the present report only results from the population perspective are copied.

The IQWiG authors reported overdiagnosis in two ways: as the risk for all screened participants to be overdiagnosed, and as the risk of overdiagnosis among women who received a breast cancer diagnosis during the screening period. The results were not pooled.[88] Their methodology is based on a comparison of the breast cancer incidence of the screened women with a control group without screening, in which only symptomatic breast cancer cases were recorded.

Gøtzsche & Jørgensen (2013; 2024) measured overdiagnosis as the difference in incidence between the screening and control arm at the latest time of follow-up in those trials that did not offer mammography screening to the control group at the end of the intervention phase.[14, 95]

ALL AGE GROUPS

Randomised controlled trials

Neither Bennett et al. (2024), nor Canelo-Aybar et al. (2021), nor the IQWiG authors provided results on overdiagnosis for all ages combined.[13, 86, 88] Gøtzsche & Jørgensen (2013; 2024) provide these results.[14, 95]

The pooled results of the trials indicate that **statistically significantly more women were diagnosed with breast cancer in the screened group compared to the control group** (RR: 1.29, 95% CI: 1.23-1.35, p<0.001; Figure 11). Importantly, the difference in cancer detection (DCIS & invasive breast cancer) between the screened group (i.e. 3906/249 025 = 1.57%) and the control group (i.e.

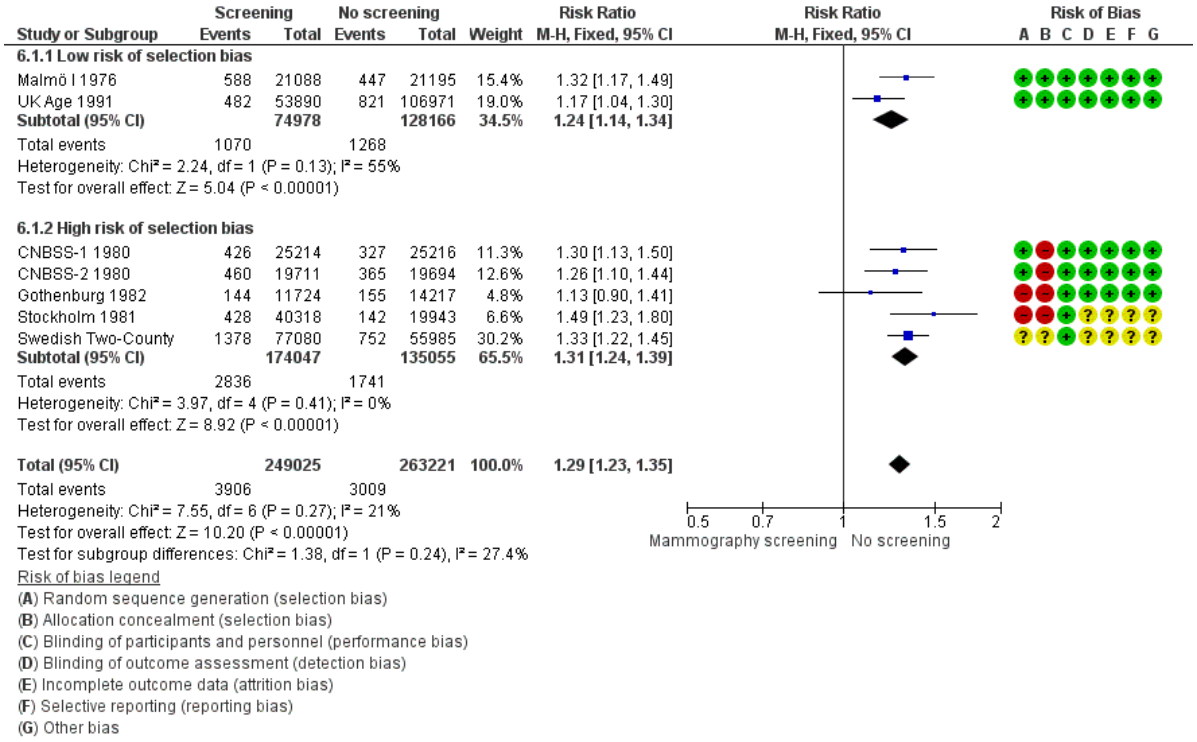
---

pp Bennett et al. (2024) argue that '*studies have suggested this to be a robust method to estimate overdiagnosis from RCT data in which there are several years of follow-up after screening stops and the control group is never screened. Cited limitations of this approach include limited external validity, diluted estimates of overdiagnosis, and a dependence on appropriate length of follow-up time*'.[13]

3009/263 221 = 1.14%) was thus 0.43%. Applying this proportion to the number of women who was invited to screening (n=249 025) results in 1071 overdiagnosed cases, or **27% of all cancers detected in the screening group** (= 1071/3906).

Based on the pooled estimate of the trials which were adequately randomised, i.e. the Malmö I study and UK Age, the same conclusion can be drawn (RR: 1.24, 95% CI: 1.14-1.34, p<0.001).

Figure 11 – Forest plot for mammography screening vs. no screening, outcome: overdiagnosis, all ages – after 7 to 9 years FU



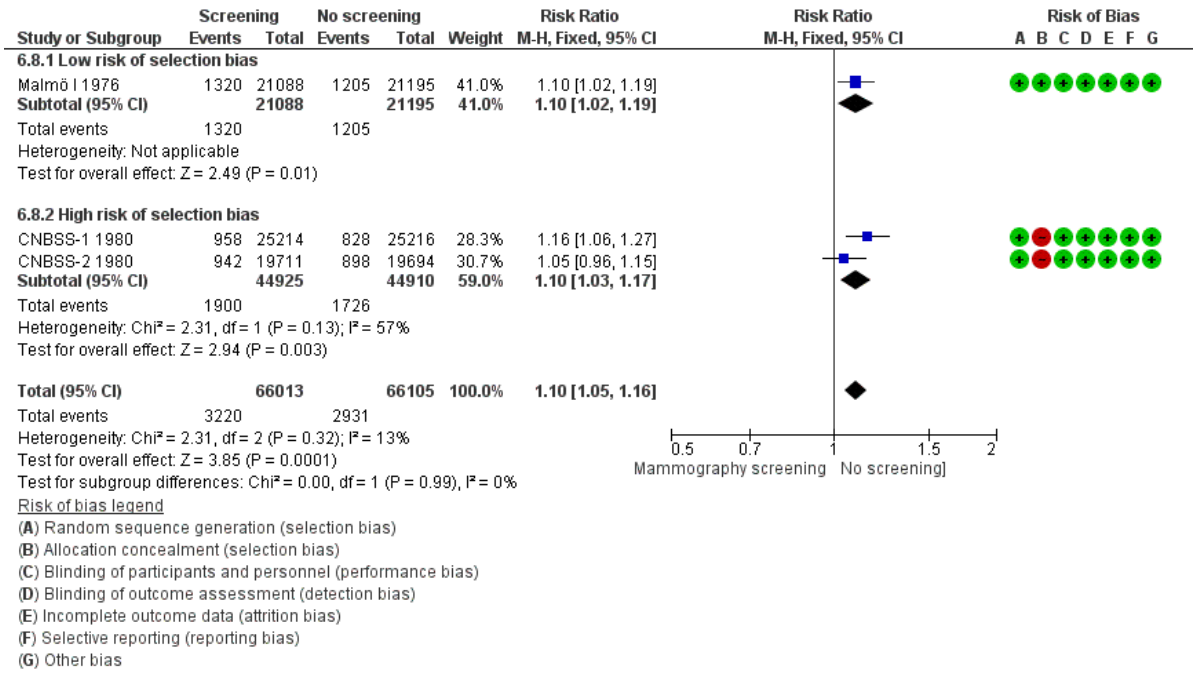
***Sources**: Gøtzsche & Jørgensen, 2013; Gøtzsche & Jørgensen, 2024,[14, 95] and Canelo-Aybar et al., 2021.[86] Data for CNBSS, Malmö I and UK Age was collected 7-9 years after randomisation, for the Stockholm and Swedish Two-County studies before the control group was offered screening, but for the Gothenburg study immediately after screening of the control group at the end of the study.*

Actually, accurate overdiagnosis assessment necessitates a sufficiently long period of follow-up after screening concludes (which corresponds to lead time) in order to allow catch-up of diagnoses in the unscreened group.[129] Since there is currently no consensus on the average lead time for breast cancer, it has been suggested that a minimum of five to ten years post-screening is required to ensure adequate longitudinal follow-up.[130, 131] Yet, to avoid contamination during this follow-up period, the control group nor the study group should be screened after the end of the trial.

For this long-term follow-up only two RCTs provide data: Malmö, CNBSS (I and II; Figure 12). The pooled results of the trials indicate that **statistically significantly more women were diagnosed with breast cancer in the screened group compared to the control group** (RR: 1.10, 95% CI: 1.05-1.16, p<0.001). Importantly, the difference in cancer detection (DCIS & invasive breast cancer) between the screened group (i.e. 3220/66013 = 4.88%) and the control group (i.e. 2931/66105 = 4.43%) was thus 0.45%. Applying this proportion to the number of women who was invited to screening (n=66013) results in 297 overdiagnosed cases, or 9.2% of all cancers detected in the screening group (= 297/3220).

When interpreting the results, one should be cautious as other scholars noted that at the end of the CNBSS I and II trial, 26% and 17%, respectively, had a mammography.[95]

Figure 12 – Forest plot for mammography screening vs. no screening, outcome: overdiagnosis, all ages – after 10 to 15 years FU



***Sources**: Gøtzsche & Jørgensen, 2024; Miller, 1993; Zackrisson et al., 2006*[14, 132, 133]

VARIOUS AGE GROUPS[qq]

A. Women aged 40-49 years

Randomised controlled trials

In women of 40-49 years old, the meta-analysis indicate that **statistically significantly more women of 40-49 years old were diagnosed with breast cancer (invasive and non-invasive breast cancer) in the screened group compared to the control group** (RR: 1.20, 95% CI: 1.11-1.31, p<0.001; Appendix Figure 84). Bennett et al. (2024) come to the same conclusion.[13] A similar result is obtained when the analysis is restricted to the UK Age trial, with a low risk of selection bias (RR: 1.17, 95% CI: 1.04-1.30, p=0.007).

Of note, the difference in cancer detection (invasive and non-invasive breast cancer) between the screened group (i.e. 1052/90 828 = 1.16%) and the control group (i.e. 1303/146 404 = 0.89%) was thus 0.27%. Applying this proportion to the number of women who was invited to screening (n=90 828) results in 245 overdiagnosed cases, or **23% of all cancers detected in the group** invited for **screening** (= 245/1052).

Canelo-Aybar et al. (2021) calculated overdiagnosis differently (cf. supra).[86] Based solely on the CNBSS 1 trial (with a high risk of selection bias), they estimated that in women aged 40-49 years,

[qq] While Bennett et al. (2024) provide in their SR results for various age groups,[13] we opted to not include their results, since they are based on longer follow-up periods (and hence difficult to compare with the "all ages group" reported supra), but also since in their results the denominators correspond to the overall randomized trial (*personal communication*).

12.4% (95% CI: 9.9-14.9%) of diagnosed breast cancers in the populations invited to screen were overdiagnosed.[86]

The IQWiG authors used the data from two trials (CNBSS and HIP), both considered as having a high risk of selection bias, and ignored the other trials as in the latter a final screening in the control group was performed which carries the risk of contamination.[88] In the CNBSS study, the risk of overdiagnosis among screened women was 0.41%, and 31.6% among women diagnosed with breast cancer. In the HIP trial, the respective estimates for overdiagnosis were much lower: 0.01% and 2.63% for the age group of 45-49 years; 0.04% and 8.06% for the age group of 40-49 years. Since all estimates, including those from trials that had a screening round for the control group at the end of the trial period, point in the same direction and are in between the results of the CNBSS study and the HIP trial, the IQWiG authors conclude that there is a hint of harm due to overdiagnosis for breast cancer screening in this age group.[88]

B. Women aged 49-52

Observational studies

When looking at overdiagnosis data from observational studies, it should be realised that although they may provide more contemporary estimates of overdiagnosis, they are also more susceptible to bias due to confounding.[64, 134]

Bennett et al. (2024) cite an observational cohort study from Norway of women screened between the age of 49 and 52 years, combining record linkage and questionnaire information. This study by Lund et al. (2018) found a 1.1% increased cumulative incidence rate or a 13.0% increased relative risk of breast cancer in screened women, but the cumulative incidence rate (CIR) was non-statistically significant and negative when only invasive cancers were included (CIR -0.2%, 95% CI: -9.1 – 8.8).[135]

C. Women aged 50-69 years

Randomised controlled trials

For this age group, we had to rely on one trial, CNBSS-2, which had a high risk of selection bias. The data illustrate that **statistically significantly more women of 50-69 years old were diagnosed with breast cancer (invasive and non-invasive breast cancer) in the screened group compared to the control group** (RR: 1.26, 95% CI: 1.10-1.44; p<0.001; Appendix Figure 85).

The difference in cancer detection (invasive and non-invasive breast cancer) between the screened group (i.e. 460/19 711 = 2.33%) and the control group (i.e. 365/19 694 = 1.85%) was thus 0.48%. Applying this proportion to the number of women who was invited to screening (n = 19 711) results in 95 overdiagnosed cases, or **21% of all cancers detected in the screening group** (= 95/460).

Canelo-Aybar et al. (2021) pooled for this age group the data of the CNBSS and the Malmö trials, and estimated for women aged 50-69 years, an overdiagnosis of 10.1% (95% CI: 8.6-11.6%).[86]

Observational studies

A large cohort study from Italy (400 000 women screened between 50-69 years) reported a 10% excess risk of in situ and invasive breast cancer (IRR: 1.10, 95% CI: 1.06-1.14) and a 5% increased risk for invasive cancers only (IRR: 1.05, 95% CI: 1.01-1.09).[123]

D. Women aged > 70 years

Randomised controlled trials

For women aged > 70 years, IQWiG found only one study for which data on this age group were available (i.e. the Swedish Two-County Study).[88] As **final screening of the control group was done** in this trial, the estimates are likely an **underestimate of the true extent of overdiagnosis** and not to be considered reliable. Keeping this in mind, the risk of overdiagnosis for all women invited to screening was 0.48%, and the risk of overdiagnosis in women diagnosed with breast cancer during the screening period was 19.45% (no 95% CIs provided).[88]

Observational studies

Mathieu et al. (2024) also looked at overdiagnosis in women > 75 years.[104] Due to lack of data from clinical trials in this age group, only observational and modelling studies were included. There was consistent evidence that screening older age groups leads to increased overdiagnosis, compared to younger age groups.[104] In a cohort study at moderate risk of bias, 47% and 54% of breast cancer cases among screened women aged 75-84 years and ≥ 85 years, respectively, were considered overdiagnoses.[122] Another cohort study at moderate risk of bias estimated that 8.3% of breast cancers diagnosed in women aged 75 years and older were overdiagnosed.[121]

DISCUSSION

Both pooled estimates from RCTs and results from observational studies suggest harms of screening due to overdiagnosis. Yet, several methods are being used to calculate overdiagnosis,[136] and the **estimated amount of overdiagnosis depends greatly on the way it is calculated**. The differences relate to which cases are included in the numerator and, especially, on the choice of denominator. Even when analysing the same trials, different meta-analyses come up with diverging estimates. As a result, comparison (and interpretation) of estimates for overdiagnosis which are reported across studies and meta-analyses, is not straightforward.

3.2.2.7 Interval cancers

Interval cancers are cancers that are detected between screening rounds, hence by definition only in the intervention arm of RCTs that compare invitation to screening vs. not being invited to screening.

Key points

All age groups

In a general population of adult women at average risk of breast cancer who were invited for a screening programme with mammography (with an interval between 13 and 24 months), on average 3.1 interval cancers per 1000 screened women (95% CI: 2.6-3.7) over a follow-up of 4.8-7.0 years (with 2.8 (95% CI: 2.4-3.3) invasive interval cancers per 1000 women and 0.2 (95% CI: 0.1-0.5) in situ interval cancers) are observed. The certainty of the evidence was considered as very low; it was downgraded one level due to serious risk of bias (Lack of reporting for how interval cancers were detected and unclear reporting on who was used in the analysis[13]), one level due to indirectness because of the age of the trials and one level due to inconsistency.

<50 years

In a general population of adult women younger than 50 years old and at average risk of breast cancer who were invited for a screening programme with mammography (with an interval between 13 and 24 months), on average 3.0 interval cancers per 1000 screened women (95% CI: 2.1-4.2) over a follow-up of 4.8-7.0 years (2.8 invasive and 0.2 in situ interval cancers) are observed, but the evidence is very uncertain. The certainty of the evidence was considered as very low; it was downgraded one level due to serious risk of bias, one level due to indirectness because of the age of the trials and one level due to imprecision as the results were derived from only one (inadequately) randomised trial.

50-69 years

In a general population of adult women of 50-59 years old and at average risk of breast cancer, who were invited for a screening programme with mammography, on average 1.9 invasive interval cancers (95% CI: 1.2-3.0; no ductal carcinoma in situ (DCIS)) per 1000 women were detected over a follow-up period of 4.8-7 years, but the evidence is very uncertain. The certainty of the evidence was considered as very low; it was downgraded one level due to serious risk of bias, one level due

to indirectness because of the age of the trials and one level due to imprecision as the results were derived from only one (inadequately) randomised trial.

≥70 years

No relevant RCT data were identified.

Cancers detected in between screening rounds are called 'interval cancers'. On average, interval cancers are thus the more aggressive and fast growing cancers, which do not benefit from screening.[14, 51, 95]

Interval cancers can, by definition, only be measured in a screened cohort. Results from meta-analyses and systematic reviews are therefore descriptive.

The results presented in this section are restricted to the biennial screening interval (and where not available, 18 months interval), as this is current practice in the various screening programmes in Belgium. For the comparison of interval cancers across different screening intervals, the reader is referred to section 3.2.3.5.

ALL AGE GROUPS

Randomised controlled trials

Bennett et al. (2024) pooled data from the Stockholm and Gothenburg trials (both considered at high risk of selection bias) with a screening interval between 13 and 24 months, and report 3.1 interval cancers per 1000 women (DCIS and interval cancers, 95% CI: 2.6-3.7) over a follow-up of 4.8-7.0 years (screening interval 18 months).[13] They estimated that 2.8 (95% CI: 2.4-3.3) interval cancers per 1000 women were invasive and 0.2 (95% CI: 0.1-0.5) were in situ interval cancers.[13]

Of note, no relevant results for the outcome interval cancers were identified in the SR by Mathieu et al. (2024), the SR by Canelo-Aybar et al. (2021), the Cochrane SR and its 2024 update, nor in the IQWiG report.[14, 86, 88, 95, 104]

VARIOUS AGE GROUPS

A. Women aged 39-49 years

Randomised controlled trials

Bennett et al. (2024) report data of the Gothenburg trial (considered at high risk of selection bias), with 3.0 (95% CI: 2.1-4.2) interval cancers (2.8 invasive and 0.2 in situ) detected per 1000 women screened over a follow-up period of 4.8-7 years.[13]

B. Women aged 50-59 years

Randomised controlled trials

Based on the Gothenburg trial, Bennett et al. (2024) report that 1.9 invasive interval cancers (95% CI: 1.2-3.0; no ductal carcinoma in situ (DCIS)) per 1000 women were detected over a follow-up period of 4.8-7 years.[13]

B. Women aged ≥60 years

No relevant data were identified.

3.2.2.8 'Treatment-related morbidity'

Key points

All age groups

In a general population of adult women at average risk of breast cancer, invitation to screening with mammography,

- results in a statistically significant increase of the probability of having breast surgery (mastectomy and lumpectomy, RR: 1.35, 95% CI: 1.26-1.44; absolute effect: 473 more per 100 000, 95% CI: from 352 more to 595 more; low certainty of evidence). The certainty of the evidence was downgraded one level because the risk of selection bias was high in four out of five RCTs and another level due to indirectness because of the age of the trials;
- results in a statistically significant increase of the probability of having radiotherapy (RR: 1.32, 95% CI: 1.16-1.50; absolute effect: 286 more per 100 000, 95% CI: from 143 more to 446 more; low certainty of evidence). The certainty of the evidence was downgraded one level because the risk of selection bias was high in one out of two RCTs and another level due to indirectness because of the age of the trials;
- does not effect the probability of having chemotherapy (RR: 0.96, 95% CI: 0.78-1.19; absolute effect: 14 fewer per 100 000, 95% CI: from 79 fewer to 68 more; very low certainty of evidence). The certainty of the evidence was downgraded one level because the risk of selection bias was high in one out of two RCTs, one level due to imprecision because the 95% CI includes a positive as well as a harmful effect and another level due to indirectness because of the age of the trials.

<50 years

In a general population of adult women younger than 50 years old and at average risk of breast cancer, invitation to screening with mammography,

- results in a statistically significant increase of the probability of having breast surgery (mastectomy and lumpectomy, RR: 1.33, 95% CI: 1.15-1.53; absolute effect: 410 more per 100 000, 95% CI: from 186 more to 658 more; very low certainty of evidence). The certainty of the evidence was downgraded one level because of the high risk of selection bias, one level due to indirectness because of the age of the trials and another level due to imprecision because the outcome was reported in only one (non-adequately) randomised trial.

No relevant RCT data were identified for the outcomes radiotherapy and chemotherapy.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

For a correct interpretation of this outcome, the **historical context** should be acknowledged. It is only in the eighties-nineties of last century that RCTs were published which compared breast conserving surgery followed by loco-regional radiation therapy with total mastectomy in stage I-II breast cancer patients.[137-141] The favourable results, indicating no difference in local tumour control, recurrence free survival and overall survival between the two treatment approaches, were later also confirmed in papers with long follow-up results,[142-146] as well as in systematic reviews and meta-analyses,[147-150] and adopted in clinical guidelines.[151-153]

Hence, **during most RCTs on breast cancer screening** (for an overview, see Appendix 9.4), **breast conserving surgery was not a treatment option yet,** and, as a consequence, mastectomy rates reflect therapy rather than morbidity.[13]

ALL AGE GROUPS

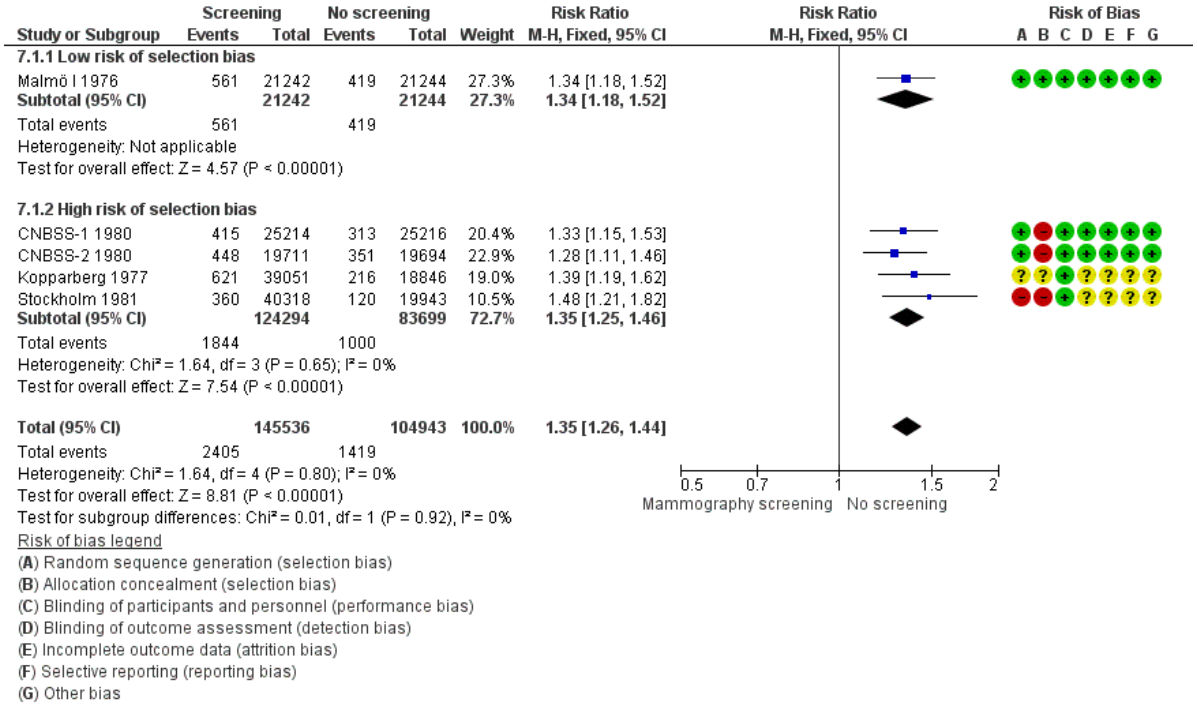
Randomised controlled trials

For this outcome we relied on the data provided by Gøtzsche & Jørgensen (2013; 2024).[14, 95] For all age groups combined, **screening with mammography results** in **statistically significantly more surgical interventions** (mastectomy and lumpectomy) compared to usual care (RR: 1.35, 95% CI: 1.26-1.44, p<0.001; Figure 13).

The same conclusion can be drawn when the results are limited to RTCs with a low risk of selection bias (RR: 1.34, 95% CI: 1.18-1.52, p<0.001).

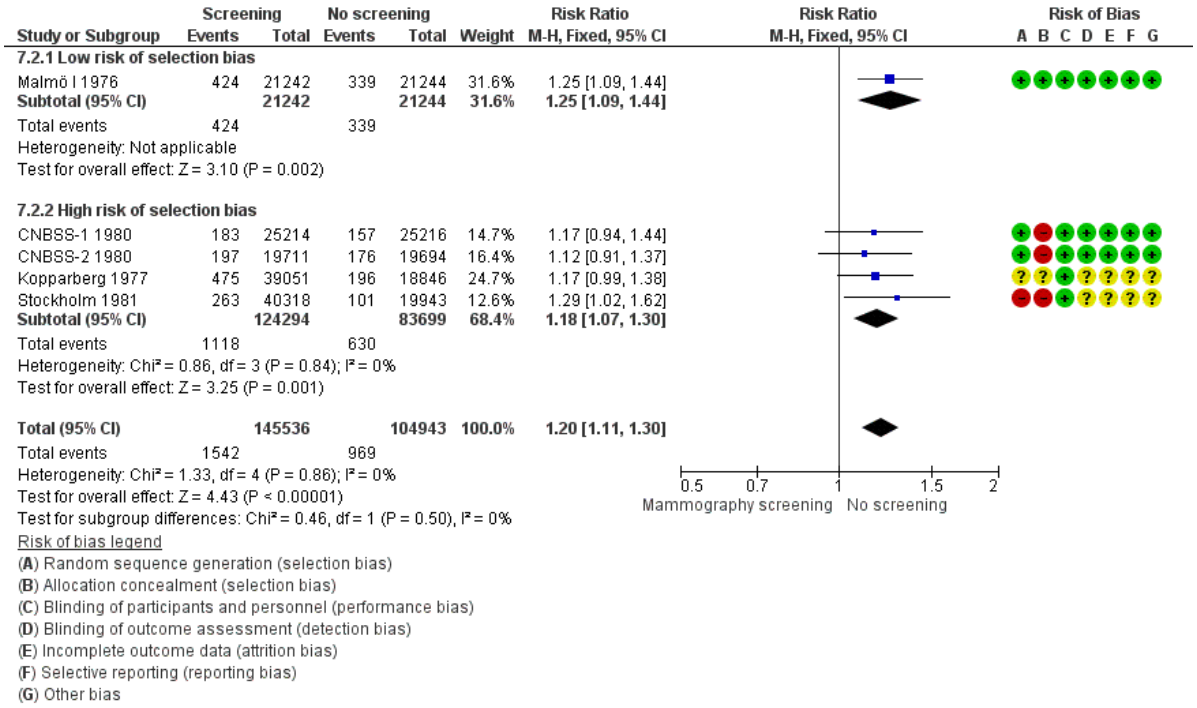
When the outcome is restricted to mastectomies, similar results were obtained (RR: 1.20, 95% CI: 1.11-1.30, p<0.001; Figure 14). Bennett et al. (2024), Canelo-Aybar et al. (2021) and Gøtzsche & Jørgensen (2013; 2024) considered in their meta-analyses only mastectomies and come to comparable results as ours (RR: 1.20, 95% CI: 1.11-1.30).[13, 86] Gøtzsche & Jørgensen (2013; 2024) noted that the increased surgery rate could not be explained by the excess of detected tumours at the first screening round, but that it seemed to persist.[14, 95]

Figure 13 – Forest plot for mammography screening vs. no screening, outcome: mastectomy and lumpectomy, all ages



***Sources**: Gøtzsche & Jørgensen, 2013; Gøtzsche & Jørgensen, 2024,[14, 95] and Canelo-Aybar et al., 2021.[86]*

Figure 14 – Forest plot for mammography screening vs. no screening, outcome: mastectomy, all ages

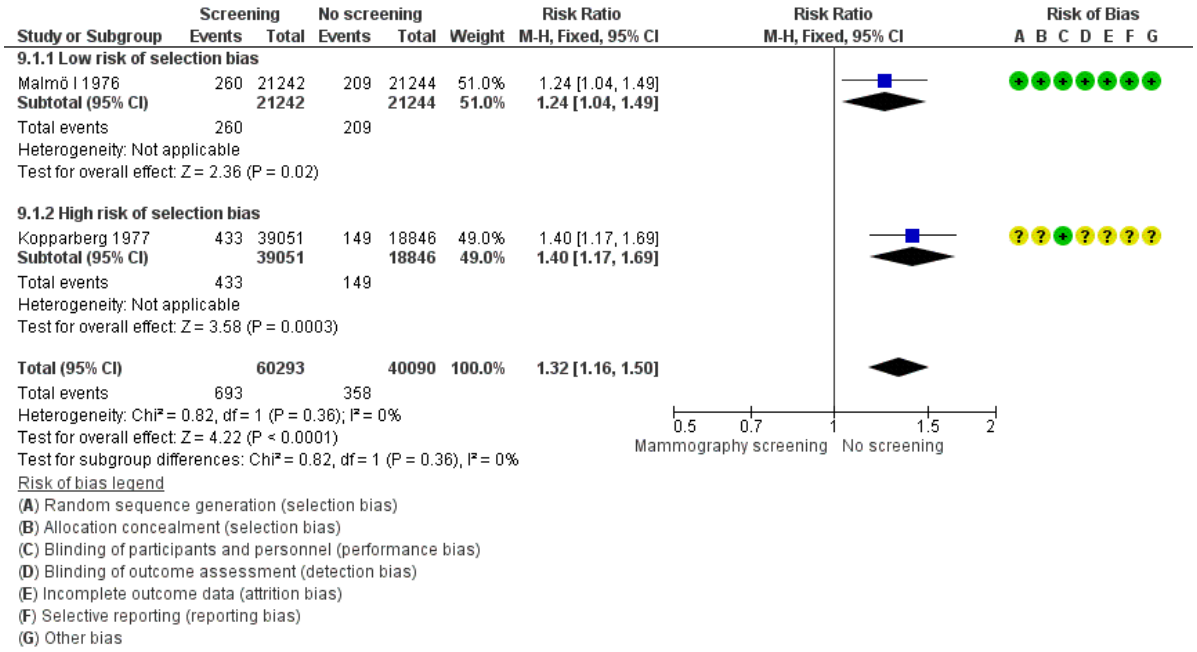


***Sources**: Gøtzsche & Jørgensen, 2013; Gøtzsche & Jørgensen, 2024,[14, 95] and Canelo-Aybar et al., 2021.[86]*

Likewise, **the probability of receiving radiotherapy is 32% higher in screened women** (RR: 1.32, 95% CI: 1.16-1.50, p<0.001; Figure 15) compared to the control group. Bennett et al. (2024) come to the same conclusion (RR: 1.32, 95% CI: 1.16-1.50).[13]

This increased risk is also noted in the single study with a low risk of selection bias, the Malmo I trial (RR: 1.24, 95% CI: 1.04-1.49, p=0.02Figure 15).

Figure 15 – Forest plot for mammography screening vs. no screening, outcome: radiotherapy, all ages



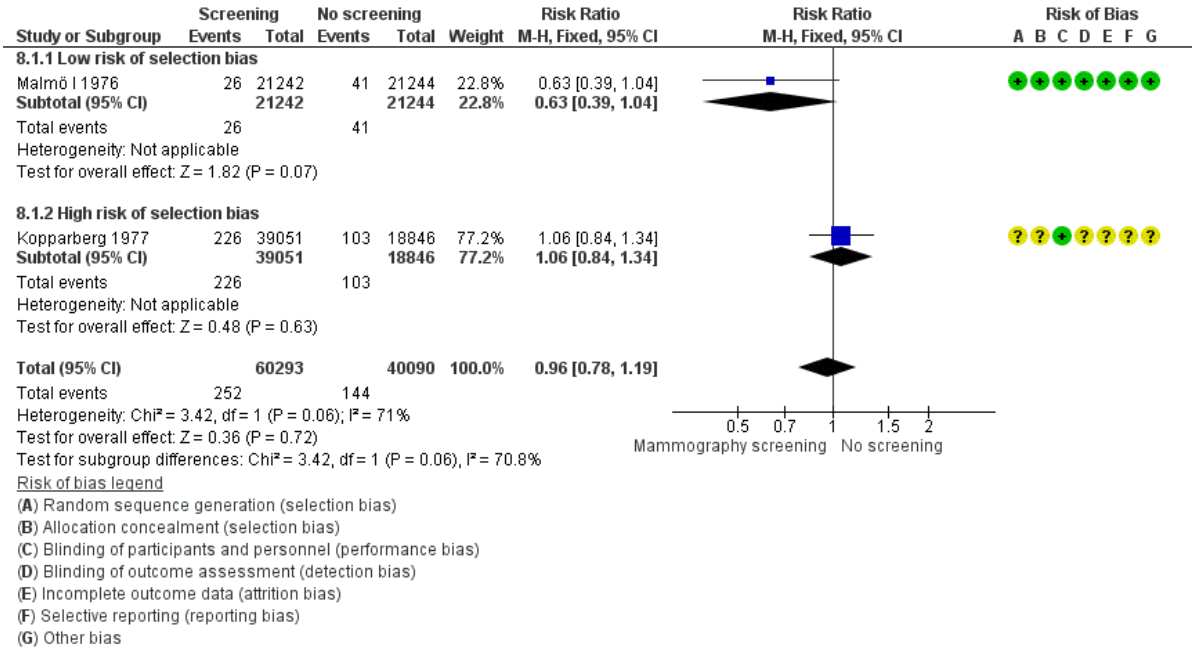
***Sources**: Gøtzsche & Jørgensen, 2013; Gøtzsche & Jørgensen, 2024,[14, 95] and Canelo-Aybar et al., 2021.[86]*

The data on **systemic treatment** are also scarce and based on the same two Swedish trials. The meta-analysis indicates no statistically significant effect of screening on the administration of chemotherapy (RR: 0.96, 95% CI: 0.78-1.19, p=0.72; Figure 16). Importantly, the forest plot illustrates **inconsistent results**: the Malmö study (low risk of selection bias) suggests that women who are screened receive less **chemotherapy**, while the Kopparberg trial suggests the opposite. However, neither of these results were statistically significant (Malmö, RR: 0.63, 95% CI: 0.39-1.04, p=0.07; Kopparberg, RR: 1.06, 95% CI: 0.84-1.34).

Bennett et al. (2024) come to the same conclusion as ours (RR: 0.96, 95% CI: 0.78-1.19).[13] Canelo-Aybar et al. (2021) pooled the same RCTs, but come to a slightly different pooled RR (RR: 0.86, 95% CI: 0.53-1.40), which may be ascribed to a different weighing[rr] of the studies. The authors noted that the result was uncertain.[86]

[rr] Canelo-Aybar et al. (2021) pooled data using a random-effects model with Review Manager v5.3,[86] while Bennett et al. used the DerSimonian and Laird random-effects method with R software.[13]

Figure 16 – Forest plot for mammography screening vs. no screening, outcome: chemotherapy, all ages

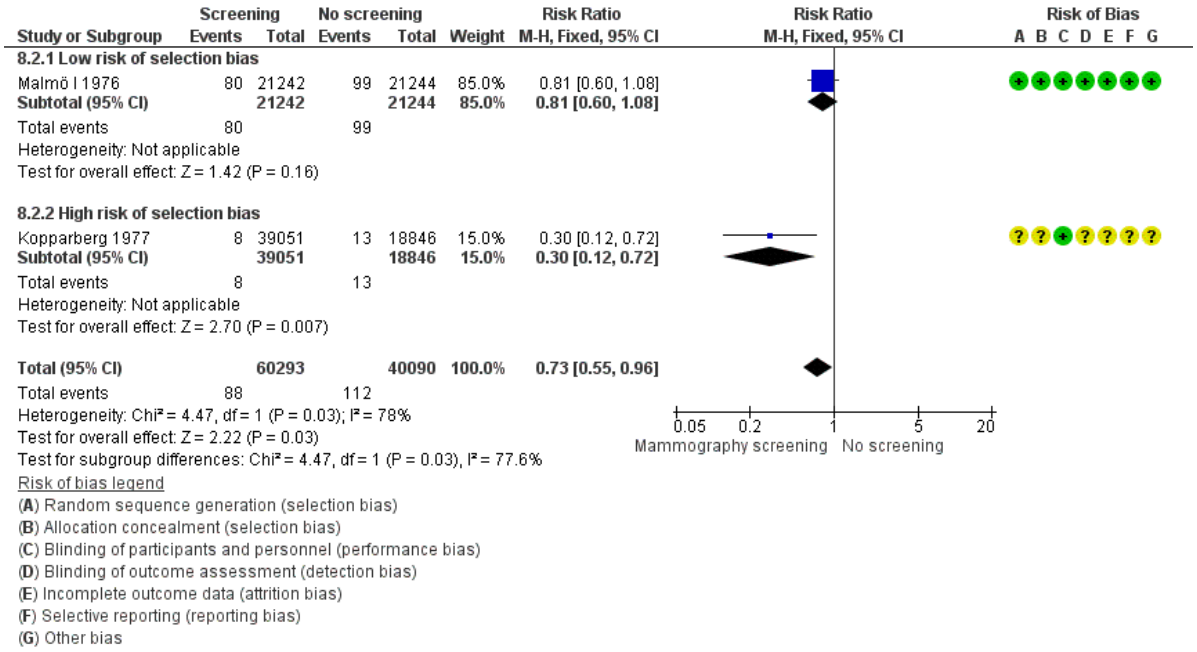


***Sources**: Gøtzsche & Jørgensen, 2013; Gøtzsche & Jørgensen, 2024,[14, 95] and Canelo-Aybar et al., 2021.[86]*

Only Gøtzsche & Jørgensen (2013; 2024) report data on hormone therapy, which are based on the same two Swedish trials.[14, 95] The meta-analysis indicate that **screening reduces the probability of receiving hormone therapy statistically significantly** (RR: 0.73, 95% CI: 0.55-0.96, p=0.03; Figure 17).

According to the Malmö trial, the single trial with a low risk of selection bias, screening has no effect on the probability to receive hormone therapy (RR: 081, 95% CI: 0.60-1.08, p=0.16; Figure 17).

Figure 17 – Forest plot for mammography screening vs. no screening, outcome: hormone therapy, all ages



***Sources**: Gøtzsche & Jørgensen, 2013; Gøtzsche & Jørgensen, 2024,[14, 95] and Canelo-Aybar et al., 2021.[86]*

Observational studies

Bennett et al. (2024) identified one cohort study (Puliti et al., 2017) providing data for all age groups, with a median follow-up of eleven years (interquartile range (IQR): 9-13 years). They provided a rate ratio for conservative surgery as treatment of breast cancer of 1.5 (95% CI: 1.4-1.6), and for mastectomy of 0.68 (95% CI: 0.63-0.72). More observational studies with data on treatment related morbidity were identified, but they did not provide relative effect estimates.[13]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials

For the surgical interventions, the IQWiG authors report the data from the CNBSS I trial (assessed as having a high risk of selection bias). The IQWiG authors argue that they used the results after seven years 'because a statistically significantly higher number of mastectomies can be expected after the first screening round[ss] as more breast cancer cases are detected through screening'.[88]

They report the results as ORs (all surgeries, OR: 1.33, 95% CI: 1.15-1.54; mastectomy, OR: 1.17, 95% CI: 0.94-1.45),[88] while the respective RRs can be found in Figure 13 (RR: 1.33, 95% CI: 1.15, 1.53) and Figure 14 (RR: 1.17, 95% CI: 0.94, 1.44), which are easier to interpret.

[ss] After the first round of screening, there was a statistically significant increase in both total surgeries (OR: 1.58, 95% CI: 1.15-2.18) and mastectomies (OR: 1.67, 95% CI: 1.10-2.52) in the screening group. The proportion of mastectomies among all breast cancer operations did not differ between groups (OR: 1.14, 95% CI: 0.60-2.18).[88]

C. Women aged 50-69 years

No relevant data were identified.

D. Women aged ≥ 70 years

Randomised controlled trials

The IQWiG authors did not identify usable data for this outcome in this age group.

Observational studies

Mathieu et al. report also for this outcome the results of the large observational study by García-Albéniz et al. (2020, n= 1 058 013 women aged 70-84 years) which applied the emulated trial[kk] methodology on US Medicare data from 1999 to 2008,[120] hence after publication of the RCTs that pointed to no difference in local tumour control, recurrence free survival and overall survival between total mastectomy and breast conserving surgery followed by radiotherapy in stage I-II breast cancer.

The authors assessed the effect of continuing screening among women aged 70 to 74 years and among those aged 75 to 84 years.[104, 120] This study reported a statistically significantly lower rate of radical mastectomy when screening was continued in both age groups (70-74 years: 13.9%, 95% CI: 13.4-14.5 versus 18.2%, 95% CI: 17.0-19.4; 75-84 years: 14.2%, 95% CI: 13.7-14.6 versus 17.0%, 95% CI: 16.0-17.9).[120] The same pattern was observed for chemotherapy (70-74 years: 15.2%, 95% CI: 14.7-15.8 versus 21.1%, 95% CI: 20.0-22.1; 75-84 years: 8.6%, 95% CI 8.3-9.1 versus 11.5%, 95% 10.6-12.3). However, when screening was continued, the rates of lumpectomy and of radiotherapy were statistically significantly higher compared to stop screening, also in both age groups. Mathieu et al. (2024) noted that this study had a serious risk of bias.[104]

DISCUSSION

Data from these RCTs are not very helpful to evaluate the effectiveness of screening in reducing mastectomy rates (in favour of lumpectomy followed by radiotherapy), as all RCTs were carried out in an era where mastectomy was the standard surgical treatment. Data from (somewhat) more recent observational studies indeed hint at a decreased risk of mastectomy in screened women and an increased rate of conservative surgery as treatment, but these estimates must be interpreted with caution as the quality of the evidence was rated by the respective review authors as very low.

3.2.2.9 Consequences of false-positive screening results

Key points

The authors cannot comment on the comparative effectiveness of breast cancer screening for interval cancers, as false-positive screening results can by definition only be derived from the intervention arm of a trial.

All age groups

No relevant RCT data were identified.

<50 years

Due to differences in reporting and completeness across studies, the results obtained in six trials could not be pooled.

50-69 years

No relevant RCT data were identified.

≥70 years

In a recent RCT on mammography screening (AgeX Pilot), 0.66% and 0.03% of women invited for screening and ≥ 70 years of age had biopsies and surgery, respectively, because of false-positive screening results (no GRADE assessment could be made).

By definition, data on false-positive results only refer to screened women and this section is therefore purely descriptive.

ALL AGE GROUPS

Randomised controlled trials

No relevant data were identified.

Observational studies

Bennett et al. (2024) reported false-positive results as 'additional imaging[tt] (no cancer)' in their systematic review.[13] They report calculations of additional imaging based on national population data from the Canadian Partnership against Cancer (CPAP). It was decided not to report these data here but calculate false-positive rates for the Belgian population based on Belgian data (see chapter xx).

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials

The IQWiG authors identified six trials[uu] with usable data to estimate false-positive results, but due to differences in reporting and completeness across studies they did not pool these data but reported instead ranges of effect estimates.[88] Overall, 0.16-3.4% of screened women of 40-49 years had a **biopsy** because of a false-positive result in the first screening round or year. In subsequent screening rounds or years, these estimates were lower (0.07%-1.8%). The proportion of biopsies with benign findings was 90% in the HIP study (which is one of the oldest breast cancer screening trials). Importantly, the AgeX Pilot study, which ran in 2009-2010 and may thus reflect better current imaging techniques, reported that after 12 months 2.31% of screened women aged 47-49 years, underwent biopsy because of a false-positive result.[88]

The proportion of screened women who underwent **surgery** because of a false-positive screening result in this age cohort ranged from 0.15% to 0.27% in the first screening round, and from 0.03% to 0.13% in subsequent rounds. The proportion of surgeries with benign findings ranged from 16% to 65%.[88]

B. Women aged 50-69 years

No relevant data were identified.

C. Women aged ≥ 70 years

Randomised controlled trials

For the age group of women ≥ 70 years of age, the data reported by IQWiG are restricted to those of the recent AgeX Pilot, hence based on current practice.[88, 117] They found that 0.66% and 0.03% of

---

tt Three different additional imaging outcomes were calculated: 1) additional imaging with or without biopsy (no cancer) calculated as the recall rate minus the cancer detection rate, 2) additional imaging with no biopsy (no cancer) calculated as the recall rate minus the sum of the cancer detection rate and non-malignant biopsy rate, and 3) additional imaging and biopsy (no cancer) which is represented as the non-malignant biopsy rate. These outcomes were calculated to approximate rates over a ten-year period.

uu AgeX Pilot, CNBSS, HIP, Malmö, UK Age & Stockholm trials

screened women had biopsies and surgery, respectively, because of false-positive screening results. The proportion of surgery with benign findings was 3%.[88, 117]

DISCUSSION

Of note, the proportion of biopsies with benign findings recorded in women aged 40-49 years during the first year/screening round in the more recent AgeX Pilot trial (i.e. 2.31%) was within the range of the older trials (0.16%-3.4%).[88] As it was a pilot trial, no results from longer follow-up were provided.[117] Yet, with regard to surgeries with benign findings, a lower proportion was found in the AgeX Pilot trial, compared to the older Gothenburg trial, a finding which has been ascribed to improved diagnostics.[88]

Further, the IQWiG authors comment that **cumulative rates per screened woman** over several screening rounds would be more meaningful for assessing the harm of screening with regard to the consequences of false-positive findings. Yet, as no data are reported on the number of women who underwent multiple biopsies or surgeries with benign findings, the calculation of a cumulative rate over the entire screening period was not possible. But, under the assumption that the events in the individual screening rounds were independent of each other, the results for biopsies with benign findings from the UK Age Study, for example, illustrate that approximately 0.51% of women aged 47-49 years receive at least one invasive diagnostic work-up with subsequent benign findings within five screening rounds/years.[88]

3.2.2.10 Consequences of false-negative results

Key points

No relevant RCT data were identified.

This outcome was not covered in any of the included systematic reviews.

3.2.2.11 Health-related quality of life (HRQoL)

Key points

No relevant RCT data were identified.

Bennett et al. (2024) report that no eligible studies were found which reported on health-related quality of life or life years gained (or lost).[13]

Canelo-Aybar et al. (2021) identified a systematic review (which included 54 primary observational studies) on psychological effects related to breast cancer screening.[86, 154] This review found no apparent increased anxiety in women given a clear mammography result and placed on routine recall. There were, however, mixed results regarding anxiety in women recalled for further testing, both in the short and long term, depending on the extent of additional examinations.

Two other reviews (including respectively 17 and 4 studies) evaluated psychological effects of false-positive results, and found that these were associated with increased breast cancer-related psychological distress.[86, 155, 156] Last, Gøtzsche & Jørgensen (2013; 2024) did not select this outcome.[14, 95]

3.2.2.12 Radiation exposure

Key points

This outcome was not covered in any of the systematic reviews.

3.2.2.13 Conclusion

The currently available RCT evidence is limited to **trials performed in the period 1960 to 1990**, and may thus not represent current screening, diagnostic and/or treatment approaches. Yet, while recent observational studies (e.g. cohort, case-control studies) comparing attendees to non-attendees represent current practices better, they are affected by selection bias ('healthy user bias'), recall bias and unmeasured confounders, resulting in overestimations of screening benefits.[157]

After 13 years of FU, inviting **adult women** to mammography screening results in a statistically significant reduction of death ascribed to breast cancer (RR: 0.81, 95% CI: 0.74-0.87; absolute effect: 81 fewer per 100 000, 95% CI: from 111 fewer to 55 fewer), while it does not have an effect on all-cause mortality (RR: 0.99, 95% CI: 0.97-1.01; absolute effect: 61 fewer per 100 000, 95% CI: from 182 fewer to 61 more). In addition, invitation to screening with mammography results in a statistically significant reduction of the probability of being diagnosed with advanced disease (stage II or higher: RR: 0.66, 95% CI: 0.61-0.72; absolute effect: 309 fewer per 100 000, 95% CI: from 354 fewer to 254 fewer - stage III or higher: RR: 0.66, 95% CI: 0.52-0.83; absolute effect: 76 fewer per 100 000, 95% CI: from 107 fewer to 38 fewer). Yet, invitation to screening with mammography results in a statistically significant increase of overdiagnosis 7-9 years after randomisation (RR: 1.29, 95% CI: 1.23-1.35; absolute effect: 332 more per 100 000, 95% CI: from 263 more to 400 more). Among adult women at average risk of breast cancer, and invited to screening, 27% of all cancers (invasive and non-invasive) detected were overdiagnosed.

The results for death ascribed to breast cancer for **women between fifty and sixty-nine years old**, are more pronounced compared to the all age group (RR: 0.74, 95% CI: 0.66-0.83; absolute effect: 169 fewer per 100 000, 95% CI: from 221 fewer to 111 fewer). For all-cause mortality no relevant data could be retrieved from the RCTs. In this age group, invitation to screening results in a statistically significantly reduced risk of being diagnosed with advanced disease (stage II or higher: RR: 0.73, 95% CI: 0.66-0.81; absolute effect: 221 fewer per 100 000, 95% CI: from 278 fewer to 155 fewer - stage III or higher: RR: 0.62, 95% CI: 0.48-0.80; absolute effect: 65 fewer per 100 000, 95% CI: from 89 fewer to 34 fewer). However, in this age group, invitation to screening with mammography results in a statistically significant risk of overdiagnosis (RR: 1.26, 95% CI: 1.10-1.44; absolute effect: 482 more per 100 000, 95% CI: from 185 more to 815 more). Among women between 50 and 69 years old and at average risk of breast cancer, and invited to screening, 21% of all cancers detected (invasive and non-invasive) were overdiagnosed.

The results for **women younger than fifty years old**, are comparable to the all age group for death ascribed to breast cancer (RR: 0.84, 95% CI: 0.73-0.96; absolute effect: 49 fewer per 100 000, 95% CI: from 82 fewer to 12 fewer), as well as for all-cause mortality (RR: 0.99, 95% CI: 0.94-1.04; absolute effect: 21 fewer per 100 000, 95% CI: from 126 fewer to 84 more). Yet, for stage shift the effect of screening is clearly less pronounced and/or no longer statistically significant (stage II or higher: RR: 0.88, 95% CI: 0.78-0.99; absolute effect: 49 fewer per 100 000, 95% CI: from 90 fewer to 4 fewer - stage III or higher: RR: 0.98, 95% CI: 0.74-1.29; absolute effect: 2 fewer per 100 000, 95% CI: from 21 fewer to 23 more). On the other hand, in this young age group, invitation to screening with mammography results in a statistically significant increase of overdiagnosis (RR: 1.20, 95% CI: 1.11-1.31; absolute effect: 178 more per 100 000, 95% CI: from 98 more to 276 more). Among women younger than 50 years old and at average risk of breast cancer, and invited to screening, 23% of all cancers (invasive and non-invasive) detected were overdiagnosed.

For **women of seventy years and older**, mammography screening does not result in a significantly decreased breast cancer related mortality risk (RR: 0.98, 95% CI: 0.93-1.03; absolute effect: 567 fewer per 100 000, 95% CI: from 1986 fewer to 851 more) nor in a reduced all-cause mortality risk (RR: 0.92, 95% CI: 0.62-1.37; absolute effect: 45 fewer per 100 000, 95% CI: from 215 fewer to 209 more). In addition, for this age group, no relevant data were retrieved from RCTs with regard to stage shift nor on overdiagnosis.

3.2.3 Findings – Various breast cancer screening intervals

In order to answer the second research question, what is in women at average risk of breast cancer the clinical effectiveness of

a. biennial screening with mammography vs. annual screening with mammography, and

b. biennial screening with mammography vs. triennial screening with mammography,

the SR by Canelo-Aybar et al. (2022) was used as the primary source, as this review of good quality focused on this research question.[100] The authors did not perform a meta-analysis because of the heterogeneity of the included studies. The SR by Canelo-Aybar et al. (2022) was preferred above the SR by Henderson et al. (2024) as the former had better AMSTAR-2 scores, yet they both included the same RCTs.[87, 100] For outcomes not covered in the SR by Canelo-Aybar et al. (2022), Henderson et al. (2024) was consulted.

Canelo-Aybar et al. (2022) identified one RCT, the United Kingdom Co-ordinating Committee on Cancer Research (UKCCR) trial of Breast Screening Frequency, which was conducted between 1989 and 1996 and which randomly allocated women aged 50-62 years to annual or triennial screening (see also Appendix 9.4).[100, 158] Additionally, 13 observational studies and 11 modelling studies were included.[100] Of note, one of the observational studies was a Finnish quasi-experimental study in which women aged 40-49 years were allocated to either annual or triennial screening based on calendar year of birth.[159] In the following sections, modelling studies are not included in the discussion of results.

The information from Canelo-Aybar et al. (2022) and Henderson et al. (2024),[87, 100] was further supplemented with results on interval cancers from the SR by Bennett et al. (2024), which used trial data to estimate interval cancers with different screening intervals.[13]

When other comparisons between screening intervals were reported (e.g. annual vs. triennial), the results are also recorded in the respective paragraphs, but not included in the key points.

3.2.3.1 All-cause mortality

Key points

No relevant RCT data were identified.

Canelo-Aybar et al. (2022) did not assess this outcome.[100]

ALL AGE GROUPS

No relevant data were identified.

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Observational studies

Henderson et al. (2024) report the results of the Finnish quasi-experimental study by Parvinen et al. (2011): thirteen years after allocation, the risk of all-cause mortality in women aged 40-49 years was not different (RR: 1.20, 95% CI: 0.99-1.46) between women who were allocated to annual screening compared to the group who had triennial screening. Yet, Henderson et al. (2024) considered the overall strength of evidence as insufficient.[87]

E. Women aged ≥50 years

No relevant data were identified.

3.2.3.2 Breast cancer-related mortality

Key points

No relevant RCT data were identified.

ALL AGE GROUPS

No relevant data were identified.

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Observational studies

For women aged 40-49 years, Canelo-Aybar et al. (2022) identified the quasi-experimental Finnish study by Parvinen et al. (2011) which found no difference in breast cancer mortality risk in annual versus triennial screening (RR: 1.14, 95% CI: 0.59-2.19). Canelo-Aybar et al. (2022) rated the certainty of the evidence as very low; they downgraded due to indirectness and imprecision.[100]

F. Women aged 50-69 years

Randomised controlled trials

Canelo-Aybar et al. (2022) reported for this outcome the results of the UKCCCR trial, which found no statistically significant difference in breast cancer mortality risk for annual versus triennial screening in women aged 50-69 years (RR: 0.93, 95% CI: 0.76-1.12).[100, 158] Canelo-Aybar et al. (2022) considered the level of evidence as moderate: they downgraded due to imprecision.[100] According to Canelo-Aybar (*personal communication*), the risk of selection bias in the UKCCCR trial was high ('*Randomisation was by month of birth in the first 2 years of the trial, after which the national screening programme software was augmented to allow individual randomisation*' and [the trial authors provided] '*No details of concealment*').

Observational studies

A Canadian before-after study[vv] also reported no difference in breast cancer mortality when comparing annual versus biennial screening for the same age group (IRR: 0.94; 95% CI: 0.68-1.31).[100, 160] Canelo-Aybar et al. (2022) rated the certainty of the evidence as very low; they downrated due to serious risk of bias.[100]

G. Women aged ≥ 70 years

No relevant data were identified.

[vv] Canadian before-after study: Coldman et al. (2008)[160]

3.2.3.3 Breast cancer stage at diagnosis

Key points

No relevant RCT data were identified.

ALL AGE GROUPS

Randomised controlled trials

Canelo-Aybar et al. (2022) did not provide data for all age groups combined.[100] Henderson et al. (2024) concluded that there are no statistically significant differences in cancer characteristics, such as tumour size, nodal status or histological grade between annual and triennial screening (no RRs nor raw data are provided in the SR, hence no forest plots were developed).[87, 158] They based their conclusion on the UKCCCR trial, which included women aged 50 to 62 years and which they rated as fair quality evidence.

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Observational studies

For the comparison between biennial vs. annual breast cancer screening, Canelo-Aybar et al. (2022) identified a prospective cohort study from the US (Miglioretti et al., 2015, 1996-2012).[100, 161] In women aged 40-49 years, there was no statistically significant difference in the risk of being diagnosed with advanced stage breast cancer[ww] (RR: 1.17, 95% CI: 0.93-1.46[xx]) nor in the risk of being diagnosed with a larger tumour (i.e. tumour size >15 mm vs. ≤15 mm; RR: 1.10, 95% CI: 0.98-1.25[xx]).[161] Canelo-Aybar et al. (2022) considered the level of evidence as very low level; they downrated due to serious risk of bias.[100]

Similar results were reported for the comparison between triennial vs. biennial breast cancer screening (IIB-IV vs. I-IIA: OR: 0.78, 95% CI: 0.54-1.11), which are based on a retrospective cohort study (O'Meara et al. (2013)[yy]).[100] Canelo-Aybar et al. (2022) rated the certainty of the evidence as very low; they downrated due to serious risk of bias.[100]

B. Women aged 50-59 years

Observational studies

Also for this age group, the prospective cohort study by Miglioretti et al. (2015) noted no statistically significant differences in the risk of being diagnosed with advanced stage breast cancer[ww] (RR: 0.98, 95% CI: 0.80-1.21[xx]) nor in the risk of being diagnosed with a larger tumour (RR: 1.09, 95% CI: 0.97-1.21[xx]) when biennial breast cancer screening is compared with annual screening.[161] Canelo-Aybar et al. (2022) rated the certainty of the evidence for both outcomes as very low; they downrated due to serious risk of bias.[100]

---

ww Advanced stage breast cancer was defined as stage IIB, III, or IV and compared with stages I or IIA.[161]

xx Canelo-Aybar et al. (2022) present these results as ORs, comparing annual vs. biennial screening,[100] while in the original publication the results are presented as RRs, comparing biennial vs. annual screening.[161] In the present report we preferred to present the original RRs, as RRs are easier to interpret compared to ORs; the ORs reported by Canelo-Aybar et al. (2022) can be found in the evidence table (Appendix 1.5.1).

yy Note: the results presented by Canelo-Aybar et al. (2022) are confined to those for white women;[100] in the original manuscript, results are also presented for black, Hispanic and Asian women, which are in line with those for white women.[162]

C. Women aged 60-69 years

Observational studies

Likewise, no statistically significant differences were reported by Miglioretti et al. (2015) for women aged 60-69 years, in the risk of being diagnosed with advanced stage breast cancer[ww] (RR: 0.99, 95% CI: 0.79-1.24[xx]) nor in the risk of being diagnosed with a larger tumour (RR: 1.13, 95% CI: 1.00-1.27[xx]) when comparing biennial vs. annual breast cancer screening.[161] Again, the certainty of the evidence for both outcomes was considered very low, as it was downrated due to serious risk of bias.[100]

D. Women aged 70-85 years

Observational studies

Miglioretti et al. (2015) did not note a statistically significant difference in women aged 70-74 years, in the risk of being diagnosed with advanced stage breast cancer[ww] (RR: 0.98, 95% CI: 0.76-1.27), nor in the risk of being diagnosed with a larger breast cancer (RR: 1.13, 95% CI: 0.99-1.29).[161] Again, the certainty of the evidence for both outcomes was considered very low; it was downrated due to serious risk of bias, inconsistency and indirectness.[100]

3.2.3.4 Overdiagnosis

Key points

No relevant RCT data were identified.

Neither Canelo-Aybar et al. (2022) nor Henderson et al. (2024) identified RCTs or observational studies for this outcome which provided relative effect estimates.[87, 100]

3.2.3.5 Interval cancers

Key points

No relevant RCT data were identified.

ALL AGE GROUPS

Randomised controlled trials

Canelo-Aybar et al. (2022) did not provide data for all age groups.[100] Henderson et al. (2024) suggested, based on the UKCCCR RCT (which they assessed as having a fair quality), that fewer interval cancers (false-negative and incident cancers) were diagnosed in the annual invitation group compared with triennial screening group (RR: 0.68, 95% CI: 0.50-0.92).[87]

Bennett et al. (2024) provided in their SR which compared mammography screening with usual care also a meta-analysis of results on interval cancers (DCIS and invasive); an overview is provided in Table 5. It is important to note that all RCTs which provided data on this outcome, were considered having a high risk of selection bias (see section 3.2.1.3).[13]

Table 5 – Interval cancers identified in RCTs, all age groups

Diagnosis	Screening interval	Sources	Result	Follow-up	Quality of evidence
DCIS & invasive BC	≤12 months	CNBSS 1 & 2	3.9/1000 (95% CI: 3.4-4.5)	5 years	Low
	13-24 months	Stockholm, Gothenburg	3.1/1000 (95% CI: 2.6-3.7)	4.8-7.0 years	Very low
	>24 months (i.e. 23-33 months)	Swedish Two County	3.9/1000 (95% CI: 3.4-4.3)	7.0 years	Low
Invasive	18 months	Stockholm, Gothenburg	2.8/1000 (95% CI: 2.4-3.3)	4.8-7.0 years	Very low
DCIS	18 months	Stockholm, Gothenburg	0.2/1000 (95% CI: 0.1-0.5)	4.8-7.0 years	Very low

***Abbreviation**: BC: breast cancer*
***Source**: Bennet et al., 2024[13]*

VARIOUS AGE GROUPS

A. Women aged 39-49 years

Randomised controlled trials

The SR by Bennett et al. (2024) provided for this age group data on interval cancers obtained in the Gothenburg trial (considered having a high risk of selection bias), where the screening interval was 18 months.[13] In women aged 39-49 years, 3.0 (95% CI: 2.1-4.2) interval cancers (2.8 invasive and 0.2 in situ) were detected per 1000 women screened over a follow-up period of 4.8-7 years.[13]

Observational studies

Canelo-Aybar et al. (2022) identified one retrospective comparative study (Hunt et al., 1999, age group 40-79 years), which reported a decreased risk of interval cancers with annual screening as compared to biennial screening for women aged 40-49 years, but the association was not statistically significant (RR: 0.46, 95% CI: 0.16-1.36).[100] They considered the level of evidence as very low; they downgraded due to risk of bias and inconsistency.[100]

The quasi-experimental study performed in Finland (1987–2003, Parvinen, 2011; Klemi, 1997) reported that in women aged 45 to 49 years, 25% (95% CI: 15-36%) of breast cancer cases detected in the annual screening group were interval cancers, whereas in the group who had triennial screening the respective proportion was 35% (95% CI: 22-50%).[159, 163] Canelo-Aybar et al. (2022) considered the level of evidence as very low; they downgraded because of indirectness.[100] Henderson et al. (2024) identified the Finnish study too and concluded that there was no statistical difference in the incidence of interval cancers diagnosed with annual compared to triennial screening.[87]

B. Women aged 50-69 years

Observational studies

In the prospective cohort study by Miglioretti et al. (1996-2012), it was reported that in women aged 50-69 years who had annual screening 22% (95% CI: 21-30%) of diagnosed breast cancer cases were interval cancers, while in women who had biennial screening the respective proportion was 27% (95% CI: 26-29%).[161]

Another cohort study provided data on interval cancers among women who had annual, biennial and triennial screening which compared triennial vs. biennial screening and annual vs. triennial (defined based on the two most recent screening examinations). The proportion of interval cancers among breast cancer cases was 30% (95% CI: 29-31%) in the annual screening group, 41% (95% CI: 39-42%) in the biennial screening group and 44% (95% CI: 41-48%) in the triennial screening group.[164]

Canelo-Aybar et al. (2022) rated the level of evidence for the above results as very low; they downgraded for risk of bias and indirectness.[100]

C. Women aged 50-79 years

Observational studies

Canelo-Aybar et al. (2022), retrieved a Canadian study by Coldman et al. (2008) which evaluated the implementation of a change in screening policy in the age group 50-79 years; i.e. in 1997 it was changed from annual to biennial mammography.[100] Canelo-Aybar et al. (2022) report that they found no difference in interval cancer (RR: 0.98, 95% CI: 0.90-1.06) comparing annual to biennial screening. The certainty of the evidence was not reported.[100] Yet, it is important to note that in the original manuscript by Coldman et al. (2008), these results were reported as the screen-detected cancers, comparing biennial versus annual screening.[160]

D. Women aged 70-74 years

Observational studies

For this age group, Canelo-Aybar et al. (2022) cite results from a retrospective cohort study by Braithwaite et al. (2013; US, 1999-2006), which noted that in women who were screened annually 23% (95% CI: 22-25%) of diagnosed breast cancers were interval cancers, while in the group who was screened on a biennial basis, the respective proportion was 33% (95% CI: 30-36%).[165] Again, Canelo-Aybar et al. (2022) rated the level of evidence for this comparison as very low.[100]

3.2.3.6 Treatment-related morbidity

Key points

No relevant RCT data were identified.

This outcome was not discussed in the SRs by Canelo-Aybar et al. (2022),[100] while Henderson et al. (2024) report that there was insufficient evidence for evaluating the effect of screening intervals on breast cancer morbidity.[87]

3.2.3.7 Consequences of false-positive screening results

Key points

No relevant RCT data were identified.

ALL AGE GROUPS

No data were retrieved across all age groups for this outcome.

VARIOUS AGE GROUPS

A. Women aged 45-49 years

Observational studies

For women aged 45-49 years, Canelo-Aybar et al. (2022) report the results of two retrospective cohort studies, which were both rated as very low level of evidence (downgraded one level for risk of bias and two levels for indirectness).[100, 162, 166] Dittus et al. (2013) noted that in women of normal weight, the 10-year probability of false-positive results with annual screening was 11.2% (95% CI: 9.8-12.8%), and 6.0% (95%CI 5.4–6.6%) with biennial screening.[100, 166] O'Meara et al. (2013) used mammography and tumour registries from the US and reported similar 10-year probabilities of false-positive results leading to biopsy recommendation in white women: 11.4% (95% CI: 10.5-12.4%) with annual screening, 5.9% (95% CI: 5.6-6.2%) with biennial screening, and 3.9% (95% CI: 3.7-4.1%) with triennial screening.[100, 162]

B. Women aged 50-69 years

Observational studies

For this age group, O'Meara et al. (2013) reported the following 10-year probabilities of a false-positive result: 55.2% (95% CI: 54.8-55.7%) with annual screening, 35.4% (95% CI: 35.0-35.7%) with biennial screening, and 24.8% (95% CI: 24.5-25.2%) with triennial screening.[100, 162] The cumulative 10-year probability of false-positive results leading to biopsy recommendations reported by the same researchers was 9.7% (95% CI: 9.3-10.1%) for annual screening, 5.4% (95% CI: 5.2-5.6%) for biennial screening and 3.7% (95% CI: 3.6-3.9%) for triennial screening.[100, 162] Canelo-Aybar et al. (2022) rated the evidence as having a very low level of certainty; they downgraded one level for risk of bias and for indirectness.[100]

C. Women aged 77-89 years

Observational studies

Braithwaite et al. (2013) estimated that the 10-year cumulative probability of false-positive results for this age group is also higher with annual screening (47%, 95% CI: 44.9–49.5%) compared to biennial screening (26.6%, 95% CI: 25.7–27.5%).[100, 165] Their cumulative 10-year probability of false-positive results leading to biopsy recommendations was 9% (95% CI: 8-11%) for annual screening and 4% for biennial screening (95% CI: 4-5%).[100, 165] The certainty of the evidence was considered very low; Canela-Aybar et al. (2022) downgraded for risk of bias.[100]

3.2.3.8 Consequences of false-negative results

Key points

No relevant RCT data were identified.

This outcome was not discussed in the SRs by Canelo-Aybar et al. (2022) or Henderson et al. (2024).[87, 100]

3.2.3.9 Health-related quality of life

Key points

No relevant RCT data were identified.

None of the included systematic reviews reported on health-related quality of life measures.[87, 100]

3.2.3.10 Radiation exposure

Key points

No relevant RCT data were identified.

Harms related to radiation exposure with different screening intervals were discussed only in modelling studies in Canelo-Aybar et al. (2022).[100]

3.2.3.11 Conclusion

Currently, there is no relevant RCT data available to answer the second research question (i.e. what is in women at average risk of breast cancer the clinical effectiveness of a) biennial screening with mammography vs. annual screening with mammography, and b) biennial screening with mammography vs. triennial screening with mammography).

3.2.4 *Findings – Invited for screening with mammography and ultrasound vs. invited for screening with mammography*

In order to answer the third research question, what is the clinical effectiveness of breast cancer screening using mammography combined with ultrasound vs mammography alone, the Cochrane SR by Glechner et al. (2023) was used, given its high quality and its relatively recent search date (spring 2021).[99]

In the review by Glechner et al. (2023) which compared mammography in combination with breast ultrasonography versus mammography alone for breast cancer screening in women at average risk of breast cancer, only one RCT (Japan Strategic Anti-cancer Randomised Trial, (J-START)) was identified; it considered **solely women aged 40 to 49**, who were enrolled between 2007 and 2011.[99] The results provided by Glechner et al. (2023) are restricted to a two-year follow-up.[99, 167, 168] This section was amended with data from Harada-Shoji et al. (2026), who provided for mixed populations of women with dense and non-dense breasts results for the J-START trial after a follow-up of eleven years (which are presented as main outcome).[89]

Glechner et al. (2023) also included two prospective cohort studies[zz] and five retrospective cohort studies[aaa], enrolling women up to age 75 with a follow-up of one to three years for the evaluation of efficacy and safety. Glechner et al. (2023) rated the overall risk of bias of the J-START trial as low, and considered the risk of selection bias also as low[bbb].[99]

Henderson et al. (2024) also covered this research question. They included the J-START trial and the retrospective cohort study by Lee et al. (2019), which were both included in the SR by Glechner et al. (2023).[87]

In this section, results are presented – whenever available – separately for mixed populations of women with dense and non-dense breasts, for women with dense breasts and for women with non-dense breasts.

It is important to mention that the J-START trial enrolled 72 717 asymptomatic women in 42 study sites in 23 of 47 prefectures in Japan, hence an Asian population where the prevalence of dense breast tissue is high.[89] Overall, 57.7% of included women were considered having dense breasts (i.e. scores of 3 or 4 in the American College of Radiology Breast Imaging Reporting and Data System).[99, 167] The **secondary analyses**, which compared the effect of mammography and ultrasound screening vs. screening with mammography alone, separately in women with dense breasts and in women with non-dense breasts **after two years follow-up**, were performed in a subpopulation of 19 213 women, enrolled in one prefecture between 2007 and 2020. In this subpopulation, 11 390 women (59.3%) were diagnosed with dense breasts.[168] The fact that the secondary analyses were performed in a subpopulation, may explain why some two-year follow-up effect estimates for the mixed population of women with dense and non-dense breasts are not situated in between the effect estimates for women with dense breasts and women with non-dense breasts.[99] So far, no 11-year follow-up data have been published for women with dense and for women with non-dense breasts.

---

zz Two prospective cohort studies: Chough et al. (2020), Giuliano et al. (2013)

aaa Five retrospective cohort studies: Buchberger et al. (2018), Chae et al. (2013), Lee et al. (2019), Starikov et al. (2016), Tohno et al. (2013)

bbb Glechner et al. (2023) used the Cochrane risk of bias tool-2 as described by the Cochrane Methods Network,[169] to assess the risk of bias in the included randomised J-Start trial. This information could not be included in the forest plots (due to limitations of Review Manager software), but can be found in Appendix 7.3.3.

3.2.4.1 All-cause mortality

Key points

All age groups[aa]

No relevant RCT data were identified.

<50 years

In a general population of adult women younger than 50 years old and at average risk of breast cancer, invitation to screening with mammography and US does not have a statistically significant effect on all-cause mortality after 11 years of FU (RR: 1.04, 95% CI: 0.88-1.22; absolute effect: 31 more per 100 000, 95% CI: from 93 fewer to 171 more; moderate certainty of evidence). The certainty of the evidence was downgraded one level because the outcome was reported in only one adequately randomised trial.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

Mixed populations of women with dense and non-dense breasts

ALL AGE GROUPS

None of the studies included in the SR by Glechner et al. (2023) assessed whether the combination of mammography screening and ultrasonography resulted in lower all-cause mortality.[99]

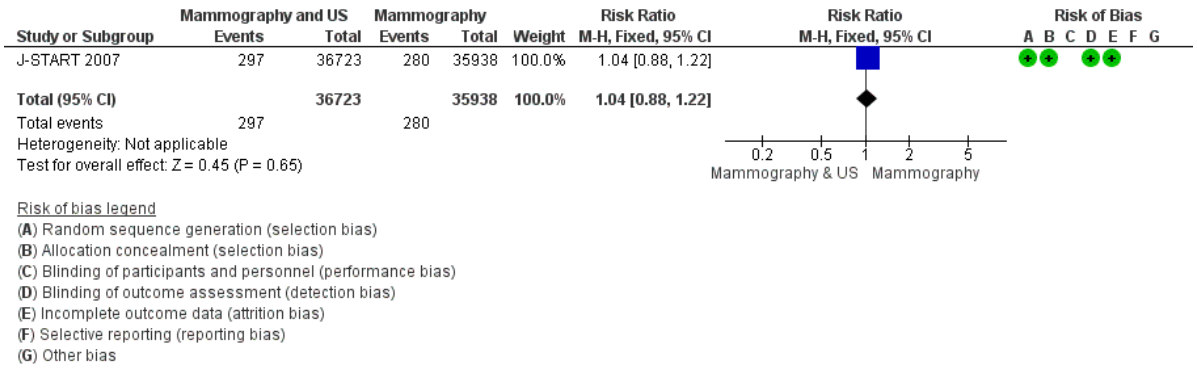
VARIOUS AGE SUBGROUPS[bb]

A. Women aged 40-49 years

Randomised controlled trials

Of note, results for this outcome were published by Harada-Shoji et al. in February 2026, and hence not included in the SR by Glechner et al. (2023). Eleven years after randomisation, screening with mammography and US **does not have an effect on all-cause mortality** in Japanese women between 40 and 49 years, compared to screening with mammography alone (RR: 1.04, 95% CI: 0.88-1.22, p=0.65; Figure 18).[89]

Figure 18 – Forest plot for mammography & US vs. mammography, outcome: all-cause mortality, 11 years FU



***Source**: Harada-Shoji et al. (2026).[89]*

B. Women aged ≥ 50 years

No relevant RCT data were identified.

3.2.4.2 Breast cancer mortality

Key points

All age groups[aa]

No relevant RCT data were identified.

<50 years

In a general population of adult women younger than 50 years old and at average risk of breast cancer, invitation to screening with mammography and US does not have a statistically significant effect on breast cancer mortality after 11 years of FU (RR: 0.98, 95% CI: 0.53-1.79; absolute effect: 1 fewer per 100 000, 95% CI: from 27 fewer to 46 more; moderate certainty of evidence). The certainty of the evidence was downgraded one level because the outcome was reported in only one adequately randomised trial.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

Mixed populations of women with dense and non-dense breasts

ALL AGE GROUPS

None of the studies included in the SR by Glechner et al. (2023) assessed whether the combination of mammography screening and ultrasonography had an impact on breast cancer mortality.[99]

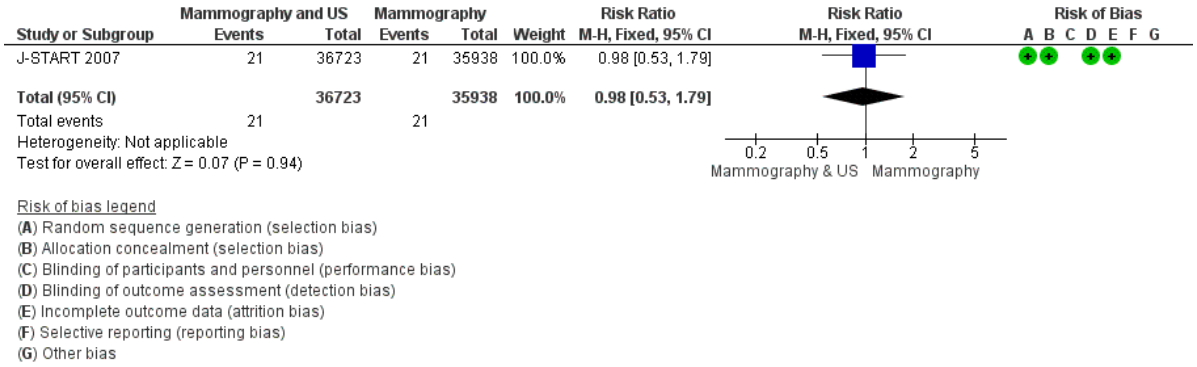
VARIOUS AGE SUBGROUPS[bb]

A. Women aged 40-49 years

Randomised controlled trials

Also for this outcome, results were published by Harada-Shoji et al. in February 2026, and hence not included in the SR by Glechner et al. (2023). Eleven years after randomisation, screening with mammography and US **does not have an effect on mortality ascribed to breast cancer** in Japanese women between 40 and 49 years, compared to screening with mammography alone (RR: 0.98, 95% CI: 0.53-1.79, p=0.94; Figure 19).[89]

Figure 19 – Forest plot for mammography & US vs. mammography, outcome: breast cancer mortality, 11 years FU



***Source**: Harada-Shoji et al. (2026).*[89]

B. Women aged ≥ 50 years

No relevant RCT data were identified.

3.2.4.3 Incremental cancer detection

Key points

All age groups

No relevant RCT data were identified.

<50 years

In a general population of adult women younger than fifty years old and at average risk of breast cancer, invitation to screening with mammography and ultrasound does not have a statistically significant effect on detection rate compared to mammography alone (RR: 1.04, 95% CI: 0.95-1.14; absolute effect: 94 more per 100 000, 95% CI: from 117 fewer to 328 more; moderate certainty of evidence). The certainty of the evidence was downgraded one level because the outcome was reported in only one adequately randomised trial.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

Mixed populations of women with dense and non-dense breasts

ALL AGE GROUPS

Observational studies

Glechner et al. (2023) pooled the data of a mixed population of 88 972 women with dense and non-dense breasts collected in two retrospective cohort studies[ccc] with moderate risk of bias for this

[ccc] Two retrospective cohort studies: Buchberger[ccc] et al. (2018), Starikov et al. (2016); *Note:* In the study by Buchberger et al. (2018) women who were invited annually or biannually for screening and who had heterogeneously dense or extremely dense breasts underwent supplemental ultrasound.[170] In addition, ultrasound could also be performed in women with non-dense breasts per specific radiologist's request.

outcome. Women screened with a combination of mammography and ultrasonography had a 35% higher risk of being diagnosed with breast cancer compared with women screened with mammography alone, yet the difference was not statistically significant (RR: 1.35, 95% CI: 0.92-1.98).

Glechner et al. considered the level of evidence provided by the two cohort studies as low; they downgraded due to risk of bias because additional ultrasonography was performed at the explicit request of a radiologist. In the study by Buchberger et al. (2018), women with non-dense breasts were specifically allocated to receive additional ultrasonography and in one of the studies, no information was provided on familial or genetic breast cancer risk.[99]

Adding a third study (Tohno et al., 2013), which carried a serious risk of bias, provided a similar risk ratio (RR: 1.43, 95% CI: 0.99-2.05).[99]

The review authors present the results of a fourth cohort study (Lee et al., 2019) separately, as it studied women who participated in one or more screening rounds, resulting in a lower pretest probability of detecting cancer at the second screening round. The study by Lee et al. (2019) was considered having a moderate risk of bias for this outcome; they found no statistically significant difference in cancer detection between the two screening approaches (RR: 0.99, 95% CI: 0.68-1.44).[99]

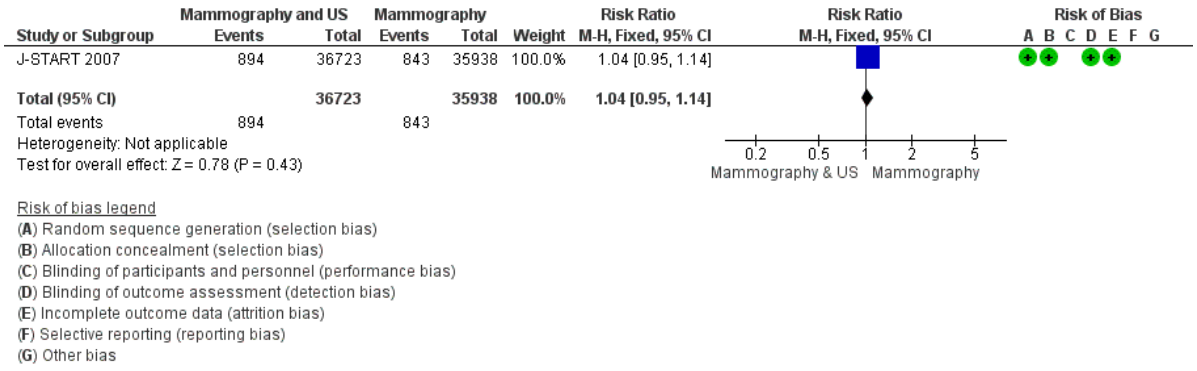
VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials

The results of the J-START trial, which enrolled 72 717 asymptomatic women between 40 and 49 years old, indicated that eleven years after randomisation, there was no significant difference in cancer detection rate between women who had the combination of mammography and ultrasound compared to women who had mammography alone (RR: 1.04, 95% CI: 0.95-1.14, p=0.43; Figure 20).[89] However, two years after randomisation, there was a 54% higher cancer detection risk when a combination of mammography and ultrasonography was used compared to mammography alone (RR: 1.54, 95% CI: 1.22-1.94; Appendix Figure 86).[99]

Figure 20 – Forest plot for mammography & US vs. mammography, outcome: incremental cancer detection, 11 years FU



***Source**: Harada-Shoji et al. (2026).[89]*

Women with a Breast Imaging Reporting and Data System (BI-RADS) 3 score (N=1255), were invited for intermediate screening in six months.[170] The review authors (Glechner et al., 2023) assumed that the intermediate mammograms at 6 months were used to calculate the specificity of screening at 12 months follow-up because the specificity was much higher (calculated: 96.0% to 98.7%) than in J-START (87.8% to 91.4%).[99]

B. Women aged ≥50 years

No relevant data were retrieved for this age group for this outcome.

Women with dense breasts

ALL AGE GROUPS

Observational studies

Glechner et al. (2023) extracted information on 71 191 women with dense breasts from four cohort studies[ddd].[99] The review authors pooled in a sensitivity analysis the results of three cohort studies with a low or unclear risk of bias[eee] for key domains (i.e. for confounding bias and study participant selection), and came to a comparable result, but with a more narrow CI (RR 1.78, 95% CI: 1.23-2.56).[99] They graded the certainty of evidence of the pooled result of these three cohort studies as moderate; they downgraded due to risk of bias.[99]

By adding the retrospective study by Chae et al. (2013), which was considered having a high risk of bias, the pooled risk ratio was higher (RR 2.17, 95% CI: 1.31-3.60);[171] it also resulted in a higher heterogeneity.[99] It is important to mention that it was not clear to the review authors whether the population of this retrospective study also included women with a personal or family history of breast cancer.[99]

The prospective cohort study by Chough et al. (2020), where women participated in one or two screening rounds, resulted in a lower and non-statistically significant RR (RR 1.33, 95% CI: 0.46-3.83).[99]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

The secondary analysis of a subpopulation of the J-START which comprised 11 390 women with dense breasts, also indicated a 65% higher cancer detection risk when a combination of mammography and ultrasonography was used compared to mammography alone (RR: 1.65, 95% CI: 1.00-2.72).[99] The authors of the Cochrane review graded the certainty of evidence as high, since the secondary analysis of the J-START was methodologically well-conducted.[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Women with non-dense breasts

ALL AGE GROUPS

Observational studies

The pooled results of two cohort studies[ggg], which included altogether 40 636 women, showed no statistically significant difference in cancer detection between both screening modalities (RR 1.13,

---

ddd Four cohort studies: Buchberger et al. (2018);[170] Chae et al. (2013)[171] Giuliano et al. (2013);[172] Starikov et al. (2016).[173]

eee Three cohort studies with a low or unclear risk of bias: Buchberger et al. (2018);[170] Giuliano et al. (2013);[172] Starikov et al. (2016).[173]

fff These results are restricted to a follow-up of two years.

ggg Two cohort studies: Buchberger et al. (2018),[170] and Starikov et al. (2016).[173]

95% CI: 0.85-1.49). The review authors considered certainty of the evidence of the pooled result of the two cohort studies as low due to risk of bias.[99]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

The secondary analysis of J-START with data from 7 823 women with non-dense breasts, came to comparable results as in the women with dense breasts (RR 1.93, 95% CI: 1.01-3.68), yet, given the imprecision of the result, the review authors graded the certainty of evidence as moderate.[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

3.2.4.4 Incremental detection of invasive cancer

Key points

All age groups

No relevant RCT data were identified.

<50 years

In a general population of adult women younger than fifty years old and at average risk of breast cancer, invitation to screening with mammography and ultrasound does not result in a statistically significantly increased probability of having an invasive cancer detected compared to mammography alone (RR: 1.01, 95% CI: 0.90-1.12; absolute effect: 19 more per 100 000, 95% CI: from 185 fewer to 222 more; moderate certainty of evidence). The certainty of the evidence was downgraded one level because the outcome was reported in only one adequately randomised trial.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

Mixed populations of women with dense and non-dense breasts

ALL AGE GROUPS

Observational studies

The pooled results of two cohort studies[hhh], with a mixed population of women with dense and non-dense breasts, pointed to no difference in terms of the proportion of invasive tumours out of all tumours detected by screening between the group who had mammography in combination with ultrasonography and the mammography alone group (RR: 1.00, 95% CI: 0.95-1.06).[99] Because of the methodological limitations of both cohort studies and the imprecise result, the review authors assessed the certainty of evidence as very low.

The cohort study by Lee et al. (2019) with one or more screening rounds, pointed to a lower rate of invasive cancer in women screened with mammography and ultrasound in comparison with mammography alone, but the result was not statistically significant (RR: 0.80, 95% CI: 0.59-1.09).[99]

[hhh] Two cohort studies: Buchberger et al. (2018),[170] and Starikov et al. (2016).[173]

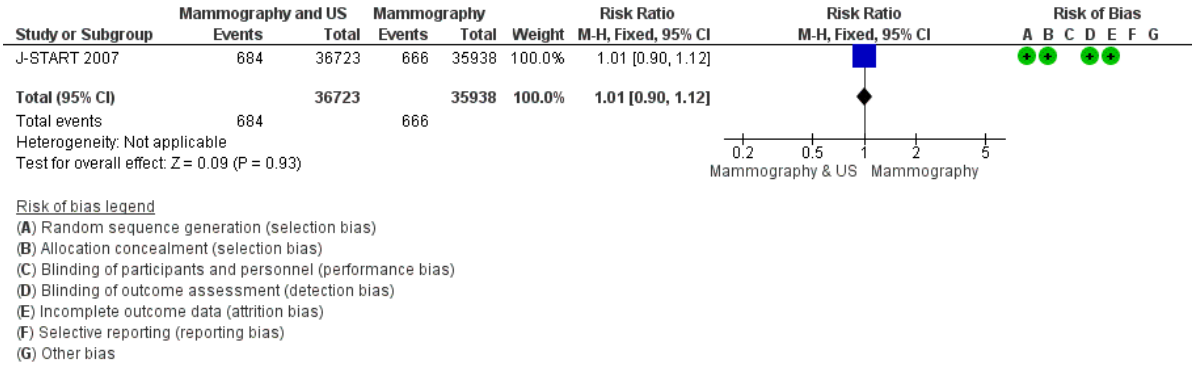
VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials

Eleven years after randomisation, there was no statistically significant difference in the detection of invasive cancers between women who had the combination of mammography and ultrasound compared to women who had mammography alone (RR: 1.01, 95% CI: 0.90-1.12, p=0.93; Figure 21).[89] However, after only two years of follow-up the risk of being diagnosed with an invasive cancer was 46% higher when screened with mammography and ultrasound compared to mammography alone (RR: 1.46, 95% CI: 1.11-1.91; Appendix Figure 87[iii]).[99, 167]

Figure 21 – Forest plot for mammography & US vs. mammography, outcome: incremental detection of invasive cancers, 11 years FU



***Source**: Harada-Shoji et al. (2026).[89]*

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Women with dense breasts

ALL AGE GROUPS

Observational studies

Glechner et al. (2023) identified three cohort studies[iii], which were exclusively performed in women with dense breasts.[99] The results of these studies were not pooled, yet they come to similar results: there was no statistically significant difference in the risk of being diagnosed with an invasive breast cancer when mammography screening was combined with ultrasonography or not. In the prospective cohort study by Giuliano et al. (2013), all included women had dense breast tissue and all detected cancer cases were invasive (RR: 1.00, 95% CI: 0.93-1.08). The results obtained in the cohort study by Chae et al. (2013) were considered at serious risk of bias (RR: 5.75, 95% CI: 0.96-34.48). Women who were enrolled in the prospective study by Chough et al. (2020) participated in one or two screening rounds (RR: 1.05, 95% CI: 0.67-1.64).[99]

[iii] It's important to mention that for these analyses Glechner et al. (2023) compared the proportion of invasive cancers among the screen-detected cancer cases between the mammography in combination with ultrasonography group and the mammography alone group.[99] As a result, the numerators (i.e. invasive cancer cases) and denominators (i.e. screen-detected cancer cases) were small and their result differed from ours (128/184 (70%) vs. 86/117 (74%); RR: 0.95, 95% CI: 0.82-1.09).[99]

[jjj] Three cohort studies: Chae et al. (2013);[171] Chough, et al. (2020);[174] Giuliano et al. (2013);[172]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

In the secondary analysis of J-START, 46 out of 65 screen-detected cancer cases among women with dense breasts were invasive (70.8%). The difference between the two screening interventions was not statistically significant (RR: 0.91, 95% CI: 0.67-1.24).[99] Glechner et al. (2023) considered the level of evidence as low: they downgraded due to imprecision as the result was not statistically significant.

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Women with non-dense breasts

ALL AGE GROUPS

No data were retrieved for this group for this outcome.

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

The secondary analysis of J-START revealed that 25 out of 41 screen-detected cancer cases among women with non-dense breasts were invasive (61.0%). The difference between the two screening interventions was, also in this subgroup, not statistically significant (RR: 0.66, 95% CI: 0.42-1.04).[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

3.2.4.5 Incremental detection of non-invasive cancers

Of note, according to the trialists, non-invasive cancers included ductal carcinoma in situ (DCIS) as well as lobular carcinoma in situ (LCIS).[167] While the former is regarded an early breast cancer, LCIS isn't cancer, but considered a benign lesion.

Key points

All age groups

No relevant RCT data were identified.

<50 years

In a general population of adult women younger than fifty years old and at average risk of breast cancer, invitation to screening with mammography and ultrasound does not result in a statistically significantly increased probability of having a non-invasive cancer detected compared to mammography alone (RR: 1.16, 95% CI: 0.95-1.42; absolute effect: 79 more per 100 000, 95% CI: from 25 fewer to 207 more; moderate certainty of evidence). The certainty of the evidence was downgraded one level because the outcome was reported in only one adequately randomised trial.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

Mixed populations of women with dense and non-dense breasts

ALL AGE GROUPS

No relevant data were retrieved.

VARIOUS AGE GROUPS

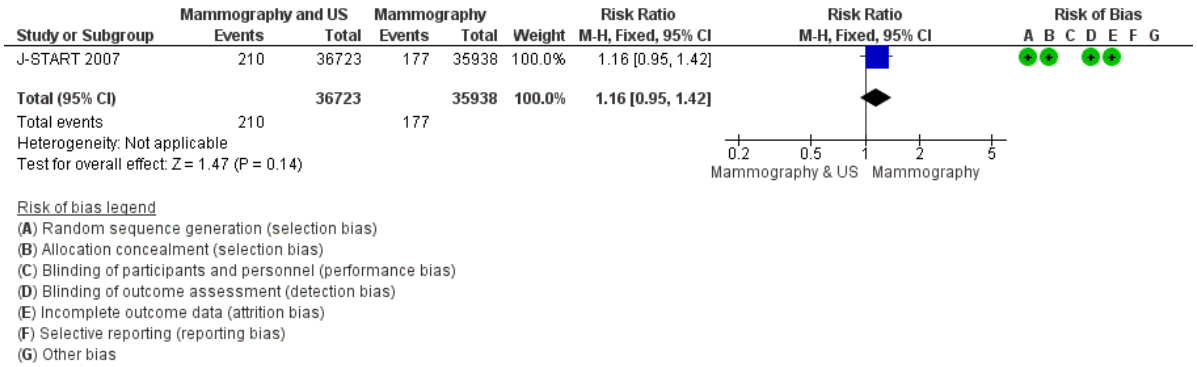
A. Women aged 40-49 years

Randomised controlled trials

After 11 years of follow-up, 23.5% and 21.0% of all breast cancers diagnosed in the intervention group (mammography and US) and control group (mammography alone), respectively, were categorised as non-invasive. The probability of being diagnosed with a non-invasive breast cancer was not statistically significant between intervention and control group (RR: 1.16, 95% CI: 0.95-1.42, P=0.14; Figure 22).[89]

Yet, after only two years of follow-up, the probability of being diagnosed with a non-invasive cancer was 77% higher when screened with mammography and ultrasound compared to mammography alone (RR: 1.77, 95% CI: 1.14-2.74; Appendix Figure 88).[99, 167]

Figure 22 – Forest plot for mammography & US vs. mammography, outcome: incremental detection of non-invasive cancers, 11 years FU



***Source**: Harada-Shoji et al. (2026).*[89]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

3.2.4.6 Incremental detection of advanced breast cancer

Key points

All age groups

No relevant RCT data were identified.

<50 years

In a general population of adult women younger than fifty years old and at average risk of breast cancer, invitation to screening with mammography and ultrasound resulted in a statistically significantly decreased probability of being diagnosed with stage II or higher breast cancer compared to mammography alone (RR: 0.83, 95% CI: 0.70-0.98; absolute effect: 131 fewer per

100 000, 95% CI: from 231 fewer to 15 fewer; moderate certainty of evidence). Yet, the effect of screening modality was not statistically significant when the outcome was restricted to stage III or higher breast cancer cases (RR: 1.07, 95% CI: 0.71-1.61; absolute effect: 9 more per 100 000, 95% CI: from 36 fewer to 75 more; moderate certainty of evidence).

For both outcomes, the certainty of the evidence was downgraded one level because they were reported in only one adequately randomised trial.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

Mixed populations of women with dense and non-dense breasts

ALL AGE GROUPS

No data were retrieved for this age group for this outcome.

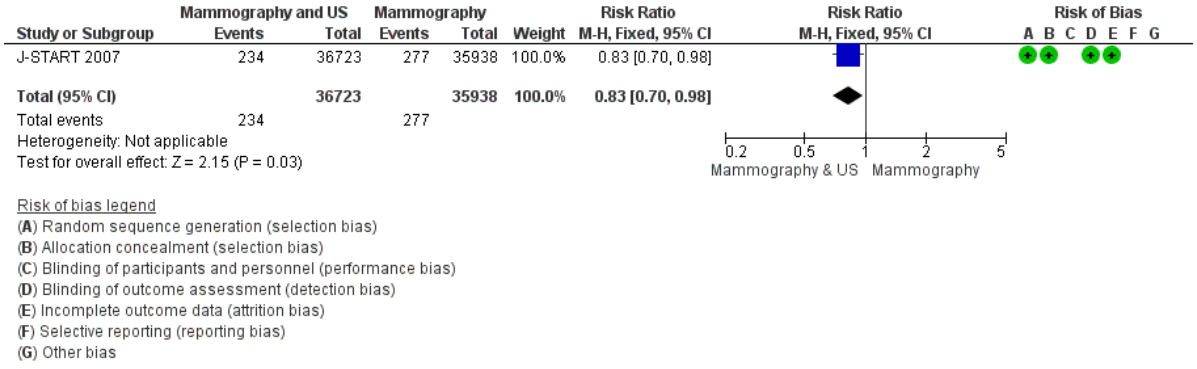
VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials

Again, the results for this outcome were published by Harada-Shoji et al. in February 2026, and hence not included in the SR by Glechner et al. (2023). After eleven years of follow-up, the results of the J-START trial pointed to a statistically significantly reduced probability of being diagnosed with a stage II or higher breast cancer when invited to screening with mammography and ultrasound compared to mammography alone (RR: 0.83, 95% CI: 0.70-0.98; Figure 23, Figure 25).[89]

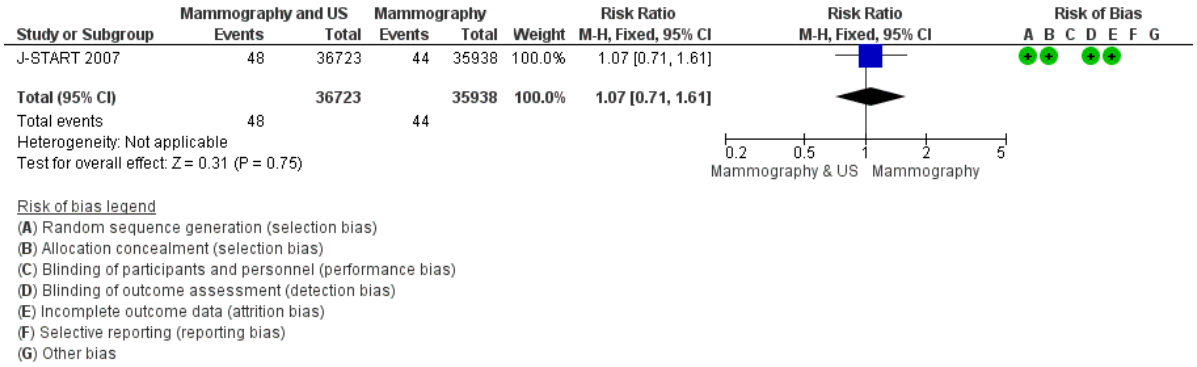
Figure 23 – Forest plot for mammography & US vs. mammography, outcome: detection of advanced breast cancer (Stage II or higher), 11 years FU



***Source**: Harada-Shoji et al. (2026).*[89]

Yet, when the outcome was restricted to stage III or higher breast cancers, there was no statistically significant difference between the group of women invited to screening with mammography and ultrasound compared to women invited to mammography alone (RR: 1.07, 95% CI: 0.71-1.61; Figure 24, Figure 25).[89]

Figure 24 – Forest plot for mammography & US vs. mammography, outcome: detection of advanced breast cancer (Stage III or higher), 11 years FU



***Source**: Harada-Shoji et al. (2026).*[89]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

3.2.4.7 Lymph node involvement at diagnosis

Key points

All age groups

No relevant RCT data were identified.

<50 years

In a general population of adult women younger than fifty years old and at average risk of breast cancer, invitation to screening with mammography and ultrasound does not have a statistically significant effect on the probability of lymph node involvement at diagnosis compared to mammography alone (RR: 0.78, 95% CI: 0.45-1.34; absolute effect: 18 fewer per 100 000, 95% CI: from 44 fewer to 27 more; moderate certainty of evidence). The certainty of the evidence was downgraded one level because the outcome was reported in only one adequately randomised trial.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

Mixed populations of women with dense and non-dense breasts

ALL AGE GROUPS

Observational studies

Glechner et al. (2023) identified four cohort studies[kkk] which provided data on this outcome. In these studies, no statistically significant differences were observed between the two screening modalities in terms of the proportion of women with positive lymph nodes amongst screen-detected invasive cancers. The review authors graded the certainty of evidence regarding this outcome as very low because of the methodological limitations and the imprecise result.[99]

VARIOUS AGE GROUPS

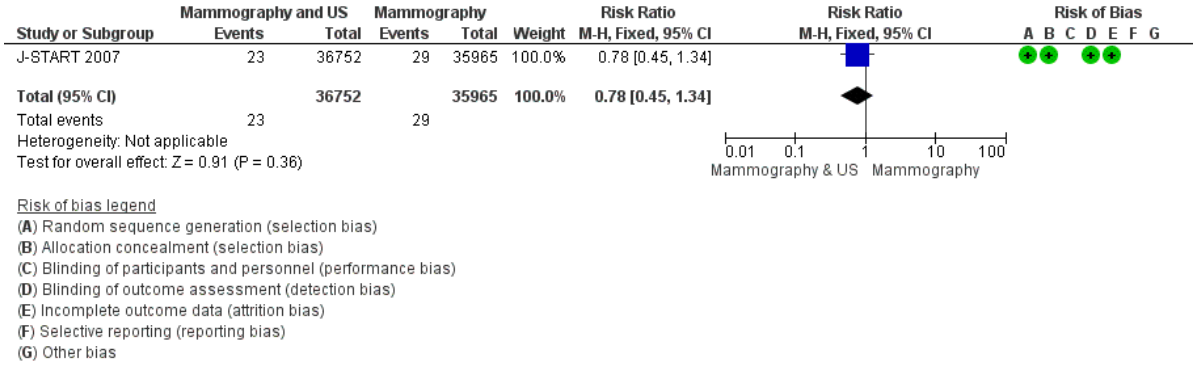
B. Women aged 40-49 years

Randomised controlled trials

In the J-START trial, the risk of being diagnosed with a screen-detected invasive breast cancer with positive lymph node status was not statistically significantly different between the mammography and ultrasonography group compared to the group who underwent mammography alone after two years of follow-up (RR: 0.78, 95% CI: 0.45-1.34; Figure 25[lll]).[167]

Again, Glechner et al. (2023) performed the analyses differently; they compared the proportion of cancers with lymph node involvement among the invasive cancer cases between the mammography and ultrasonography group and the mammography alone group.[99] Moreover, they reversed the data for this outcome between intervention and control group, hence their result is not reliable.[99]

Figure 25 – Forest plot for mammography & US vs. mammography, outcome: lymph node status at diagnosis, 2 years FU



***Source**: Ohuchi et al., 2016[167]*

C. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

---

kkk Four cohort studies: Buchberger et al. (2018);[170] Chough et al. (2020);[174] Giuliano et al. (2013);[172] Lee et al. (2019).[175]

lllll We relied for this outcome on the results presented in the original manuscript by Ohuchi et al., 2016.[167]

Women with dense breasts

ALL AGE GROUPS

Observational studies

Two cohort studies provided data for the SR by Glechner et al. (2023).[99] In the cohort study by Giuliano et al. (2013), only one woman in each intervention group was node-positive (RR: 0.45, 95% CI: 0.03-6.86), while in the study by Chough et al. (2020), where women participated in one or two screening rounds, all women with dense breasts were node-negative.[99]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

In the secondary analysis of the J-START trial that focused on women with dense breasts, after two years of follow-up comparable results were obtained in the group who had the combination of mammography and ultrasonography (18%, i.e. 5 out of 28), yet, in the mammography alone group a lower proportion of positive lymph node status was noted in women with screen-detected invasive breasts (11%, i.e. 2 out of 18).[99] However, the numbers were small, resulting in a non-statistically significant difference between the two intervention groups (RR: 1.61, 95% CI:0.35-7.42).[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Women with non-dense breasts

ALL AGE GROUPS

No data were retrieved for this outcome.

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

Likewise, in the secondary analysis of the J-START trial that focused on women with non-dense breasts, numbers were small and the differences between subgroups not statistically significant after two years of follow-up (RR: 0.47, 95% CI: 0.14-1.56).[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

3.2.4.8 Interval cancer

Key points

All age groups

No relevant RCT data were identified.

<50 years

In a general population of adult women younger than fifty years old and at average risk of breast cancer, invitation to screening with mammography and ultrasound results in a statistically significantly reduced detection rate for interval cancers compared to mammography alone after two years of follow-up (RR: 0.50, 95% CI: 0.29-0.89; absolute effect: 49 fewer per 100 000, 95% CI:

from 69 fewer to 11 fewer; moderate certainty of evidence). The certainty of the evidence was downgraded one level because the outcome was reported in only one adequately randomised trial.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

Mixed populations of women with dense and non-dense breasts

ALL AGE GROUPS

Observational studies

In the large retrospective cohort study by Buchberger et al. (2018), all women whose data were evaluated in the study had breast cancer screening with a combination of mammography and ultrasonography. Overall, in 28 out of 66 680 (0.04%) women interval cancers were diagnosed.[99]

The retrospective cohort study by Lee et al. (2019) with 36 143 screening exams in 18 562 women (who attended one or more screening rounds), fewer interval cancers were diagnosed in the mammography plus ultrasonography group, but the result was not statistically significant (RR: 0.79, 95% CI: 0.39-1.61).[99] Similarly, 1.5 interval cancer cases per 1000 screening examinations were reported in the mammography with ultrasonography group, compared to 1.9 in the mammography alone group (RR: 0.67, 95% CI: 0.33-1.37). The review authors assessed the certainty of the evidence for this outcome as low, given the methodological limitations of cohort studies.[99]

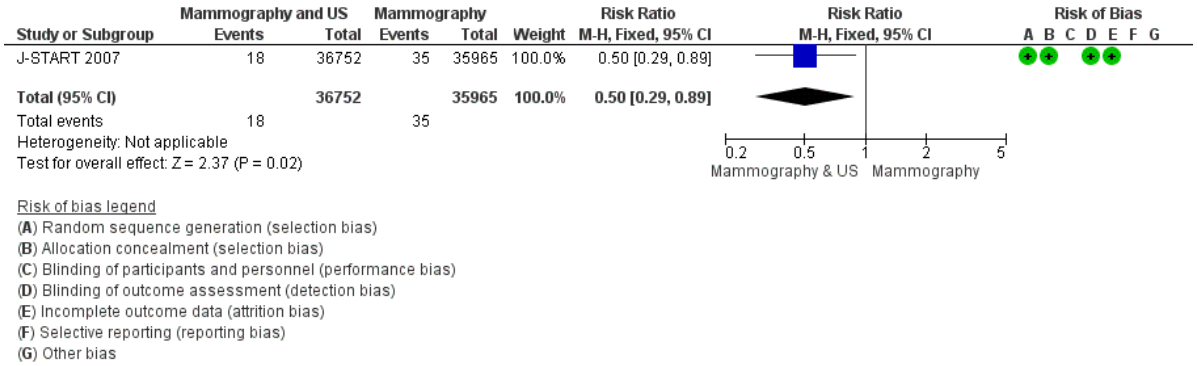
VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials

After two years of follow-up, interval cancers occurred in the J-START trial in 0.05% (18 out of 36 752) of women screened by mammography in combination with ultrasonography and 89% (16 out of 18) of these interval cancers were invasive, whereas in the group screened by mammography alone, the incidence of interval cancers was double (i.e. 0.10%, 35 out of 35 965; RR: 0.50, 95% CI: 0.29-0.89; Figure 26), with 77% (27 out of 35) of these being invasive.[99]

Figure 26 – Forest plot for mammography & US vs. mammography, outcome: interval cancers, 2 years FU



***Source**: Glechner et al. (2023).*[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Women with dense breasts

ALL AGE GROUPS

Observational studies

Glechner et al. (2023) obtained results from two cohort studies (Chae et al., 2013; Chough, et al., 2020), which were exclusively performed in women with dense breasts.[99] In the prospective cohort study by Chough et al. (2020), in which women attended one to two screening rounds, no interval cancers were identified. In the retrospective cohort study by Chae et al. (2013), which had according to the review authors a serious risk of bias, no interval cancers were diagnosed in women screened by mammography and ultrasonography, whereas in the control group which received only mammography, five interval cancers were seen (RR: 0.14, 95% CI: 0.01-2.46).[99]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

The secondary analysis of the J-START trial came after two years of follow-up to comparable results in women with dense and non-dense breasts. In women with dense breasts, the proportion of interval cancers was lower in the group of women who received mammographic screening in combination with ultrasonography compared to the group that received mammographic screening alone, yet the difference was not statistically significant (RR: 0.29, 95% CI: 0.08-1.05).[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Women with non-dense breasts

ALL AGE GROUPS

No data were retrieved for this age group for this outcome.

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

The secondary analysis of J-START indicated that in women with non-dense breasts, the proportion of interval cancers was after two years of follow-up lower in the group of women who received mammographic screening in combination with ultrasonography compared to the group that received mammographic screening alone, yet the difference was again not statistically significant (RR: 0.22, 95% CI: 0.05-1.03).[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

3.2.4.9 False-negative rate

Key points

All age groups

No relevant RCT data were identified.

<50 years

In a general population of adult women younger than fifty years old and at average risk of breast cancer, invitation to screening with mammography and ultrasound results after two years of follow-up in a statistically significantly reduced false-negative rate compared to mammography alone (RR: 0.39, 95% CI: 0.23-0.66; absolute effect: 14 046 fewer per 100 000, 95% CI: from 17 730 fewer to 7 829 fewer; moderate certainty of evidence). The certainty of the evidence was downgraded one level because the outcome was reported in only one adequately randomised trial.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

Mixed populations of women with dense and non-dense breasts

ALL AGE GROUPS

Observational studies

Glechner et al. (2023) identified two observational studies for this outcome.[99] In the retrospective cohort study by Buchberger et al. (2018), a statistically significantly lower false-negative rate was observed when the assessment was based on the combination of mammography and ultrasonography compared to mammography alone (9.4% (28 of 298) vs. 21.5% (64 of 298); RR: 0.44, 95% CI: 0.29-0.66). In the smaller retrospective study by Lee et al. (2019), in which women participated in one or more screening rounds, no statistically significant difference was observed between the two screening methods (21% (9 of 42) vs. 25% (56 of 221); RR: 0.85, 95% CI: 0.45-1.57). The review authors considered the certainty of evidence based on this cohort study as very low because of the methodological limitations and the inconsistent results.[99]

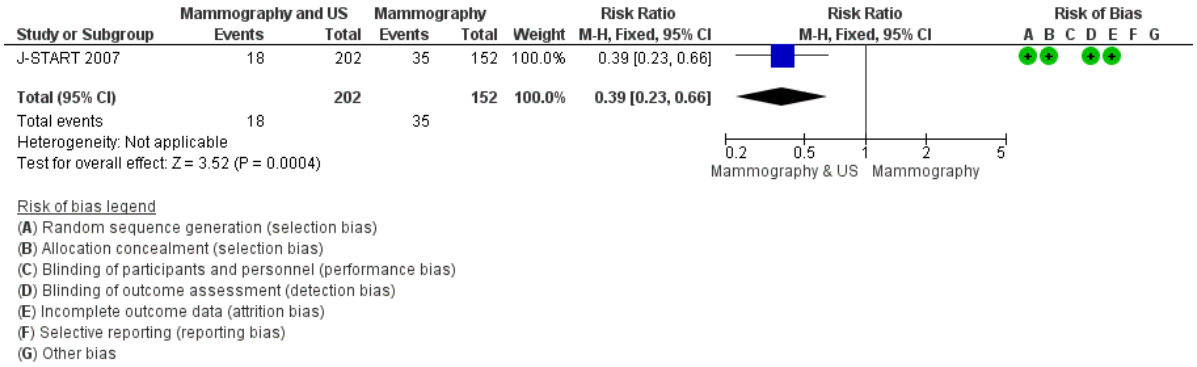
VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials

In the Japanese J-START trial, after two years of follow-up 202 women in the group who was screened with a combination of mammography and ultrasonography had cancer, of whom 18 (8.9%) were not detected by screening, hence classified as false negatives.[99] Among women with cancer who had screening with mammography alone, 35 out of 152 (23%) were misclassified as not having cancer. In other words, false-negative results among women who had cancer, were 61% less frequently observed in women who were screened with a combination of mammography and ultrasonography compared with women who underwent mammography alone (RR: 0.39, 95% CI: 0.23-0.66; Figure 27).[99]

Figure 27 – Forest plot for mammography & US vs. mammography, outcome: false-negative rate, 2 years FU



***Source**: Glechner et al. (2023).*[99]

In the original manuscript of the J-START trial, the sensitivity was reported[mmm]. Yet, the trialist underline that sensitivity and specificity were calculated with data from the first screening round. Because characteristics of breast cancer, like distribution of tumour size or sojourn time (i.e. the period during which a woman is asymptomatic but the breast cancer is detectable by mammography), differ between first and later rounds of screening, the authors notify that their findings should not be extended beyond the first round.[167]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Women with dense breasts

ALL AGE GROUPS

Observational studies

The cohort studies by Buchberger et al. (2018) and Chae et al. (2013) came to similar results. When the evaluation in the first study was based on the combination of mammography and ultrasonography the chance of being misclassified as not having cancer was smaller than when it was based on mammography alone (18.8% (18 of 96) vs. 38.5% (37 of 96); RR: 0.49, 95% CI: 0.30-0.79). In the study by Chae et al. (2013), no false-negative results were recorded in the group of women with dense breasts who had mammography and US, while five in eleven women (45.5%) who were screened with mammography alone received a false-negative result.[99] Again, the review authors graded the certainty of evidence based on these cohort studies as very low because of the methodological limitations and the heterogeneous results.[99]

[mmm] The J-START trialists report a significantly higher sensitivity of screening in the group who had mammography plus ultrasonography compared to the group with mammography alone (91.1%, 95% CI: 87.2–95.0 vs. 77.0%, 95% CI: 70.3–83.7; p=0.0004). To calculate the sensitivity of the first-round screening, screen-detected breast cancers were defined as those categorised as probably benign but further assessment needed, probably malignant or malignant.[167]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

The secondary analysis of the J-START trial, indicated that also in women with dense breasts who had cancer, the risk of false-negative results was statistically significantly lower with the combination of mammography and ultrasonography compared with mammography screening alone (6.8% vs. 29.4%; RR: 0.23, 95% CI: 0.07-0.78).[99] The review authors classified the certainty of evidence for this outcome also as moderate due to imprecision (i.e. wide CI).[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Women with non-dense breasts

ALL AGE GROUPS

Observational studies

Similar results were recorded for women with non-dense breasts: when assessment in women with non-dense breasts who had cancer was done with the combination of mammography and ultrasonography, the false-negative rate was lower than with mammography alone (10 of 202 (5.0%) vs 27 of 202 (13.4%); RR: 0.37, 95% CI: 0.18-0.74).[99] The review authors graded the certainty of evidence based on these cohort studies as very low because of the methodological limitations inherent to cohort studies and because of the heterogeneous results.[99]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

In the secondary analysis of J-START, a similar conclusion was made: the risk of false-negative results among women who had cancer was lower in the group screened with a combination of mammography and ultrasonography than in the subgroup who had mammography alone (6.9% vs. 39.1%; RR: 0.18, 95% CI: 0.04-0.74).[99] Again, because of the wide CI, the certainty of evidence was scored as moderate by the review authors.[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

3.2.4.10 False-positive rate

Key points

All age groups

No relevant RCT data were identified.

<50 years

In a general population of adult women younger than fifty years old and at average risk of breast cancer, invitation to screening with mammography and ultrasound results in a statistically significantly increased false-positive rate compared to mammography alone (RR: 1.43, 95% CI: 1.37-1.50; absolute effect: 3 706 more per 100 000, 95% CI: from 3 189 more to 4 310 more; moderate certainty of evidence). The certainty of the evidence was downgraded one level because the outcome was reported in only one adequately randomised trial.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

Mixed populations of women with dense and non-dense breasts

ALL AGE GROUPS

Observational studies

Glechner et al. (2023) based their conclusions for this outcome on the same two observational studies.[99] Buchberger et al. (2018) reported that in women who had breast cancer screening with mammography and ultrasonography the risk of a false-positive result was 11% higher compared to the group who had mammography alone (3.0% vs. 2.7%; RR: 1.11, 95% CI: 1.04-1.19). Similar results were obtained in the retrospective cohort study by Lee et al. (2019), where the proportion of false-positive results were 5.2% in the group who had the combined screening versus 2.2% in those who had mammography (RR: 2.34, 95% CI: 2.05-2.66).[99] Yet, the certainty of evidence of these studies was considered as very low because of the methodological limitations of cohort studies and the heterogeneous results.[99]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials

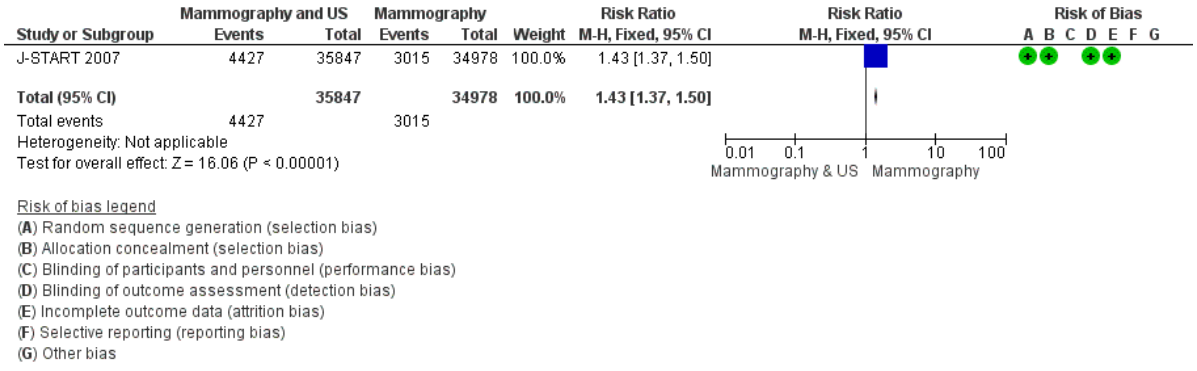
Comparable conclusions can be drawn from the J-START trial: the proportion of false-positive results among women with no cancer was higher in the group screened by mammography and ultrasonography compared to the group who had mammography alone (12.3% vs. 8.6%; RR: 1.43, 95% CI: 1.37-1.50; Figure 28).[99]

The J-START trialists suggest that the separate categorisation of the mammographic and ultrasound images may have led to the reduced specificity and an increased recall rate. In addition, they put forward that the low specificity may be due to the fact that in women with dense breasts ultrasonography can detect some lesions that are not visible on mammography.[167]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Figure 28 – Forest plot for mammography & US vs. mammography, outcome: false-positive rate, 2 years FU



***Source**: Glechner et al. (2023).*[99]

Women with dense breasts

ALL AGE GROUPS

Observational studies

Also in women with dense breasts, the proportion of false-positive results recorded in the retrospective cohort study by Buchberger et al. (2018), was lower than in J-START, with a statistically significantly higher risk of false-positive results when the result is based on mammography and ultrasonography (i.e. 2.1% (666 of 31 822) vs. 1.2% (385 of 31 822); RR: 1.73, 95% CI: 1.53-1.96).[99] Likewise, the retrospective study by Chae and colleagues (2013) pointed also to a higher false-positive rate in the group of women with dense breasts who were screened with mammography and ultrasonography in comparison to the group who had mammography alone (5.2% (432 of 8335) vs. 4.1% (517 of 12,494); RR: 1.25, 95% CI: 1.11-1.42).[99] The review authors graded the certainty of evidence of differences in terms of the rate of false-positive results between the two screening methods in women with dense breasts to be very low given the heterogenous results, risk of bias, and the methodological limitations inherent to cohort studies.[99]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

The secondary analysis of the J-START showed that in women with dense breasts, the risk of receiving a false-positive result after a screening with mammography and ultrasonography was 76% higher compared to screening with mammography alone (14.6% versus 8.3%; RR: 1.76, 95% CI: 1.58-1.96). The certainty of the evidence was considered high.[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Women with non-dense breasts

ALL AGE GROUPS

Observational studies

The largest retrospective cohort study (Buchberger et al., 2018), indicated that the proportion of false-positive results was similar between the two screening modalities (3.7% (1293 of 34 560) vs. 4.0% (1375 of 34 560); RR: 0.94, 95% CI: 0.87-1.01). Yet, it should be kept in mind that in this study, women with non-dense breasts were referred at the specific request of a radiologist, hence suspicious or unclear mammography results may have led to additional ultrasonography being performed.[99]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

Also in women with non-dense breasts, the secondary analysis of the J-START pointed to an elevated risk of getting a false-positive result after being screened with mammography and ultrasonography compared to screening with mammography alone (11.0% versus 8.1%; RR: 1.36, 95% CI: 1.18-1.56).[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

3.2.4.11 Rate of biopsies

Key points

All age groups

No relevant RCT data were identified.

<50 years

In a general population of adult women younger than fifty years old and at average risk of breast cancer, invitation to screening with mammography and ultrasound results in a statistically significantly increased rate of biopsies compared to mammography alone (RR: 2.49, 95% CI: 2.28-2.72; absolute effect: 2 714 more per 100 000, 95% CI: from 2 331 to 3 132 more; moderate certainty of evidence). The certainty of the evidence was downgraded one level because the outcome was reported in only one adequately randomised trial.

50-69 years

No relevant RCT data were identified.

≥70 years

No relevant RCT data were identified.

Mixed populations of women with dense and non-dense breasts

ALL AGE GROUPS

Observational studies

In the retrospective study by Lee et al. (2019), which was performed among women across the spectrum of breast cancer risk, the risk of receiving a biopsy after screening was twice as high among women who had an additional ultrasound compared to women who only had a mammography (5.7% vs. 2.8%; RR: 2.05, 95% CI: 1.79-2.34). In the cohort study by Buchberger et al. (2018), biopsy rates in both subgroups were much lower than in J-START and the cohort study by Lee et al. (2019): 623 of 66 680 (0.9%) in women who were screened with mammography and ultrasonography and 422 of 66 680 (0.6%) in women who had mammography alone. The lower biopsy rate may (in part) be attributed to the intermediate mammograms (at six months) that were taken in a subgroup of patients.[99] Yet, also here the rate was statistically significantly higher in those who had mammography and ultrasonography (RR: 1.48, 95% CI: 1.31-1.67). The review authors graded the certainty of evidence regarding biopsy rates as very low due to the heterogeneous results and methodological limitations inherent to cohort studies.

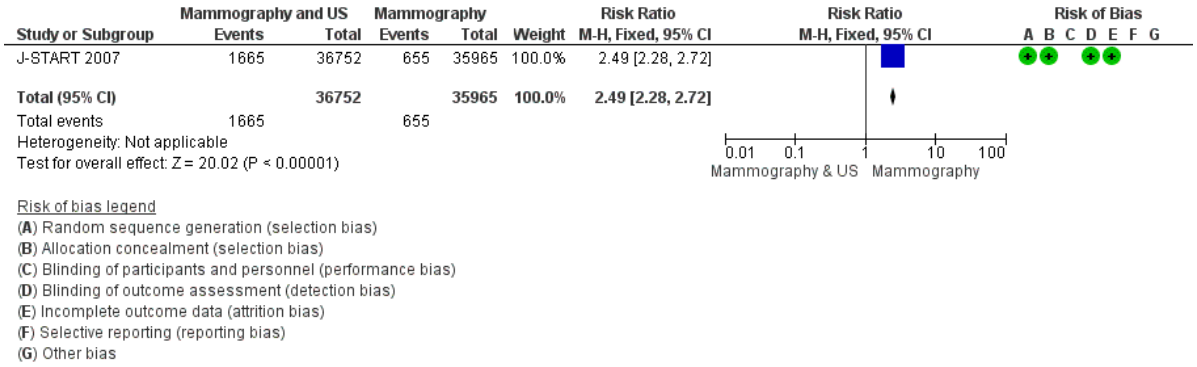
VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials

Among the 40-49 year-old Japanese women included in the J-START, 1665 out of 36 752 (4.5%) women who had mammography screening and additional ultrasonography underwent a biopsy, versus 655 out of 35 965 (1.8%) women who had mammography only. The biopsy rate was thus 2.5 times higher in the mammography plus ultrasound group compared to the mammography group (RR: 2.49, 95% CI: 2.28 - 2.72; Figure 29).[99]

Figure 29 – Forest plot for mammography & US vs. mammography, outcome: rate of biopsies, 2 years FU



***Source**: Glechner et al. (2023).*[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Women with dense breasts

ALL AGE GROUPS

Observational studies

The cohort study by Buchberger et al. (2018) obtained similar results: the risk of receiving a biopsy was 85% higher when the assessment was done based on the combination of mammography and ultrasonography compared to mammography alone (RR: 1.85, 95% CI: 1.47-2.32).[99]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

The secondary analysis of the J-START trial, indicated that young women with dense breast had a similar higher risk of undergoing a biopsy when ultrasonography was added to the screening compared to screening with mammography alone (360 of 5797 (6.2%) vs. 127 of 5593 (2.3%); RR: 2.73, 95% CI: 2.24-3.34).[99] The review authors classified the certainty of evidence for this outcome also as high.[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

Women with non-dense breasts

ALL AGE GROUPS

Observational studies

The cohort study by Buchberger et al. (2018) came to comparable conclusions. When the assessment was done based on the combination of mammography and ultrasonography the chance of receiving a biopsy was higher than when it was based on mammography alone (RR: 1.34, 95% CI: 1.16-1.55).[99]

VARIOUS AGE GROUPS

A. Women aged 40-49 years

Randomised controlled trials[fff]

Also in women with non-dense breasts, the secondary analysis of J-START pointed to a higher risk of having to undergo a biopsy when screening was based on the combination of mammography and ultrasonography compared to screening with mammography alone (RR: 2.58, 95% CI: 1.96-3.40).[99] The review authors classified the certainty of evidence for this outcome also as high.[99]

B. Women aged ≥50 years

No data were retrieved for this age group for this outcome.

3.2.4.12 Health-related quality of life

Key points

No relevant RCT data were identified.

None of the studies included in systematic review by Glechner et al. (2023) assessed whether the combination of mammography screening and ultrasonography had an effect on quality of life.[99]

3.2.4.13 Conclusion

The currently available **RCT evidence is limited**. So far only one large and methodologically very well performed RCT in Japanese **women younger than fifty years** old has been published.[89, 99, 167]

After eleven years of FU, invitation to screening with the combination of mammography and ultrasonography **does not have a statistically significant effect on all-cause mortality** (RR: 1.04, 95% CI: 0.88-1.22; absolute effect: 31 more per 100 000, 95% CI: from 93 fewer to 171 more), **nor on breast cancer mortality** (RR: 0.98, 95% CI: 0.53-1.79; absolute effect: 1 fewer per 100 000, 95% CI: from 27 fewer to 46 more) compared to being invited to mammography alone.[89]

Yet, the proportion of **false-positive results was statistically significantly higher** than with mammography screening alone (RR: 1.43, 95% CI: 1.37-1.50; absolute effect: 3 706 more per 100 000, 95% CI: from 3 189 more to 4 310 more), which leads to unnecessary harm caused by subsequent investigations while there is no gain in health benefits.[89] Therefore, we conclude with Glechner et al. (2023) that currently there is no methodologically sound evidence justifying the routine use of US as an adjunct to mammography in breast cancer screening in women who are not at increased risk of breast cancer.[99]

4 SYSTEMATIC REVIEW OF ECONOMIC EVALUATIONS ON BREAST CANCER SCREENING

Key points

This systematic review includes 24 studies, consisting of 7 systematic reviews and 17 primary evaluations. The 17 primary evaluations were analysed in depth in the critical assessment.

While the majority of studies conducted a **cost-utility analysis**, four studies unfortunately solely performed cost-effectiveness analyses and therefore did not take into account the impact of breast cancer screening, treatment as well as screening-related harms on QoL.

All included economic evaluations opted for a **stage shift approach** to model screening effectiveness. Instead of using the TNM stage classification, numerous studies chose to use tumour size alone as a proxy for stage. In addition, transparency in methodology, input parameters and sources to model screening effectiveness (the stage shift) was in general relatively poor. None of the evaluations was purely based on a RCT or RCT evidence, meaning other sources had to be used to model screening effectiveness. In general, two approaches could be distinguished: (1) Use of registry-based stage distributions and (2) Simulation of tumour growth or size at various ages combined with tumour detection probability by screening. In the former approach, studies examining the cost-effectiveness of an existing screening programme relied on the stage distribution before the screening programme had started to mimic the situation without screening (comparator). Those studies examining extensions of screening programmes in terms of age eligibility criteria employing the first approach generally used the stage distributions of the closest screened age (group) to mimic the stage distribution in the age group to which screening is extended (intervention). The studies employing the second approach generally simulated tumour growth or size at various ages as well as tumour detection by screening based on screening sensitivity and specificity parameters or a detection probability function compared to detection without screening. Within this approach, multiple studies used some level of RCT evidence as inputs for their respective models. Unfortunately, especially this second approach was characterized by a lack of transparency in methodology, input parameters and sources used to model screening effectiveness.

The underlying **Quality of Life (QoL) studies** used by the various evaluations are characterized by methodological strengths and shortcomings of their own. For example, a substantial number of studies estimated QoL values by asking the general public or clinical/public health experts to evaluate QoL associated with hypothetical descriptions of health states, events, treatments, and harms instead of asking breast cancer patients to evaluate the QoL they currently experience in their specific health state. Finally, only a few QoL studies chose to measure stage-specific QoL values. Most studies focused instead on estimating QoL for specific treatments (e.g. radiotherapy, chemotherapy, etc.) or health states not linked to a specific classification system (e.g. inoperable locally advanced breast cancer). Most QoL studies are performed in non-European populations.

Irrespective of which modelling approach was used, it is important to correct for **screening-related harms** and to transparently report how these were taken into account in the models. A substantial number of included studies did not meet these requirements. A lack of transparency was primarily present in the way models handled **false positives** and overdiagnosis. Only half of the evaluations reported the number of false positives resulting from the examined screening strategies and it was unclear in the majority of evaluations which costs were assigned to false positives. With regard to QoL, one study assigned a specific disutility to false-positive cases while the other evaluations generally assigned the disutility related to the screening mammography and/or the disutility associated with diagnostic work-up to these cases.

While the majority of the included studies does claim taking into account **overdiagnosis**, only about half of the studies reported the magnitude of overdiagnosis in their models either as an input or

output with considerable variation between them, likely stemming from the uncertainty around the true overdiagnosis rate in the clinical literature. From a methodological point of view, overdiagnosed cases cannot benefit from the screening in any way and should be assigned the costs they bring forward (i.e. screening, treatment, follow-up). Unfortunately, there is a considerable lack of transparency in how overdiagnosis was handled methodologically in the evaluations, for example in which stage(s) these cases were diagnosed and on which costs were assigned to overdiagnosed cases. With the exception of two studies, all other studies with transparent reporting make the assumption in the base case that overdiagnosed cases are solely DCIS. In a scenario analysis, one evaluation showed that in case overdiagnosis would not be limited to DCIS the impact on the ICER could be highly substantial (see 4.3.2). With regard to QoL, it is not clear whether the studies assigned an additional disutility resulting from the unnecessary treatment and the associated consequences (e.g. complications, anxiety and distress) related to unnecessarily labelling someone as a cancer patient.

While in general the included costs and their sources were reported transparently, **costs related to inviting the screen-eligible population and especially the costs associated with the organization of the screening programme** were insufficiently taken into account in most studies. Many studies downsized the costs of screening to solely the cost of a mammography, which underestimates the screening costs and subsequently leads to an overly optimistic, but incorrect estimation of the cost-effectiveness of breast cancer screening. In addition, many studies implicitly or explicitly assumed an unrealistic participation rate of 100%.

Finally, three studies taking a **societal perspective** also included indirect costs. All three studies incorporate the loss in productivity due to premature death from breast cancer, which favors screening over no screening in terms of cost-effectiveness. However, none of the studies take into account the productivity loss associated with participating in the screening programme and overdiagnosis, which would have an unfavourable impact on the cost-effectiveness.

4.1 Methods

4.1.1 Research questions

The aim of this chapter is to systematically review the available evidence on the cost-effectiveness of breast cancer screening modalities and strategies in an adult female population not at high risk for breast cancer. The purpose of this review is twofold. First, to check whether one or more health economic evaluations can be identified that can be transposed to or are performed in the Belgian health care context. In this case, conducting a primary analysis ourselves may not be required. Second, by systematically reviewing the literature, the methodological approaches taken by other evaluations (e.g. handling of false positives and overdiagnosis) as well as the chosen inputs (e.g. on QoL) can provide input for a de novo economic evaluation. The following research questions are addressed in this chapter:

RQ1: What is the cost-effectiveness of invitation to breast cancer screening with mammography in a population of adult women

- aged 50-69 years (i.e. current screening programme in all Belgian regions)
- aged 40-49 years
- aged ≥70 years

without suspected breast cancer and without a specifically increased risk of breast cancer?

RQ2: What is the cost-effectiveness of invitation to

- biennial vs. annual, and

- biennial vs. triennial

breast cancer screening with mammography in a population of adult women without suspected breast cancer and without a specifically increased risk of breast cancer?

RQ3: What is the cost-effectiveness of breast cancer screening using mammography combined with ultrasound compared to mammography alone in a population of adult women without suspected breast cancer and without a specifically increased risk of breast cancer?

A systematic literature review was conducted to provide an answer to the abovementioned research questions and to systematically review the methodologies used by the included studies.

4.1.2 Search strategy

The systematic search for relevant health economic literature was conducted in two stages:

First, a search for Health Technology Assessments (HTAs) on breast cancer screening was performed on the websites of HTA institutes which are members of the International Network of Agencies for Health Technology Assessment (INAHTA) and, by extension, the institutes evaluating preventive health policies. The full list of consulted institutes is available in Appendix 10. The search terminology used to search the websites was 'breast cancer screening', 'mammography screening', 'dépistage du cancer du sein' (French-speaking sources only), and 'borstkankerscreening' (Dutch-speaking sources only). The first stage was conducted in April 2024.

Second, a systematic search for full primary economic evaluations and systematic reviews of these evaluations was launched in EMBASE, Medline (Ovid), PsycInfo (Ovid), and EconLit (Ovid). The search strategy was based on the search strategy of HTAs identified in the first stage, supplemented with economic search terminology of KCE Report 379 on lung cancer screening. More details can be found in Appendix 11. The search for the second stage was performed in April 2024. Systematic reviews were included if published in the past 10 years (i.e. since 2014), whereas the primary evaluations were included if published from 2016 onwards. This time constraint was imposed to reflect as much as possible the current clinical practices and to build further upon the already existing systematic reviews on the topic. Due to the substantial number of systematic reviews identified in the past ten years, it was not necessary to extend the search with an earlier start date. As the systematic review of Schiller-Früwirth et al. (2017b), which included studies up to November 2016, is the only review analysing general modelling techniques and the approaches used to model screening effectiveness, 2016 was chosen as cut-off for the primary evaluations.[176] For studies performed in Belgium or the Netherlands, publications from 2000 onwards were included.

The economic evaluations identified by the search strategy were assessed for eligibility using pre-defined inclusion and exclusion criteria as presented in Appendix Table 91.

4.1.3 Quantity of available evidence

The first stage of the search identified 70 potentially relevant publications. After screening the titles and abstracts, eight HTAs with both a clinical and economic evaluation were retained. After full text evaluation, seven HTAs were excluded (see Appendix Table 96). The remaining HTA performed a systematic literature review on the cost-effectiveness of ultrasound in adjunct to mammography in breast cancer screening but only included one primary study. In addition, a relevant report was retrieved on a study recently performed by the Interuniversity Centre for Health Economics Research (I-CHER) which was requested by the Flanders Department of Care.

The selection process of the second stage search strategy is summarized in Appendix Table 92 and Appendix Figure 89.

The initial search yielded 6115 results from which 1308 duplicates were removed. Of the remaining 4807 citations, 3243 were published before 2016, resulting in 1564 potentially relevant citations to be reviewed based on title and abstract. The remainder of the selection process was based on the

inclusion & exclusion criteria in Appendix Table 91 and is depicted in Appendix Figure 89 using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart standards. In the first round, 1481 citations were excluded based on title and abstract and 83 were retained. In the second round, another 51 citations were excluded based on the setting of the study and restrictions in scope related to the population and intervention. The reference and reason for exclusion for these excluded studies is provided in Appendix Table 97. Based on full text evaluation, 11 additional studies were excluded based on the inclusion & exclusion criteria (see Appendix Table 97 for reason of exclusion). In the next phase, two relevant studies from the first stage were added as well as one relevant primary study from the Netherlands published before 2016 (i.e. 2015) previously identified by the included systematic reviews. Finally, 24 studies were included in this chapter consisting of seven systematic reviews and 17 primary evaluations (see Appendix Table 95). Consequently, these 17 primary evaluations were analysed in depth in the critical assessment. An update was carried out in December 2024, which did not identify any additional relevant studies published between April and December 2024.

As a check for the performance of the search strategy employed in this report, Appendix Table 94 provides an overview of the included studies in the retrieved systematic reviews (including those performed within an HTA) from 2000 onwards. As shown in the second column of Appendix Table 94, the search strategy was able to identify all relevant economic evaluations identified by previously performed systematic reviews with the exception of a study by IMS Health Pty Ltd Australia, which originates from an archived web page and was therefore not included in the search engines we used.[177] As the study was published before 2016 in another country than Belgium or the Netherlands, it was not eligible for inclusion in this report. The exclusion criteria for studies included by other systematic reviews but excluded in this report are available in Appendix 12 (Table 97).

4.1.4 Data extraction

An in-house developed data extraction sheet was used to extract the relevant information from the selected publications. The template of the data extraction sheet is available in Appendix 14. The information gathered in the data extraction sheets served as the basis for the summary tables in the report.

4.2 Overview of primary economic evaluations

4.2.1 General characteristics

Key points

- One study was conducted for the setting of Flanders (Belgium) and three for the Netherlands. The other studies represent a variety of countries, including six other European countries, the USA (5 studies), Canada (1 study), and Singapore (1 study).
- While the majority of studies opted for a cost-utility analysis, combined (or not) with cost-effectiveness analysis (13/17), four studies solely conducted a cost-effectiveness analysis by not taking into account quality of life.
- Various types of models were used, with state-transition modelling being the most commonly chosen analytic approach, followed by discrete event simulation. Rarely chosen alternatives include the decision tree and a flexible stage distribution model. Four studies were based on the previously validated MISCAN model.
- The most frequently taken perspective in the selection of studies was the health care payers perspective (9/17), sometimes in addition to the societal perspective (2/17).
- Except for one study, which modelled costs and effects over a time horizon of 20 years, all other studies modelled them over a lifetime horizon.

- The majority of studies opted for constant discounting of costs and health effects at a rate of 3% or 3.5%. In line with their national guidelines, only two studies discounted costs and health outcomes at different discount rates in the base case. Two studies failed to report the discount rates used for future costs and/or effects.
- With the exception of five studies, all other studies provided conflict of interest statements of which two disclosed a potential conflict of interest.

The general characteristics of the included primary evaluations are summarized in Table 6. While the selection criteria ensured including studies from comparable countries in terms of the human development index and national income via the world bank classification, the included studies were performed in a considerable variety of **countries**. Five studies were conducted for the setting of the United States of America (USA), three for the Netherlands and two for the United Kingdom (UK). The remaining studies represent the settings of Austria, Flanders (Belgium), Canada, Finland, Norway, Singapore and Spain with one study for each country. While six studies solely conducted a cost-utility analysis (CUA), four studies opted to solely conduct a cost-effectiveness analysis (CEA) by not taking into account quality of Life (QoL). In line with the Belgian guidelines for economic evaluations, four studies conducted both a CUA and CEA. An additional three studies expressed the incremental health effects in life years gained (LYG) and quality-adjusted life years (QALYs) but only calculated the Incremental Cost-Effectiveness Ratio (ICER) in terms of incremental cost/QALY gained.

As shown in Table 6 (column 4), a variety of **modelling techniques** were used by the selected studies. The most common methodology to model the cost-effectiveness or cost-utility of breast cancer screening is a state-transition model, opted for by eight of the seventeen studies.[180, 181, 190-193, 197, 200] Of these eight models, seven were analysed using individual level microsimulation while I-CHER (2024) is the only study preferring cohort simulation.[200] Six studies employed discrete event microsimulation modelling to evaluate the cost-effectiveness of breast cancer screening.[183, 184, 187, 189, 195, 196] Morton et al. (2017) and Blankenburg et al. (2023) built a decision tree to evaluate the cost-effectiveness of breast cancer screening, with Blankenburg et al. (2023) combining the decision tree with a state-transition model.[179, 194] Finally, Shafik et al. (2023) opted for a mathematical flexible stage distribution model, which estimates the stage distributions at diagnosis of breast cancer under different screening strategies and attaches stage-specific survival and costs without explicitly modelling transitions between the states.[178] Interestingly, four studies used the MIcro-simulation SCreening Analysis (MISCAN) model, a previously validated health economic model originating from the Netherlands, (as the basis) for their models.[181, 184, 192, 197] While the original MISCAN and MISCAN-Breast models use tumour stages, MISCAN-Fadia is an adapted version using a natural history component simulating a continuously growing tumour until it reaches a fatal diameter (FaDia). More information on how screening effectiveness was modelled in the various studies is available in section 4.2.3.1 and Table 8.

The **perspective** from which each evaluation is performed, which determines which cost categories should be included in the analyses, is reported in column 5 of Table 6. Slightly more than half of the studies (9/17) opted to conduct the analyses solely from the health care payers perspective by including the costs incurred by the health care system as a whole (including costs paid by the health care budget and patients' out-of-pocket expenses).[179, 181, 189, 190, 193-197] Tina Shih et al. (2019) and Tina Shih et al. (2021), on the other hand, chose to conduct their analyses solely from the broadest 'societal perspective', implying that they also included the direct and indirect non-medical costs (such as productivity loss and transportation costs).[183, 187] van Luijt et al. (2017) and I-CHER (2024), who conducted their analyses from the societal perspective, also report the ICER(s) from the health care payers perspective. While three studies only included costs related to productivity loss,[183, 187, 200] van Luijt et al. (2017) additionally included travel costs for screening and follow-up (for more information on costs, see section 4.2.4).[192] In contrast to the perspectives mentioned above, Schousboe et al.

(2022) conducted their analyses from the national health payer perspective by not taking into account the out-of-pocket payments of the patients.[180] Chootipongchaivat et al. (2021) mention opting for the healthcare provider perspective, implying only the costs incurred by those delivering health care (such as hospitals, physicians, etc.) are included. However, the authors do seem to take into account out-of-pocket payments of the patients for e.g. the digital screening mammography.[184] It is therefore not clear for which actors the incurred costs were taken into account in their analysis. Finally, both Shafik et al. (2023) and Carter et al. (2018) did not report the perspective out of which their respective evaluations were carried out.[178, 191]

Table 6 – Model characteristics of primary evaluations

	Country	Analytic technique*	Type of model	Perspective	Time horizon	Discount rate costs	Discount rate benefits	Conflict of interest
Journal publications								
Shafik, 2023[178]	Finland	CEA	Flexible stage distribution model	*Not reported*	Lifetime (until death or 100y)	*Not reported*	*Not reported*	No
Blankenburg, 2023[179]	United States of America	(CEA &) CUA	Decision tree & Markov model	Health care payers	Lifetime	3%	3%	Potential COI declared
Schousboe, 2022[180]	United States of America	CUA	State-transition Markov micro-simulation model	National health payer	Lifetime (until death or 100y)	3%	3%	No
Kregting, 2022[181]	The Netherlands	CUA	MISCAN-Breast: state-transition microsimulation model	Health care payers	Lifetime (until death or 100y)	3.5%	3.5%	No
Tina Shih, 2021[183]	United States of America	(CEA &) CUA	Discrete event microsimulation model	Societal	Lifetime	3%	3%	No
Chootipongchaivat, 2021[184]	Singapore	(CEA &) CUA	MISCAN-Fadia: discrete event microsimulation model	"Health care provider"	Lifetime	3%	3%	No
Tina Shih, 2019[187]	United States of America	CUA	Discrete event microsimulation model	Societal	Lifetime	3%	3%	*Not reported*
Mittmann, 2018[189]	Canada	CEA & CUA	University of Wisconsin Breast Cancer Epidemiology Simulation Model: Discrete event microsimulation model	"Publicly funded health care system"	Lifetime	1.5%	1.5%	Potential COI declared

	Country	Analytic technique*	Type of model	Perspective	Time horizon	Discount rate costs	Discount rate benefits	Conflict of interest
Koleva-Kolarova, 2018[190]	The Netherlands	CEA	SiMRiSc: State-transition Markov microsimulation model	Health care payers	Lifetime	4%	1.5%	No
Carter, 2018[191]	United States of America	CEA	State-transition microsimulation model	*Not reported*	Lifetime	3%	3%	*Not reported*
van Luijt, 2017[192]	Norway	CUA	MISCAN: state-transition micro-simulation model	Societal & health care payers	Lifetime	3.5%	3.5%	*Not reported*
Schiller-Frühwirth, 2017[193]	Austria	CEA	State-transition Markov micro-simulation model	Health care payers	Lifetime	3%	3%	*Not reported*
Morton, 2017[194]	United Kingdom	CUA	Decision tree (cohort-level)	NHS (health care payer)	20 years	3.5%	*Not reported*	No
Rafia, 2016[195]	United Kingdom	CEA & CUA	Discrete event micro-simulation model	NHS (health care payer)	Lifetime	3.5%	3.5%	No
Arrospide, 2016[196]	Spain	CUA	Discrete event micro-simulation model	Health care payers	Lifetime	3%	3%	No
Sankatsing, 2015[197]	The Netherlands	CEA & CUA	MISCAN: state-transition micro-simulation model	Health care payers	Lifetime	3.5%	3.5%	No
Other reports								
I-CHER, 2024[200]	Flanders (Belgium)	CEA & CUA	State-transition model (cohort-level)	Societal & health care payers	Lifetime (until death or 100y)	3%	1.5%	*Not reported*

***Notes**: *Analytic technique between brackets means incremental costs and incremental effects were reported, but no ICER was calculated using this information.*
***Abbreviations**: CEA = Cost-Effectiveness Analysis, CUA = Cost-Utility Analysis, MISCAN = MIcro-simulation SCreening Analysis, NHS = National Health Service, y = years, COI = Conflict of Interest*

In addition to the perspective, the **time horizon** over which the health economic evaluations are performed determines to what extent future costs and health effects are taken into account in the analyses. As shown in Table 6, with the exception of Morton et al. (2017)[194], all evaluations adopted a lifetime horizon with four studies reporting having capped the upper age limit at 100 years. In contrast, Morton et al. (2017) selected a considerably shorter time horizon of 20 years, without providing justification for this choice.[194] Only I-CHER (2024) explored the impact of the time horizon on the outcomes in the sensitivity analyses with scenarios for 20 and 50 years.[nnn200]

Regarding **discount rates,** which are used to bring future costs and health effects to their present value, approximately half of the studies applied a discount rate of 3% to both future costs and health effects (8/17).[179, 180, 183, 184, 187, 191, 193, 196] Four studies opted for a slightly more elevated discount rate of 3.5%, meaning future costs and health effects reduce in weight compared to a lower discount rate.[181, 192, 195, 197] Only Mittman et al. (2018) conducted their analyses using a considerably lower discount rate of 1.5% for both costs and effects in line with the Canadian guidelines for health economic evaluations.[189] Only two studies used a different discount rate for costs and effects following their respective national guidelines, with Koleva-Kolarova et al. (2018)[190] using 4% for costs and 1.5% for effects (cfr. Dutch guidelines for economic evaluation at the time)[ooo] and I-CHER (2024)[200] 3% and 1.5% respectively (cfr. Belgian guidelines for economic evaluation). While Morton et al. (2017) only reported a discount rate for the costs of 3.5%, no discount rate was reported for the health effects.[194] In addition, Shafik et al. (2023) did not report any discount rate for costs or effects.[178] Failing to provide this information (or not discounting costs and effects) can be considered as a shortcoming of these particular studies. Only five studies reported their results using other discount rates than the ones used in the base case analysis.[179, 189, 195, 197, 200]

While twelve out of seventeen studies disclosed whether potential **conflicts of interest** might be present, five unfortunately did not (see Table 6). Of these twelve studies, two disclosed a potential conflict of interest. More specifically, the study of Mittman et al. (2018) was funded by the Canadian Breast Cancer Foundation and the study of Blankenburg et al. (2023) was funded by Bayer AG with multiple authors also being employed by Bayer AG at the time the study was carried out.[179, 189]

4.2.2 Population, intervention, comparator, and outcome characteristics

Key points

Population:

- The most frequently examined age windows include 50-74/75 years and 50-69/70 years, which corresponds to the age eligibility criteria of currently organized screening programmes in the examined countries.
- Other studies mainly focus on the age windows 45/46-74 years and 40-74/75 years, which in turn corresponds to recently issued recommendations by ECIBC and IQWIG for the former and by USPSTF for the latter.
- The earliest examined starting age is 40 years and the latest stopping age is lifelong screening.

Intervention:

- Almost all studies (15/17) evaluated biennial breast cancer screening.
- Other frequently examined screening intervals include annual and triennial screening. Rarely examined alternatives are hybrid screening intervals and quadrennial screening.

[nnn] The results of these sensitivity analyses are discussed in section 4.3.2.

[ooo] In the most recent version of the Dutch guidelines for economic evaluation, these discount rates were changed to 3% for costs and 1.5% for health effects.

- While the large majority of studies (14/17) examined a screening protocol solely based on age, three studies evaluated a screening protocol based on age and breast density (once known).
- All but two studies evaluated mammography screening as imaging technique for breast cancer screening. The two remaining studies included a supplemental ultrasound for women with dense breasts.

Comparator:

- No screening (11/17) and/or the screening programme currently in place in the examined country (7/17) were the most frequently used primary comparators.

Outcome:

- Most studies express the ICER in terms of incremental cost/QALY gained (13/17), with four of these studies additionally expressing the ICER in incremental cost/LYG.
- Four studies did not take into account the impact of breast cancer (screening) on QoL, by solely calculating the ICER in terms of incremental cost/LYG.

4.2.2.1 Population

With age being the primary eligibility criterium in breast cancer screening for the female average risk population, columns 2 and 3 of Table 7 show the **starting and stopping ages** of the screening strategies (either as intervention or comparator) examined in the evaluations. The most frequently examined age windows include 50-74/75 years (10 studies as intervention and 1 as primary comparator) and 50-69/70 years (7 studies as intervention and 3 as primary comparator), followed by 45/46-74 years (8 studies as intervention), 40-74/75 (7 studies as intervention) and 45/46-69 years (5 studies as intervention). The earliest examined starting age is 40 years and the latest stopping age is lifelong screening.

4.2.2.2 Intervention

An important characteristic of any screening programme, which can largely influence health effects as well as costs, is the frequency with which an eligible women is screened (i.e. **screening interval**). As shown in column 4 of Table 7, the included studies examined various (combinations of) screening intervals. Biennial screening, the current interval in many countries (including Belgium), was included by all but two studies.[194, 195] Annual and triennial screening were each examined by seven studies. Only Kregting et al. (2022) evaluated the cost-effectiveness of quadrennial screening.[181] In addition, seven studies chose to evaluate hybrid screening intervals, in which a subset of the screen-eligible population is screened at a different frequency than other subsets. Six studies defined these subsets based on age, meaning women are screened at frequency x until they reach a certain age after which they are screened at frequency y. For example, the study by I-CHER (2024) evaluates a scenario where women are screened biennially between 45 and 69 years after which the frequency changes to triennial screening between 70 and 74 years, which is in line with the current recommendations of the European Commission Initiative on Breast Cancer (ECIBC).[200] The other five studies evaluated annual screening for the younger women and biennial screening for the older subset of women.[184, 187, 189, 191, 197] Only Tina Shih et al. (2021) defined the screening frequency based on breast density, where screen-eligible women with dense breasts are screened annually and those with non-dense breasts are screened either biennially or triennially.[183] Nine studies compared the cost-effectiveness of various screening intervals in scenario analyses.[179-181, 183, 184, 187, 189, 191, 197] For the impact of the screening interval on the cost-effectiveness of breast cancer screening, we refer to section 4.3 of the report.

As a consequence of the inclusion and exclusion criteria (see Appendix Table 91), two **types of screening** were included in this report: (1) screening protocol & eligibility is solely based on age and (2) screening protocol & eligibility is based on age and breast density. While the former was evaluated

by the large majority of studies (14/17), three studies opted to evaluate the latter. In the study by Blankenburg et al. (2023), the authors modelled that women with dense breasts who had a negative screening mammography additionally received an ultrasound.[179] Schiller-Frühwirth et al. (2017), on the other hand, assumed that for every screening mammography 0.35 ultrasounds were performed to incorporate adjacent ultrasound for women with dense breasts into their model.[193] These two studies are also the only studies in our selection evaluating other **imaging** techniques than mammography alone. In contrast to different imaging modalities, Tina Shih et al. (2021) adapted the screening interval for women with dense breasts (see above).[183] While I-CHER (2024) evaluated mammography as screening tool, they did take into account that for 35% of the screenings an additional MRI is necessary due to unclear imaging of the mammography screening.[200]

Table 7 – Population, Intervention, Comparator, and Outcome(s) considered in primary evaluations

	Population eligibility criteria		Intervention*			Comparator*		Outcome**
	Starting ages	Stopping ages	Screening interval(s)	Type of screening	Imaging	Strategy	Imaging	
Journal articles								
Shafik, 2023[178]	46, 50	69, 74	Biennial	Age-based	Mammography	Current screening strategy *(biennial 50-69)*	Mammography	ICER (€/LYG)
Blankenburg, 2023[179]	40	74	Annual Biennial Triennial	Age- & density-based	Mammography (+ ultrasound for women with dense breasts)	Current screening *(only age-based)*	Mammography alone	*ICER (€/LYG)* ICER (€/QALY gained)
Schousboe, 2022[180]	65	75, 80, 85, 90	Biennial Annual	Age-based	Mammography	Current screening strategy *(biennial 65-74)*	Mammography	*ICER (US$/QALY gained)*
Kregting, 2022[181]	40-60	64-84	Quadrennial Triennial Biennial Annual	Age-based	Mammography	Current screening strategy *(biennial 50-74)* No screening	Mammography	ICER (€/QALY gained)
Tina Shih, 2021[183]	40, 50	75	Triennial Biennial Annual *Hybrid: annual & biennial/ triennial (density-based)*	Age- & density-based	Mammography	No screening	/	*ICER (€/LYG)* ICER (€/QALY gained)
Chootipongchaivat, 2021[184]	40, 45, 50, 55	64, 69, 74, 79	Triennial Biennial	Age-based	Mammography	No screening	/	*ICER (SGP/LYG)*

	Population eligibility criteria		Intervention*			Comparator*		Outcome**
	Starting ages	Stopping ages	Screening interval(s)	Type of screening	Imaging	Strategy	Imaging	
			Annual *Hybrid: annual & biennial (age-based)*					ICER (SGP$/QALY gained)
Tina Shih, 2019[187]	40, 45, 50, 55	75, 80, no upper limit	Biennial Annual *Hybrid: annual & biennial (age-based)*	Age-based	Mammography	No screening	/	ICER (US$/QALY gained)
Mittmann, 2018[189]	40, 50	69, 74	Triennial Biennial Annual *Hybrid: annual & biennial (age-based)*	Age-based	Mammography	No screening	/	ICER (CA$/LYG) ICER (CA$/QALY gained)
Koleva-Kolarova, 2018[190]	46, 48, 50	74	Biennial	Age-based	Mammography	Current screening strategy *(biennial 50-74)*	Mammography	ICER (€/YOLS)
Carter, 2018[191]	45, 50	74, 80	Biennial *Hybrid: annual & biennial (age-based)*	Age-based	Mammography	No screening	/	*ICER (€/LYG)*
van Luijt, 2017[192]	50	69	Biennial	Age-based	Mammography	No screening	/	ICER (NOK/QALY gained)
Schiller-Frühwirth, 2017[193]	45	69	Biennial	Age- and density-based	Mammography (+ ultrasound for women with dense breasts)	No screening Opportunistic screening	/ Mammography	ICER (€/LYG)

	Population eligibility criteria		Intervention*			Comparator*		Outcome**
	Starting ages	Stopping ages	Screening interval(s)	Type of screening	Imaging	Strategy	Imaging	
Morton, 2017[194]	50	70	Triennial	Age-based	Mammography	No screening	/	ICER (£/QALY gained)
Rafia, 2016[195]	50	70, 72, 75, 78, 81, 84, 87, 90	Triennial	Age-based	Mammography	Current screening strategy *(triennial 50-70)*	Mammography	ICER (£/LYG) ICER (£/QALY gained)
Arrospide, 2016[196]	50	69	Biennial	Age-based	Mammography	No screening	/	ICER (€/QALY gained)
Sankatsing, 2015[197]	40, 45, 48, 49, 50	74, 76	Biennial *Hybrid: annual & biennial (age-based)*	Age-based	Mammography	No screening	/	ICER (€/LYG) ICER (€/QALY gained)
Other reports								
I-CHER, 2024[200]	45, 50	69, 74	Biennial *Hybrid: biennial & triennial (age-based)*	Age-based	Mammography	Stop screening Current screening strategy *(biennial 50-69)*	/ Mammography	ICER (€/LYG) ICER (€/QALY gained)

***Notes**: *Based on the examined literature and to achieve consistency, the comparator is formulated as the entire screening strategy to which the entire intervention screening strategy is compared, even when part of the age eligibility criteria overlap between comparator and intervention (e.g. biennial screening in women aged 50-69 vs biennial screening in women 50-74). An alternative way of formulating intervention & comparator in this example could be biennial screening in women aged 70-74 vs no screening in women aged 70-74. **Outcomes in Italic mean the ICER can be estimated based on information provided by the authors, but was not calculated by the authors themselves.*
***Abbreviations**: ICER = Incremental Cost-Effectiveness Ratio, QALY = Quality-Adjusted Life Year, LYG = Life Year Gained, LYS = Life Years Saved, CESM = Contrast-Enhanced Spectral Mammography, YOLS = Years of Life Saved, US$ = United States Dollar, SGP = Singaporean Dollar, CA$ = Canadian Dollar, NOK = Norwegian Krone, £ = Pound Sterling, DKK = Danish Krone.*

4.2.2.3 Comparator

The primary comparators used in each study are summarized in columns 7 and 8 of Table 7. To gain more insight into the cost-effectiveness of the various screening strategies, the inclusion and exclusion criteria did not limit the comparators to the current situation in Belgium (i.e. biennial screening between 50 and 69 years). As evaluating the currently implemented screening programme was the objective of numerous included studies, no screening was most often defined as (one of) the primary comparator(s) in the selected studies (11/17).[181, 183, 184, 187, 189, 191-194, 196, 197] In addition, as multiple studies sought to explore the cost-effectiveness of extending the existing screening programmes, the current screening programme was taken as (one of the) primary comparator(s) by seven (predominantly more recent) studies.[178-181, 190, 195, 200] Of these studies, only the comparator screening strategy of Shafik et al. (2023) and I-CHER (2024) corresponds to the one currently in place in Belgium.[178, 200] In addition, Schiller-Frühwirth et al. (2017) also defined opportunistic screening as a comparator besides no screening at all and I-CHER (2024) defined stopping the screening programme as a comparator when evaluating the current screening programme following the reasoning that the effects of having screened part of the population will continue some time into the future.[193, 200] In this section the focus was primarily on what the authors defined and reported as their primary comparators. As international guidelines recommend to calculate the cost-effectiveness of an intervention relative to its next-best alternative (i.e. the non-dominated, less costly and less effective option), other screening strategies (shown in the intervention columns) were rightfully used as relevant comparators in the analyses. These will be touched upon further in section 4.3 if relevant for the conclusions.

4.2.2.4 Outcome

As shown in the final column of Table 7, due to the inclusion & exclusion criteria, all studies used the incremental cost-effectiveness ratio (ICER) as primary outcome. In line with the majority of studies conducting cost-utility analysis, most studies express the ICER in terms of incremental cost/QALY gained (13/17).[201] As recommended in the Belgian guidelines for economic evaluation, four of these studies additionally expressed the ICER in incremental cost/LYG.[189, 195, 197, 200] While having all required information to calculate the ICER in terms of incremental cost/LYG in addition to incremental cost/QALY gained, three studies solely reported incremental costs and LYG without calculating the ICER with these outcomes. For these studies, an estimation was reported in the results section of this report, calculated by dividing the reported incremental costs by LYG. Four studies solely conducted a CEA, meaning QoL was not taken into account, with the incremental health effects reported in LYG by three studies[178, 191, 193] and in years of life saved (YOLS (*~equivalent to LYG)*) by Koleva-Kolarova et al. (2018).[190]

4.2.3 Model inputs: effectiveness of breast cancer screening & quality of life

Key points

Screening effectiveness:

- Out of 17 studies, with the exception of Morton et al. (2017),[194] all studies evaluating (an extension of) breast cancer screening opted for the stage shift approach to model screening effectiveness.
- None of the evaluations used stage shift or distribution data directly derived from an RCT.
- Within the stage shift modelling technique, generally one of two approaches was used:
  - Current stage distribution vs stage distribution before screening (for existing screening programme) OR stage distribution of closest screened age groups (for extension of screening programme),

- Simulation of tumour growth or size at various ages combined with tumour detection by screening based on screening sensitivity and specificity parameters or a detection probability function vs detection without screening.

- In general, multiple studies lack transparency on the parameters, sources and/or techniques used to model screening effectiveness.

Health states:

- Studies used different ways of defining health states. The most commonly chosen formulations are:
  - DCIS – Stage I – Stage II – Stage III – Stage IV,
  - DCIS – Localized – Non-localized/Regional – Metastasis/Distant.
- While employing the stage shift technique, nine studies solely used tumour size in their models as a proxy for breast cancer stage.

(Dis)Utilities

- A large range of QoL studies was used, with each QoL study used by a maximum of three included studies.
- Multiple studies (partially) relied on expert opinion or non-patients for determining health (dis)utilities.
- Ten studies used disutilities to take into account breast cancer screening-related harms.
- Morton et al. (2017) derived a QALY gain corrected for screening-related harms from the literature instead of specifically including health utilities and life years in their decision tree.[194]

As discussed previously, all included studies used health economic modelling to evaluate the cost-effectiveness of breast cancer screening. This section focuses on the sources and methods used to embed the clinical effectiveness of screening and quality of life in these health economic models, which is summarized for each individual study in Table 8. In addition, a summary of the various quality of life studies referenced by the economic evaluations is shown in Table 9.

4.2.3.1 Effectiveness, health states and (dis)utilities

Multiple approaches are available to incorporate the clinical effectiveness of breast cancer screening in economic models. Column 2 of Table 8 summarizes which approach the various studies have taken and which primary sources they used to **model the clinical effectiveness**. As described by Schiller-Frühwirth et al. (2017b), studies generally use one of two ways to model the clinical effectiveness of breast cancer screening: (1) model stage shift or (2) use screen-related mortality reduction.[176] In line with the findings of Schiller-Frühwirth et al. (2017b), the studies evaluating screening versus a situation without screening or an extension of screening to younger/older women use the former approach in which a shift in stage distribution due to screening is modelled. An exception is Morton et al. (2017), who take a gain in QALYs due to screening from the literature.[176, 194]

Specifically, the stage shift approach assumes that screening leads to earlier diagnosis of the breast tumour and these tumours are therefore on average diagnosed in a lower stage compared to symptomatic detection (i.e. no screening). This stage shift in turn generally leads to improved survival and lower costs. Surprisingly, rather than using the breast cancer stages as defined by the TNM classification, two thirds of the studies employing the stage shift approach (9/15) used tumour size (e.g. predicted by a tumour growth model, from trial or registry data, etc.) as a proxy for stage in their respective models. None of the studies transparently reporting their methodology derive the stage shift directly from randomized clinical trials. For example, three studies used the clinical stage distribution before the introduction of organized screening (i.e. data dating back to the 70s,[180] 80s,[193] and 90s[196]

respectively) to simulate the situation without screening. In contrast, Shafik et al. (2023) used the stage distribution of the closest screened age group when evaluating an age extension of the breast cancer screening programme in Finland.[178] Other studies approached the stage shift from a different angle and simulated tumour growth or size at various ages and subsequently simulated tumour detection by screening based on screening sensitivity and specificity parameters or a detection probability function.[183, 187, 189-191, 195, 200] Tina Shih et al. (2021) and Tina Shih et al. (2019), for example, simulated tumour growth and sojourn time based on the CNBSS (i.e. Canadian National Breast Screening Study) and Nijmegen trials.[183, 187] The SiMRiSc model, employed by Koleva-Kolarova et al. (2018), was developed for and based on a high-risk population. However, the authors did not disclose how they adapted the model to fit a general average-risk screening population. This model also takes radiation-induced breast cancers due to the screening strategy under evaluation into account, and is the only study reporting having included this factor in their model.[190] Unfortunately, in general, transparency on the sources used to model screening effectiveness and the modelling approaches was lacking in multiple evaluations. For example, Mittmann et al. (2018) used age-, density- and screening round-specific sensitivity and specificity, but did not provide the values, formulas or sources used for these parameters.[189] In addition, the studies using (an adapted version of) the MISCAN model generally referred to a set of previous publications using the model but did not transparently disclose which specific parameters were used (or adapted) with their original sources. In a publication referred to by Chootipongchaivat et al. (2021), the authors do mention that all screening parameters of the MISCAN-Fadia model originate from the Swedish Two-County Trial.[184] In addition, van Luijt et al. (2017)[192] state that “breast cancer mortality is stage-dependent and based on the data from the Swedish breast cancer screening trials” and Sankatsing et al. (2015)[197] indicate that “survival rates after screen-detection were estimated using data from the Swedish randomised controlled trials”. Unfortunately, however, none of these studies transparently provide the specific parameters used to model screening effectiveness in their publications, which makes an adequate assessment of their models ultimately difficult. Blankenburg et al. (2023) altered screening sensitivity and specificity in women with dense breasts resulting from adding an ultrasound to the screening mammography for this subpopulation when evaluating the cost-utility of an adjacent ultrasound for women with dense breasts.[179] As the focus of this section is primarily on the methodological approach to model screening effectiveness, sensitivity and specificity input values of the individual studies are shown in Table 10 and discussed elaborately in section 4.2.5.

The included economic evaluations used a considerable variety in the way they define the **health states** in their respective models. Seven distinct formulations could be identified. The most frequently used formulations are ‘DCIS/Stage 0 – Stage I – Stage II – Stage III – Stage IV’ (6 studies)[183, 184, 187, 195, 196, 200] on the one hand and ‘DCIS – Localized – Non-localized/Regional – Metastasis/Distant’ (5 studies)[178, 180, 181, 189, 193] on the other hand. As discussed above, nine studies used tumour size as a proxy for stage or as primary health states in their models. For example, Koleva-Kolarova et al. (2018) defined health states based on tumour size alone and two other studies defined the health states based on solely the extent of the tumour (T) out of the TNM classification (i.e. van Luijt et al. (2017) & Sankatsing et al. (2015)).[190, 192, 197]

To take into account the impact of breast cancer screening, diagnosis, treatment, etc. on QoL, thirteen studies used literature-based and/or expert opinion-based **utility and disutility** estimates in their respective models. The four remaining studies conducted a CEA, in which QoL is not taken into account. A large number of different QoL studies was used, with each QoL study used by a maximum of three included studies (i.e. QoL study of Stout et al. (2006))[202]. Mittmann et al. (2018) is the only study solely relying on QoL estimations made by experts in breast cancer and public health, but unfortunately disclosed the utility values in an unavailable appendix.[189] Other studies, more specifically Blankenburg et al. (2023) and Rafia et al. (2016), used a combination of utilities derived from QoL studies and estimations based on expert opinion or assumptions.[179, 195] The remaining majority of studies relied on previously published utilities. While most studies assigned a state-specific health utility to the health states they defined in their model, a selection of studies used disutilities related to specific procedures or treatments for breast cancer (e.g. chemotherapy, surgery, hormonal therapy, etc.) in their modelling.[183, 187, 192]

Table 8 – Effectiveness and Quality of Life measures used in primary evaluations

Study	Primary source of clinical effectiveness*	Health states and associated utilities	QoL source or methodology	Disutilities
Journal articles				
Shafik, 2023[178]	Stage shift approach: Stage distribution of current screening programme (closest screened age group) Incidence of BC due to screening: Incidences of younger age groups were increased while incidence in 50-51 was decreased Incidences of elder age groups follow the pattern observed in a steady-state (Sweden)	Unknown Localized Non-localized/Regional lymph nodes metastases Metastasized farther than regional lymph nodes or invades adjacent tissues In situ carcinoma	*Not applicable (CEA)*	*Not applicable (CEA)*
Blankenburg, 2023[179]	Mammography + ultrasound vs mammography alone: Literature-based sensitivity & specificity of **mammography + ultrasound vs mammography alone** in high breast density population Scenario: expert opinion	No breast cancer: 0.99 Undetected breast cancer: 0.98 Non-invasive DCIS: 0.97 Invasive breast cancer with tumour size T1mi, T1a, T1b, T1c: 0.91 Invasive breast cancer with tumour size T2: 0.75 Invasive breast cancer with tumour size T3: 0.51 Post-treatment A (localized tumours): 0.99 Post-treatment B (large tumours): 0.95 Death: 0	Assumption Assumption Schleinitz et al. (2006)[203] Schleinitz et al. (2006)[203] Schleinitz et al. (2006)[203] Schleinitz et al. (2006)[203] Assumption Assumption /	Screening: 0.01 False negatives: 0.01 *(no references reported)*
Schousboe, 2022[180]	Stage shift approach: Current stage distribution from registry: - annual: stage distribution of women with mammo 9-17 months prior to diagnosis - biennial: stage distribution of women with mammo 18-30 months prior to diagnosis)	DCIS ○ Age 64 ▪ Year 1: 0.683 ▪ Later years: 0.756 ○ Age 74 ▪ Year 1: 0.667 ▪ Later years: 0.738 ○ Age 84 ▪ Year 1: 0.631 ▪ Later years: 0.698	Adapted from Lidgren et al. (2007)[204] using linear interpolation between age categories	Mammography: 0.006 for 1 week (Li et al. (2019)[205]) Diagnostic work-up positives: 0.105 for 5 weeks (Li et al. (2019)[205])

Study	Primary source of clinical effectiveness*	Health states and associated utilities	QoL source or methodology	Disutilities
	vs stage distribution before screening (1973-1979)	○ Age 94 ▪ Year 1: 0.595 ▪ Later years: 0.658 Localized invasive breast cancer ○ Age 64 ▪ Year 1: 0.640 ▪ Later years: 0.741 ▪ Last year: 0.629 ○ Age 74 ▪ Year 1: 0.624 ▪ Later years: 0.723 ▪ Last year: 0.614 ○ Age 84 ▪ Year 1: 0.591 ▪ Later years: 0.684 ▪ Last year: 0.581 ○ Age 94 ▪ Year 1: 0.557 ▪ Later years: 0.645 ▪ Last year: 0.547 Regional breast cancer ○ Age 64 ▪ Year 1: 0.569 ▪ Later years: 0.684 ▪ Last year: 0.629 ○ Age 74 ▪ Year 1: 0.556 ▪ Later years: 0.668 ▪ Last year: 0.614 ○ Age 84 ▪ Year 1: 0.526 ▪ Later years: 0.632 ▪ Last year: 0.581 ○ Age 94 ▪ Year 1: 0.495 ▪ Later years: 0.595 ▪ Last year: 0.547 Distant breast cancer ▪ Year 1: 0.569 ▪ Later years: 0.629 ○ Age 74		

Study	Primary source of clinical effectiveness*	Health states and associated utilities	QoL source or methodology	Disutilities
		▪ Year 1: 0.556 ▪ Later years: 0.614 ○ Age 84 ▪ Year 1: 0.526 ▪ Later years: 0.581 ○ Age 94 ▪ Year 1: 0.495 ▪ Later years: 0.547 *Linear interpolation for ages in between*		
Kregting, 2022[181]	MISCAN-Breast Stage shift approach: *source for clinical effectiveness not reported*	No breast cancer: 0.858 DCIS/localised: 0.772 for 2 years Regional: 0.644 for 2 years Metastasis: 0.515 until death	Stout et al. (2006)[202]	Mammography: 0.006 for 1 week (de Haes et al. (1991)[206]) Referral: 0.105 for 5 weeks (de Haes et al. (1991)[206])
Tina Shih, 2021[183]	*Screening vs no screening:* Stage shift approach: Age of onset combined with sojourn times & tumour growth rate, and age-specific & tumour size-specific sensitivity (age and tumour size at detection, both assumed lower due to screening, determine survival) Sojourn times & growth rate are based on CNBSS and the Nijmegen Trial *Dense vs non-dense breasts*: Literature-based sensitivity & specificity with age weighting	Cancer-free Asymptomatic or preclinical screening-detectable cancer (stage 0-IV) Symptomatic or clinical cancer (stage 0-IV) Death due to breast cancer or competing risks Surgery: 0.87 (3 months) Radiation therapy: 0.80 (3 months) Chemotherapy: 0.74 (1 year) Hormonal therapy with tamoxifen: 0.99 (5 years) Terminal stage from breast cancer: 0.29 (last 3 months of life) Terminal stage from other diseases: 0.375 (last 3 months of life)	Ahern et al. (2014): Tengs et al. (2000)[207] from De Koning et al. (1991)[208] Tengs et al. (2000)[207] from De Koning et al. (1991)[208] Tengs et al. (2000)[207] from Jansen et al. (1998)[209] Tengs et al. (2000)[207] from Carter et al. (1998)[210] Tengs et al. (2000)[207] from De Koning et al. (1991)[208] Ahern et al. (2014)[211]	Mammography: 0.006 for 7 days (Stout et al. (2014)[212] -> de Haes et al. (1991)[206]) Diagnostic work-up: 0.105 for 35 days (Stout et al. (2014)[212] -> de Haes et al. (1991)[206])
Chootipongchaivat, 2021[184]	**MISCAN-Fadia** adapted with Singaporean data on attendance, specificity and treatment	DCIS: ○ Initial care: 0.786 ○ Continuous care: 0.903 ○ Terminal care: 0.352	Adapted from Kim et al. (2017)[213]	Mammography: 0.006 for 1 week (de Haes et al. (1991)[206])

Study	Primary source of clinical effectiveness*	Health states and associated utilities	QoL source or methodology	Disutilities
	Stage shift approach: Screening parameters originate from the Swedish Two-County Study: - detection rates at successive screening rounds - interval cancer rates - tumour diameter distribution of screen-detected cancers, of interval cancers, and of cancers diagnosed in the control group - survival by mode of detection and time since diagnosis - survival by tumour diameter and time since diagnosis	Stage I & II ○ Initial care: 0.731 ○ Continuous care: 0.841 ○ Terminal care: 0.352 Stage III ○ Initial care: 0.599 ○ Continuous care: 0.688 ○ Terminal care: 0.352 Stage IV ○ Initial care: 0.352 ○ Continuous care: 0.405 ○ Terminal care: 0.352		Diagnostic work-up: 0.105 for 5 weeks (de Haes et al. (1991)[206])
Tina Shih, 2019[187]	Stage shift approach:: Age of onset combined with sojourn times & tumour growth rate, and age-specific & tumour size-specific sensitivity (age and tumour size at detection, both assumed lower due to screening, determine survival & life years) Sojourn times & growth rate are based on CNBSS and the Nijmegen Trial	DCIS Stage I Stage II Stage III Stage IV Surgery: 0.87 (3 months) Radiation therapy: 0.80 (3 months) Chemotherapy: 0.74 (1 year) Hormonal therapy with tamoxifen: 0.99 (5 years) Terminal stage from breast cancer: 0.29 (last 3 months of life) Terminal stage from other diseases: 0.375 (last 3 months of life)	Ahern et al. (2014)[211]	*None reported*
Mittmann, 2018[189]	University of Wisconsin Breast Cancer Epidemiology Simulation Model adapted to Canadian context Stage shift approach: Age of onset combined with tumour growth rate (tumour size as surrogate for stage) & thresholds for clinical or	In situ Localized Regional metastasis Distant metastasis Tumour size is used as a surrogate for stage *(utility values not reported: Appendix not available)*	Expert opinion	Screening: 0.006 for 1 week *(incorrect reference provided by authors)* Diagnostic work-up: 0.105 for 5 weeks *(incorrect reference provided by authors)*

Study	Primary source of clinical effectiveness*	Health states and associated utilities	QoL source or methodology	Disutilities
	screening detection with age-, density-, and screening round-specific sensitivity and specificity			
Koleva-Kolarova, 2018[190]	Based on SiMRiSc: - Originally developed for high-risk population (BRCA) and MRI but adapted and externally validated to general population - DCIS not included Stage shift approach: Age-specific risk of developing breast cancer, age-specific tumour growth model & age-specific breast density combined with detection probability based on screening specificity (Oslo II study) & density-specific sensitivity (non-referenced/-discussed meta-analysis) Other characteristics: - Includes radiation-induced breast cancers via radiation induction model - Includes a systematic error: "Based on expert opinion, we assumed 10% of all cases that should be detected based on tumour volume would not be detected due to their characteristics (lobular carcinomas, dense breast tissue and tumours located close to the thorax wall)"	No states reported, model is solely based on tumour size	*Not applicable (CEA)*	*Not applicable (CEA)*
Carter, 2018[191]	Stage shift approach: Age of onset & tumour growth combined with age-specific &	No breast cancer Local non-invasive Local invasive Latent nodal	*Claims taking into account the health utilities provided by de Haes, 1991 but the authors did not report these utilities & reported the outcome in LYG only*	*Not reported*

Study	Primary source of clinical effectiveness*	Health states and associated utilities	QoL source or methodology	Disutilities
	tumour size-specific screening sensitivity	Latent metastatic Detected, staged and treated Detected with latent metastatic cancer Metastatic terminal cancer Cured Dead, breast cancer Dead, non-breast cancer		
van Luijt, 2017[192]	MISCAN Stage shift approach: "Early detection has an improved prognosis and therefore reduced mortality. Breast cancer mortality is stage-dependent and based on the data from the Swedish breast cancer screening trials.21, 22 Screening is superimposed on the natural history."	No breast cancer DCIS T1A T1B T1C T2+ Dead	Disutilities: de Haes et al. (1991)[206] (see final column)	Screening: 0.01 Diagnostic phase: 0.11 Initial surgery: 0.13 Initial radiotherapy: 0.20 Initial chemotherapy: 0.28 Initial hormonal therapy: 0.18 Terminal illness: 0.71 Palliative therapy + chemotherapy: 0.47 Palliative therapy + radiotherapy: 0.42 Palliative therapy + surgical therapy: 0.38 Palliative therapy + hormonal therapy: 0.34 Disease-free 2m-1y mastectomy: 0.16 Disease-free 2m-1y breast conserving Disease-free >1y mastectomy: 0.05 Disease-free >1y breast conserving: 0.04
Schiller-Frühwirth, 2017[193]	Stage shift approach: Stage distribution based on situation before opportunistic screening (i.e. 1987–1991) was used for no screening	No breast cancer DCIS Localized Regional Distant Death from breast cancer	*Not applicable (CEA)*	*Not applicable (CEA)*

Study	Primary source of clinical effectiveness*	Health states and associated utilities	QoL source or methodology	Disutilities
	Screening vs opportunistic screening vs no screening: Prevalence of DCIS and invasive breast cancer stages at age 40, the incidence rates of DCIS and invasive breast cancer per 5-year age group, breast cancer–specific and all-cause mortality combined with stage-specific sojourn times, sensitivity and specificity of the (opportunistic) screening tests per 5-year age group	Death from other causes		
Morton, 2017[194]	Gain in BC-specific mortality & QALYs taken from the literature, adjusted for screening-related harms	*Stages not taken into account*	Gain in QALYs taken from the literature	Loss in QALYs due to screening-related harms (Raftery et al. (2011)[214])
Rafia, 2016[195]	Natural history model combined with postdiagnosis model Model structure: - Simulates time at which cancer presents - Simulates tumour characteristics (size, nodal involvement, grade, ER status) at detection with & without screening - Simulates whether screen detection occurs or fails (non-attendance or false negative) - Prognostic profile at diagnosis determines survival (based on registry data) Stage shift approach: - Shift in prognosis profile at diagnosis	Disease free: general population QoL In situ disease: QoL between disease free & stage I for 1y Stage I: 0.91 for 1y Stage II: 0.75 for 2y Stage III: 0.51 for 3y Stage IV: 0.36 for lifetime	/ Expert opinion Schleinitz et al. (2006)[203] Schleinitz et al. (2006)[203] Schleinitz et al. (2006)[203] Schleinitz et al. (2006)[203]	Mammography: reduction of 20% in QoL for 2 hours (Bonomi et al. (2008)[215]; Rijnsburger et al. (2004)[216]; expert opinion) Anxiety after recall: reduction of 35% in QoL for 3 weeks (Bonomi et al. (2008)[215]; Rijnsburger et al. (2004)[216]; expert opinion)

Study	Primary source of clinical effectiveness*	Health states and associated utilities	QoL source or methodology	Disutilities
	(Link to full report with evidence-base no longer functional)			
Arrospide, 2016[196]	Natural history model with 4 health states: - Disease-free or undetectable BC - Asymptomatic BC detectable by screening - Symptomatic BC - Death from BC Stage shift approach: Age-& stage-specific BC incidence from registry data & time-to-event distributions combined with screening sensitivity & specificity and age- and stage-specific survival Vs time-to-event distributions combined with detection stages before introduction of screening (1995) with age- and stage-specific survival	Healthy ○ 50-64y: 0.824 ○ 65-74y: 0.770 ○ 75-84y: 0.682 ○ >84y: 0.563 In situ ○ 50-64y: 0.742 ○ 65-74y: 0.693 ○ 75-84y: 0.614 ○ >84y: 0.507 Stage I ○ *Idem as in situ* Stage II ○ 50-64y: 0.618 ○ 65-74y: 0.578 ○ 75-84y: 0.512 ○ >84y: 0.422 Stage III ○ *Idem as stage II* Stage IV ○ 50-64y: 0.495 ○ 65-74y: 0.462 ○ 75-84y: 0.409 ○ >84y: 0.338	Specific percentages were applied to general population utilities in order to estimate the potential negative effects of a BC diagnosis during the first year of treatment and end of life Methodology of Stout et al. (2006)[202] (proportion of healthy)	*Not reported*
Sankatsing, 2015[197]	MISCAN Stage shift approach: Survival rates after screen-detection were estimated using data from the Swedish randomised controlled trials	Stages: DCIS T1AN- T1AN+ T1BN- T1BN+ T1CN- T1CN+ T2+N- T2+N+ QoL: In situ/localized: 0.9 for 2 years Regional or distant: 0.75 for 2 years	Stout et al. (2006)[202] (adjusted)	Screening participation: 0.006 for 1 week (de Haes et al. (1991)[206]) Positive result: 0.105 for 5 weeks (de Haes et al. (1991)[206])

Study	Primary source of clinical effectiveness*	Health states and associated utilities	QoL source or methodology	Disutilities
		Breast cancer death: 0.6 until death		
Other reports				
I-CHER, 2024[200]	Model structure: Determination of 'susceptible to breast cancer' sub-population based on incidence (lifetime risk calculation) for separate risk groups with state transitions to undetected health states which can be detected through organized screening or other testing (opportunistic screening, symptomatic testing) Stage shift approach: Probability of screening event taking place (based on age, risk group, stage, participation rate) combined with screening sensitivity & specificity	Stage 0 (DCIS) ○ Unidentified: -0.085 ○ Treatment: -0.171 ○ Follow-up: -0.066 Stage I ○ Unidentified: -0.171 ○ Treatment: -0.341 ○ Follow-up: -0.066 Stage II ○ Unidentified: -0.171 ○ Treatment: -0.341 ○ Follow-up: -0.083 Stage III ○ Unidentified: -0.217 ○ Treatment: -0.434 ○ Follow-up: -0.460 Stage IV ○ Unidentified: -0.305 ○ Treatment: -0.610 ○ Follow-up: -0.673	In text of report: Kaur et al. (2022)[217] & Global Burden of Disease data In ShinyApp: Kim et al. (2017)[213] & Global Burden of Disease data	False positive: 0.02 *(duration not reported, but false positives are picked up during multidisciplinary oncological consult after which they return to the respective health state)* (Li et al. (2019)[205])

***Notes**:*Input values for screening sensitivity and specificity are shown in Table 10.*
***Abbreviations**: BC = breast cancer, CEA = Cost-Effectiveness Analysis, CNBSS = Canadian National Breast Screening Studies, DCIS = Ductal Carcinoma in Situ, ER = Emergency Room, HIP = Health Insurance Plan of Greater New York, MISCAN = MIcrosimulation SCreening Analysis, m = months, QALY = Quality-Adjusted Life Year, QoL = Quality of Life, SiMRiSc = Simulation Model on Radiation Risk and Breast Cancer Screening, y = years.*

A second means of using disutilities, applied by ten studies, is to take into account the harms related to breast cancer screening which affect patients' QoL.[179-181, 183, 184, 189, 192, 195, 197, 200] With the exception of I-CHER (2024), all of these studies assigned a disutility to screening participants as a consequence of the physical and psychological burden related to mammography screening.[200] Two thirds of these studies used a disutility of 0.006 for the duration of one week, which was rounded to 0.01 by Blankenburg et al. (2023) and van Luijt et al. (2017) in their analyses.[179, 192] In contrast, Rafia et al. (2016) opted for a 20% reduction in QoL for two hours due to screening participation.[195] Seven studies additionally assigned a disutility of 0.105 for 5 weeks to participants with a (false-) positive result requiring diagnostic work-up,[180, 181, 183, 184, 189, 192, 200] which was rounded to 0.11 in the study of van Luijt et al. (2017).[192] I-CHER (2024), on the other hand, used a disutility of 0.02 and assigned it solely to participants with a false-positive screening result.[200] Unfortunately, no duration and reference for the source was reported by the authors. Rafia et al. (2016) chose to reduce the QoL by 35% for three weeks due to anxiety after recall.[195] Finally, Blankenburg et al. (2023) is the only study applying a disutility (0.01) to false-negative results.[179] More information on the underlying QoL studies used by the various included economic evaluations is summarized in Table 9.

Worth noting with regard to QoL are the studies of Morton et al. (2017) and Carter et al. (2018).[191, 194] While most studies used (dis)utilities and applied these to the duration of health states or procedures, Morton et al. (2017) used a literature-based estimated gain in QALYs due to breast cancer screening and applied a correction due to screening-related harms. The study on which this QALY gain was based (i.e. Raftery et al. (2011)), claims that breast cancer screening in the UK might have caused net harm (i.e. QALY loss) up to ten years after entry to screening when QoL losses due to harms (e.g. false-positives, overtreatment, etc.) are taken into account in their life table model.[214] While Carter et al. (2018) conduct a cost-effectiveness analysis, thus not taking into account QoL, the authors mention utilizing the (dis)utilities of de Haes et al. (1991) but do not report these utilities and do not use them in any outcome measure.[191, 206]

4.2.3.2 Quality of life studies used by economic evaluations

Table 9 summarizes the QoL studies of which the results were used by the primary economic evaluations in our selection. The respective QoL studies are grouped in two categories: (1) studies used for QoL estimates of the health states defined in the models and (2) QoL studies employed to incorporate disutilities related to screening-related harms in the models. The studied sample, methods, results and limitations of individual studies are shown in Table 9. In general, multiple studies were based on a sample of non-breast cancer patients (e.g. general population) or experts who evaluated hypothetical scenarios, which raises questions on the validity of their results. Furthermore, a diversity of methods was used to capture quality of life, including Time Trade-Off (TTO), Standard Gamble (SG), Visual Analogue Scale (VAS) and the EQ-5D questionnaire. Of the studies estimating QoL of breast cancer health states, only one study used the latter approach. In the second set of studies, estimating (dis)utilities associated with breast cancer screening, EQ-5D was used more often. The main limitations included hypothetical bias, selection bias, shortcomings inherent to the methods used and the timing of measuring QoL.

Table 9 – Overview of studies used as reference for Quality of Life

Study (First author, year)	Population & sample	Methods	Results	Limitations
Studies used for quality of life estimates of health states				
Kaur, 2022[217]	Systematic review: 79 QoL studies	Systematic review	*Unclear how the utilities used by I-CHER (2024) were derived from this publication (see Table 8 for utilities used by I-CHER (2024))*	
Kim, 2017[213]	509 randomly selected individuals from the Korean general population: male & female >19 years old	Visual Analogue Scale (VAS) (8 scenarios/ respondent) Standard Gamble (SG) (1 scenario/respondent)	1) non-invasive BC with mastectomy only • VAS: mean of 0.669 (SD: 0.199) and median of 0.694 • SG: mean of 0.790 (SD: 0.265) and median of 0.900 2) non-invasive BC with mastectomy and followed by reconstruction • VAS: mean of 0.681 (SD: 0.199) and median of 0.700 • SG: mean of 0.804 (SD: 0.260) and median of 0.900 3) non-invasive BC with breast-conserving surgery & radiation therapy • VAS: mean of 0.662 (SD: 0.201) and median of 0.684 • SG: mean of 0.779 (SD: 0.261) and median of 0.900 4) invasive BC with surgery (mastectomy or breast-conserving surgery), radiation therapy, and/or chemotherapy • VAS: mean of 0.579 (SD: 0.202) and median of 0.600 • SG: mean of 0.731 (SD: 0.255) and median of 0.800 5) locally advanced BC with radical mastectomy & radiation therapy • VAS: mean of 0.435 (SD: 0.178) and median of 0.451 • SG: mean of 0.610 (SD: 0.261) and median of 0.650 6) inoperable, locally advanced BC • VAS: mean of 0.415 (SD: 0.173) and median of 0.438 • SG: mean of 0.587 (SD: 0.259) and median of 0.650 7) loco-regional recurrent BC • VAS: mean of 0.333 (SD: 0.176) and median of 0.357 • SG: mean of 0.496 (SD: 0.260) and median of 0.500 8) metastatic BC • VAS: mean of 0.170 (SD: 0.220) and median of 0.167 • SG: mean of 0.352 (SD: 0.275) and median of 0.300	Low number of scenarios Response rate not recorded SG not applied for states worse than death *Hypothetical bias* *No respondent completed more than one scenario with SG*

Study (First author, year)	Population & sample	Methods	Results	Limitations
Lidgren, 2007[204]	361 consecutive breast cancer patients attending outpatient clinic at Karolinska University hospital (Sweden)	EQ-5D-3L dimensions VAS Time trade-off (TTO)	First year after primary BC • EQ-5D: mean of 0.696 and median of 0.725 • TTO: mean of 0.901 and median of 1.000 First year after recurrence • EQ-5D: mean of 0.779 and median of 0.725 • TTO: mean of 0.842 and median of 0.973 2nd & following years after primary BC or recurrence • EQ-5D: mean of 0.779 and median of 0.796 • TTO: mean of 0.889 and median of 1.000 Metastatic disease • EQ-5D: mean of 0.685 and median of 0.725 • TTO: mean of 0.820 and median of 0.850	No random selection of patients Selection bias: non-metastatic patients with more severe symptoms more likely treated in outpatient clinic Selection bias: metastatic patients with less severe symptoms more likely treated in outpatient clinic
Schleinitz, 2006[203]	Age- and race-stratified sample of 156 women not undergoing BC treatment above 25 years old in the USA	SG (for BC stages) TTO (for therapeutic modalities)	Stage I: Mean: 0.68 (SD: 0.06); Median 0.91 (IQR: 0.5-1.0) Stage II: Mean: 0.61 (SD: 0.06); Median: 0.75 (IQR: 0.26-0.99) Stage III: Mean: 0.56 (SD: 0.06); Median: 0.51 (IQR:0.25-0.94) Stage IV (oestrogen receptor positive): Mean: 0.41 (SD: 0.06); Median: 0.36 (IQR: 0-0.75) Stage IV (oestrogen receptor negative): Mean: 0.42 (SD: 0.06); Median: 0.40 (IQR: 0-0.79) Chemotherapy: Mean: 0.48 (SD: 0.06); Median: 0.50 (IQR: 0-0.92) Hormonal therapy: Mean: 0.54 (SD: 0.07); Median: 0.58 (IQR: 0-1) Radiation therapy: Mean: 0.61 (SD: 0.07); Median: 0.83 (IQR: 0.1-1)	Hypothetical bias Only black and white women included (other races excluded) Women at high risk for BC not included
Stout, 2006[202]	/	Age-specific utility value of healthy population derived from Medical Expenditure Panel Survey (2000) Utilities of other health states assumed to be proportion of healthy utility	Healthy (30-34y; >84): 0.856; 0.590 In situ or localized BC: 90% of healthy (2 years) Regional BC: 75% of healthy (2 years) Distant BC: 60% of healthy (remainder of life) Mammography with negative result: 25% of healthy (7 days) Mammogram with false-positive result: 25% of healthy (25 days)	QoL of BC health states & screening are assumption-based
Carter, 1998[210]	/	Basic reference gamble method (expert opinion)	Radiation: 0.99 Tamoxifen: 0.999	*None reported*

Study (First author, year)	Population & sample	Methods	Results	Limitations
			Mastectomy: 0.99 Local recurrence: 0.80 Metastatic disease: 0.50 Endometrial cancer: 0.95 Reversible complication: 0.99	*Expert opinion*
Jansen, 1998[209]	70 early-stage BC patients who completed primary treatment (lumpectomy or mastectomy) in the Netherlands	TTO SG	Radiotherapy (6 months): • TTO: 0.89 (SD: 0.13) • SG: 0.87 (SD: 0.19) Chemotherapy (6 months): • TTO: 0.74 (SD: 0.26) • SG: 0.75 (SD: 0.27)	Relatively low participation rate (64%) *Only partially reflects BC treatment & population*
de Koning, 1991[208]	27 clinicians and public health experts in the Netherlands	VAS (transformation of median values to TTO values using unretrievable formula)	Screening: 0.99 (1 week) Breast conserving therapy: 0.91 (10 months) Breast conserving therapy instead of mastectomy: 0.93 (10 months) Primary surgical treatment: 0.88 (2 months) Primary radiation treatment: 0.80 (2 months) Adjuvant hormonal therapy: 0.82 (2 or 5 years) Treatment of advanced disease: 0.63 (20 months) Terminal stage: 0.29 (1 month) Year in follow-up (life-year gained): 0.95 Year in follow-up (earlier diagnosis): 0.94 Biopsy outside programme with benign result: 0.89 (5 weeks) False positive leading to biopsy: 0.89 (5 weeks)	*None reported* *Expert opinion* *Assumption-based durations without justification* *Non-transparent conversion to TTO values*
Studies used as reference for disutilities associated to screening-related harms				
Li, 2019[205]	Systematic review: 27 studies (10 BC studies)	Screening: VAS False positive: VAS EQ-5D TTO Procedure-wise:	Screening: Bonomi, 2008 (low-quality study): - 0.2 de Haes, 1991 (low-quality study): - 0.006 de Koning, 1991 (low-quality study): - 0.01 False positive (period: approximately 12 months): Tosteson, 2014 (high-quality study): ○ EQ-5D: 0 ○ VAS: - 0.01 to - 0.02	Quality criteria need further validation No meta-analysis *Does not consistently use the same comparator to report utility decrements (e.g. utility decrement of false positive reported by Tosteson et al.*

Study (First author, year)	Population & sample	Methods	Results	Limitations
		VAS EQ-5D SF-6D	Bonomi, 2008 (low-quality study): - 0.19 (VAS) Gerard, 1999 (medium-quality study): o EQ-5D: - 0.15 o TTO: - 0.26 Johnston, 1998 (medium-quality study): o TTO: - 0.25 o VAS: - 0.23 de Koning, 1991 (low-quality study): - 0.11 (VAS) Procedure-wise disutilities in diagnostic work-up: Bonomi, 2008 (low-quality study): - 0.45 (VAS) Rijnsburger, 2004 (high-quality study): o EQ-5D: 0 o SF-6D: 0 de Haes, 1991 (low-quality study): - 0.105 (VAS) Overtreatment: *no studies retrieved*	*(0.02) is difference in utility between true negative and false positive while the reported utility decrement of Bonomi et al. (0.19) is the utility of false positive in comparison with perfect health. If compared to true negative, the utility decrement of Bonomi et al. would be 0.081 instead of the reported 0.19)*
Tosteson, 2014[218]	1028 randomly selected screening trial participants (positive & negative results) Women with diagnosed breast cancer in trial excluded	Telephone survey (shortly after baseline mammogram + 1y later): EQ-5D-3L with US scoring 6-question short form of the STAI-6 Health rating scale *(VAS)*	*Differences between true-negative (TN) and false-positive (FP) groups:* After baseline mammography: EQ-5D (not significant): o Before work-up showed negative result in FP group (woman unaware of not having BC): 0 o After work-up showed negative result in FP group (woman aware of not having BC): + 0.01 VAS (not significant): o Before work-up showed negative result in FP group (woman unaware of not having BC): - 0.01 o After work-up showed negative result in FP group (woman aware of not having BC): - 0.02 STAI-6: significantly higher anxiety in false-positive group (significance level 0.05) 1y later: EQ-5D (not significant): o Before work-up showed negative result in FP group (woman unaware of not having BC): 0 o After work-up showed negative result in FP group (woman aware of not having BC): + 0.01	Part of the sample was interviewed after work-up meaning they already knew they did not have BC Trial participants might not be representative for full eligible population

Study (First author, year)	Population & sample	Methods	Results	Limitations
			VAS (not significant): ○ Before work-up showed negative result in FP group (woman unaware of not having BC): - 0.02 ○ After work-up showed negative result in FP group (woman aware of not having BC): 0 STAI-6: no significant difference in anxiety Conclusion: false-positive mammograms associated with increased short-term anxiety (no long-term effect) no measurable health utility decrement on EQ-5D	
Bonomi, 2008[215]	131 randomly selected women between 50 & 79 years participating in BCS in the USA (10% history of BC)	VAS (11 scenarios/respondent)	Screening mammography: Mean: 80.4 (SD: 14.3); Median: 80 True negative: Mean: 89.1 (SD: 9.8); Median: 90 False negative: Mean: 48.5 (SD: 20.7); Median: 50 Diagnostic mammography (screen-positive): Mean: 55.3 (SD: 19.7); Median: 50 True positive after diagnostic work-up: Mean: 45.7 (SD: 20.5); Median: 50 False positive after diagnostic work-up: Mean: 81.0 (SD: 14.7); Median: 80 Lumpectomy: Mean: 53.0 (SD: 21.3); Median: 50 Mastectomy: Mean: 48.3 (SD: 22.0); Median: 50 Adjuvant radiation: Mean: 46.2 (SD: 23.4); Median: 40 Adjuvant chemotherapy: Mean: 39.7 (SD: 20.6); Median: 40 Antiestrogen therapy: Mean: 52.0 (SD: 21.6); Median: 50 Disease free at 1 year: Mean: 76.8 (SD: 13.4); Median: 80 Recurrence at 1 year: Mean: 33.0 (SD: 19.4); Median: 30 Palliation/end of life: Mean: 35.8 (SD: 27.4); Median: 30	No background information on participants With the exception of the end-of-life scenario, none of the scenarios included prognostic or life-expectancy information *Hypothetical bias*
Rijnsburger, 2004[216]	329 Dutch participants of magnetic resonance imaging screening (MRISC) study (high-risk population) without screen-detected or interval cancer	2 months prior to screening: SF-36 EQ-5D-3L (UK value set) VAS Symptom Checklist-90 (SCL-90) Day of screening (before screening):	2 months prior to screening: SF-36 (25^{th}-75^{th} percentile): ○ Physical component: 52.5 (49.8-57.7) ○ Mental component: 51.2 (48.3-57.8) VAS (25^{th}-75^{th} percentile): 81.9 (73.0-90.0) EQ-5D-3L (25^{th}-75^{th} percentile): 0.88 (0.80-1.00) SCL-90: 17.5 (14.0-19.0) Day of screening (before screening): VAS (25^{th}-75^{th} percentile): 79 (70.0-90.0)	*Not reported* *High-risk population* *No random selection* *Intervention = biennial CBE + annual MRI + annual mammography + monthly BSE*

Study (First author, year)	Population & sample	Methods	Results	Limitations
		VAS After message 'no breast cancer': SF-36 EQ-5D-3L (UK value set) VAS SCL-90: (score 12-60)	○ Significant difference with 2 months prior to screening at significance level 0.01 After message 'no breast cancer': SF-36 (25th-75th percentile): ○ Physical component: 52.3 (49.2-57.7) ○ Mental component: 52.2 (48.6-58.0) VAS (25th-75th percentile): 80.7 (70.0-90.0) ○ Significant difference with day of screening at significance level 0.05 EQ-5D-3L (25th-75th percentile): 0.88 (0.80-1.00) SCL-90: 17.1 (13.0-19.0) Conclusion: BCS has no unfavourable impact on short-term QoL.	*No analysis for (false-)positive result* *UK value set for Dutch setting (Dutch value set not yet available at time of publication)*
Gerard, 1999[219]	440 women eligible for BCS selected from registers of 4 local UK BCS centres	EQ-5D-3L (only 'usual activity', 'pain/discomfort', 'anxiety/depression' dimensions with other assumed non-impacted by BCS; UK value set) TTO Sequence: true negative, false positive, true positive, false negative Period: true negatives and false positives last for 12 months True positives and false negatives last for the remaining life expectancy	True negative (no significant difference with full health) EQ-5D-3L: 0.94 (SD: 0.14) TTO: 0.91 (SD not correctly reported, but indicated as 'wide') False positive (significant difference with full health, but significance level not reported) EQ-5D-3L: 0.79 (SD: 0.21) TTO: 0.65 (SD not correctly reported, but indicated as 'wide') True positive (significant difference with full health, but significance level not reported) EQ-5D-3L: 0.48 (SD: 0.30) TTO: 0.66 (SD: 0.29) False negative (significant difference with full health, but significance level not reported) EQ-5D-3L: 0.45 (SD: 0.30) TTO: 0.66 (SD: 0.29)	Not all EQ-5D dimensions were scored Not representative for UK population No states worse than death allowed in TTO interview *Scenario-based approach (hypothetical bias)*
Johnston, 1998[220]	440 randomly selected women aged 40-44y	TTO	True negative TTO: Mean: 0.91 (SD: 0.21); Median: 0.98 (IQR: 0.96-0.99)	*Not reported*

Study (First author, year)	Population & sample	Methods	Results	Limitations
	and 50-64y who were invited for BCS at least 6 months ago from 4 trial centres in UK	VAS	VAS: Mean: 0.92 (SD: 0.18); Median: 1.0 (IQR: 0.90-1.0) False positive (significant difference with true negative at significance level 0.05) TTO: Mean: 0.66 (SD: 0.38); Median: 0.83 (IQR: 0.22-0.96) VAS: Mean: 0.67 (SD: 0.28); Median: 0.75 (IQR: 0.50-0.90) True positive TTO: Mean: 0.66 (SD: 0.29); Median: 0.75 (IQR: 0.55-0.95) VAS: Mean: 0.75 (SD: 0.25); Median: 0.80 (IQR: 0.60-0.98) False negative TTO (no significant difference with true positive at significance level 0.05): ○ Mean: 0.66 (SD: 0.29); Median: 0.75 (IQR: 0.45-0.95) VAS (significant difference with true positive at significance level 0.05): ○ Mean: 0.60 (SD: 0.27); Median: 0.63 (IQR: 0.40-0.80)	
de Haes, 1991[206]	15 employees of Department of Public Health and Social Medicine and 12 experts in breast cancer treatment & epidemiology in the Netherlands	VAS (horizontal line without subdivision) *15 health states/respondent in random order* *Converted VAS scores to TTO (function & original source unretrievable)*	Screening participation: Mean: 92 (SD: 6.9); Median: 94; Derived utility: 0.994 Duration: 1 week (assumed to be period between mammography and result) Diagnostic phase: Mean: 71 (SD: 15.5); Median: 67; Derived utility: 0.867 Duration: 5 weeks (assumed to be average time between referral by GP to biopsy) Initial surgery: Mean: 62 (SD: 14.2); Median: 71; Derived utility: 0.895 Duration: 2 months (assumed time of hospitalization and waiting time for RT) Initial radiotherapy: Mean: 60 (SD: 15.2); Median: 59; Derived utility: 0.803 Duration: 2 months (assumed duration of morbidity due to RT) Initial chemotherapy: Mean: 50 (SD: 17.6); Median: 50; Derived utility: 0.717 Duration: 6 months Initial hormonal therapy: Mean: 63 (SD: 16.8); Median: 61; Derived utility: 0.820 Duration: 2 years 3 months – 1 year after mastectomy: Mean: 65 (SD: 16.5); Median: 64; Derived utility: 0.844	Difficult questionnaire for non-experts Experts & patients have different preferences *Non-transparent conversion to TTO values* *Scenario-based approach (hypothetical bias)*

Study (First author, year)	Population & sample	Methods	Results	Limitations
			Duration: 10 months (does not include the utilities above if at same time) 3 months – 1 year after breast-conserving therapy: Mean: 71 (SD: 15.5); Median: 74; Derived utility: 0.914 Duration: 10 months (does not include the utilities above if at same time) Disease-free > 1 year after mastectomy: Mean: 77 (SD: 12.7); Median: 80; Derived utility: 0.947 Duration: Life expectancy Disease-free > 1 year after breast-conserving therapy: Mean: 82 (SD: 11.5); Median: 83; Derived utility: 0.960 Duration: Life expectancy Palliative + chemotherapy: Mean: 36 (SD: 17); Median: 34; Derived utility: 0.531 Duration: 4 months Palliative + radiotherapy: Mean: 43 (SD: 16.4); Median: 38; Derived utility: 0.591 Duration: 1 month Palliative + surgery: Mean: 46 (SD: 16.4); Median: 41; Derived utility: 0.617 Duration: 5 weeks Palliative + hormonal therapy: Mean: 47 (SD: 19.1); Median: 45; Derived utility: 0.663 Duration: 14 months Terminal illness: Mean: 19 (SD: 15.3); Median: 17; Derived utility: 0.288 Duration: 1 month	

***Abbreviations**: BC = breast cancer, BCS = breast cancer screening, BSE = breast self-examination, CBE = clinical breast examination, IQR = Interquartile Range, QoL = Quality of Life, RT = radiotherapy, SCL-90 = Symptom Checklist-90, SD = Standard Deviation, SG = Standard Gamble, STAI-6 = Spielberger State-Trait Anxiety Inventory state scale, TTO = Time Trade-Off, USA = United States of America, VAS = Visual Analogue Scale, y = year.*

4.2.4 Model inputs: costing

Key points

The relevant costs related to breast cancer screening can be categorized in the following groups:

- Breast cancer screening-related costs (e.g. invitation, organization, harms, etc.);
- Costs of breast cancer diagnosis (e.g. needle biopsy, MRI, etc.);
- Costs of breast cancer treatment (including follow-up & palliative care).

The following trends related to the costing methodology were observed in the literature:

- Only a handful of studies (5/17) reported costs associated with inviting the screen-eligible candidates.
- Only I-CHER (2024) included the costs related to the organization of the screening programme (e.g. quality assurance, media campaigns, etc.).
- A specific cost for false positives or unnecessary treatment due to overdiagnosis was reported by three studies as inputs to their models.
- Three studies included productivity costs and one study additionally included travel costs in their analyses from a societal perspective.
- In general, studies transparently reported the cost elements taken into account and their sources.

The costs taken into account by each of the included studies as well as their sources are reported in Appendix Table 99. In general, three main cost categories could be distinguished: (1) breast cancer screening-related costs, (2) costs related to breast cancer diagnosis and (3) costs of breast cancer treatment.

4.2.4.1 Breast cancer screening-related costs

The **breast cancer screening-related costs** are defined as all costs arising from the breast cancer screening programme. This cost category includes the costs of organizing the screening programme and the screening mammographies as well as the costs associated with the potential harms of breast cancer screening (e.g. false positives, overdiagnosis, etc.). The principal cost of breast cancer screening is considered to be the cost of the screening mammographies, used to date as the primary breast cancer screening tool. Therefore, with the exception of three studies,[178, 192, 194] all other studies reported the cost for the screening mammography separately. Costs for the screening mammography ranged between £14.9 (~€17) and US$151.19 (€146.65) with an average of €91 and a median of €78. As reported by I-CHER (2024), the cost of a screening mammography in Belgium was €70.89 in 2024.[200] Besides the cost of the screening mammography, a second screening-specific cost is the cost related to the organization of the screening programme. This cost includes inviting the screen-eligible population, quality assurance, communication of the test results, media campaigns, etc. While only five studies[181, 193, 195, 197, 200] included a cost for inviting the screen-eligible population, implying the other studies downsized the cost of the screening to solely the cost of the mammographies, I-CHER (2024) is the only study additionally taking into account the cost of the entire organization of the screening programme.[200] Finally, Shafik et al. (2023) and van Luijt et al. (2017) did report a cost of screening, but do not disclose which cost elements are included in this overall cost.[178, 192] Morton et al. (2017) is the only study not separately reporting a cost related to screening as the authors opted to report all costs related to screening, diagnosis and treatment in a single summed cost.[194] Three studies varied the cost of screening in sensitivity analyses to estimate the impact of (the uncertainty in) the magnitude of this cost on the cost-effectiveness of the screening programme.[180, 195, 200]

One way of taking into account **screening-related harms** in health economic models is through the costs. While most studies included a general cost in their models (e.g. the cost of diagnostic work-up)

assigned to people who experience the harm (e.g. the people with a false-positive result), three studies opted to report harm-specific costs (e.g. cost of a false-positive result). More specifically, Schousboe et al. (2022) and Chootipongchaivat et al. (2021) reported and used a specific cost for false-positive cases.[180, 184] In addition, with regard to overdiagnosis, Morton et al. (2017) included a cost for unnecessary treatment and Blankenburg et al. (2023) defined a cost for overtreatment which they assigned to overdiagnosed cases.[179, 194] More information on the extent to which the included studies took into account screening-related harms, is available in section 4.2.5.

4.2.4.2 Costs of diagnosis

The majority of studies (15/17) reported taking into account the **costs for diagnosis** following a positive screening result. While I-CHER (2024) and van Luijt et al. (2017) included the costs related to diagnosis, both studies did not report these costs separately from the treatment costs.[192, 200] In contrast, Morton et al. (2017) and Shafik et al. (2023) do not report including any cost for diagnosis in their analyses.[178, 194] While eight studies provided the cost of the various diagnostic procedures separately,[181, 183, 187, 191, 193, 195-197] three studies only reported a general cost for diagnostic work-up without detailing which procedures and costs this summed cost entails.[180, 184, 189] The most frequently reported diagnostic procedures are (ultrasound-guided) needle biopsy and MRI.

4.2.4.3 Costs of breast cancer treatment, follow-up and palliative care

The final identified cost category includes all **costs related to breast cancer treatment**, including follow-up and palliative care. Twelve out of the seventeen included studies reported a cost for each health state in their model of which four subdivided this cost in different phases of care (e.g. initial phase and follow-up). Four studies provided a considerably more detailed overview of breast cancer treatment costs by reporting the costs on a procedure-level, sometimes specifically for each stage. The most frequently reported procedures in breast cancer treatment are: surgery (mastectomy & lumpectomy), chemotherapy, radiotherapy and hormonal therapy. Costs related to follow-up were reported by twelve studies, whereas palliative care costs were separately reported by eleven studies. Morton et al. (2017) only calculated and reported an average cost for treatment of a screened person and subsequently for an unscreened person without providing more details on what these costs entail.[194] Three studies examined the increase or decrease in treatment costs on the cost-effectiveness of breast cancer screening, of which two increased/decreased treatment costs in general and one separately increased the costs of stages 0-I-II on the one hand and the costs of stages III-IV on the other hand.[178, 180, 200]

4.2.4.4 Other costs & sources

As discussed previously, four studies conducted their analyses from a societal perspective implying that both direct and **indirect costs** were included. While I-CHER (2024), Tina-Shih et al. (2019) and Tina-Shih et al. (2021) only included indirect costs related to productivity loss due to premature death from breast cancer,[183, 187, 200] van Luijt et al. (2017) additionally included travel costs for screening and follow-up.[192]

As shown in Appendix Table 99, studies used a considerable variety of sources for the costs included in their respective models. The majority of studies used costs originating from a combination of sources, including official fee schedules, registry data, previously published literature and screening organizations.

4.2.5 Breast cancer screening-specific factors

Key points

Participation rate:

- Approximately half of the studies (8/17) explicitly or implicitly used an unrealistic participation rate of 100%.
- Participation rates in studies using a non-100% participation rate ranged between 60% and 82%.
- A minority of studies (5/17) included participation rate in the sensitivity analyses.

Sensitivity & false negatives:

- Only half of the studies (9/17) reported the sensitivity parameters used in the respective models with considerable variety between them.
- Most of these studies report multiple sensitivity values, each applicable to subgroups (e.g. age, tumour size, breast density) or specific years in the analyses.
- While false-negatives were modelled implicitly by assuming these cases were picked up in a following round or present themselves as interval cancers in all studies, one study opted to apply a specific disutility to these cases.

Specificity & false positives:

- Slightly more than half of the studies (10/17) reported the specificity parameters used in the respective models ranging from 92% to 99.5%.
- Three studies preferred the positive predictive value (PPV) of screening over specificity with values ranging from 12% to 35% (for subgroups).
- Slightly more than half of the studies (10/17) disclosed the number of false positives associated with the examined screening strategies.
- Overall, the number of false positives ranged between 18 per 1000 women screened for quadrennial screening and 1410 per 1000 women screened for hybrid annual-biennial screening.

Overdiagnosis:

- The majority of the studies (14/17) claim taking into account overdiagnosis in the analyses.
- The magnitude of overdiagnosis, reported by slightly more than half of the studies (10/17), varied considerably between studies and screening strategies.
- Input values for overdiagnosis ranged between 2% and 31% of screen-detected breast cancers.
- In general, the studies lack transparency on the methodology used to take into account overdiagnosis in the respective models.
- Of the studies transparently reporting the stage of overdiagnosed cancers (4/17), two studies modelled invasive cancers to be overdiagnosed in addition to DCIS.
- None of the studies applied a specific disutility to overdiagnosed cases (on top of mammography, diagnosis and treatment disutilities).

Other relevant factors:

- Only one study specifically reported incorporating the effect of radiation-induced breast cancer.
- None of the studies reported correcting for lead time or length bias.

Based on KCE report 379 on lung cancer screening and the identified literature for the purpose of this report, a specific set of factors important to be taken into account in economic evaluations of (breast cancer) screening programmes was identified. These factors, shown in Table 10 for the individual studies, include the participation rate, screening sensitivity and specificity as well as the magnitude and methodological approach to incorporate the impact of false positives and overdiagnosis. Additional relevant factors for breast cancer screening discussed in KCE report 379 include lead-time bias, length bias and the handling of false negatives. While not included in Table 10, the handling of false negatives will be discussed in the text below. As lead-time bias and length bias were not discussed in any of the included evaluations, the choice was made not to include these factors in Table 10.

4.2.5.1 Participation rate

The first important factor to be considered in an economic evaluation of breast cancer screening is the **participation rate** of the screen-eligible population in the organized screening programme. The importance of using a realistic participation rate lies in the notion that, in contrast to full participation (i.e. 100% participation rate), lower participation rates still incur fixed costs related to the screening programme as well as the costs for inviting the entire screen-eligible population to be spread over fewer participants. As a consequence, all else equal, lower participation rates lead to an unfavourable impact on the cost-effectiveness of screening. Twelve out of the seventeen included evaluations clearly reported the participation rate(s) used in the base case of their analyses.[178, 179, 181, 184, 187, 189, 192-194, 196, 197, 200] Participation rates used in the base case analyses ranged between 60% and 100%. While two studies applied age (group)-specific participation rates in their base case analysis,[192, 200] Schiller-Frühwirth et al. (2017) opted to use a 60% participation rate applied to all ages for the organized screening programme and use age-specific participation rates for opportunistic screening and Kregting et al. (2022) only used age-specific participation rates in a scenario analysis assuming full participation in the base case (see section 4.3.2).[181, 193] In contrast, Chootipongchaivat et al. (2021) did not prefer one specific participation rate for the base case and therefore presented their base case results separately for individual participation rates.[184] Other participation rates applied in the analyses were 77%,[194] 80%,[196, 197] 82%[178] and 100%[179, 181, 187, 189]. Carter et al. (2018) first report a participation rate of 30% and later in the publication a rate of 100%, making it unclear which rate was actually applied.[191] Four studies did not report any participation rate in the base case, implying they implicitly assumed full participation to the screening programme.[180, 183, 190, 195] This implies that approximately half of the studies (8/17) either explicitly or implicitly assumed a participation rate of 100%. Remarkably, Rafia et al. (2016) report participation rates of 65% and 100% in the sensitivity analyses but did not disclose which participation rate was applied in the base case.[195] An additional four studies did explicitly opt to use the rather unrealistic participation rate of 100%, with Tina Shih et al. (2019) claiming their choice for 100% participation is not of significance as they assess relative costs and benefits between screening strategies.[187] However, this argument does not apply here as the authors do calculate an ICER compared to no screening, which would illustrate a potentially overoptimistic estimation of the cost-effectiveness of that strategy. Unfortunately, only five studies did vary the participation rate in the sensitivity analyses to estimate the impact of a higher/lower participation among the screen-eligible population on the result of the evaluation.[ppp184, 193, 195, 196, 200]

[ppp] The results of these sensitivity analyses are discussed in section 4.3.2.

4.2.5.2 Screening sensitivity

A second important factor in evaluating screening programmes is screening **sensitivity**. The higher the sensitivity of the screening, the more likely it is that someone who actually has the disease will get a positive test result and the fewer false-negative cases are missed by the screening protocol. A high number of false-negative cases (i.e. low screening sensitivity) would not lead to additional costs on the short term as no diagnostic procedures and treatments will take place until the missed breast tumour(s) will be picked up in a next screening round or when symptoms become visible. However, in the longer run, false-negative cases may lead to later diagnosis and subsequently more sophisticated, radical, and/or more expensive procedures with an impact on both costs as well as survival and quality of life. Only half of the included studies (9/17) reported the sensitivity parameters used as inputs in the respective models with considerable variety between them.[179, 183, 187, 190, 193, 194, 196, 197, 200] Two studies reported one single value for screening sensitivity applicable to the full population, more specifically 59%[179] and 80%.[194] Other studies reported multiple sensitivity values each applicable to a specific subgroup in their models based on tumour size[187] (combined with age),[197] breast density[190] (combined with age)[183] or age[193]. In contrast, I-CHER (2024) and Arrospide et al. (2016) reported year-specific sensitivity values ranging between 65% - 68% and 85% - 95% respectively, but it was not clear how these were incorporated in the models (e.g. weighted average, specific value for specific year, etc.).[196, 200] Besides Blankenburg et al. (2023) who applies a disutility of 0.01 for an undisclosed time period to false negatives, none of the studies specifically discussed how they handled false-negative cases in their analyses.[179] Likely the assumption was made by other studies that false negatives were picked up in a subsequent screening round or were detected symptomatically in between screening rounds (i.e. interval cancers). With regard to mammography combined with an ultrasound in women with dense breasts, Blankenburg et al. (2023) used a sensitivity of 55% in the base case and 50% in an expert opinion scenario.[179] While an important parameter in many models, only three studies varied the sensitivity of screening in sensitivity analyses.[190, 193, 197]

4.2.5.3 Screening specificity

Screening **specificity** is the third important factor to consider when evaluating (breast cancer) screening programmes. Specificity is defined as the percentage of true negative test results among non-sick persons. This means that lower rates of specificity are related to higher numbers of false-positive cases. Similar to sensitivity, screening specificity was reported by ten of the seventeen included studies. Three studies reported a single rate applicable to the entire population: 93% by Carter et al. (2018), 94% by Blankenburg et al. (2022) and 96.5% by Koleva-Kolarova et al. (2018).[179, 190, 191] In addition, three studies reported age-dependent specificity values[183, 187, 193] and an additional three reported separate specificity values for different years.[184, 196, 200] However, as in the case of year-specific sensitivity, it was unclear how these studies incorporated year-specific specificity values in the model. Schousboe et al. (2022) reported proportions of mammographies leading to a false-positive result ranging between 6.24% and 9.28% depending on screening interval, age and Charlson Comorbidity Score (CCS).[180] While not reporting specificity, three studies did report a positive predictive value (PPV) of screening.[181, 192, 197] van Luijt et al. (2017) reported an overall PPV of 12.5%, Sankatsing et al. (2015) reported PPV for two age categories (40-49: 12% and 50-74: 30%) and Kregting et al. (2022) reported PPV for below 50 years and above 49 years for each examined screening interval separately (<50: 10%-15% and >49: 24%-35%).[181, 192, 197] For mammography combined with an ultrasound in women with dense breasts, Blankenburg et al. (2023) used the same specificity as mammography alone (94%) in the base case and a value of 85% in an expert opinion scenario.[179] Four studies reported to have conducted sensitivity analyses by varying the specificity or PPV to estimate the impact of these parameters on the ICER.[180, 183, 190, 197]

Table 10 – Overview of factors considered in primary evaluations

Study	Participation rate	Sensitivity	Specificity	False positives	Overdiagnosis*
Journal publications					
Shafik, 2023[178]	82%	*Not reported*	*Not reported*	*Magnitude not reported* Costs: *Methodology not reported* QoL: CEA (no QoL included)	*Magnitude not reported* *Methodology not reported*
Blankenburg, 2023[179]	100%	Mammography: Literature: 59%[221] Experts: 30% Ultrasound as add-on to mammography: Literature: 55%[222] Experts: 50%	Mammography: Literature: 94%[221] Experts: 90% Ultrasound as add-on to mammography: Literature: 94%[222] Experts: 85%	Mammography (per 1000): Average risk: 1790 Intermediate risk: 1720 Ultrasound as add-on to mammography (per 1000): Average risk: 3473 Intermediate risk: 3336 Costs: Cost of false positives (output) separately reported QoL: Disutility of screening, *no additional QoL adjustment*	*Magnitude not reported* Costs: Cost of overtreatment included (not reported) QoL: *no specific QoL adjustment reported*
Schousboe, 2022[180]	*Not reported*	*Not reported*	Proportion of mammographies leading to false-positive result: Biennial: 7.23% – 9.28% *(depending on CCS & age)* Annual: 6.24% – 7.59% *(depending on CCS & age)* *Source*: Breast Cancer Surveillance Consortium	Biennial: 107 – 389 per 1000 women screened *(depending on CCS & stop age)* Annual: 199 – 661 per 1000 women screened *(depending on CCS & stop age)* Costs: Costs up to work-up included QoL: Disutilities for screening & diagnostic evaluation	Biennial: 2.2 – 7 per 1000 women screened *(depending on CCS & stop age)* Annual: 3.3 – 9.9 per 1000 women screened *(depending on CCS & stop age)* Costs: Included QoL: *no specific QoL adjustment reported (QoL of DCIS assumed)*

Study	Participation rate	Sensitivity	Specificity	False positives	Overdiagnosis*
					Scenario with 100% DCIS and 5% invasive excess incidence
Kregting, 2022[181]	100% *(scenario with age-specific participation rates)*	*Not reported*	*Not reported* *PPV (annual – biennial – triennial – quadrennial):* <50: 10%[223] - 12%[224-226] - 13%[223] - 15% (extrapolated) >49: 24%[223] - 28%[224-226] - 31%[223] - 35% (extrapolated)	Per 1000 women screened: Annual: 259 - 275 Biennial: 89 - 168 Triennial: 83 - 109 Quadrennial: 18 - 62 *(depending on age range)* Costs: *Methodology not reported* QoL: *No specific QoL adjustment reported*	Per 1000 women screened: Annual: 8.6 – 10.1 Biennial: 5.8 – 7.3 Triennial: 4.9 – 5.7 Quadrennial: 1.7 – 4.3 *(depending on age range)* Costs: Included QoL: *No specific QoL adjustment reported*
Tina Shih, 2021[183]	*Not reported*	Dense breasts:[227] 40-49: 0.801 50-64: 0.868 >64: 0.905 Non-dense breasts:[227] 40-49: 0.897 50-64: 0.933 >64: 0.953	0.922-0.995[228] *(age-dependent)*	Triennial 50-75: 141.2 per 1000 women screened Biennial 50-75: 202.5 per 1000 women screened Hybrid (dense vs non-dense breasts): 567.3 per 1000 women screened Costs: *Methodology not reported* QoL: Disutility for mammography & diagnostic work-up included (scenario)	Triennial 50-75: 10.5% Biennial 50-75: 12.4% Hybrid (dense vs non-dense breasts): 17.2% Costs: Included QoL: *No specific QoL adjustment reported*
Chootipongchaivat, 2021[184]	ICER calculated for various rates of participation	Sensitivity is not a direct input in the model	0.9505-0.9583 *(depending on calendar year)* *Source*: BreastScreen Singapore (BSS) and National Registry of Diseases Office (NRDO)	For 100% attendance & current screening strategy (annual 40-49 and biennial 50-69): 56 133 per 100 000 women alive at start of screening Costs: *Methodology not reported*	For 100% attendance: 19.4% Costs: Included

Study	Participation rate	Sensitivity	Specificity	False positives	Overdiagnosis*
				QoL: Disutility for mammography & diagnostic work-up included	QoL: *No specific QoL adjustment reported*
Tina Shih, 2019[187]	100%	40-49 years:[228] 0.5 cm: 0.113-0.245 1 cm: 0.353-0.549 2 cm: 0.881-0.961 >49 years:[228] 0.5 cm: 0.169-0.626 1 cm: 0.454-0.875 2 cm: 0.917-0.998	40-49 years: 0.922-0.967[228] >49 years: 0.968-0.995[228]	*Magnitude not reported* Costs: up to treatment + additional follow-up mammography 6 months after screening QoL: No disutility for false positives taken into account	Non-progressive DCIS: 9% of screen-detected cancers Non-progressive DCIS is assumed to be overdiagnosis Costs: Included QoL: *No specific QoL adjustment reported*
Mittmann, 2018[189]	100%	Sensitivity based on age, density & screening round *(not disclosed)* Sensitivity calibrated by adjusting model parameters to match empirical data	Specificity function based on age, density & screening round *(not disclosed)*	*Magnitude not reported* Costs: *Methodology not reported* QoL: Disutility for mammography & diagnostic work-up included	*Not reported*
Koleva-Kolarova, 2018[190]	*Not reported*	BI-RADS 1: 0.87 BI-RADS 2: 0.84 BI-RADS 3: 0.73 BI-RADS 4: 0.65 *Source*: unreferenced meta-analysis of 9 primary studies[229-237]	96.5%[225]	Biennial: 375.4 - 408.5/1000 women screened *(depending on age)* Costs: *Methodology not reported* QoL: CEA (no QoL included)	Overdiagnosis not taken into account: DCIS not included in model
Carter, 2018[191]	Unclear: 30% or 100%	"Modeled on tumour size and menopausal status, increasing with enlarging tumour diameter and advancing age"	93%[238]	ACS 2015: 1410 per 1000 women screened USPSTF 2016: 790 per 1000 women screened Costs: *Methodology not reported* QoL: CEA (no QoL included)	31% of screen-detected cancers Costs: Included QoL: CEA (no QoL included)

Study	Participation rate	Sensitivity	Specificity	False positives	Overdiagnosis*
van Luijt, 2017[192]	Age- and year-specific participation rate (based on Norway Cancer Registry data)	*Not reported*	*Not reported* *Positive predictive value: 12.5%*[239]	Biennial 50-69y: 317.43 per 1000 women Costs: Up to diagnosis QoL: Disutility for mammography & diagnostic work-up included	Biennial 50-69: Estimated at 2-3% Costs: Included QoL: *No specific QoL adjustment reported*
Schiller-Frühwirth, 2017[193]	Organized screening: 45-69y: 60% Opportunistic screening: 40-44y: 22.5% 45-69y: 27.7% 70-74y: 15%	Organized screening: 40-49y: 84.7%[240] 50-54y: 87.2%[241] 55-59y: 92.9%[241] 60-64y: 91.7%[241] 65-69y: 91.8%[241] 70+: 91.2%[240] Opportunistic screening: 40-44y: 71.9%[242] 45-49y: 76%[241] 50-54y: 80%[241] 55-59y: 83.6%[241] 60-64y: 86.9%[241] 65-69y: 83.1%[241] 70-74y: 83.9%[242] 75+: 85.1%[242]	Organized screening: 40-49y: 93.9%[240] 50-54y: 97.5%[241] 55-59y: 97.7%[241] 60-64y: 98%[241] 65-69y: 98.1%[241] 70+: 93.4%[240] Opportunistic screening: 40-44y: 89.1%[242] 45-49y: 89.6%[241] 50-54y: 90%[241] 55-59y: 90.3%[241] 60-64y: 90.9%[241] 65-69y: 91.5%[241] 70-74y: 93.1%[242] 75+: 93.6%[242]	*Magnitude not reported* Costs: M*ethodology not reported* QoL: CEA (QoL not included)	*Rate not reported* Costs: Included QoL: CEA (QoL not included)
Morton, 2017[194]	77%	80% *Source not clear*	*Not reported*	*Magnitude not reported* Costs: *Methodology not reported* QoL: QALY gain multiplied by 0.253	*Rate not reported* Accounted for by multiplying the QALY gain by a literature-based ratio to correct for harms Costs: unnecessary treatment cost included QoL: QALY gain multiplied by 0.253
Rafia, 2016[195]	*Not reported*	*Not reported*	*Not reported*	*Magnitude not reported*	6.2% at age of 72

Study	Participation rate	Sensitivity	Specificity	False positives	Overdiagnosis*
	(SA: 65% and 100%)			Costs: *Methodology not reported* QoL: Disutility for mammography & recall included	37.9% at age of 90 Costs: Included QoL: *No specific QoL adjustment reported*
Arrospide, 2016[196]	80% (SA: 30% and 50%)	1996-1999: 95.20% 2000-2005: 83.40% 2006-2008: 83.52% 2009-2011: 85.86% *Source*: BC Screening Programme in the Basque Country	1996-1999: 90.44% 2000-2005: 90.61% 2006-2008: 93.67% 2009-2011: 94.13% *Source*: BC Screening Programme in the Basque Country	*Magnitude not reported* Costs: *Methodology not reported* QoL: *No relevant QoL adjustments reported*	Biennial 50-69: 4% Costs: Included QoL: *No specific QoL adjustments reported*
Sankatsing, 2015[197]	80%	DCIS: 100% T1AN- & T1AN+ <50y: 70% ≥55y: 94% T1BN- & T1BN+ <50y: 83% ≥55y: 100% T1CN- & T1CN+ <50y: 80% ≥55y: 100% T2+N- & T2+N+ <50y: 93% ≥55y: 100% *Source*: [226, 243, 244]	*Not reported* *Positive predictive value:* *50-74: 30%*[226] *40-49: 12% (assumption: 40% of 50-74y based on literature*[224, 225, 245]*)*	124 - 229 per 1000 women *(depending on age & interval)* Costs: *Methodology not reported* QoL: CEA (no QoL included)	4.84-5.97% *(depending on age & interval)* Costs: Included QoL: *No specific QoL adjustment reported*
Other reports					
I-CHER, 2024[200]	Organized screening: 45-49y: 36.99% 50-54y: 45.66%	64.7% - 68% depending on year (2017-2022)[84, 246-250]	98% - 98.3% depending on year (2017-2022) [84, 246-250]	Stop screening: 3932 (2.48% of opportunistic screenings)	*Rate not reported*

Study	Participation rate	Sensitivity	Specificity	False positives	Overdiagnosis*
	55-59y: 51.98% 60-64y: 55.92% 65-69y: 57.50% 70-74y: 56.71% (SA: 25% increase/decrease) Opportunistic screening: 40-44y: 9.83% 45-49y: 9.33% 50-54y: 8.83% 55-59y: 8.34% 60-64y: 7.84% 65-69y: 7.34% 70-74y: 6.84%			Current screening: 5861 (1.97% of opportunistic + organized screenings) Extended screening: 7237 (1.94% of opportunistic + organized screenings) Costs: *Methodology not reported* QoL: Additional disutility of 0.02 *(period not reported)*	*Methodology not reported (worded as 'overscreening')* Costs: Included QoL: *No specific QoL adjustment reported*

***Notes**: *Percentages of overdiagnosis are relative to the number of screen-detected cancers unless otherwise specified. Due to a lack of transparency, the assumption was made that costs relevant for overdiagnosed cancers were included if overdiagnosis was taken into account in the model (label: 'included).*
***Abbreviations**: ACS = American Cancer Society, CSS = Charlson Comorbidity Score, DCIS = Ductal Carcinoma in Situ, LYG = Life Years Gained, QALY = Quality-Adjusted Life Year, QoL = Quality of Life, SA = sensitivity analysis, USPSTF = United States Preventive Services Task Force, y = years.*

4.2.5.4 *False positives*

A direct consequence of the assumed specificity in the model is the number of **false positives** that are unnecessarily identified by the screening protocol. False-positive results can impact the cost-effectiveness of lung cancer screening in several ways. First, the unnecessary diagnostic procedures following a false-positive result lead to additional unnecessary costs as well as unnecessary adverse events and/or complications resulting from the diagnostic procedures. Second, multiple studies argue that a (false-) positive screening result has a negative impact on the quality of life of participants due to increased anxiety levels (see 4.2.3.2). Column 5 of Table 10 summarizes the number of false positives reported by each of the included studies as well as which costing methodology was used for false-positive cases and whether a QoL adjustment was applied. Out of the seventeen included studies, ten reported the number of false positives associated with the examined screening strategies. Due to the variety in screening protocols (in terms of intervals, age eligibility criteria, and different protocols for women with dense breasts), the reported magnitudes of false-positive cases also varied considerably between the included studies. Overall, the number of false positives ranged between 18 per 1000 women screened for quadrennial screening and 1410 per 1000 women screened for hybrid annual-biennial screening.[qqq] In general, studies lacked transparency on which costs were applied to false-positive cases as only three studies clearly reported this information. While van Luijt et al. (2017) state that all costs are included up to the point of the diagnostic work-up, Tina Shih et al. (2019) additionally take into account the cost of a recall examination and a portion of the treatment cost if women with a false-positive result received (part of the) treatment before the screening result was confirmed as false.[187, 192] With regard to QoL, only I-CHER (2024) assigned a specific disutility (i.e. 0.02) to false positives in the base case analysis but fail to provide the period for which this disutility was applied.[200] The other studies performing a cost-utility analysis did not assign one specific disutility to false-positive cases in the base case, but eight of them did take into account some kind of QoL loss. Six studies applied a disutility of the screening mammography and an additional disutility for the diagnostic work-up to false-positive cases, as they did to negative (only disutility of screening) and true positive cases (both disutilities).[rrr180, 183, 184, 189, 192, 195] In contrast, Blankenburg et al. (2023) only assigned a disutility of screening to false positives and Morton et al. (2017) multiplied the QALY gain due to screening by 0.253 to correct for all screening harms.[179, 194] While Tina Shih et al. (2019) specifically reported not taking into account any disutilities for false positives, Kregting et al. (2022) and Arrospide et al. (2016) did not provide sufficient information to be able to conclude on their QoL adjustments.[181, 187, 196]

4.2.5.5 *Overdiagnosis*

An additional harm specifically related to cancer screening is **overdiagnosis**. Overdiagnosed cancers are those (often small and slow-growing) cancers detected by the screening programme that would never have become symptomatic over the patient's remaining life span. By detecting these overdiagnosed cancers, unnecessary, but costly diagnostic procedures and treatments are performed without additional health benefits compared to a no-screening situation. Both these procedures as well as the potential complications could be considered as harms of screening. While all costs related to overdiagnosed cancers should always be taken into account and patients with an overdiagnosed cancer cannot benefit from the screening in any way, modelling overdiagnosis often poses methodological challenges. These challenges arise from the difficulties in measuring the magnitude of overdiagnosis in clinical trials and lack of evidence concerning the impact on QoL, leading to

---

[qqq] The number of false positives can exceed the number of women screened as the numbers are expressed over the lifetime of the women. For example, 1410 false positives per 1000 women corresponds to an average of 1.41 false-positive results over a lifetime in which women were screened on average 20.5 times over their lifetime.

[rrr] For the specific disutilities used by the studies, we refer to section 4.2.3.1. Tina Shih et al. (2021) only assigned a disutility in a scenario.[183]

uncertainty on which overdiagnosis rate to use and whether a specific QoL decrement should be taken into account or not (and if so, which disutility for which time period). Fourteen out of the seventeen included evaluations introduced the concept of overdiagnosis in their articles.[179-181, 183, 184, 187, 190-197] While Koleva-Kolarova et al. (2018) explain the concept, a substantial methodological limitation of their model is the exclusion of DCIS which in turn prevents taking into account overdiagnosis.[190] Of the remaining thirteen studies, ten reported the magnitude of overdiagnosis either as an input (6/10) or output (4/10) of their model. Input values ranged between 2%[192] and 31%[191] and varied considerably depending on age eligibility criteria and screening interval. Tina Shih et al. (2021), for example, used an overdiagnosis rate of 12.4% for biennial screening and 10.5% for triennial screening for women between 50 and 75.[183] Interestingly, Rafia et al. (2016) reported age-specific overdiagnosis rates with an estimated overdiagnosis of 6.2% at the age of 72 and 37.9% at the age of 90.[195] van Luijt et al. (2017) reported finding an 11% overdiagnosis rate in a UK review of screening harms and a rate of 2-3% in a publication on the Norwegian screening programme for screening biennially between 50 and 69, and opted to only use the latter.[192] Similarly, Arrospide et al. (2016) considered 4% of screen-detected cancers to be overdiagnosis when evaluating the same interval in the same age category, a value which was based on the Basque screening programme.[196] Carter et al. (2018) reported a flat rate of 31% based on an estimation from Surveillance, Epidemiology, and End Results Program (SEER) data over 30 years of screening.[191] Tina Shih et al. (2019) assumed all non-progressive DCIS (9% of screen-detected cancers) to be overdiagnosis based on a publication of van Ravesteyn et al. (2018),[251] reporting that approximately 17% of patients with breast cancer were diagnosed with DCIS, and of those, 51.3% had non-progressive DCIS.[187] In contrast, four studies reported overdiagnosis solely as an output of their model. Chootipongchaivat et al. (2021) reported overdiagnosis rates between 15.5% and 25.6% of screen-detected cancers depending on age eligibility criteria and interval, with broader age windows and more frequent screening logically leading to more overdiagnosis.[184] While evaluating similar screening strategies as Chootipongchaivat et al. (2021), Sankatsing et al. (2015) found far less overdiagnosis resulting from their model. For example, for biennial screening between 40 and 74 years, Chootipongchaivat et al. (2021) found a value for overdiagnosis of 20.5% of screen-detected cancers in contrast to the 4.84% reported by Sankatsing et al. (2015).[184, 197] From the respective articles, it is not clear where the difference originates from. Finally, Schoesbou et al. (2022) and Kregting et al. (2022) report highly similar overdiagnosis estimates from their respective models ranging between 2.2 and 7.3 per 1000 women screened for biennial screening and between 3.3 and 10.1 per 1000 women screened for annual screening (depending on which age window was evaluated).[180, 181] In general, the included evaluations poorly documented and explained how overdiagnosis was treated in their respective models. This lack of transparency made it virtually impossible to evaluate whether and how overdiagnosis was taken into account in the models. For example, only four studies clearly reported in which stages the overdiagnosed cancers were assumed to be diagnosed.[180, 183, 184, 187] Three out of the four studies transparently reporting the stages of overdiagnosed cancers assumed all overdiagnosis to be DCIS.[180, 183, 187] In contrast, Chootipongchaivat et al. (2021) do assume part of the overdiagnosed cancers to be invasive and separately report overdiagnosed DCIS and overdiagnosed invasive breast cancers of each examined screening strategy. For example, when screening triennially between 50 and 64 years with a participation rate of 10%, the results indicate that 3 out of the 42 overdiagnosed cancers are invasive breast cancers. Unfortunately, the authors do not specify the stages of the overdiagnosed invasive cancers.[184] In addition, while assuming all overdiagnosis to be DCIS in the base case, Schousboe et al. (2022) did run a scenario analysis in which 5% of excess invasive breast cancers were included in the group of overdiagnosed cancers on top of all excess DCIS incidence.[180]

With regard to costs, two studies reported a separate cost for overtreatment or unnecessary treatment.[179, 194] All other studies did not specifically describe which costs were taken into account for the overdiagnosed cancers. However, it is highly likely these studies assigned the same costs to overdiagnosed cases as assigned to non-overdiagnosed cases without specifically reporting doing so

in the costing methodology.[sss] With the exception of Morton et al. (2017), correcting for all screening-related harms by multiplying the QALY gain by 0.253, none of the other studies used a specific disutility for overdiagnosis due to the unnecessarily labelling someone as cancer patient with all related consequences having a potential impact on QoL.[194] As described above, the systematic review of Li et al. (2019) did not find any study examining the impact of overdiagnosis on QoL, which could explain why utility decrements were not taken into account in the included evaluations.[205] Only Schiller-Frühwirth et al. (2017) explicitly stated that a disutility could not be taken into account, and that the lack of evidence was the reason.[193]

4.3 Results of primary economic evaluations

Key points

Critical assessment of economic evaluations is necessary before interpreting results and conclusions. In this part, results as reported by the authors are presented. For some critical remarks, we refer to the discussion.

Extending breast cancer screening to younger women (i.e. start ages between 40 and 50 years) is found to be considerably more cost-effective than extending screening to older women (i.e. stop ages between 70 and 90).

While annual screening leads to more health gains, biennial screening is found to be more cost-effective. Similarly, triennial screening is considered more cost-effective than biennial screening.

Only one study was performed for a Belgian (Flemish) setting, which found the following results in terms of cost-utility:

- Current screening (biennial screening: 50-69 years): ICER = €10 033/QALY gained
- Extension to younger women (biennial screening: 45-69 years): ICER = €10 740/QALY gained
- Extension to older women (biennial screening: 50-69 years + triennial screening: 70-74 years): €35 752/QALY gained

4.3.1 Base case results

Due to the considerable variations in study setting, screening modalities, model inputs and approaches, as well as the main focus of this review being on the methodology used by the studies, this section will only focus on the main trends visible in the results of the included studies. For more detailed results, we refer to Appendix Table 100 and the individual publications of the included studies. It is important to note that we refer to conclusions as made by the authors of the study. Nevertheless, study results and conclusions should be critically approached. Interpretations on whether or not something can be considered to be cost-effective depend, among others, on the reliability of the applied input values and assumptions made for (the most important) variables, applied willingness-to-pay (WTP) thresholds, etc.

Several of the included studies focused specifically on examining the cost-effectiveness and/or cost-utility of **extending breast cancer screening eligibility criteria to younger and/or older age groups**. In general, studies concluded that extending breast cancer screening to ages below 50 and above 69 is cost-effective. Taking a closer look at the studies that examine the downwards extension and upwards extension separately, some important insights appear. First, extending the age eligibility criteria either downward or upward is associated with both an increase in costs for the health care payers as well as health gains in the eligible female population. Second, extending breast cancer

[sss] As a consequence, we labelled costs related to overdiagnosis as ‘included’ without further information in Table 10.

screening to lower ages (i.e. between 40 and 50 years) is generally associated with a lower ICER than extending breast cancer screening to older ages (i.e. between 70 and 90). More specifically, the former leads on average to ICERs around or below €20 000/QALY gained or LYG while the latter is associated with ICERs above €35 000/QALY gained or LYG and several even above €50 000/QALY gained or LYG. This shows the importance of presenting results for the extension to the younger and older age groups separately instead of solely reporting the ICER for the combination of lowering and increasing the age eligibility criteria. Interesting in this regard is the **only study reflecting a Belgian (Flemish) setting** (i.e. I-CHER (2024)), which examines three strategies: (1) biennial screening between 45 and 69 years combined with triennial screening between 70 and 74, (2) biennial screening between 45 and 69 years, and (3) biennial screening between 50 and 69 years combined with triennial screening between 70 and 74 years. Compared to current screening in Belgium (i.e. biennial screening between 50 and 69 years), the strategies led to ICERs of €16 244, €10 740 and €35 752/QALY gained respectively.[200] Interestingly, Shafik et al. (2023) arrived at the same conclusion for the Finnish situation.[178] The general trend thus seems to indicate that extending breast cancer screening to younger age groups is (considerably) more cost-effective than extending screening to older age groups. This phenomenon could be explained by the lower remaining life expectancy as well as the level of overdiagnosis in these older age groups.

With regard to the **screening interval** some clear trends are visible in the results. In general, screening more frequently (i.e. shorter screening intervals) leads to higher ICERs. More specifically, this means that shifting to a shorter screening interval (e.g. from biennial to annual screening) is associated with a proportionally larger increase in the costs than in the health effects (i.e. QALYs or LYG). This trend leads to the conclusion that biennial screening for breast cancer is more cost-effective than annual screening and that triennial screening at its turn is found to be more cost-effective than biennial screening.[179-181, 183, 184, 187, 189] While still considered to be cost-effective by most studies, which is also influenced by the WTP thresholds applied in these studies, shifting from biennial to annual screening was decision-altering (i.e. no longer cost-effective) in only two studies.[180, 181] As mentioned previously, one study examined an adapted screening interval for women with dense breasts (i.e. annual screening from 40 to 75 for women with dense breasts and biennial screening from 50 to 75 for women with non-dense breasts) in comparison with biennial screening between 50 and 75 for all women. While this former strategy led to additional health gains, it also induced a steep increase in the costs, which resulted in an ICER above US$50 000/QALY gained. A second study examining a specific screening protocol for women with **dense breasts** is the study of Blankenburg et al. (2023), which evaluated a **combination of mammography and ultrasound** for this specific population. Compared with screening all women using mammography only, the adjusted screening protocols led to health gains but also additional costs resulting in ICERs between US$10 797/QALY gained (for triennial screening) and US$23 394/QALY gained (for annual screening).[179]

4.3.2 Sensitivity analyses

Key points

The ICER of breast cancer screening was found to be **highly sensitive to**:

- Screening interval;
- Overdiagnosis;
- Screening sensitivity;
- Variables at the basis of the modelling approach (e.g. incidence, prevalence, stage shift/distributions, etc.)
- Cost of mammography and cost of screening as a whole;
- Treatment cost of stages I and II;
- Treatment cost of stages III and IV;

- Time horizon.

A **moderate to minor impact** on the ICER was found when changing:

- Treatment cost as a whole in all stages;
- Disutility associated with mammography;
- Disutility associated with diagnostic work-up;
- Disutility associated with false-positive result.

Studies found **mixed results** on the sensitivity of the ICER to:

- Screening specificity;
- Participation rate.

Of the seventeen included studies, only van Luijt et al. (2017) did not perform sensitivity analyses.[192] Approximately half of the remaining studies solely performed deterministic sensitivity analyses[178, 181, 184, 189, 190, 194, 195] and the other half combined deterministic with probabilistic sensitivity analyses.[179, 180, 183, 187, 191, 193, 196, 197, 200] Table 101 in Appendix 16 lists the variables included in the sensitivity and scenario analyses of each study. In the remainder of this section, the main results of these analyses are summarized.

In general, longer **screening intervals** (i.e. less frequently screening, e.g. biennial) were found to be more cost-effective than shorter screening intervals (e.g. annual). The screening intervals were shown to have a substantial impact on the ICER of the screening programme (see 4.3.1 for more information).

While only included in sensitivity analyses by one study, the amount of **overdiagnosis** was found to have a substantial unfavourable effect on the cost-effectiveness of breast cancer screening. In the base case analysis, where all overdiagnosis was assumed to originate solely from excess DCIS incidence, the ICER of biennially screening until 80 was US$54 000 for women with a CCS of 0 and US$85 000/QALY gained for women with a CCS above 1. In a scenario analysis, overdiagnosis was assumed to be 100% of excess DCIS incidence (as in the base case) in addition to 5% of invasive cancers. As a consequence, ICERs increased to US$69 000/QALY gained and US$117 000/QALY gained respectively.[180]

All **variables that form the basis of the modelling approach** were proven to have a major impact on the resulting ICERs. Examples of these variables include incidence, prevalence, stage distributions, and stage shift.

With regard to the impact of **specificity** on the ICER, the studies found mixed results. Tina Shih et al. (2021) concluded that reducing specificity by 8% had a considerable impact on the ICER with an increase from US$22 276 to US$33 176/QALY gained.[183] In contrast, Koleva-Kolarova et al. (2018) concluded that the impact of changing specificity by two times its standard deviation on the ICER was very minor.[190] Similarly, Sankatsing et al. (2015) conducted a scenario analysis in which the PPV for women aged <50 was brought to 50% of the PPV for women aged ≥50 and also concluded the impact on the ICER was minor.[197]

In contrast, studies consistentently found screening **sensitivity** to be highly influential on the cost-effectiveness of breast cancer screening. For example, Koleva-Kolarova et al. (2015) found sensitivity of the screening mammography to be the third most influential factor on the ICER, with lower sensitivity values leading to higher ICERs.[198]

With regard to costs, especially increasing the **cost of the screening mammography or the screening cost as a whole (mammography, invitations, etc.)** has a substantial unfavourable impact on the ICER of the screening programme.[180, 195, 200] I-CHER (2024) also solely varied the organizational cost of the screening programme and found this factor to affect the ICER mildly.[200]

Three studies examined the increase or decrease in **treatment costs** on the cost-effectiveness of breast cancer screening, of which two increased/decreased treatment costs in general and one separately increased the costs of stages 0-I-II on the one hand and the costs of stages III-IV on the other hand.[178, 180, 200] The studies following the former approach found a decrease in the ICER when treatment costs were increased.[178, 180] When the treatment costs were increased by 50%, Shafik et al. (2023) even found breast cancer screening to be cost-saving.[178] In contrast, Schousboe et al. (2022) found only a mild effect of changing the treatment costs on the ICER.[180] Due to the stage shift approach, resulting in more breast cancers diagnosed in the earlier stages of breast cancer, which are typically characterized by lower treatment costs than the advanced stages, I-CHER (2024) found a mild to moderate decrease in the ICER when the costs of stages III and IV were increased and vice versa for the costs of stages 0, I and II.[200]

Studies found mixed results on the sensitivity of the ICER to the **participation rate**. Arrospide et al. (2016), for example, stated finding a proportional decrease in incremental costs and effects when the participation rate was lowered.[196] However, while the authors state that incremental costs and effects decrease proportionally, the ICER does halve in magnitude when the participation rate was lowered from 80% to 30%. Rather counter-intuitively (see discussion), this would mean that the screening programme would become more cost-effective if less people participated (i.e. when the participation rate was lowered from 80% to 50% and 30%). Rafia et al. (2016) and Kregting et al. (2022), who in contrast to Arrospide et al. (2016) did include a cost for the invitations, also found that the screening programme would become more cost-effective (lower ICER) if the participation was lower and less cost-effective (higher ICER) if participation was increased, albeit with a relatively small impact in magnitude on the ICER.[181, 195] Chootipongchaivat et al. (2021) found mixed results with a positive effect (i.e. higher participation leads to higher ICER) for most of the screening strategies. However, for some screening strategies (e.g. screening annually between 40 and 79 years) the ICER decreases when the participation rate is increased from 10% to 30% and 30% to 50%, but increases when the participation rate is further increased from 50% to 70% or from 70% to 100%.[184] Finally, I-CHER (2024), which is the most exhaustive in terms of including all costs related to the screening programme, increased age-specific participation rates by 25% and found an increase of approximately 8.5% in the ICER, which is mainly attributable to a 20% increase in costs and approximately a 10% increase in effects.[200] In contrast to the results of these studies, Schiller-Frühwirth et al. (2017) consistently found lower ICERs when screening attendance was systematically increased. Interestingly, the authors concluded that the attendance rate had a stronger impact on the ICER than screening sensitivity in a two-way sensitivity analysis.[193]

While the separate impact of either in- or excluding a **disutility related to the screening mammography** or a **disutility related to diagnostic work-up** on the ICER seems relatively mild,[180, 195] the impact of in- or excluding both disutilities could be more substantial, especially in screening strategies with shorter screening intervals (e.g. annual screening or a combination of biennial and annual screening).[183] As mentioned previously, I-CHER (2024) does not take into account a disutility for the mammography and diagnostic work-up, but instead opted to assign a disutility to false-positive cases. The authors concluded that (not) taking into account this specific QoL loss due to a false-positive result had only a minor effect on the cost-utility of breast cancer screening.[200]

Finally, while only varied by I-CHER (2024), the **time horizon** over which the costs and health effects were modelled was shown to be of importance. The sensitivity analyses showed substantially higher ICERs for the shorter time horizons when comparing the current screening programme to stopping with screening and mixed results (i.e. higher ICER with time horizon of 20 years and lower ICER for 50 years compared to ICER of lifelong time horizon) for extending screening to older and younger women compared to the current screening programme.[200]

4.4 Discussion & conclusions

While the majority of studies conducted a **cost-utility analysis**, four studies unfortunately solely performed cost-effectiveness analyses and therefore did not take into account the impact of breast cancer screening, treatment as well as screening-related harms on QoL. Including these impacts on QoL in the analyses can be considered a major strength. In contrast, for this specific intervention, only evaluating breast cancer screening in terms of length of life might lead to a misrepresentation of its true impact as, among others, the consequences of screening-related harms (such as false positives) will not be fully taken into account in the analyses. The underlying QoL studies used by the various evaluations are summarized in Table 9. It should be noted that also these studies are characterized by methodological strengths and shortcomings of their own. For example, a substantial number of studies estimated QoL values by asking the general public or clinical/public health experts to evaluate QoL associated with hypothetical descriptions of health states, events, treatments, and harms instead of asking breast cancer patients to evaluate the QoL they currently experience in their specific health state. With regard to the QoL elicitation methods, indirect methods (primarily the EQ-5D questionnaire) were used by multiple studies but several also relied on direct QoL elicitation methods, such as the Visual Analogue Scale (VAS) as well as the Time Trade-Off (TTO) and Standard Gamble (SG) methods. The studies using both the EQ-5D and one of the latter alternatives show that resulting QoL estimates for the same health state can differ greatly depending on which method was used. For example, Lidgren et al. (2007) found a utility of 0.685 for metastatic breast cancer using the EQ-5D and 0.820 when using the TTO.[204] Direct methods like TTO or SG are generally known to produce deviating utility estimates for a variety of reasons (e.g. risk aversion, time preference, coping, etc.). Finally, only a few QoL studies chose to measure stage-specific QoL values. Most studies focused instead on estimating QoL for specific treatments (e.g. radiotherapy, chemotherapy, etc.) or health states not linked to a specific classification system (e.g. inoperable locally advanced breast cancer). Of specific interest, due to the highly similar setting (Flanders, Belgium) of this report, are the QoL estimates used by ICHER (2024).[200] However, the (dis)utilities used by the authors could not be retrieved in the reported references and the authors do not disclose how these values were transformed to fit their model.

In line with the systematic review of Schiller-Frühwirth et al. (2017b), this update shows that all included economic evaluations opted for a stage shift approach to model **screening effectiveness**.[176] Remarkably, instead of using the TNM stage classification, numerous studies chose to use tumour size alone as a proxy for stage. In addition, transparency in methodology, input parameters and sources to model screening effectiveness (the stage shift) was in general relatively poor; an observation also made previously in the systematic reviews of Koleva-Kolarova et al. (2015) and Mandrik et al. (2019) which is unfortunately still valid.[188, 198] None of the evaluations was purely based on a RCT or RCT evidence, meaning other sources had to be used to model screening effectiveness. In general, two approaches could be distinguished: (1) Use of registry-based stage distributions and (2) Simulation of tumour growth or size at various ages combined with tumour detection probability by screening. In the former approach, studies examining the cost-effectiveness of an existing screening programme relied on the stage distribution before the screening programme had started to mimic the situation without screening (comparator). While this approach is pragmatic and transparent, it requires the assumption that today's stage distribution without screening would be the same as the stage distribution back in the seventies and eighties, which is susceptible to criticism. Those studies examining extensions of screening programmes in terms of age eligibility criteria employing the first approach generally used the stage distributions of the closest screened age (group) to mimic the stage distribution in the age group to which screening is extended (intervention). The assumption that age groups eligible for an extended screening programme have the same stage distribution as the nearest screened age group becomes increasingly difficult to justify with an increasing length of the extension. For example, when the stopping age is extended with one year (e.g. from 69 to 70), it is highly plausible that the stage distribution with screening of the 70 year olds would be highly similar to the one of the 69 year olds. In contrast, when the stopping age is extended with a longer period of time (e.g. from 69 years to 80), it is far less plausible that the 80 year olds would have the same stage distribution with screening as the 69 year olds and the assumption therefore becomes harder to defend. The studies

employing the second approach generally simulated tumour growth or size at various ages as well as tumour detection by screening based on screening sensitivity and specificity parameters or a detection probability function compared to detection without screening. Within this approach, multiple studies used some level of RCT evidence as inputs for their respective models. For example, the MISCAN model primarily relies on evidence originating from the Swedish breast cancer screening trials. Unfortunately, especially this second approach was characterized by a lack of transparency in methodology, input parameters and sources used to model screening effectiveness. This limitation makes it very difficult to fully assess the validity and completeness of the models.

Irrespective of which modelling approach was used, it is important to correct for **screening-related harms** and to transparently report how these were taken into account in the models. Unfortunately, a substantial number of included studies did not meet these requirements. A lack of transparency was primarily present in the way models handled false positives and overdiagnosis. Only half of the evaluations reported the number of **false positives** resulting from the examined screening strategies and it was unclear in the majority of evaluations which costs were assigned to false positives. Most studies seem to assign all costs up to diagnostic work-up to false positives. However, as discussed previously in the clinical evidence review, some positive screenings results are only discovered to be false during surgery, implying that part of the treatment costs should also be assigned to these cases. Unfortunately, of those studies with transparent reporting, this seems to be the case in only one evaluation. With regard to QoL, one study assigned a specific disutility to false-positive cases while the other evaluations generally assigned the disutility related to the screening mammography and/or the disutility associated with diagnostic work-up to these cases. The QoL literature clearly indicates a significant difference between the QoL of true negative and false-positive cases (see Section 4.2.3), meaning that assigning no additional disutility to false positives on top of the disutility of the screening mammography (also assigned to true negatives) seems not to be the correct approach. As most of the false positives are identified in the diagnostic work-up phase, applying the according disutility seems appropriate. However, as a minority of the false positives are identified during surgery, an additional QoL loss for this group could be justified.

In their systematic review, Mühlberger et al. (2021) observed that only half of the screening studies modelled **overdiagnosis**, often without reporting their magnitude.[182] This observation is partially in line with the included evaluations in this systematic review. While the majority of the included studies does claim taking into account overdiagnosis, only about half of the studies reported the magnitude of overdiagnosis in their models either as an input or output with considerable variation between them, likely stemming from the uncertainty around the true overdiagnosis rate in the clinical literature. From a methodological point of view, overdiagnosed cases cannot benefit from the screening in any way and should be assigned the costs they bring forward (i.e. screening, treatment, follow-up). Unfortunately, there is a considerable lack of transparency in how overdiagnosis was handled methodologically in the evaluations, for example in which stage(s) these cases were diagnosed and on which costs were assigned to overdiagnosed cases. With the exception of two studies, all other studies with transparent reporting make the assumption in the base case that overdiagnosed cases are solely DCIS. In a scenario analysis, one evaluation showed that in case overdiagnosis would not be limited to DCIS the impact on the ICER could be highly substantial (see 4.3.2). With regard to QoL, one could argue that overdiagnosed cases should be assigned an additional disutility resulting from the unnecessary treatment and the associated consequences (e.g. complications, anxiety and distress) related to unnecessarily labelling someone as a cancer patient. Due to a lack of transparency in methodology, it was unclear whether the studies included in this review have followed this line of reasoning or not. Modelling this disutility poses challenges and adds uncertainty due to the lack of evidence on magnitude and duration of the disutility. In their systematic review on the impact of breast cancer screening on QoL, Li et al. (2019) did not identify a single QoL study attempting to examine the latter.[205] While considerable uncertainty remains present in the correct overdiagnosis rate and whether a QoL adjustment should be incorporated, overdiagnosis likely has a substantial impact on the cost-effectiveness of breast cancer screening. This further illustrates the importance of including the most reliable overdiagnosis rate in an evaluation of breast cancer screening, supplemented with additional scenarios (potentially also including a disutility) to handle this uncertainty.

While in general the included **costs** and their sources were reported transparently, costs related to inviting the screen-eligible population and especially the costs associated with the organization of the screening programme were insufficiently taken into account in most studies. Many studies downsized the costs of screening to solely the cost of a mammography, which underestimates the screening costs and subsequently leads to an overly optimistic, but incorrect estimation of the cost-effectiveness of breast cancer screening. In addition, many studies implicitly or explicitly assumed an unrealistic **participation rate** of 100%. While full participation might be included as a scenario, for example to show the programme's full potential or impact in terms of cost-effectiveness, a realistic participation rate should be applied in the base case to show a more reliable result. In addition, in reality, participation rates vary with age. Applying the average participation rate of the current screening programme to the groups to which the programme might be extended could lead to an overly optimistic estimation of the true participation rate in these groups. Hence the importance of presenting multiple scenarios ranging from low to high (up to full participation) in the sensitivity analyses and/or using age-specific participation rates when available. The results of these sensitivity analyses on the magnitude of the impact of participation rate on the ICER are mixed, but rather counter-intuitively most studies find that increasing the participation rate has an unfavourable effect on the cost-effectiveness of breast cancer screening. One would expect the screening programme to become more cost-effective when more people participate as the fixed costs can be divided over more participants. Unfortunately, none of the studies offer a plausible explanation for these findings. One potential explanation could be that the fixed costs of the screening programme were not or incorrectly taken into account in the models. A second explanation could be that the authors assumed that the characteristics of the participants at a low participation rate are inherently different from those at a high participation rate, and subsequently incur different costs and gain different benefits from the screening.

Finally, three studies taking a societal perspective also included **indirect costs**. All three studies incorporate the loss in productivity due to premature death from breast cancer, which favors screening over no screening in terms of cost-effectiveness. However, none of the studies take into account the productivity loss associated with participating in the screening programme and overdiagnosis, which would have an unfavourable impact on the cost-effectiveness. While the time needed to participate seems limited on an individual basis (e.g. up to 4 hours), it does apply to a large population (all screening participants). As shown by a previously published version of the evaluation by ICHER (2024), the impact of including the productivity costs associated with participating in the screening programme (4 hours) is substantial with an increase in the ICER from €23 063/QALY gained to over €40 000/QALY gained.[ttt252] This shows the importance of incorporating the full impact on productivity when taking a societal perspective, at least in a scenario analysis, to avoid overly optimistic estimations of the cost-effectiveness.

[ttt] The publication of ICHER (2015) was not included in this review,[252] as we opted to include its update (2024).[200]

5 BELGIAN DATA

5.1 Data sources and data protection

To carry out this study, three databases were provided by the Belgian Cancer Registry (BCR); they are briefly described in the following sections. The data transfer was authorised by the Information Security Committee, Social Security and Health Chamber in December 2024.[253] Data pseudonymization by the eHealth platform as a trusted third-party ensured the individuals' data privacy before data delivery and enabled the linkage between the three datasets.

In addition to the usual data privacy precautions (e.g. pseudonymization of individual identifiers by the eHealth Platform, publication of anonymous results, duty of confidentiality of researchers), the BCR replaced all dates in the databases by dates relating to a random date (between January 1, 1900 and December 31, 2001). The date of completeness of mortality registration was not disclosed, in order to prevent the re-identification of absolute calendar dates from relative dates.

Last, the Flemish Institute for Care Quality (Vlaams Instituut voor Kwaliteit van Zorg, VIKZ) performed an identification risk analysis, including a small cell risk analysis, on 30 April 2025. They concluded that no security restriction was indicated.

5.1.1 BCR – IMA – AIM database

The main data source for this study was the nationwide **BCR database**, covering over 98% of all incident cancer diagnoses in Belgium.[254] This high level of completeness is reached thanks to mandatory registration for hospitals and laboratories for anatomical pathology.[255, 256]

The BCR linked its database with:

- **Health insurance data**, obtained via the Intermutualistic Agency (IMA – AIM). The database[uuu] comprises information on (cancer-related) diagnostic and therapeutic procedures as well as pharmaceuticals provided from one year before until five years after cancer incidence, and which are reimbursed by the national health insurance. Inherently, only billed procedures were taken into account in the analyses. Of note, procedures performed within the context of clinical trials are not (routinely) billed and hence not included in the IMA – AIM data.
- **Vital status data**, obtained via the Crossroads Bank for Social Security ('Kruispuntbank van de Sociale Zekerheid' – 'Banque Carrefour de la Sécurité Sociale', KSZ-BCSS).

The linkage of the above databases was based on the patients' unique social security number (INSZ – NISS), and has been approved by the Sector Committee of Social Security and of Health (Health Section) of the Belgian Privacy Commission (Sectoraal comité van de Sociale Zekerheid en van de Gezondheid, afdeling gezondheid/Comité sectoriel de la Sécurité Sociale et de la Santé, section santé).[257]

At the time of data extraction, tumour data were available for women diagnosed between 2004 and 2022, health insurance data were available from 2014 until 30th June 2024 and the vital status from 2004 until an undisclosed date in 2025 (for data protection reasons).

5.1.2 Screening database (SD database)

The three regional screening organisations (CvKO, BruPrev, CCRef, see also Chapter 2) transmit on a regular basis screening-related data to the BCR. For breast cancer screening, these data include the invitation date, the screening date, the result of the screening (after possible necessary follow-up examinations) and the possible exclusion periods to being invited to the organised screening

[uuu] In Belgium, the codes of the fee schedule are called 'nomenclature codes'.

programme (e.g. two years after most recent screening mammography[vvv], ten years after a breast cancer diagnosis, permanently after a bilateral mastectomy). We refer to Chapter 2 for more information on the invitation criteria and process in each region. Data were available for the study from screening year 2015 until invitation year 2023 (screenings performed in 2024 after an invitation in 2023 were not available yet).

5.1.3 Medical acts in screening population database (DD database)

Since 2002, the BCR receives from the IMA – AIM data on a limited list of diagnostic and therapeutic procedures (i.e. medical imaging[www], biopsy, breast surgery; see Appendix 24.2) that are necessary for an efficient organisation, evaluation and monitoring of the organised breast cancer screening programmes. The transfer of these codes allows the analysis of additional diagnostic procedures following an abnormal mammography, the establishment of exclusion lists for invitations to the screening programmes, and to identify eligible women who were never screened. Last but not least, these data enable the BCR to support the continuous improvement of the organised screening programme by calculating performance and quality indicators.

For the present study, the DD data were available from 2004 up to the first quarter of 2024. Of note, the last complete year was 2021. The variable "age" was available per age group of five years.

More details on the SD and DD datasets can be found in the three deliberations of the Information Security Committee on the exchange of personal data relating to health between the each of the regional screening organisation, the Cancer Registry Foundation, the sickness funds and the Intermutualistic Agency in the context of organised breast cancer screening.[258-260]

5.2 Methods

5.2.1 The study cohorts

A primary data extraction from the BCR database included all breast cancers (i.e. invasive breast tumours (ICD-10: C50) or in situ breast tumours (ICD-10: D05)) diagnosed between 1 January 2004[xxx] and 31 December 2022, in women aged 40 years or more with an official residence in Belgium at the time of diagnosis (see Appendix 18). In this first step 222 493 breast tumours, corresponding to 208 162 unique patients, were identified.

5.2.1.1 The cohort selected for the survival analyses

The extraction was further **refined** to enhance data quality (see flow chart in Appendix 18). First, tumours with unknown "combined stage"[yyy] ("X"; n=18 313, 8.2%) or "is"[zzz] (n=41, <0.1%) were excluded. In the few cases where two breast tumours were registered in the same patient on the same

---

vvv In the Flemish and Walloon region, women with a (diagnostic or opportunistic) mammography outside the regional screening programmes are also not invited for two years after that mammography.

www The invoicing codes for imaging are available for age (at time of performance) group 40-72 years, those for biopsies, punctures and tumourectomy for age group 48-72 years and those for mastectomy for all ages.

xxx The BCR database is considered complete at the national level from the year 2004 onwards.

yyy In order to overcome to some extent missing stage data, BCR developed the variable "combined stage", where known pathological stage prevails over known clinical stage, except when there is clinical proof of distant metastasis (cM1). When only the clinical stage is known, this stage determines the combined stage. Obviously, when neither the pathological stage nor the clinical stage are known, the combined stage is unknown too. The TNM classification was performed according to the 6th (incidence years 2003-2009), 7th (incidence years 2010-2016) and 8th edition (as from incidence year 2017) of the TNM Classification of Malignant Tumours, International Union Against Cancer.[261, 263]

zzz "Is" corresponds to an unconfirmed combined stage registered in the BCR database; it is kept until further information becomes available.

date (n=3711, 1.7%), the tumour with the highest combined stage was kept. In cases where the combined stage was identical, the tumour with the most complete data was selected. As women who are diagnosed with breast cancer are not invited for the population screening during the ten subsequent years (as they are considered being offered tailored follow-up), 7136 breast cancers (3.6% of the remaining tumours) preceded by another invasive breast cancer within the last ten years were also excluded. For this purpose, we used an ad hoc variable created and added by BCR to the extracted dataset based on previously available data (including cancers diagnosed before the age of 40 years). At this stage, **192 809 breast tumours**, corresponding to **191 439 unique patients** aged 40 years or more who were diagnosed between 2004 and 2022 formed the **cohort for the survival analyses presented in this chapter**.

In the main survival analysis used for the economic model presented in Chapter 6 and for Figure 34, all women who were diagnosed between 2004-2022 with (non-invasive or invasive) breast cancer were included, irrespective of individual breast cancer risk. To create a cohort that resembled as closely as possible the cohort which is invited for the current biennial screening programmes, only women between 50 and 69 years old who did not have breast cancer during the preceding ten years were included (cf. supra). This cohort (n = 98 025 for 2004-2022 and n = 79 134 for 2016-2022) is referred to as the "all risk group" in the remainder of the report.

5.2.1.2 The cohorts selected for the cost calculations (based on administrative data)

The cohort of breast cancer patients aged 50-69 years

Cost inputs for the economic model were calculated on a more restricted cohort (see Appendix 18). Given the availability of health insurance (IMA – AIM) data (i.e. 1 January 2014 - 2024 Q2), the costs were calculated for (invasive and non-invasive) breast tumours diagnosed between 1 January 2018 and 31 December 2022. Moreover, as IMA – AIM data do not include the specific indication for which procedures are performed, women who had any other invasive tumour in the five years prior to or following the diagnosis of breast cancer, were excluded from the cost analyses, in order to maximally ensure that recorded procedures were indeed performed to diagnose or treat the breast tumour under study and not another malignancy.[aaaa] Finally, patients for whom no IMA – AIM data were available, were removed. The selection for the cost calculations comprised **26 572 breast tumours**, corresponding to **26 572 unique patients between 50 and 69 years old at the time of diagnosis.**

The cohort with true-positive screening results

In the SD database (see section 5.1.2), **true-positive results** are defined as abnormal screening mammography findings that are followed by a cancer diagnosis within a period of 24 months. Women with suspicious screening (positive) but negative work-up who developed an interval cancer are recorded as true positive, but they were not included in the true-positive cohort of the present project to calculate costs related to a true-positive diagnosis. This was done to avoid costs related to the diagnosis of the interval cancer would be attributed to the costs of a true-positive finding. After linkage between screening data and tumour data, the "true-positive cohort" comprised **6704 women** aged 50 to 69 years old, who were diagnosed with breast cancer between 2018 and 2022 (i.e. the five most recent years for which linked data were available).

Given the high number of small cells (no or too few patients diagnosed per stage, incidence year, and calculation period), we decided to use the costs of the 26 572 unique patients aged 50 to 69 years at the time of diagnosis, without restricting the analysis to true-positive results. When differences were observed, mean costs tended to be higher in the overall group than in the true-positive group. These differences were also more frequent in higher stages. Therefore, using the mean costs from the overall group rather than those from the true-positive group as inputs for the economic model supports the

[aaaa] As an example, it is not possible to differentiate radiotherapy for breast cancer from radiotherapy for another cancer.

choice of a conservative model (i.e., assumptions that are optimistic in favour of organized breast cancer screening).

The cohort with false-positive screening results

Costs were also calculated for women who had a false-positive screening result (cf. section 5.2.3.2). They were calculated based on a **cohort of 19 158 women** who received between 2020 and 2022 a false-positive screening result.

The cohort with negative screening results

No specific costs beside the screening cost (e.g. additional work-up) were calculated for women with a negative screening result, nor for women who were diagnosed with an interval cancer (regardless of whether the preceding screening result was negative or abnormal).

5.2.2 Survival analyses

The date of diagnosis of the first tumour was used as the start of survival follow-up. Survival probability was examined descriptively using Kaplan Meier curves, stratified by age, WHO performance status, combined tumour stage and year, all of which were determined at the time of diagnosis.

Besides the "all risk group", a secondary group, called "average risk group", was established through the linkage with the screening and the screening follow-up databases (SD and DD; see section 5.1.2). The latter is restricted to women having participated in the population breast cancer screening in Belgium. This average risk group comprises 14 476 unique patients, diagnosed with breast cancer between 2016 and 2022. The linkage process allowed for a time window of no more than two years between screening date and incidence date, which excluded very few cases. To assess whether survival estimates derived from the "all risk group" (n = 79 134 between 2016 and 2022) were comparable to those of the "average risk group", survival curves were compared side by side for women aged 50–69 years diagnosed from 2016 onwards (n = 14 476). In addition, an exploratory comparison of survival during the first ten years after diagnosis was performed for all women diagnosed between 2004 and 2022 (and followed to somewhere in 2025; see section 5.1.1), allowing assessment of potential changes in survival over time when compared to the group of women diagnosed between 2016 and 2022.

5.2.3 Costs of breast cancer care and the management of false-positive cases

5.2.3.1 Costs of breast cancer care

Costs categories

As the health economic evaluation will be performed from the healthcare payers' perspective, costs were defined as the amount reimbursed by the national health insurance as well as the patient's share for diagnostic, therapeutic and follow-up procedures identified through billing codes and Anatomical Therapeutic Chemical (ATC) product codes in the IMA – AIM database. In an alternative scenario, the patient's supplements were also included.

We delineated **seven different cost categories**: costs linked to the organised screening programme, the diagnostic and follow-up procedures, the surgical interventions, radiotherapy, systemic therapy, palliative care and the miscellaneous group. For the identification of the billing (nomenclature) codes and ATC codes to be included in each group, various sources were consulted (e.g. recent KCE report on lung cancer screening, recent KCE report on the quality of breast cancer care).[264, 265] Additionally, lists of billing and ATC codes frequently billed in the days preceding hospital admission for breast cancer related surgery, radiotherapy or systemic therapy allowed to expand the lists with other codes frequently billed in breast cancer care (e.g. blood tests, preoperative consultations or chemotherapy supportive therapy). The lists were then submitted to clinical experts (e.g. medical oncologists,

radiation oncologists, radiologists, surgeons) for amendment, correction and validation. Tables listing the billing and ATC codes included in the different cost categories can be found in Appendix 24.1.

Accounting for potential uncertainties in the exact date of incidence or the exact date of certain health care interventions, the **time window** considered for eligibility was seven days before up to five years after incidence date, except for the category 'diagnostic and follow-up costs'. The time window considered to calculate the latter cost group span from three months before the date of cancer incidence until one week after the last therapeutic service, i.e., the last date of chemotherapy session, targeted or immunotherapy administration, radiotherapy session or the last discharge after surgical hospitalisation. Extending diagnostic and follow up costs beyond this limit was avoided because many of the corresponding billing codes (e.g. CT scans) lack disease specificity and may capture healthcare utilisation unrelated to cancer once cancer-related therapeutic services had ceased. For some subcategories of costs (e.g. preoperative costs, radiotherapy preparation and supervision, systemic therapy supportive medication and supervision) dedicated time windows were specified, based on clinical expert consultation (cf. infra).

To avoid double counting, each billing/ATC code was allocated to a single, mutually exclusive category using a predefined prioritisation rule. Surgical interventions, radiotherapy, palliative care and miscellaneous (e.g. patient travel costs, see Appendix 24.1) were identified, followed by the reconstruction of surgical episodes to capture all associated costs. Subsequently, all costs related to systemic and supportive therapy were then assigned. The cost items for the category 'Diagnostics and follow-up' were finally derived from the remaining claims once the start and end dates of the different therapeutic modalities were determined. Likewise, as the costs of the organised screening programme were estimated based on other sources than the administrative databases, the invoicing codes related to the screening mammographies were excluded from the calculation to avoid double counting.

Specific methodological choices applied for each cost category are presented in Appendix 19.

Cost parameters for the economic evaluation

The economic evaluation was supported by both published evidence and real-world data. By linking the different data sources, we identified and quantified the different categories of costs associated with the management of breast cancers. **Mean costs and standard deviations** were calculated for each cost category and used to populate the economic model.

To capture the temporal distribution of healthcare expenditures with appropriate granularity, costs were aggregated over a series of predefined, mutually exclusive intervals following the incidence date. The end of the first interval was set at one month after incidence date and its start depended on the type of costs: with the exception of diagnostic costs being included from 3 months before incidence date, all other healthcare costs were included from 7 days prior to the incidence date onward. The successive time intervals were defined as follows: -7 days –31 days, 31–61 days, 61–92 days, 92–183 days, 183–274 days, 274–365 days, and subsequently annually up to five years. These cut-offs correspond to intervals of one month, one month, one month, three months, three months, three months, followed by four consecutive one-year periods, covering a total follow-up of five years. These cut-offs were chosen in function of the evolution of survival and costs over time. To calculate the mean cost, the total cost over a period of time should be divided by the number of patients. However, the number of patients in the sample in the beginning of the period is different from the number at the end of the period due to mortality. As a consequence, the total cost was divided by the number of patients alive halfway the time period. Costs for the category 'diagnostic and follow-up' of the first time interval were identified from 91 days prior to incidence date until 31 days post-incidence.

The mean cost for each cost category and each time interval T was computed using the following formula:

$$\text{Mean}_T = \frac{\Sigma\text{Costs}_T}{N_{T_{1/2}}} \qquad \text{where} N_{T_{1/2}} \text{is the number of patients alive halfway through time interval T}$$

For the ten time intervals, the value of $N_{T1/2}$ corresponded to the number of patients alive at the midpoint of each interval: day 16, day 46, day 77, day 138, day 229, day 320, day 548, day 913, day 1278, and day 1643. For diagnostic costs, two means were calculated for the first interval: the first covering the period up to and including the incidence date, and the second covering the month following that date. For the first mean of diagnostic costs, $N_{T1/2}$ was then defined as the number of patients alive on day 0 or at incidence date, i.e., the full cohort at baseline. For all other categories of costs the 7 days preceding the incidence days belonged to the first interval.

This choice of denominator accounts for attrition due to mortality occurring within each interval, ensuring that cost estimates reflect the population at risk at mid-interval.

The standard deviation for each cost category for each interval *T* was calculated as:

$$\mathrm{SD}_T = \sqrt{\frac{\Sigma\left(\mathrm{Costs}_T - \mathrm{Mean}_T\right)^2}{\mathrm{N}_{T_0} - 1}} \quad \text{where} N_{T_0} \text{is the total number of patients alive at the start of time interval T}$$

The most recent costs available were used. Since IMA – AIM data were available until 30th June 2024, the most recent year with complete registration of healthcare interventions was 2023. Consequently, the parameters for the first year of costs were calculated on women diagnosed in 2022, those for the second year of costs were calculated on women diagnosed in 2021, etc. Only women diagnosed in 2018 had complete five-year costs data, their costs were therefore used to calculate the parameters for the fifth year of costs.

5.2.3.2 Costs in case of false-positive screening results

For women with a false-positive screening result, i.e. in whom no breast cancer was diagnosed and hence no BCR – IMA-AIM data were available, the database of medical acts in screening population (DD database, see section 5.1.3) was used to determine the costs related to the false-positive screening result[bbbb].

Importantly, unlike the IMA-AIM database, the DD database does not contain the fees and the amounts reimbursed, but rather the dates (relative to the screening date) on which the procedures were performed. To calculate the costs, the frequency with which codes were invoiced within three months after the screening date, was multiplied by the tariffs of 2026.

The number of screened women who underwent surgery following a false-positive screening result was very small (242/19 158; 1.26%). Therefore, we did not reconstruct the costs of the entire surgical episode, unlike cancer cases, as surgical costs were expected to have a negligible impact on overall costs for false-positive findings.

5.2.4 Statistical and data processing software

The analyses were performed using SAS software™ (SAS/BASE, SAS/STAT and SAS/GRAPH), version 9.4 M5 for Microsoft Windows and r version 4.5.2 with addition of packages ‘survival’ and ‘survminer’.[266, 267]

---

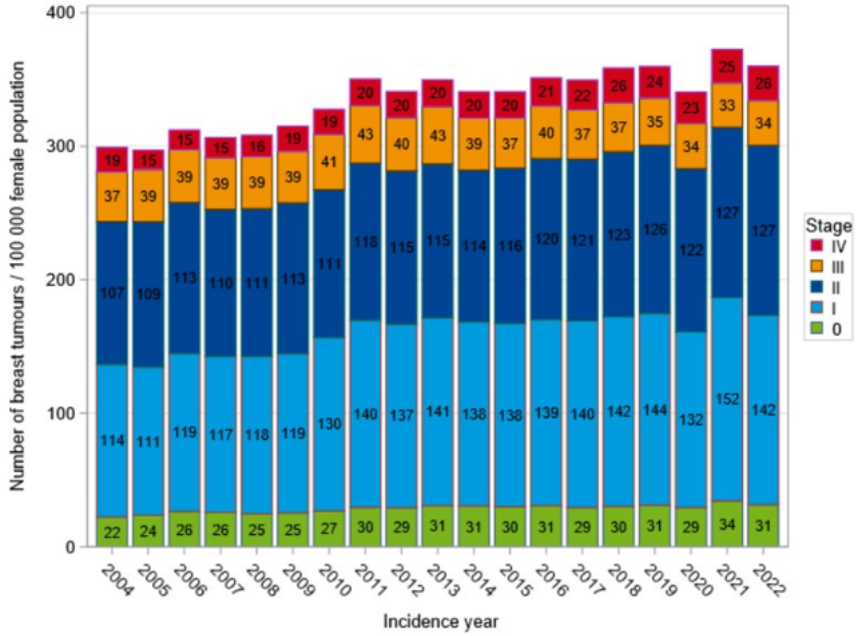
bbbb Of note, only screening results recorded on a screening date corresponding to a screening date in the DD dataset were retained to investigate the health interventions performed in women who were screened. Some screening results could not be linked to records in the DD dataset. Presumably it relates to women who are not affiliated with any of the seven sickness funds (which transfer their data to IMA - AIM) and for whom no invoicing data are thus available, but who reside in Belgium and are invited to participate in the organised screening programme.

5.3 Characteristics of the study cohort

5.3.1 Evolution of breast cancer incidence over 2004-2022

In the remainder of this chapter, the study cohort selected for the survival analysis (i.e. 191 439 women aged forty years or more at time of breast cancer diagnosis) is described (see also Appendix 18). As Figure 30 shows, the absolute number of breast cancers diagnosed annually, has increased from 2004 to 2022. A dip is observed for 2020, reflecting the temporary suspension of organised cancer screening programmes in Belgium during the COVID 19 pandemic.[268]

Figure 30 – Evolution of number of breast tumours per 100 000 women aged ≥ 40 years per combined stage (2004–2022)



5.3.2 Patient characteristics

The number of women aged 40 or older and diagnosed with breast cancer in 2015-2022 (the eight most recent study years), is not evenly distributed among the various regions (Table 11). A substantially lower number can be noticed for Brussels (Capital region) as opposed to the other regions, which may (in part) be attributed to the age distribution of their respective populations, as these data are not age-standardised.

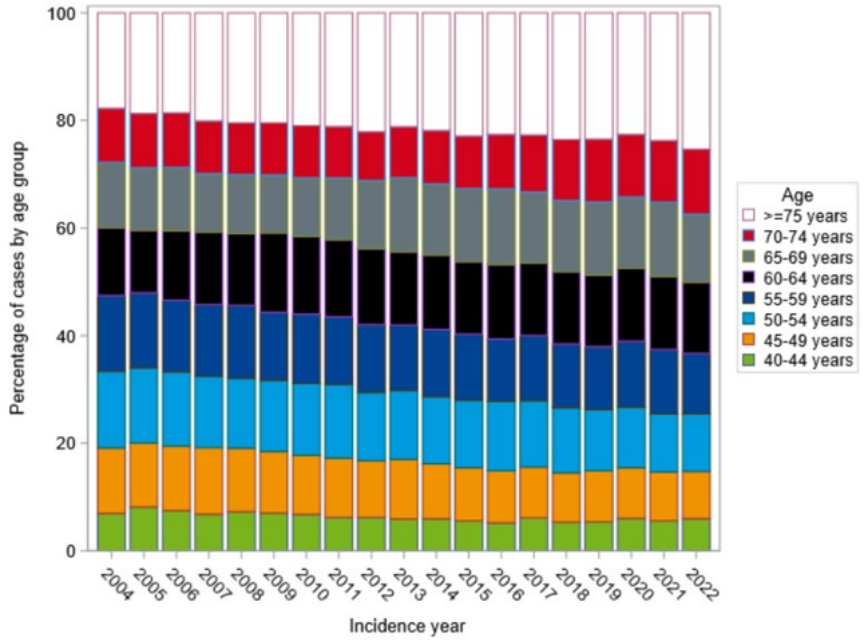
Table 11 – Unadjusted annual incidence of breast cancer per 100 000 women aged 40 years of more in the three Belgian Regions, 2015-2022

	2015	2016	2017	2018	2019	2020	2021	2022
Brussels (Capital region)	335	335	318	364	348	334	349	318
Flemish region	343	352	354	361	370	355	387	368
Walloon region	325	345	343	367	359	329	358	357
Belgium	337	348	348	363	364	345	375	360

Source of population characteristics *(by place of residence aged 40 years or more): Statbel (Directorate General Statistics – Statistics Belgium)*[cccc]

Between 2015 and 2022, the mean age at diagnosis was 64.2 years old (SD: 13). Women aged between 50 and 69 years account for about half of the cases with breast cancer (from 47.8% to 51.9%, Table 102 in Appendix 20), while the group older than 75 years old for about a quarter of cases. The data suggest a tendency towards an increasing proportion of patients in this oldest age group (Figure 31).

Figure 31 – Evolution of age distribution in women aged ≥ 40 years (2004–2022)



Over 90% of women with breast cancer had either a WHO performance status[dddd] equal to zero (i.e. no complaints or symptoms of disease, able to perform daily activities without restrictions) or symptomatic but completely ambulatory (WHO performance status 1). The proportion of women with WHO performance status equal to zero seems to increase somewhat over time (from 57.3% to 62.3% between 2015 and 2022; see Table 102 in Appendix 20). However, the absence of a consistent directional pattern across severity categories, combined with year-to-year fluctuation in missing performance status data, suggests caution when interpreting these trends.

[cccc] Downloaded from be.STAT

[dddd] WHO performance status was assessed at diagnosis.

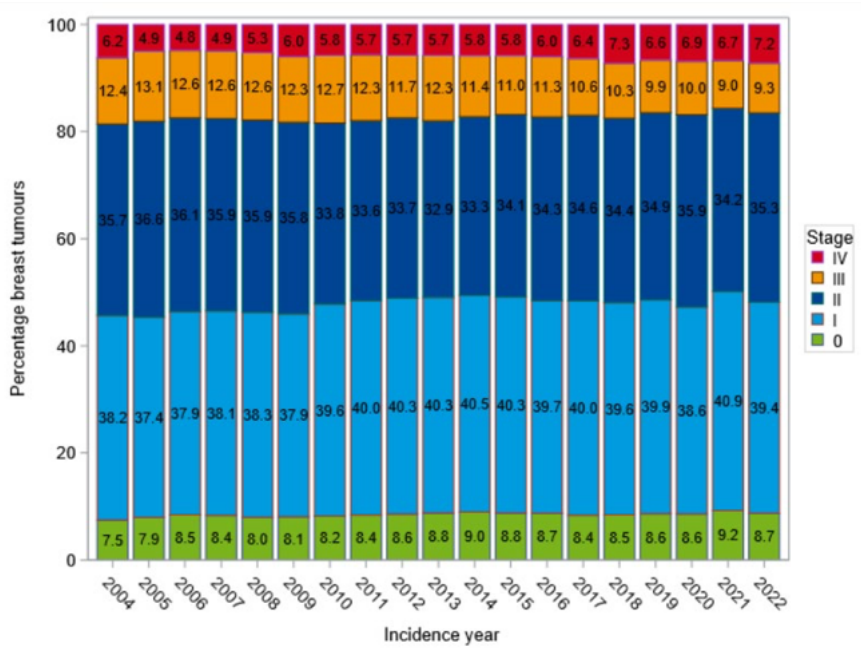
5.3.3 Tumour characteristics

In the period 2004-2022, the relative proportion of women diagnosed with non-invasive breast cancer (i.e. stage 0; based on the combined cancer stage[yyy] information) ranged between 7.5 and 9.2% (Figure 32). Most invasive tumours were diagnosed as stage I or II, together accounting for roughly three quarters of the cases. There were no major shifts in the distribution of stages since 2004.

Grade (differentiation) data are presented in Table 103 (Appendix 20). In 2015-2022, these data were missing for 26.4% - 37.2% of in situ breast cancers and for 4.9% - 7.0% of invasive breast cancers. The decreasing proportion of missing values over time suggest improvements in data recording practices. In the most recent year (2022), half of the invasive breast cancers were assigned grade II. By contrast, the majority of in situ tumours were classified as poorly differentiated (31.8%; Grade III).

The majority of in situ breast cancers are situated intraductally (DCIS; 76.4% - 75.7%; Table 103 in Appendix 20). For invasive breast cancers, about one third is situated in the upper-outer quadrant of the breast (32.5% - 35.4%). Of note, the proportion of invasive breast tumours assigned code C50.9 (i.e. malignancy without any specification of the site) has continued to decline from a third to a quarter of the cases over the same period.

Figure 32 – Evolution of combined breast cancer stage distribution in women aged ≥ 40 years (2004–2022)



5.3.4 Observed survival

Five years after the diagnosis of breast cancer, the observed survival probability for women 40 years or more diagnosed with breast cancer between 2004 and 2022 (see section 5.2.1.1), was 83.4% (95% CI: 83.2-83.6; Table 12) and dropped to 70.8% (95% CI: 70.6-71.0) after ten years and to 49.0% (95% CI: 48.5-49.4) after twenty years. Patients with non-invasive breast cancer consistently demonstrated higher survival across all follow-up intervals, with nearly three quarters (73.1; 95% CI: 71.8-74.6) of patients alive 20 years after diagnosis. In contrast, survival among patients with invasive disease, who

constituted the majority of the cohort, was markedly lower, with slightly less than half (46.8%; 95% CI: 46.3-47.2) of patients alive 20 years after diagnosis.

Table 12 – Unadjusted observed survival for patients ≥40 years, diagnosed with (invasive and non-invasive) breast cancer in 2004-2022

Unadjusted observed survival (%, 95% CI)						
Subgroup	**1-year**	**3-year**	**5-year**	**10-year**	**15-year**	**20-year**
Total N=191 438[eeee]	96.2 [96.2, 96.3] (N=184 060)	89.6 [89.4, 89.7] (N=163 576)	83.4 [83.2, 83.6] (N=132 990)	70.8 [70.6, 71.0] (N=74 266)	59.8 [59.5, 60.1] (N=2076)	49.0 [48.5, 49.4] (N=4970)
Non-invasive N=16 268	99.5 [99.3, 99.6] (N = 16 151)	98.0 [97.8, 98.2] (N = 15 161)	96.3 [96.0, 96.6] (N = 12 943)	89.9 [89.4, 90.5] (N = 7 899)	81.8 [81.0, 82.7] (N = 3 578)	73.1 [71.8, 74.6] (N = 563)
Invasive N=175 170	95.9 [95.9, 96.0] (N = 167 909)	88.8 [88.6, 88.9] (N = 148 415)	82.2 [82.0, 82.4] (N = 120 047)	69.0 [68.8, 69.3] (N = 66 367)	57.8 [57.5, 58.1] (N = 28 498)	46.8 [46.3, 47.2] (N = 4 407)

***Notes**: N in each cell denotes the number of patients with invasive breast cancer at risk at the specified follow-up time, defined as those alive and not censored before that time. Censoring can be due to either loss to follow-up or too recent to have long enough follow-up.*

Observed survival is very similar in the two youngest age groups (40-44 and 45-49 years), whereas a clear decline in observed survival is observed with increasing age (Figure 33, A; Table 104 in Appendix 21). Similarly, better WHO performance status at diagnosis is consistently associated with higher survival probabilities, with the largest difference observed between performance status categories 1 and 2; the two most severe categories show very similar survival patterns (Figure 33, B; Table 105 in Appendix 21). Observed survival by combined stage demonstrates also a clear ordering of the survival curves, with more advanced stages associated with poorer outcomes (Figure 33, C). Finally, across the incidence years 2004-2022, no clear differences in survival probabilities are observed (Figure 33, D). However, Figure 90 in Appendix 21, which stratifies these patterns by stage, shows that this overall stability masks important variation by combined stage. Up until stage II the observed survival is largely unaffected by the year of diagnosis, but a clear shift in observed survival across time can be noticed for stages III and IV. As stages III and IV only account for 17.4% of women with breast cancer and therefore have a minor impact on the curves in figure 33D.

Figure 34 presents a side-by-side comparison of survival probabilities stratified by age and combined tumour stage across three progressively restricted sub-cohorts: from the full survival cohort shown in the leftmost column (described in section 5.2.1.1), over the cohort restricted to women diagnosed with breast cancer from 2016 onwards (second column) to the most narrowly defined cohort, i.e. the average risk cohort (described in section 5.2.1.1), shown in the rightmost column. In contrast to the data presented above, Figure 34 is limited to women aged 50-69 to make a correct comparison.

For some subgroups of age and combined stage, differences can be noticed between the first and second column (i.e. when excluding the subgroup that was diagnosed between 2004 and 2014). In addition, survival probabilities are better for the “average risk cohort” compared to the 2016-2022 “all risk cohort”, especially for women with stage III and IV. This is even more pronounced when comparing to the 2004-2022 “all risk cohort”. Furthermore, the stage-specific survival curves are more widely separated in the all-risk population than in the average-risk population, indicating larger differences in survival probability between stages when not restricted to “average risk”. Again, for some strata of age

[eeee] One woman was excluded because of a data anomaly (last known vital status precedes breast cancer incidence date).

and combined stage, this difference is even more pronounced when compared with the 2004-2022 “all risk cohort”.

Figure 33 – Unadjusted observed survival of patients with breast cancer, by age at diagnosis (A), WHO performance status (B), combined stage (C) and year of diagnosis (D)

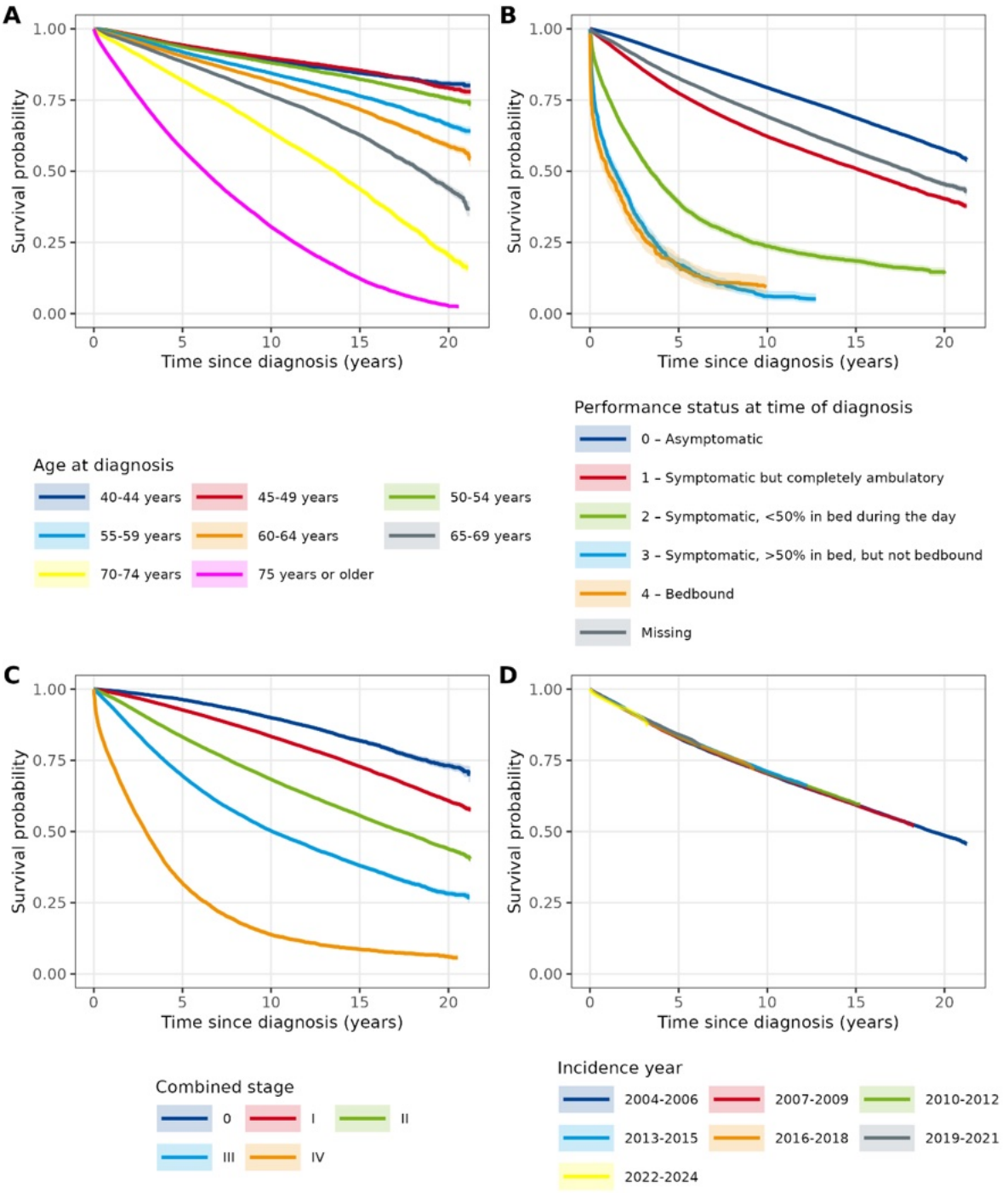
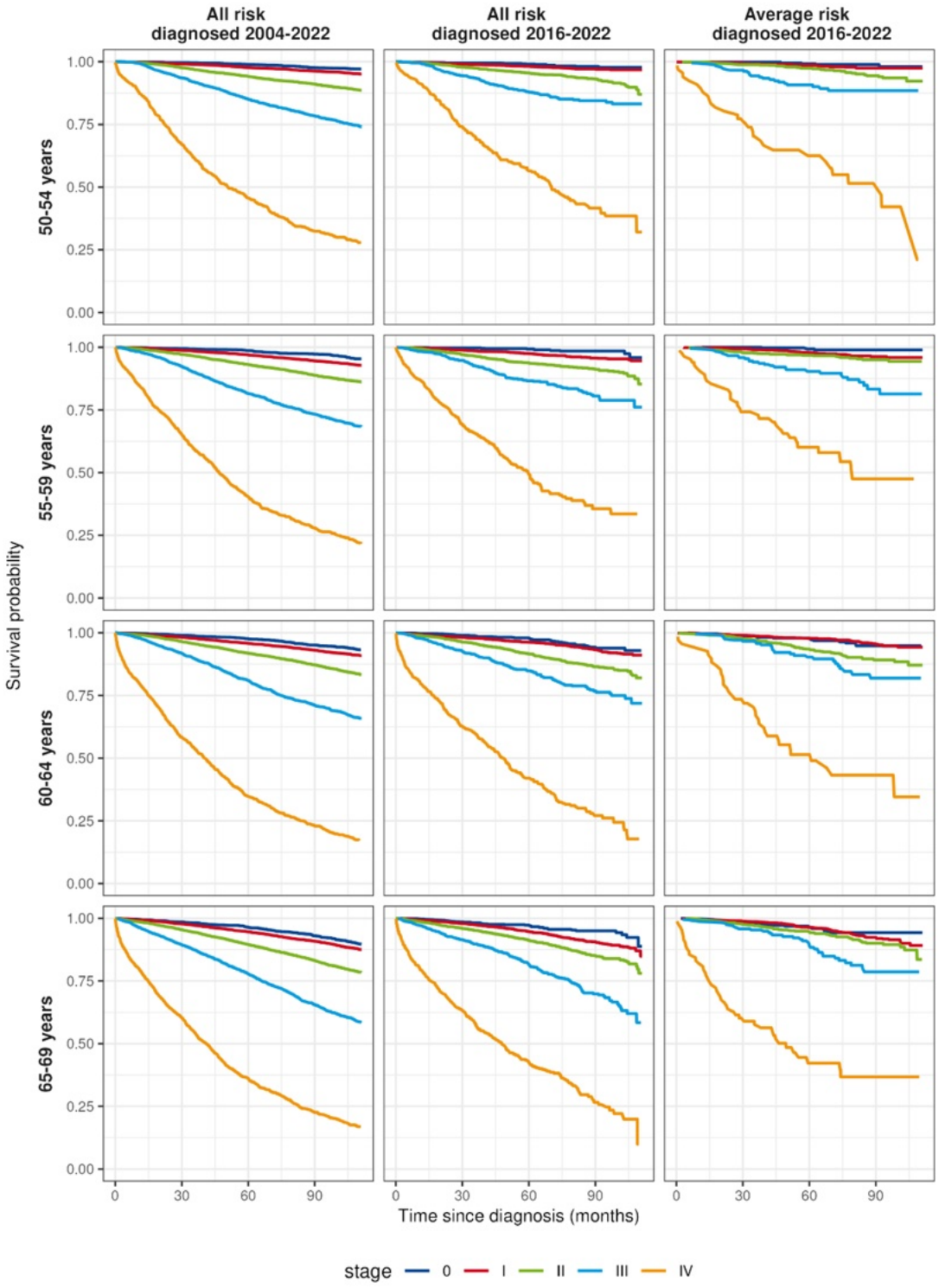


Figure 34 – Side by side comparison of stage-specific survival probability by risk group, timeframe and age



6 DE NOVO HYPOTHETICAL ECONOMIC EVALUATION OF BREAST CANCER SCREENING

Key points

The development of an evidence-based economic evaluation was hampered by the available evidence, which consists of outdated RCTs, the majority of which are associated with a high risk of bias.

Availability of Belgium-specific real-world data from the Belgian Cancer Registry (BCR), the InterMutualistic Agency (IMA), and the Crossroads Bank for Social Security enables us to construct a snapshot of the current situation in Belgium, reflecting the number of women invited and attending the population breast cancer screening, positive and negative screening results, true and false positives, the number of interval cancers, stages of screen-detected and interval cancers by age category, age- and stage-specific (long-term) survival, and stage-specific treatment costs subdivided in various cost categories.

Based on a "what-if" analysis, a hypothetical economic evaluation was conducted in which the current situation of population-based screening among asymptomatic women aged 50–69 years without an increased risk of breast cancer (intervention) was transformed into a scenario without screening (comparator). This evaluation made use of the stage shift, i.e. the potential of screening to detect cancers at an earlier stage.

The most influential variables are the risk ratio for being diagnosed with advanced breast cancer (i.e. stage II, III, or IV) when attending the organised screening programme compared to not attending and the proportion of overdiagnosis (among women who attended screening).

Under a set of optimistic assumptions, in combination with a risk ratio for being diagnosed with advanced breast cancer of 0.66 (based on five RCTs with low and high risk of selection bias) and an overdiagnosis of 20.7%, the average ICER for screening is approximately €13 700 per QALY gained. However, when a risk ratio for being diagnosed with advanced breast cancer of 0.83 is assumed (solely based on one RCT with a low risk of selection bias), the corresponding ICER increases substantially, reaching an average of €51 900 per QALY gained. In the latter scenario, assuming an overdiagnosis of 27.1%, drives the ICER to €59 900 per QALY gained. Uncertainty surrounding the clinical input parameters translates directly into uncertainty regarding the cost-effectiveness of screening.

When interpreting the findings of the hypothetical economic evaluation, one should realize that the above cited RR of 0.66 is based on a meta-analysis of old RCTs, both with low and high risk of selection bias. In addition, this magnitude of a stage shift attributed to breast cancer screening is not observed in current real-world data. More precisely, it is very likely more modest. Consequently, the most optimistic ICERs might not be realistic and the ICERs calculated using a lower RR for being diagnosed with advanced breast cancer (like the scenarios applying a RR of 0.83) reflect most probably better the cost-effectiveness of current screening programmes in Belgium.

With regard to the extension of the current breast cancer screening programme to younger and older age groups, we can deduce that; a) due to the lower incidence of breast cancer among younger women (aged 40–49 years), and b) the higher likelihood of overdiagnosis among older women (aged 70 years and older), the cost-effectiveness of both extensions is expected to be worse than that of the current organised screening programme.

6.1 Methods

6.1.1 Research questions

The aim of this chapter is to calculate the cost-effectiveness of breast cancer screening in asymptomatic women without a specifically increased risk of breast cancer. The original research question for this report focused on three age groups: 1) the current target population (women aged 50–69 years); 2) women aged 40-49 years; and 3) women aged 70-74 years. However, the available evidence regarding the expansion of screening to the latter two age groups is limited (see Chapter 2). We are therefore initially focusing on the current target population. The research question has therefore been revised as follows:

> What is the cost-effectiveness of breast cancer screening in asymptomatic women aged 50-69 years, when carried out in accordance with the current Belgian breast cancer screening guidelines?

The impact of extending the age range for screening or changing the screening interval will be outlined in the discussion.

6.1.2 Approach

6.1.2.1 Hypothetical health economic model

The clinical section of this HTA report shows that there are no recent RCTs on which we can reliably base a calculation of the cost-effectiveness of breast cancer screening. Many economic evaluations of screening have already been published. Rather than limiting ourselves to commenting on these existing evaluations and their shortcomings, we have also carried out a hypothetical de novo economic evaluation. We explicitly refer to this evaluation as 'hypothetical' to emphasise that shortcomings in the reliability of the clinical literature automatically affect the reliability of the economic evaluation.

Economic evaluations compare the costs and effects of an intervention and a comparator. Breast cancer screening is already in use in Belgium for women aged 50–69. A major advantage of our calculations is that we have access to several real-world Belgian databases made available by the BCR (see Chapter 5). Based on this data, we can provide a clear picture of the current organised screening programme. Information is available on the current screening programme, including details on invitation process, participation, and diagnostic results. Additionally, we can establish links with age- and stage-specific survival rates, as well as the costs associated with breast cancer treatment. Therefore, in our approach, we will calculate cost-effectiveness by applying the following steps:

1. Modelling the costs and effects of the current situation for asymptomatic women participating in the organized population-based breast cancer screening programme (i.e. the intervention).
2. Modelling the costs and effects if these women were not to participate in this programme (i.e. the comparator).

This chapter provides a detailed overview of the method used to model the effect of screening, along with all the input variables and assumptions involved. The analysis is carried out in accordance with the Belgian guidelines for economic evaluations and budget impact analyses.[3] The Consolidated Health Economic Evaluation Reporting Standards (CHEERS) guidelines[269] are used to support reporting.

6.1.2.2 Non-systematic search for HRQoL and stage shift values

In our review of the literature on economic evaluations, we found that a wide variety of approaches had been taken to modelling HRQoL. Often, it was unclear which primary study the values had been derived from, and values based on expert opinion or non-patients were also used. Due to the limitations of the values identified in existing studies, we decided to conduct a pragmatic, non-systematic search for systematic reviews and meta-analyses. This search was conducted in Ovid

Medline and Google Scholar using the search terms 'utility' or 'utilities' or 'quality of life' combined with 'breast cancer'. In Ovid Medline, the search was restricted to titles containing the search terms. The results were limited to stage-specific health utilities and screening-specific (dis)utilities (e.g. screening participation, false-positive findings, and overdiagnosis). Primary studies reporting only the utility of one specific stage were excluded.

6.2 Overview of de novo hypothetical economic evaluation

6.2.1 Population

In Belgium, asymptomatic women are invited for screening in the year they become 50 up till the age of 69. Women in Flanders and Wallonia do not receive an invitation if they: (1) have had a bilateral mastectomy; (2) have been diagnosed with breast cancer in the past 10 years; (3) have had a screening mammography in the previous year; (4) have had a bilateral mammography outside the population screening programme; (5) have specifically declared that they do not want to be invited; or (6) expressed their wish to no longer be invited. In Brussels, only points (1), (2), (3), (5) and (6) apply. This reflects our target population.

As this population is already being screened, we have the advantage of being able to depict the situation accurately based on the available information. It is as if we are taking a snapshot of the current situation using the most recent available data. This therefore includes women participating in screening for the first time, as well as those who have already undergone one or more screening rounds. There is an inflow and outflow of women participating in screening every year. Assuming no major demographic changes makes the snapshot of the current situation valid to make an estimate of the intervention's cost-effectiveness.

For the purposes of this analysis, we only consider women aged 50–69 who, according to the billing codes, underwent a mammogram as part of the organised screening programme. In other words, the base-case analysis does not include opportunistic screening.

6.2.2 Intervention and comparator

The intervention reflects the current breast cancer screening programme in Belgium, which invites asymptomatic women aged 50–69 to undergo a screening mammogram every two years.

In this hypothetical calculation, the comparator reflects the situation for these women if they were not to undergo a screening mammogram. It is also assumed that they do not participate in opportunistic screening.

6.2.3 Type of economic evaluation

Quality of life is of major importance to cancer patients and can be affected by the disease itself, as well as its treatments. Therefore, a cost-utility analysis (CUA) is performed to express the results as additional costs (€) per QALY gained. The results of the cost-effectiveness analysis (CEA), which does not adjust for quality of life, are also presented. The results section presents both incremental costs (IC) and incremental effects (IE) in life years and QALYs gained, as well as incremental cost-effectiveness ratios (ICER).

The results are presented in tabular format, on the cost-effectiveness plane and in tornado graphs. As there is no official cost-effectiveness threshold in Belgium, the results are also presented on the cost-effectiveness acceptability curve.[3]

6.2.4 Perspective

In accordance with the Belgian guidelines for economic evaluations,[3] the reference case analysis is performed from the healthcare payer's perspective and includes direct healthcare costs. Following these guidelines, payments from the government's healthcare budget, as well as patients' co-payments (excluding supplements), are included. In a scenario analysis, supplements are also included (scenario 'Supplements').

No extension to a societal perspective was modelled. Therefore, the potential impact on productivity and travel costs is excluded. We come back to this in the discussion.

6.2.5 Time horizon and discount rate

In order to capture the full potential impact on relevant incremental costs and effects, the results are extrapolated over a lifetime horizon. The Belgian life tables present survival data up to the age of 105, which is also the upper age limit in our model (see section 6.2.7.7).

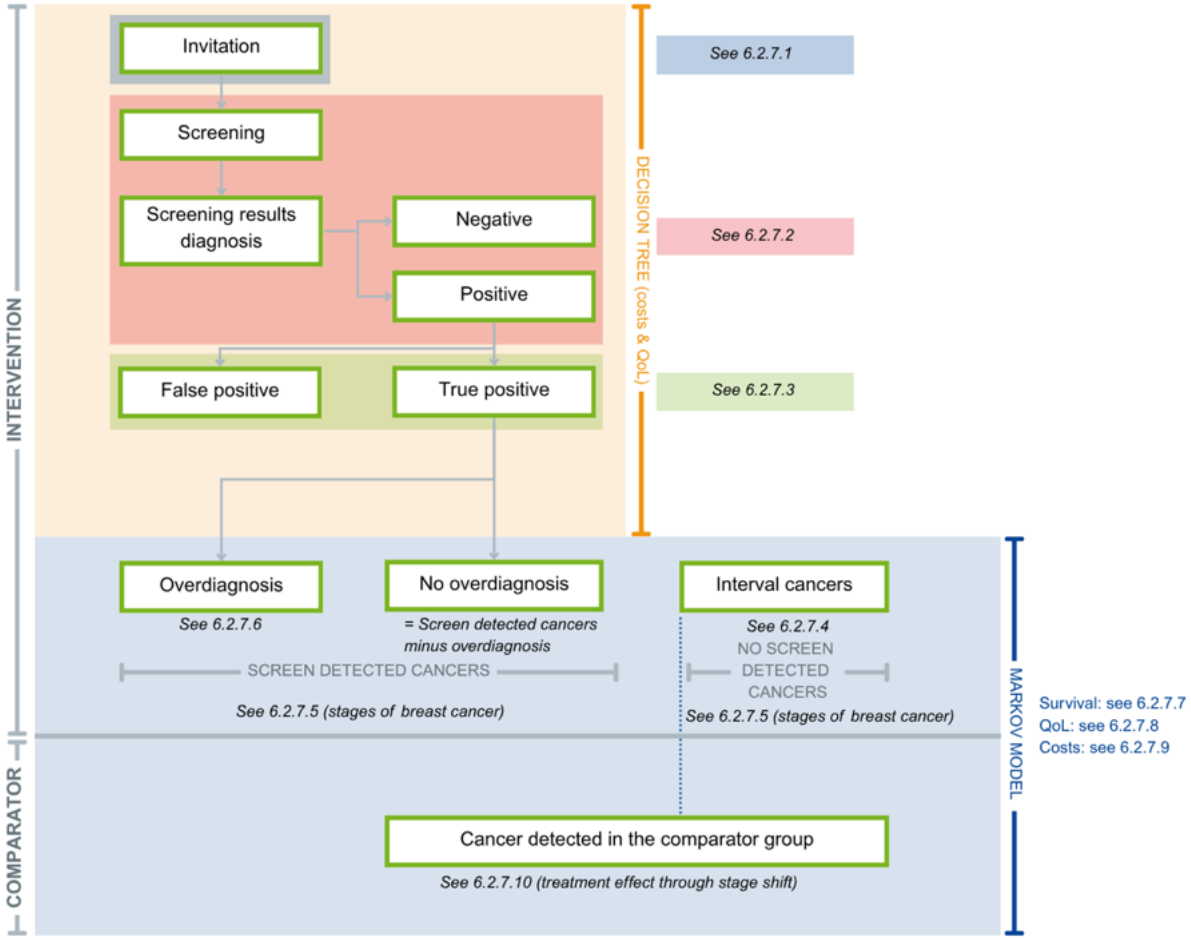
A shorter time horizon is less influenced by the uncertainty surrounding the extrapolation phase of the economic model. In line with the Belgian guidelines,[3] results for multiple shorter time horizons (in 10-year increments) will be presented alongside the lifetime horizon in scenario analyses (scenario 'Time horizon'). The same time horizon is applied to both costs and effects. A shorter time horizon is also used to validate the model's results (see section 6.2.7.13).

According to the Belgian guidelines for economic evaluations,[3] a yearly discount rate of 3% and 1.5%, respectively, is applied to costs and effects in the base case. Following these guidelines, 0%, 3% and 5% are applied to both costs and effects in scenario analyses (scenario 'Discount rate'). Information on costs for the first three months after a breast cancer diagnosis is presented on a monthly basis. As a result, the model uses monthly cycles. Consequently, the annual discount rates are converted into monthly percentages (1.5% per year ➔ 0.1241% per month; 3% per year ➔ 0.2466% per month; 5% per year ➔ 0.4074% per month).

6.2.6 Model structure

The model combines a decision tree and a Markov model (see Figure 35). In the first stage, the costs of invitations, screening and diagnosis are modelled using a decision tree, which includes QoL loss due to false-positive findings. In the Markov model, diagnosed age- and stage-specific breast cancers are linked to costs and effects in terms of life years, adjusted for QoL. After the number of overdiagnosed cancers has been excluded, an inverse stage shift is applied to reflect the situation as if no cancer screening had occurred.

Figure 35 – Model structure (combination of a decision tree and Markov model)



Abbreviations: *QoL: quality of life.*

6.2.7 Input parameters

6.2.7.1 Invitation and participation

In Belgium, breast cancer screening is organised at the regional level by three screening organizations, each responsible for a specific region in Belgium. In Flanders, this is the Centre for Cancer Screening (Centrum voor KankerOpsporing – CVKO, borstkanker.bevolkingsonderzoek.be), in Wallonia the Centre de Coordination et de Référence pour le dépistage des cancers (CCRef, http://www.ccref.org), and in Brussels it is BruPrev (www.bruprev.be). Information about the organisation of breast cancer screening was obtained from these organisations. We rely on Belgian data obtained via the BCR for information on the number of invitees and participants.

Table 13 provides an overview of the data related to the (costs of) invitations and participation, which are explained in more detail below.

Table 13 – Overview of input variables (invitation and participation)

Variable	Value	Uncertainty	Source
Invited persons (n)	633 471*	Fixed	BCR (data 2023)
Participation rate (%)	40.8%	Beta (222 007; 322 529)** Scenario 'Participation': 10, 20, 30, 40, 50, 60 and 70% (Beta distributions)	BCR (data 2022)
Invitation cost (€) • Via eBox • Letter	 • €0 • €0.608	Uniform (0.8; 1.2)** Scenario 'Invitation cost letter': €0.76 for all invitations (i.e. no use of eBox) Scenario 'Invitation cost excluded': €0 Scenario 'Invitation cost alternative': €0***	Flanders Department of Care and personal communication
Usage of the eBox (%)	30-40%	Uniform (30-40%)	Flanders Department of Care and personal communication
Cost result letter (€) • To participant • To GP	 • €0.8924 • €0.1	Uniform (0.8; 1.2) Scenario 'Invitation cost excluded': €0 Scenario 'Invitation cost alternative': €0***	Flanders Department of Care and personal communication
Lump sum costs related to screening (€)	€3.44 per invited person	Uniform (0.8; 1.2) Scenario 'Invitation cost excluded': €0 Scenario 'Invitation cost alternative': €5.52 (€3.44 + €2.08***) or €5.99 or €3.69	Flanders Department of Care and personal communication

Abbreviations: *BCR: Belgian Cancer Registry; GP: general practitioner.*
Notes: ** The number of invited persons may vary depending on the year and the source used. Changes in this number do not affect cost-effectiveness, as all consequences – both costs and effects – are proportionally impacted in the same way.*
*** For further information on the type of distribution we refer to section 6.2.7.12. A beta distribution is determined by two parameters. The alpha parameter equals the number of events (e.g., in this case 218 295 + 3712 = 222 007). The beta parameter equals the size of the sample minus the number of events (e.g., in this case 544 536 – 222 007 = 322 529). The numbers are explained in the text of this section. The uniform (0.8; 1.2) indicates the minimum and maximum values of this distribution are -/+ 20% of the average value.*
**** To avoid double counting, no separate costs are allocated for the invitation or communication of results in this scenario, as these are covered by the variable subsidy (i.e. included in the €2.08).*

INVITATION & PARTICIPATION

In 2023, the most recent year for which data is available, 633 471 people were invited to participate in breast cancer screening. Of these, 77 965 were invited for the first time. Data on participation for the year 2024 are not yet available. Consequently, complete participation data for women invited in 2023 cannot currently be obtained, and estimates are therefore based on participation data from the

preceding year. In 2022, 544 536[ffff] people were invited, 40.8%[gggg] of whom participated in the population screening in 2022 and 2023 (218 295 and 3712, respectively). This percentage is applied in the baseline scenario as the participation rate.

This figure is influenced by participation in the three regions. The highest participation rate is recorded in Flanders, being 54% over the years 2022-2024.[85] Furthermore, due to the structure of the Belgian billing codes, the BCR data cannot distinguish between individuals participating in opportunistic screening and those who have a mammogram because of symptoms. Therefore, this percentage only includes individuals who have had a mammography of both breasts as part of the population screening programme (Nomenclature codes 450192-450203). Participation would be even higher if opportunistic screening outside the population-based screening programme were to be included. To analyse the impact of participation (e.g. lower in Wallonia or Brussels or higher in Flanders and/or if opportunistic screening is taken into account), the following participation rates were modelled in scenarios: 10%, 20%, 30%, 40%, 50%, 60% and 70% (scenario 'Participation').[hhhh] We elaborate on this point in the discussion section.

COSTS INVITATIONS (& COMMUNICATION OF RESULTS)

For the invitation costs associated with breast cancer screening, most transparent information could be obtained from personal communication with the Flanders Department of Care, supplemented with public information on population screening financing (Financiering_bevolkingsonderzoeken_2025_mqeb1h.pdf). CvKO sends invitation letters to a selection of women aged 50-69 residing in Flanders who do not meet at least one of the exclusion criteria to be invited (see section 6.2.1). In general, when the time between sending the invitation and the proposed appointment for the mammography is sufficient,[iiii] the invitations will first be sent via eBox,[jjjj] if active. Using the eBox for communication is always free of charge. If the eBox invitation is not opened within three days or the time between sending the invitation and the proposed appointment is between seven and fourteen days, the invitation will be sent additionally by post. It is estimated that overall 30-40%[kkkk] of invited women open their eBox invitation. The cost for sending invitations (including a brochure with information about the screening) by post is €0.608 per invitation (personal communication[llll]). According to information received from CCRef and BruPrev, this is €0.667 and

---

[ffff] The number of women invited varies from year to year. Nevertheless, this number is not altered in the scenario analyses. This is because the calculated cost-effectiveness remains unaffected by the initial number of women invited, since the incremental costs and benefits are applied proportionally to those participating.

[gggg] This percentage varies by age group, ranging from 33.9% for 50–54-year-olds to 45.6% for 60–64-year-olds. However, this has no impact on the allocation of costs for invitations and participation, since the cost-effectiveness of the target population of women aged 50–69 years is calculated. This is why general figures are used for this variable in the model.

[hhhh] This range includes the coverage of 58% reported for Belgium (2023), ranging from 45% in Brussels to 65% in Flanders.[270] This coverage rate includes opportunistic screening.

[iiii] If the time between sending the invitation and the appointment is less than 7 days, CvKO will contact the invitee personally. If the interval is between 7 and 14 days, both an eBox and postal invitation are sent.

[jjjj] eBox is Belgium's secure digital mailbox where citizens receive official messages and documents from government institutions online instead of by paper mail.

[kkkk] Around 60% of people have activated their eBox, of whom around 60% view the messages. This equates to 36% of invited individuals opening their eBox.

[llll] In a public document, the cost is stated as €0.407 per invitation (Financiering_bevolkingsonderzoeken_2025_mqeb1h.pdf). However, based on more recent information from the Flanders Department of Care, it is considered more appropriate to use a cost of €0.608.

€0.76 respectively, with the eBox not currently being used.[mmmm] A scenario is applied with the latter highest cost (scenario 'Invitation cost letter'). No reminder letters are sent for this population screening. The cost of the initial invitation is also the same as for an invitation in a subsequent round.

Population screening involves more than just identifying and inviting the target group and carrying out the screening. It also includes the following components: registration and communication of screening results; diagnostic feedback to the Belgian Cancer Registry (BCR); quality control; evaluation of the screening program; and awareness campaigns. Based on experience in Flanders with population screening, we have included costs for several of these activities. Various costs apply to actions that will be addressed later in this chapter (e.g. communication costs following diagnosis). These costs are already covered here, as they originate from budgets that are described collectively in this section.

1) Registration screening results: CvKO has developed the Heracles system, used for registration and data processing to support the invitation process and follow-up for screening. The maintenance costs of the Heracles registration and invitation system are included in CvKO's lump-sum subsidy obtained from the Flanders Department of Care (see below). Over the years, CvKO has spent approximately €470 000 on this system each year.[264]

2) Communication of screening results: The results of the screening mammography must be communicated to both the participants and their general practitioner (GP). This can be done via the eBox, which is free of charge, and/or by post. Negative results are sent in first instance to participants via the eBox if active. When the eBox message is not opened within three days, the negative result is sent by post. Positive results are always sent by post and eBox (if active). According to information from the Department of Care, the average cost of sending the results letter is €0.8924 per participant. Notifying the GP incurs a cost of €0.1 per results letter (personal communication[nnnn]). The GP receives this information electronically via eHealth. The cost of €0.1 is taken into account for communication to GPs since, according to information received from the Department of Care, this applies to over 99%[oooo] of results letters. Similar information with somewhat higher costs was also received by BruPrev and CCRef.[pppp]

3) Diagnostic feedback: The BCR collects data on all new cancer diagnoses in Belgium and their subsequent management. In addition, it collects all anatomopathological test results for the early detection of cervical, breast and colorectal cancers. The data from these registries are relevant for the

---

mmmm In Brussels, plans are in place to start using the eBox within the next two years. The same will apply in Wallonia from May 2026, with an invitation being sent by post if there is no response within two weeks. The system has already been tested, and initial findings suggest that around 30% of recipients open the eBox. We have used this figure as the lower limit for this variable (see above).

nnnn In a public document, the cost is stated as €0.18 per electronic communication (Financiering_bevolkingsonderzoeken_2025_mqeb1h.pdf). However, based on more recent information from the Flanders Department of Care, it is considered more appropriate to use a cost of €0.1.

oooo In 2025, approximately 400 result letters were sent by post to general practitioners at an estimated cost of €1.342 per letter (personal communication). For 2026, this figure is estimated to be 780 letters. As this figure represents only around 0.25% of the total number of letters sent, it has not been taken into account here.

pppp Information was also received from CCRef and BruPrev:
CCRef: 99% of negative results is communicated electronically to the GPs (cost: €0.1). Only 1% occurs through post (cost: €1.47). There is also an extra cost of €0.1 to provide the information electronically to the health network. This amounts to €0.21 for negative cases.
Positive results are always sent to the GP by post at a cost of €1.47, and in around 50% of cases, a copy of the letter is sent to another healthcare provider at an additional cost of €1.47. In virtually all cases (>99%), communication also takes place electronically (€0.1 to GPs and €0.1 in around 50% of cases to other healthcare providers). Additionally, electronic transmission to the health network costs €0.1. In total, this amounts to €2.46 for positive cases.
BruPrev: the cost per positive result letter is €1.34.

efficient organisation, monitoring and evaluation of the three population-based cancer screening programs (Financiering_bevolkingsonderzoeken_2025_mqeb1h.pdf). The BCR receives an annual subsidy for the three programs. In 2025, this subsidy amounted to €244 307 (Financiering_bevolkingsonderzoeken_2025_mqeb1h.pdf). Additionally, IMA received a subsidy of €5000 (personal communication). As these costs are a direct consequence of organizing population cancer screening and are also paid by the regional governments, these costs are added to the lump-sum subsidy for CvKO. In a later step, these costs will be allocated to breast cancer screening via a distribution key (see below).

4) Quality control of execution: Mammography units (radiology services) must be recognised by the Flemish government in order to be allowed to perform screening mammograms. This involves a personnel cost at the administration. This unknown cost falls outside the lump-sum costs included in the model. Thus, there is a conservative underestimation of costs.

5) Evaluation of screening programme (periodic working group): The Flemish government's population screening is continuously monitored and evaluated by multidisciplinary working groups, among others. There is one working group dedicated to each population screening, as well as two general working groups. There are about three yearly meetings per working group. The average expense allowance for experts in these groups is €120 per expert per meeting and approximately 10 participants receive an expense allowance. This would amount to an annual cost of approximately €18 000.[qqqq] Similarly to the costs related to diagnostic feedback, this cost is not included in the lump-sum subsidy and added separately to be allocated to breast cancer screening via a distribution key (see below).

6) Awareness campaigns: In 2024, the last representative year, €360 000 was taken from the reserves to fund a campaign. An additional €50 000 was spent on creating and distributing information and awareness-raising materials for the website. From 2027 onwards, 10% of the lump-sum subsidy will be allocated to communication activities aimed at raising awareness. This amounts to around €500 000 per year. As it is part of the lump-sum subsidy, no additional costs need to be added separately.

The Flanders Department of Care provides an annual subsidy to the CvKO. The total cost of the three population-based cancer screenings (breast, colon and cervix) in 2025 was €8 782 662 (personal communication with the Flanders Department of Care), consisting of a lump-sum subsidy of €5 374 077 and a variable subsidy of €3 408 585. The costs associated with the variable subsidy, such as invitations and notification of results to doctors and participants, are already included in the economic model using the procedure and costs described above. The costs associated with the lump-sum subsidy, as well as the costs related to BCR, IMA and the fees paid to the participants of the periodic working groups, are allocated via a distribution key.

We use the number of people invited as the distribution key. This information is obtained from a recent annual report of the CvKO.[271] In 2023, 432 556 invitations were sent to women as part of the breast cancer screening program. For colorectal cancer screening, 832 424 invitation letters were sent out that year. For cervical cancer screening, 313 360 invitation letters were sent by post, and an additional 61 838 invitations were sent by email to women who had previously participated. This amounts to a total of 1 640 178 invitations. If we divide the lump-sum subsidy and other identified costs of approximately €5.6 million by the number of people invited, we obtain a cost of approximately €3.44 per invited person.[rrrr]

[qqqq] 5 working groups x 3 meetings per year x 10 participants/meeting x €120/participant = €18 000. We have chosen to aggregate the costs of all working groups across all population-based screening programmes, allocating them to the various programmes using an allocation formula (see below). Given that these working groups account for a relatively small proportion (<0.5%) of the total costs discussed in this section, a different approach would have no significant impact.

[rrrr] €5 374 077 + €244 307 + €5000 + €18 000 = €5 641 384. Divided over 1 640 178 invited persons = €3.44 per invited person.

Four extra scenario analyses will be applied to the above costs. In the first scenario, the invitation costs will be excluded to demonstrate the impact of not taking this cost into account (scenario 'Invitation cost excluded'). In the second scenario, the variable subsidy will also be allocated to invited individuals using the same distribution key as for the lump-sum subsidy (see above). This amounts to an additional cost of €2.08 per person,[ssss] giving a total of €5.52 per person (scenario 'Invitation cost alternative'). To avoid double counting, no additional costs are allocated for invitations and communication of results in this scenario, as these are covered by the variable subsidy. Finally, CCRef and BruPrev also communicated the results of their calculations, which provided a total cost of €5.99 and €3.69 per invited person,[tttt] respectively (personal communication) (scenario 'Invitation cost alternative').

6.2.7.2 Screening

Table 14 provides an overview of the costs associated with mammography and the test results in the context of population screening for breast cancer.

Table 14 – Overview of input variables (screening and test results)

Variable	Value	Uncertainty	Source
Costs screening: • Mammography • Second reading	 • €74.59 • €7.46	Fixed	NIHDI Nomenclature (2026)
Screening results: • Negative • Positive	 • 96.8% • 3.2%	 • x1 = Beta (204 024; 6717) • 1 – x1	BCR (data 2020)

***Abbreviations**: BCR: Belgian Cancer Registry; NIHDI: National Institute for Health and Disability Insurance.*

Participants undergo mammography as part of a population screening programme. Each mammogram is reviewed independently by two radiologists (second reading). If the conclusions of the first and second reader differ, a third radiologist examines the images. No separate reimbursement is provided for the latter. The following nomenclature numbers apply to these services (2026 values):

- 450192-450203[uuuu]: Mammography of both breasts, as part of a government-organised screening programme (€74.59)[vvvv]

---

ssss €3 408 585 divided over 1 640 178 invited persons = €2.08 per invited person.

tttt CCReF communicated a total cost of €839 109 divided over 140 000 invitations (€5.99 per invited person). BruPrev communicated a total cost of €220 289 divided over 59 683 invitations (€3.69 per invited person).

uuuu There is also a separate nomenclature number (450354-450365) for 'mammography of both breasts as part of breast cancer screening in asymptomatic women with a significantly increased risk profile'. This number has not been selected, as the focus of this report is on women at average risk.

vvvv A second mammogram may be performed in cases of uncertainty or poor image quality. Among women with a negative screening result, this occurred in only 0.03% of cases, which is negligible and therefore not included in the model.
The study also examined how many women with a negative result had undergone an MRI scan within one week or one month of the mammogram. This was the case for 9 (<0.01%) and 34 (0.01%) women out of a total of 238 418, respectively. Consequently, costs for such a scan were not added in the model for these women.

- 450214-450225: Second reading of a screening mammogram of both breasts, as part of a government-organised population screening programme (€7.46)[wwww]

We use data from 2020 for the screening results, as this data set is complete.[xxxx] Out of 210 741 tests conducted that year, 204 024 (96.8%) were negative, while 6717 (3.2%) were positive. These percentages vary according to age group (see Appendix 26). This is consistent with observations in Flanders, where 6.5% of people have abnormal results after the initial screening and 2.4% have abnormal results after mammograms in a subsequent screening round.

Section 6.2.7.1 provides information on how these results are communicated to the participant and GP. If the result is negative, people are invited to the next screening round until the end of the programme. No additional costs or effects are modelled following a negative test result. If the test result is positive, the diagnostic procedure is started (see section 6.2.7.3).

6.2.7.3 Diagnosis (true- and false-positive findings)

Table 15 presents the percentage of true- and false-positive findings and the costs associated with a false-positive diagnosis.

Table 15 – Overview of input variables (true and false-positive findings)

Variable	Value	Uncertainty	Source
True and false positives: • TP • FP	 • 19.9% • 80.1%	 • x1 = Beta (1337; 5380) • 1 – x1 Scenario 'True positives': x_1 = Beta (1418; 5299)	BCR (data 2020)
Costs diagnosis: • TP • FP	 • See part 0 • €125.61+ €52.34	 • See part 0 • Beta distributions based on frequencies + uniform (33.74; 70.94)	BCR – IMA data (2020-2022) + NIHDI Nomenclature (2026)

***Abbreviations**: BCR: Belgian Cancer Registry; FP: false positive; IMA: Intermutualistic Agency; TP: true positive.*

Of the 6717 positive test results, 1418 were originally categorized as true positives (TP) (a positive predictive value (PPV) of 21.1%), accounting for 0.67% (1418/210 741) of all tests performed, while 5299 (78.9%) were false positives (FP), accounting for 2.51% (5299/210 741) of all tests performed. The last figure is based on data from 2020 for Belgium and corresponds to the most recent data for the year 2022 published for Flanders, with a specificity of 97.5% (false-positive rate = 1 - specificity).[85] However, when analyzing the number of interval cancers (see applied definitions in section 6.2.7.4),

[wwww] The nomenclature code for a second reading was present in 99.5% of cases. A third reading was recorded in just 0.01% of cases. In the model, a second reading was assigned to all cases.

[xxxx] A false positive is defined as an abnormal mammogram result during population screening, followed by no cancer diagnosis within 24 months. The available data does not yet include 24 months of complete follow-up for screenings carried out in 2021 and 2022. This is why we are using data from 2020. However, we note that using percentages from other years is not problematic since: 1) the probabilities, and not the underlying absolute figures, are combined with the other data in the model; and 2) our analysis of Belgian data shows that this percentage does not vary over the years for which data are available and complete (see Appendix 25).

we found that 81 cancers were classified as an interval cancer following a positive test result.[yyyy] In our opinion, these tests should be classified as FP at the time of the initial test results. The adjusted number of true positives and FPs becomes 1337 (19.9%) and 5380 (80.1%), respectively. FP results account for 2.55% (5380/210 741) of all tests performed. Although the difference between the two interpretations is minimal, we have decided to test the impact this has on the intervention's cost-effectiveness in a separate scenario (scenario 'True positives').

The diagnostic procedures following a positive result lead to additional costs. These costs associated with a true positive test are presented in section 0. The costs associated with FP tests are based on a frequency table of examinations carried out on a cohort of FP patients between 2020 and 2022. The fees for 2026 were then applied to these frequencies (see Appendix 23). To avoid interaction with examinations linked to the diagnosis of interval cancer, this analysis was carried out on false-positive diagnoses where no interval cancer was subsequently diagnosed. The total cost of these examinations was on average €125.61. In addition, the cost of an outpatient visit to a medical specialist (gynaecologist or breast surgeon, €33.74 or €37.20,[zzzz] respectively), possibly in combination with a visit to a GP (€33.74)[aaaaa], was also included. Combining these costs, a minimum or maximum extra cost of €33.74 and €70.94, respectively, was included.

We note that it is not necessary to model false-negative (FN) test results separately. If a woman receives a negative result, no further steps are taken, and she is invited to attend a subsequent screening round at a later date, provided that she still meets the conditions for population screening. A FN test result can be identified in a subsequent screening round or as an interval cancer (see section 6.2.7.4).

6.2.7.4 Interval cancers

Table 16 shows the information related to interval cancers.

Table 16 – Overview of input variables (interval cancers)

Variable	Value	Uncertainty	Source
Interval cancers	0.31%	Beta (647; 210 094)	BCR (2020)

***Abbreviations**: BCR: Belgian Cancer Registry.*

A screen-detected cancer is defined as a primary breast cancer that is diagnosed within 24 months of an abnormal mammogram screening result, with or without a follow-up examination. An interval cancer is defined as a primary breast cancer detected after a negative mammogram or after an abnormal mammogram for which no malignancy was found during follow-up examinations. This cancer must be diagnosed before the next screening, or within 2 years of the last screening, if the woman has exceeded the maximum age for participation or does not participate in the next screening.[272] Figure 36 shows the algorithm used by BCR to determine interval cancers.

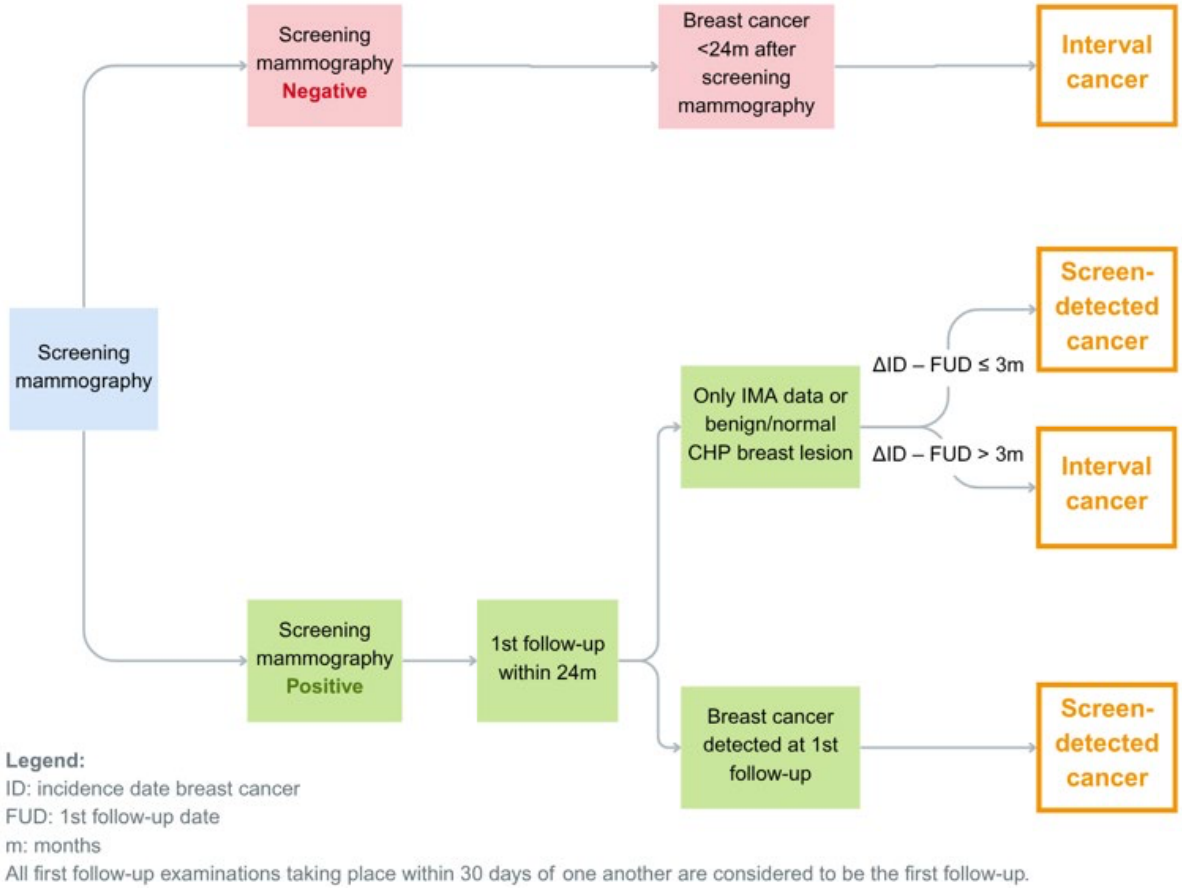
---

yyyy These are most likely positive test results. Initially, they were regarded as false positives, but were later confirmed as interval cancers (see section 6.2.7.4 for the definition of an interval cancer).

zzzz Nomenclature code 105593 ('Raadpleging in de spreekkamer door een geaccrediteerde arts-specialist in de gynaecologie, inclusief een eventueel schriftelijk verslag aan de behandelende arts') or 101290 ('Raadpleging in de spreekkamer door een geaccrediteerde arts-specialist in de heelkunde, inclusief een eventueel schriftelijk verslag aan de behandelende arts').

aaaaa Nomenclature code 101076 ('Raadpleging in de spreekkamer door een geaccrediteerde huisarts').

Figure 36 – Schematic representation of the algorithm used to determine screen-detected and interval cancers



***Source**: CvKO (2024)*[272]

***Notes**: Interval cancer: 1) After a negative screening mammogram: If breast cancer is found in the cancer registry within 24 months of the date of the screening mammogram, it is considered an interval cancer. 2) After an abnormal mammogram: First, the abnormal screening mammograms are linked to the nomenclature data for the breast, the CHP (cyto-histopathology)-breast and cancer registry database, in order to identify the first follow-up examination. All examinations that took place within 30 days of each other are considered to be the first follow-up. 2a) If an incidence date for breast cancer is found within three months of the first follow-up, it is considered to be screen-detected. 2b) If an incidence date for breast cancer is found more than three months after the first follow-up, it is considered an interval cancer.*[272]

The number of screen-detected cancers was 1337 and is already presented in section 6.2.7.3. Based on Belgian cancer data from 2020 (the last year for which complete 24-month follow-up data is available), on a total of 210 741 participants, 647 (0.31%) interval cancers were diagnosed. Of these, 566 were detected following a negative screening result and 81 following a positive result. In other words, around two-thirds (67.4%) of cancers are detected through screening, while the remaining third are detected between screening rounds. This aligns with the reported sensitivity of 67.5% in Flanders in 2022.[272]

6.2.7.5 Stages of screen-detected and interval breast cancers

Table 17 and Table 18 show an overview of the input variables related to the staging of cancers per age category, for both true-positive and interval cancers, respectively.

Table 17 – Overview of input variables (cancer stages at diagnosis – true positive cancers)

	Value	Uncertainty		Source
Number of cancers per age category				BCR (2020-2022)*
50-54y	951	π1 = Beta (951; 3401)	(21.9%)	
55-59y	980	π2 = (1 - π1) x [Beta (980; 2421)]	(22.5%)	
60-64y	1271	π3 = (1 - π1 - π2) x [Beta (1271; 1150)]	(29.2%)	
65-69y	1150	π4 = (1 - π1 - π2 - π3)	(26.4%)	
50-54y		Conditional beta distributions:		BCR (2020-2022)*
Stage 0	207/951	π1 = Beta (207; 744)	(21.8%)	
Stage I	457/951	π2 = (1 - π1) x [Beta (457; 287)]	(48.1%)	
Stage II	232/951	π3 = (1 - π1 - π2) x [Beta (232; 55)]	(24.4%)	
Stage III	44/951	π4 = (1 - π1 - π2 - π3) x [Beta (44; 11)]	(4.6%)	
Stage IV	11/951	π5 = (1 - π1 - π2 - π3 - π4)	(1.2%)	
55-59y		Conditional beta distributions:		
Stage 0	163/980	π1 = Beta (163; 817)	(16.6%)	
Stage I	529/980	π2 = (1 - π1) x [Beta (529; 288)]	(54.0%)	
Stage II	242/980	π3 = (1 - π1 - π2) x [Beta (242; 46)]	(24.7%)	
Stage III	34/980	π4 = (1 - π1 - π2 - π3) x [Beta (34; 12)]	(3.5%)	
Stage IV	12/980	π5 = (1 - π1 - π2 - π3 - π4)	(1.2%)	
60-64y		Conditional beta distributions:		
Stage 0	194/1271	π1 = Beta (194; 1077)	(15.3%)	
Stage I	729/1271	π2 = (1 - π1) x [Beta (729; 348)]	(57.4%)	
Stage II	302/1271	π3 = (1 - π1 - π2) x [Beta (302; 46)]	(23.8%)	
Stage III	41/1271	π4 = (1 - π1 - π2 - π3) x [Beta (41; 5)]	(3.2%)	
Stage IV	5/1271	π5 = (1 - π1 - π2 - π3 - π4)	(0.4%)	
65-69y		Conditional beta distributions:		
Stage 0	182/1150	π1 = Beta (182; 968)	(15.8%)	
Stage I	656/1150	π2 = (1 - π1) x [Beta (656; 312)]	(57.0%)	
Stage II	260/1150	π3 = (1 - π1 - π2) x [Beta (260; 52)]	(22.6%)	
Stage III	43/1150	π4 = (1 - π1 - π2 - π3) x [Beta (43; 9)]	(3.7%)	
Stage IV	9/1150	π5 = (1 - π1 - π2 - π3 - π4)	(0.8%)	

Abbreviations: *Y: years.*
Remark: ** While calendar years refer in the preceding sections to the year of screening, they refer in this table to the year of breast cancer diagnosis (or incidence year).*

Table 18 – Overview of input variables (cancer stages at diagnosis – interval cancers)

	Value	Uncertainty		Source
Number of cancers per age category				BCR (2020)*
50-54y	**152**	**π1 = Beta (152; 528)**	**(22.4%)**	
55-59y	**168**	**π2 = (1 - π1) x [Beta (168; 360)]**	**(24.7%)**	
60-64y	**191**	**π3 = (1 - π1 - π2) x [Beta (191; 169)]**	**(28.1%)**	
65-69y	**169**	**π4 = (1 - π1 - π2 - π3)**	**(24.9%)**	
50-54y		Conditional beta distributions:		BCR (2020)*
Stage 0	9/152	π1 = Beta (9; 143)	(5.9%)	
Stage I	46/152	π2 = (1 - π1) x [Beta (46; 97)]	(30.3%)	
Stage II	67/152	π3 = (1 - π1 - π2) x [Beta (67; 30)]	(44.1%)	
Stage III	24/152	π4 = (1 - π1 - π2 - π3) x [Beta (24; 6)]	(15.8%)	
Stage IV	6/152	π5 = (1 - π1 - π2 - π3 - π4)	(3.9%)	
55-59y		Conditional beta distributions:		
Stage 0	13/168	π1 = Beta (13; 155)	(7.7%)	
Stage I	64/168	π2 = (1 - π1) x [Beta (64; 91)]	(38.1%)	
Stage II	61/168	π3 = (1 - π1 - π2) x [Beta (61; 30)]	(36.3%)	
Stage III	22/168	π4 = (1 - π1 - π2 - π3) x [Beta (22; 8)]	(13.1%)	
Stage IV	8/168	π5 = (1 - π1 - π2 - π3 - π4)	(4.8%)	
60-64y		Conditional beta distributions:		
Stage 0	7/191	π1 = Beta (7; 184)	(3.7%)	
Stage I	82/191	π2 = (1 - π1) x [Beta (82; 102)]	(42.9%)	
Stage II	64/191	π3 = (1 - π1 - π2) x [Beta (64; 38)]	(33.5%)	
Stage III	25/191	π4 = (1 - π1 - π2 - π3) x [Beta (25; 13)]	(13.1%)	
Stage IV	13/191	π5 = (1 - π1 - π2 - π3 - π4)	(6.8%)	
65-69y		Conditional beta distributions:		
Stage 0	9/169	π1 = Beta (9; 160)	(5.3%)	
Stage I	73/169	π2 = (1 - π1) x [Beta (73; 87)]	(43.2%)	
Stage II	65/169	π3 = (1 - π1 - π2) x [Beta (65; 22)]	(38.5%)	
Stage III	15/169	π4 = (1 - π1 - π2 - π3) x [Beta (15; 7)]	(8.9%)	
Stage IV	7/169	π5 = (1 - π1 - π2 - π3 - π4)	(4.1%)	

Abbreviations: *Y: years.*
Remark: ** While calendar years refer in the preceding sections to the year of screening, they refer in this table to the year of breast cancer diagnosis (or incidence year).*

For interval cancers, the data from 2020 were the most recent complete figures available, since two years of follow-up are required. For TP cancers, however, we have complete data up to and including 2022. Therefore, we chose to base our findings on the complete 2020 data for interval cancers. For information related to TP diagnoses, we combined the data for the period 2020–2022. This enables us to base our analysis on a larger, as recent as possible sample.

The data are presented in 5-year age categories. The original data on women who participated in breast cancer screening included a relatively small group of women in the 45–49 and 70–74 age categories (see Table 19 and Table 20). The first category comprises women aged 49 who were invited for the organized population screening in the year they turned 50, participated, and who were diagnosed with cancer before their 50th birthday. The second category comprises 69-year-old women who were invited for the organized population screening, participated, and diagnosed with cancer at

70. In the model, to reduce complexity, the 49-year-olds are included in the 50–54 age group, and the 70-year-olds in the 65–69 age group (see Table 17 and Table 18).

Table 19 – Stage distribution true-positive cancers (2020-2022, original age-categories)

Stages	45-49y	50-54y	55-59y	60-64y	65-69y	70-74y	Total
0	16 (15.1%)	191 (22.6%)	163 (16.6%)	194 (15.3%)	182 (15.9%)	/	**746 (17.1%)**
I	49 (46.2%)	408 (48.3%)	529 (54.0%)	729 (57.4%)	654 (57.0%)	2 (66.7%)	**2371 (54.5%)**
II	31 (29.2%)	201 (23.8%)	242 (24.7%)	302 (23.8%)	260 (22.7%)	/	**1036 (23.8%)**
III	8 (7.5%)	36 (4.3%)	34 (3.5%)	41 (3.2%)	43 (3.7%)	/	**162 (3.7%)**
IV	2 (1.9%)	9 (1.1%)	12 (1.2%)	5 (0.4%)	8 (0.7%)	1 (33.3%)	**37 (0.9%)**
Total	**106**	**845**	**980**	**1271**	**1147**	**3**	**4352**

Abbreviations: *Y: years.*
Notes: *Number of observations: 1315 (2020), 1697 (2021) and 1340 (2022).*
Remark: *There are slight differences in the data provided because data relating to screening is based on the year participants were screened (e.g. 1337 true positives in 2020), while data on breast cancer stages is based on the year of incidence (e.g. 1315 cases in 2020). This does not pose a problem as the percentages are applied in the model.*

Table 20 – Stage distribution interval cancers (2020, original age-categories)

Stages	45-49y	50-54y	55-59y	60-64y	65-69y	70-74y	Total
0	/	9 (6.0%)	13 (7.7%)	7 (3.7%)	6 (4.4%)	3 (8.8%)	**38 (5.6%)**
I	/	46 (30.5%)	64 (38.1%)	82 (42.9%)	55 (40.7%)	18 (52.9%)	**265 (39.0%)**
II	1 (100.0%)	66 (43.7%)	61 (36.3%)	64 (33.5%)	55 (40.7%)	10 (29.4%)	**257 (37.8%)**
III	/	24 (15.9%)	22 (13.1%)	25 (13.1%)	13 (9.6%)	2 (5.9%)	**86 (12.6%)**
IV	/	6 (4.0%)	8 (4.8%)	13 (6.8%)	6 (4.4%)	1 (2.9%)	**34 (5.0%)**
Total	**1**	**151**	**168**	**191**	**135**	**34**	**680**

Abbreviations: *Y: years.*
Notes: *Number of observations: 680 (2020).*
Remark: *There are slight differences in the data provided because data relating to screening is based on the year participants were screened (e.g. 647 interval cancers in 2020), while data on breast cancer stages is based on the year of incidence (e.g. 680 interval cancers in 2020). This does not pose a problem as the percentages are applied in the model.*

6.2.7.6 *Overdiagnosis*

Table 21 shows the information related to the percentage of overdiagnoses, expressed as the proportion of cancers diagnosed in women invited for screening.

Table 21 – Overview of input variables (overdiagnosis)

Variable	Value	Uncertainty	Source
Overdiagnosis	20.7% (to be applied on the total number of identified cancers in the screening group, i.e. screen-detected and interval cancers)	Beta (95; 365) Scenarios 'Overdiagnosis': 27.1% (Beta(1059; 2847) 0, 10, 20, and 30%	Meta- and subgroup analysis (see section 3.2.2.6)

As explained in the background, overdiagnosis in screening is the over-detection of a disease or abnormal condition that would not have caused the person any harm if left undiscovered.[45, 53, 54] It refers to 1) detection of a tumour that would have regressed spontaneously; 2) detection of a slow or non-growing tumour which would not have posed a problem for the patient during his or her lifetime; or 3) detection of a tumour at a time when the patient is more likely to die from other diseases or old age (see section 2.2.3). Overdiagnosis cannot be measured at the level of the individual screened person.[14] Nevertheless, it is important to explicitly include overdiagnosis in economic evaluations of population screening. These patients do not benefit from the cancer diagnosis in terms of survival, and induce additional costs linked to overtreatment, which might also be linked to loss of quality of life.

It is important to clearly define how overdiagnosis is expressed. This issue is addressed in the publication by the Independent UK Panel on Breast Cancer Screening in The Lancet: "*although the definition of an overdiagnosed case, and thus the numerator in a ratio, is clear, the choice of denominator has been the source of further variability in reported estimates.*"[129] The authors demonstrate the impact of using a different denominator by presenting estimates of overdiagnosis from three RCTs in which the control group was not systematically screened at the end of the trial. These estimates were calculated using four different methods (Table 22).

Table 22 – Estimates of overdiagnosis (an example of the impact of applying different denominators)

RCT	Method A	Method B	Method C	Method D
Malmö I 55-69 years	11.7% (82/698)	10.5% (82/780)	18.7% (82/438)	29.1% (82/282)
Canada I*	14.1% (82/581)	12.4% (82/663)	22.7% (82/361)	29.4% (82/279)
Canada II*	10.7% (67/626)	9.7% (67/693)	16.0% (67/420)	19.8% (67/338)

***Important remark**: these numbers are a copy of Table 4 in Independent UK Panel on Breast Cancer Screening (2012).[129] These numbers are not up to date and are only used to illustrate the point. For an update, we refer to Table 23.*
***Notes**: Numbers of excess cancers are expressed as a percentage of different denominators. A = excess cancers as a proportion of cancers diagnosed over whole follow-up period in unscreened women. B = excess cancers as a proportion of cancers diagnosed over whole follow-up period in women invited for screening. C = excess cancers as a proportion of cancers diagnosed during screening period in women invited for screening. D = excess cancers as a proportion of cancers detected by screening in women invited for screening.[129]*
** Canada I and II refer to the CNBSS-1 and CNBSS-2 studies in the clinical part of this report.*

With the same absolute number of overdiagnoses, method B results in the lowest percentage (10–12%), while method D results in the highest percentage (20–29%). In method B, excess cancers are expressed as a proportion of those diagnosed during the entire follow-up period in women invited for screening (i.e. including interval cancers). In method D, excess cancers are expressed as a proportion of those detected by screening in women invited for screening.

The UK Panel believes that "*there is no single optimum way to estimate overdiagnosis but that the two most useful estimates are: from the population perspective, the proportion of all cancers ever diagnosed in women invited to screening that are overdiagnosed (method B […]); and from the perspective of a woman invited to screening, the probability that a cancer diagnosed during the screening period represents overdiagnosis.*[273, 274] *Because the number of interval cancers depends on screening frequency,*[158] *interval cancers should be included in the denominator (method C […]).*"[129]

The evidence presented in the clinical part of this report includes both screening-detected and interval cancers in the screening group as an event over the full follow-up period. This reflects method B. According to this evidence, based on information from seven RCTs among adult women at average risk of breast cancer and invited to screening, 27.1% (1059/3906) of all cancers (DCIS and invasive) detected (i.e. including both screen-detected and interval cancers) were overdiagnosed.[bbbbb]

As mentioned by the UK Panel on Breast Cancer Screening and IQWIG,[88, 129] the most reliable estimates of overdiagnosis come from RCTs in which women in the control group were not offered screening at the end of the trial. This is the case in the Malmö and the Canadian CNBSS trial (I and II). In the other RCTs, "*all the women in the control group were offered screening at the end of the active period of the trial, which itself might be expected to lead to some overdiagnosis and thus to an overall underestimate of overdiagnosis,*"[129] i.e. bias due to the risk of contamination. Combining the results of the Malmö and CNBSS trials results in an overdiagnosis rate of 22.8% (337/1474) (see Table 23). Excluding the CNBSS-1 study, because of an age at randomization outside of the target population, did not meaningfully impact this percentage (22.7%, 238/1048). The lowest percentage was observed in CNBSS-2 (20.7%, 95/460).

In line with our approach to make optimistic assumptions in the base case, we chose the lowest percentage of overdiagnosis of 20.7% from the CNBSS-2 trial and applied the 27.1% in a scenario analysis (scenario 'Overdiagnosis').

Table 23 – Updated estimates of overdiagnosis (Method B)

	Screening		No screening		
RCT	**Cancers (n/%)**	**Participants**	**Cancers (n/%)**	**Participants**	**% Overdiagnosis**
7 RCTs	3906 (1.57%)	249 025	3009 (1.14%)	263 221	1059/3906 (27.1%)
Malmö I	588 (2.79%)	21 088	447 (2.11%)	21 195	143/588 (24.4%)
CNBSS-1	426 (1.69%)	25 214	327 (1.30%)	25 216	99/426 (23.2%)
CNBSS-2	460 (2.33%)	19 711	365 (1.85%)	19 694	95/460 (20.7%)
3 studies	*1474 (2.23%)*	*66 013*	*1139 (1.72%)*	*66 105*	*337/1474 (22.8%)*
2 studies*	1048 (2.57%)	40 799	812 (1.99%)	40 889	238/1048 (22.7%)

Abbreviations: *n: number; RCT: randomized controlled trial.*
Notes: ** The CNBSS-1 includes women of 40-49 years at randomisation. In- or excluding this trial from the above analysis did not impact the percentage of overdiagnosis. The UK Age trial also did not offer screening at the end of the active period of the trial, but is excluded for the same reason (age at randomisation of 39-41 years).*

bbbbb As mentioned by the UK Panel on Breast Cancer Screening,[129] "*Because screening advances detection of breast cancer, follow-up should continue after the screening period to allow catch-up of diagnoses in the unscreened group. In principle, the extended period of follow-up should correspond to the lead time, but the average lead time is debated and is not equal for all cancers. An adequate follow-up is a minimum of 5–10 years after the end of the intervention period.*[130,131]

Due to the high level of bias caused by contamination, we have excluded any data from individual RCTs in which screening was offered to the 'no screening' group at the end of the active period of the trial. This practice leads to some degree of overdiagnosis in the comparator group and thus to an overall underestimation of overdiagnosis in the screening group.[129]

To assess the impact of including or excluding overdiagnosis, four additional hypothetical scenarios are presented, in which an overdiagnosis rate of 0%, 10%, 20% or 30% is applied to the total number of cancers diagnosed in the screening group (scenario 'Overdiagnosis').

Finally, in the model, we assume that the number of overdiagnosed cancers are primarily stage 0 breast cancers. If the number of overdiagnosed cancers exceeds the number of detected stage 0 cases, we assume that the remaining number of overdiagnosed cancers are stage I.

All costs associated with cancers detected through screening are taken into account, i.e. including overdiagnosed cancers. Further information on costs can be found in section 0. For information on the impact on HRQoL, we refer to section 6.2.7.8.

6.2.7.7 Survival and extrapolation

Table 24 shows the information related to the age and stage-specific breast cancer survival.

Table 24 – Overview of input variables (survival)

Variable	Value	Uncertainty	Source
Survival	Age (50-54, 55-59, 60-64, 65-69 years) and stage-specific (stage 0, I, II, III, and IV): see Figure 41 to Figure 44	/	Based on real-world data from the Belgian Cancer Registry (~stages of breast cancer) and the Crossroads Bank for Social Security (~vital status).

In Belgium, it is compulsory for hospitals and pathology laboratories to report all newly diagnosed cancer cases to BCR. The completeness of the incidence data has been estimated to cover at least 98% of all cancer cases in Belgium since 2004.[254] The cancer registry database contains the following patient and tumour characteristics, among others: sex; age at diagnosis; date of incidence (i.e. the date of the first histopathological confirmation of the tumour, or if this is unavailable, the date of the technical procedure or clinical investigation that led to the cancer diagnosis); topography and histology of the tumour (ICD-O-3); and clinical and pathological TNM stage at diagnosis.[275] The BCR is legally entitled to use the patient's unique social security number.[275] As such BCR data can be linked to the individual vital status, transmitted by the Crossroads Bank for Social Security.

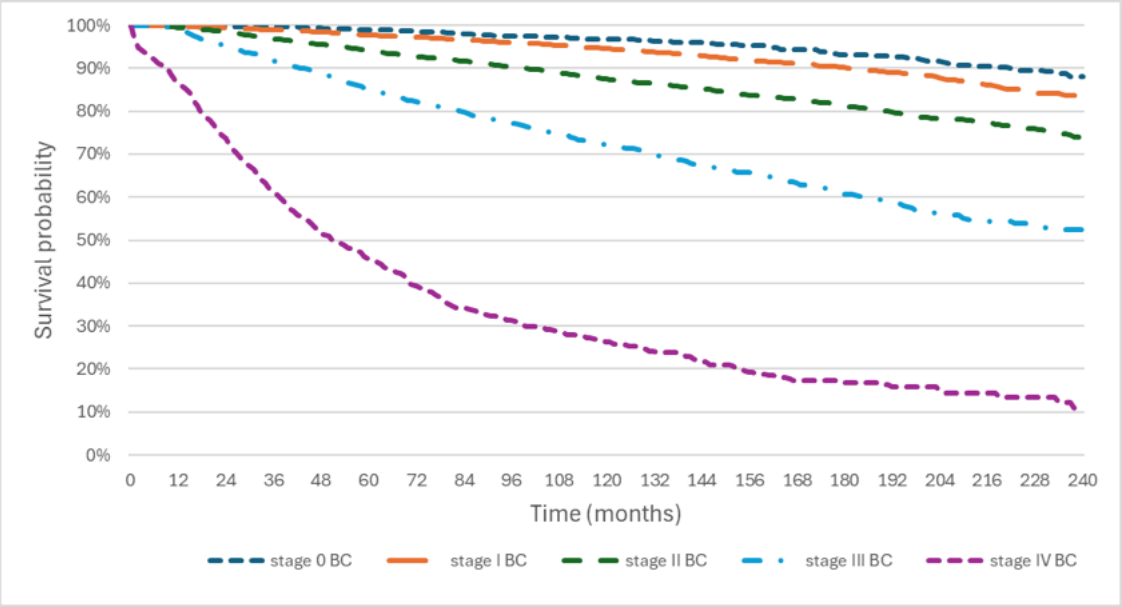
SURVIVAL CURVES

The selected BCR data were restricted to women aged 50–69 years with breast cancer, reflecting the target population. We selected survival data for all patients diagnosed with stage 0, I, II, III or IV breast cancer, not just patients for whom breast cancer was detected after population screening. There are three reasons for this choice: 1) It provides us with the largest possible sample for estimating survival; 2) Survival is primarily determined by the stage of breast cancer, rather than the pre-assessed risk of breast cancer; And 3) the survival data for breast cancer diagnosed after screening are influenced by lead-time bias. However, we cannot rule out the latter, even if we would exclude cancers diagnosed after screening, as our data cannot distinguish between diagnoses following the onset of symptoms and opportunistic screening in the remaining cancer cases. Lead time is therefore present in the survival data, which leads to a more optimistic result in favor of screening. We will address this further in the discussion.

Furthermore, we have chosen to present the survival curves with a follow-up up to 20 years. This minimises the influence of extrapolation assumptions. If we were to limit ourselves to patients with a recent breast cancer diagnosis, we would only be able to produce short-term survival curves. Including

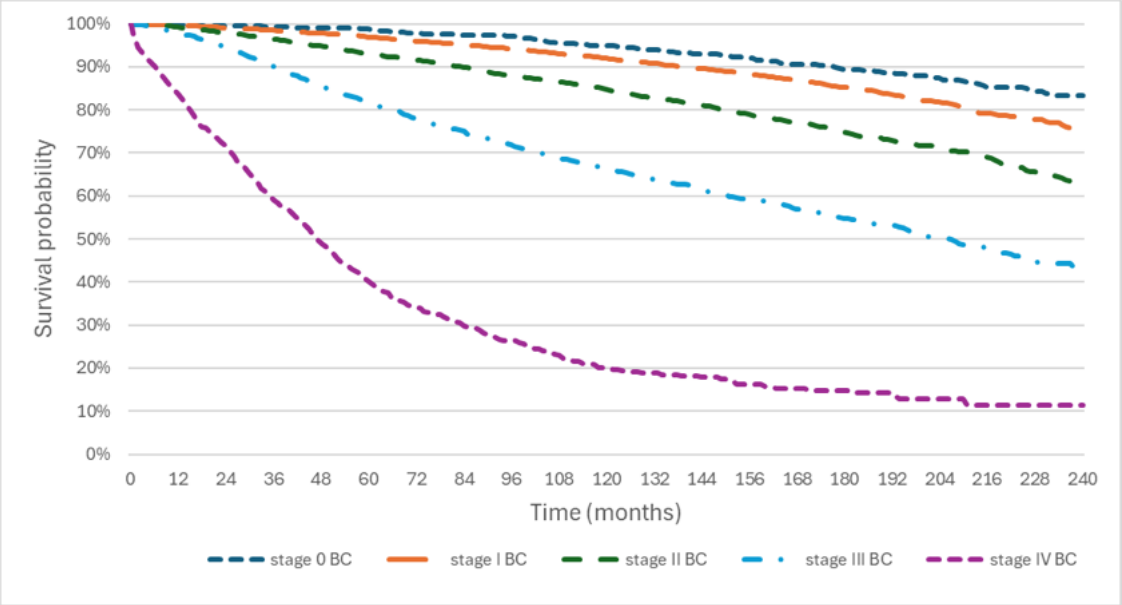
data of patients in whom cancer is diagnosed a long time ago might ignore the impact of recent advances. This shortcoming favours screening. We will also come back to this in the discussion.

Figure 37 – Stage-specific breast cancer survival (age at diagnosis: 50-54 years)



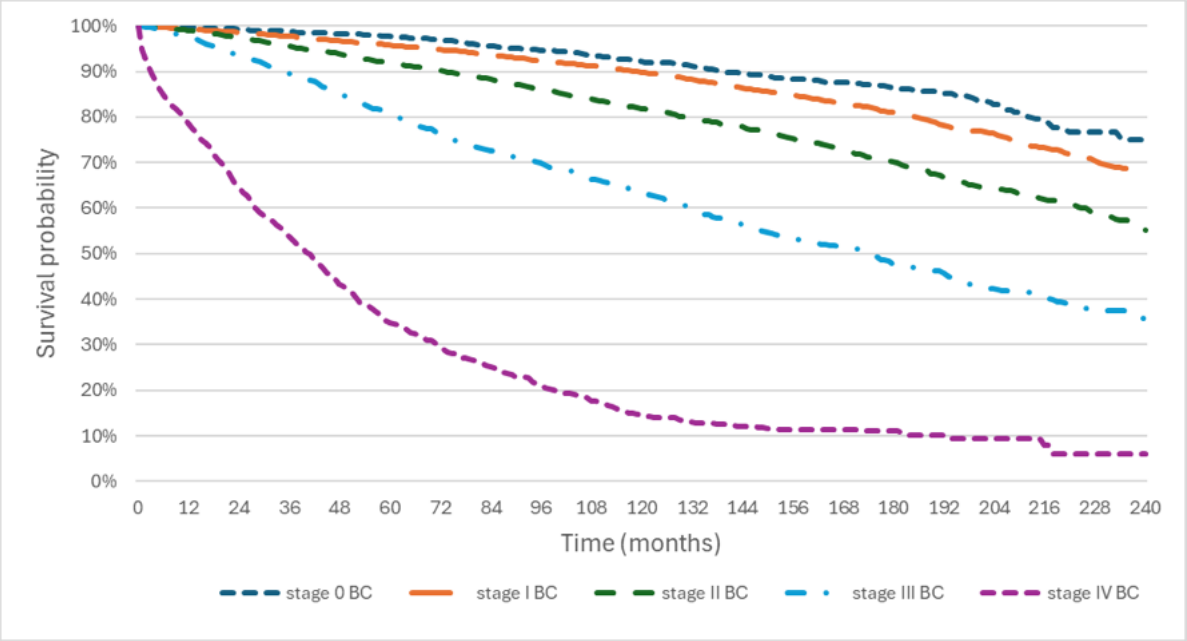
Abbreviations: *BC: breast cancer.*
Notes: *Based on information from 2918, 10 415, 7545, 2325 and 885 patients, for stage 0-IV breast cancer, respectively.*

Figure 38 – Stage-specific breast cancer survival (age at diagnosis: 55-59 years)



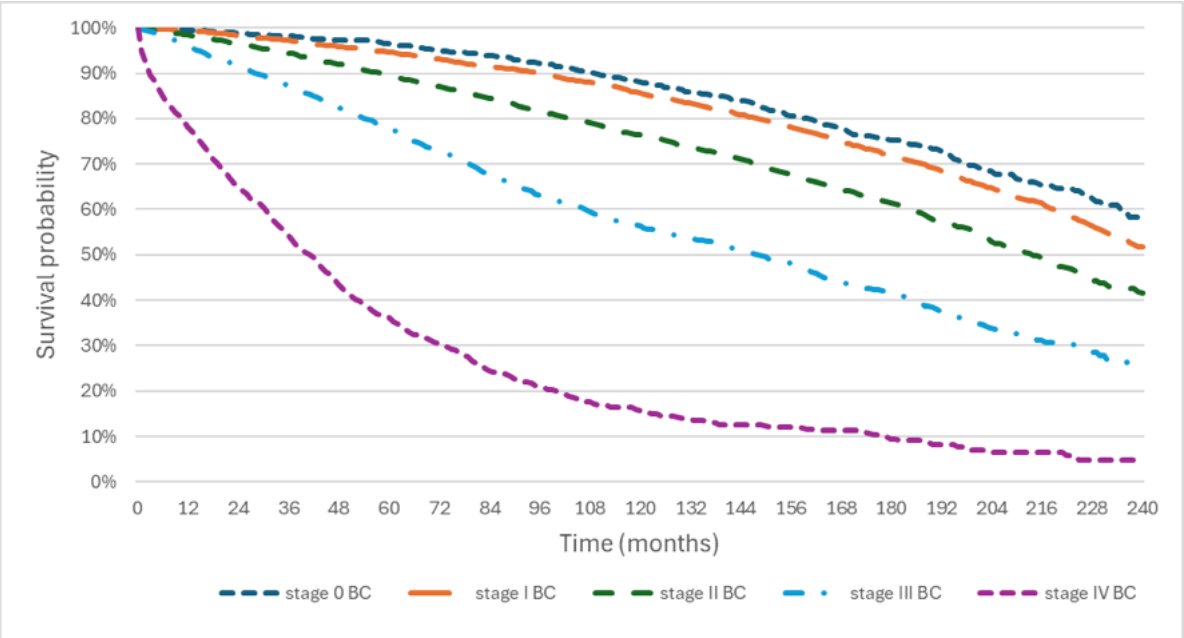
Abbreviations: *BC: breast cancer.*
Notes: *Based on information from 2547, 10 618, 7324, 2264, and 1119 patients, for stage 0-IV breast cancer, respectively.*

Figure 39 – Stage-specific breast cancer survival (age at diagnosis: 60-64 years)



Abbreviations: *BC: breast cancer.*
Notes: *Based on information from 2564, 11 851, 7864, 2308, and 1284 patients, for stage 0-IV breast cancer, respectively.*

Figure 40 – Stage-specific breast cancer survival (age at diagnosis: 65-69 years)



Abbreviations: *BC: breast cancer.*
Notes: *Based on information from 2226, 11 093, 7229, 2304, and 1341 patients, for stage 0-IV breast cancer, respectively.*

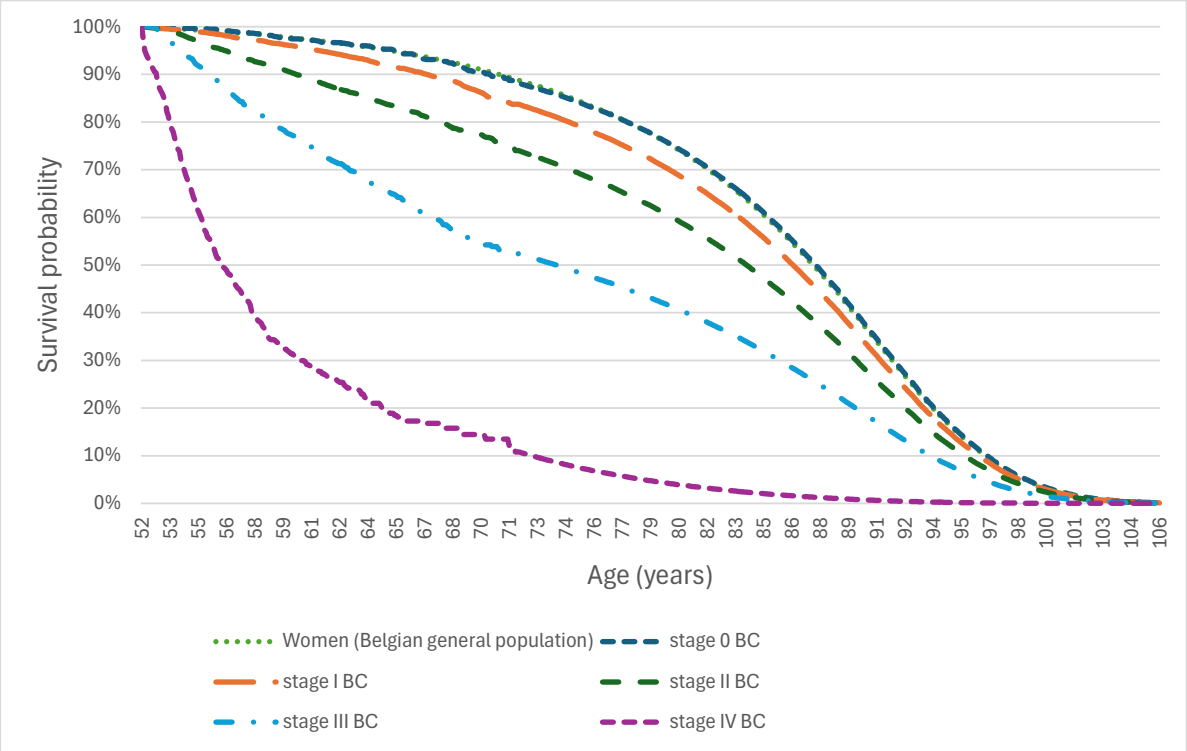
EXTRAPOLATION

The above survival curves are used in the economic model. Adopting a lifetime time horizon is necessary to capture the full possible impact of all relevant incremental costs and effects. Therefore, the 20-year survival curves are extrapolated to the end of life.

Following the Belgian guidelines for economic evaluations, three extrapolation scenarios were considered: 1) the mortality risk remains constant over time; 2) the mortality risk increases with the absolute age-specific increase in the mortality risk of the general female Belgian population; and 3) the mortality risk increases with the relative age-specific increase in the mortality risk of the general female Belgian population. The baseline mortality risk used for extrapolation is determined by the last three years of the 20-year survival curves presented in Figure 37 to Figure 40.[ccccc] We apply the average age for each 5-year age category as the starting age: 52, 57, 62, and 67 years, respectively.

We checked the face validity of these three approaches by comparing the extrapolated survival curves with the age-specific survival curves of the general Belgian female population. The first approach was not considered appropriate because it yielded an unrealistically high life expectancy, with too many patients living past 100 years. Similarly, the third approach was not appropriate for populations diagnosed with a lower stage of breast cancer. Therefore, based on face validity, the second approach was deemed the most appropriate and was applied to the model. Figure 41 through Figure 44 present the extrapolated survival curves that are used in the model. We note that these figures also present the age-specific survival curve (i.e. dotted line) of the female Belgian population. This allows the reader to check the face validity of all applied survival curves.

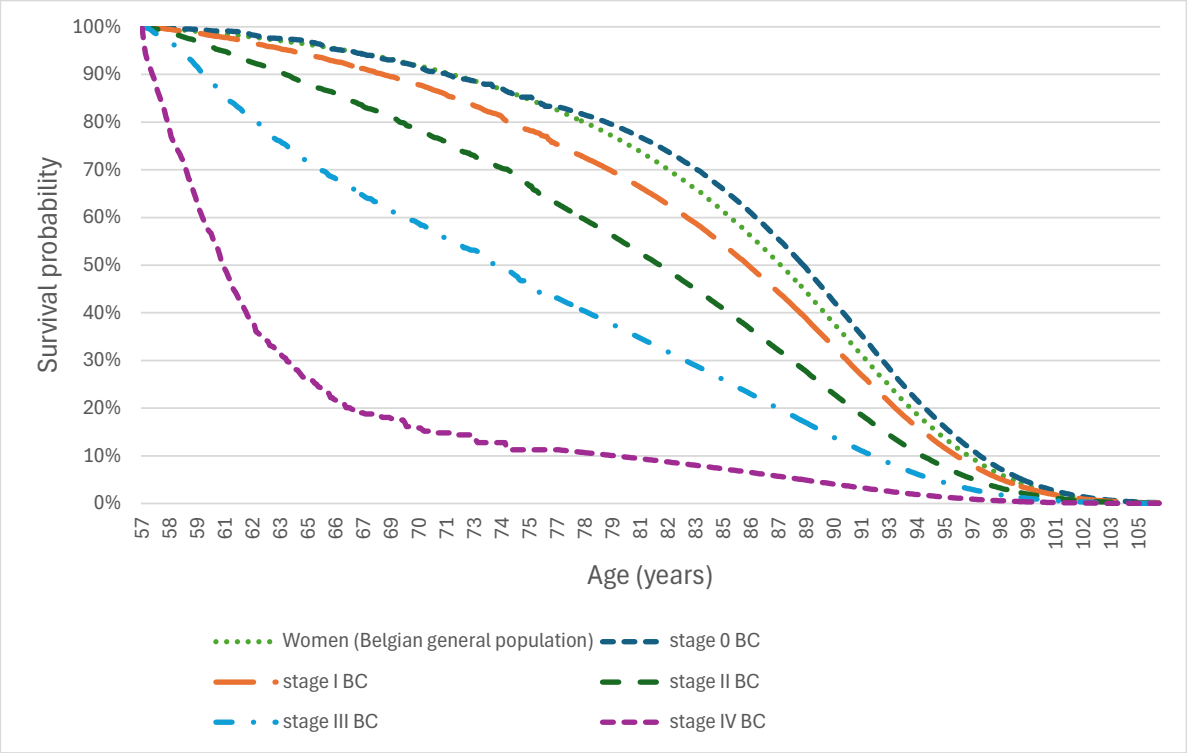
Figure 41 – Stage-specific breast cancer survival – extrapolated (age at diagnosis: 50-54 years)



***Abbreviations**: BC: breast cancer.*
***Remark**: The survival curve for the general female population in Belgium is difficult to distinguish because it is virtually identical to the survival curve for stage 0 breast cancer and lies underneath it in the above figure.*

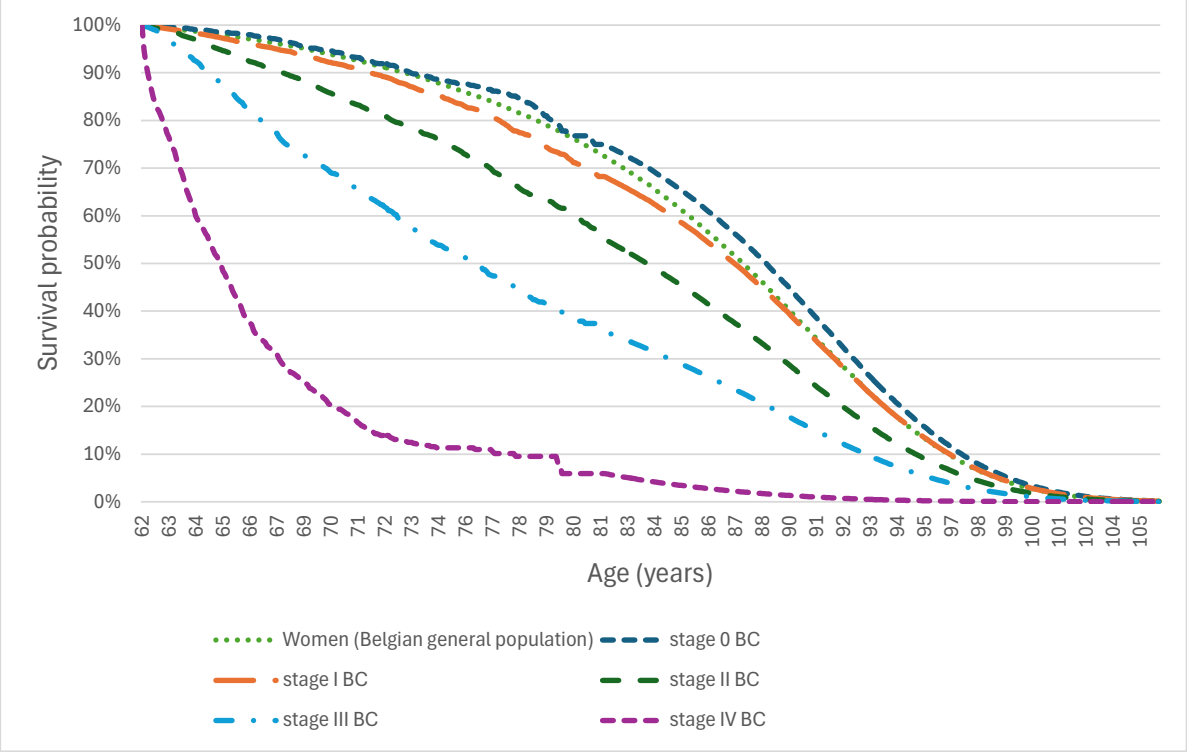
[ccccc] The last three years were chosen to have a relatively stable period on which the extrapolations could be applied. The following formula is applied: monthly mortality risk during the last three years of the 20-year survival curve = 1 – [(1-((Survival month 204 – Survival month 240)/Survival month 204)^(1/36 months)]

Figure 42 – Stage-specific breast cancer survival – extrapolated (age at diagnosis: 55-59 years)



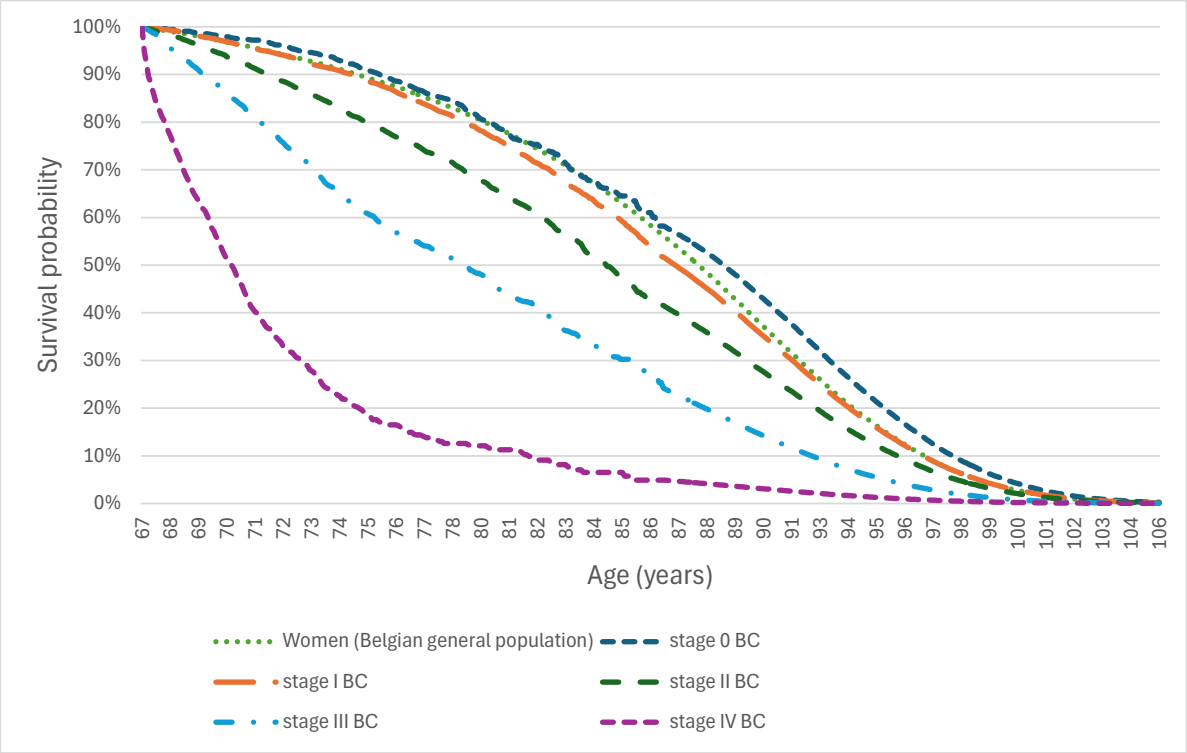
Abbreviations: *BC: breast cancer.*

Figure 43 – Stage-specific breast cancer survival – extrapolated (age at diagnosis: 60-64 years)



Abbreviations: *BC: breast cancer.*

Figure 44 – Stage-specific breast cancer survival – extrapolated (age at diagnosis: 65-69 years)



***Abbreviations**: BC: breast cancer.*

6.2.7.8 Health-related quality of life (utilities)

Table 25 provides an overview of the quality of life-related input variables.

Table 25 – Overview of input variables (utilities)

Variable	Value	Uncertainty	Source
Stage-specific utilities:			
Stage 0	0.85	x_1 = Beta (40.8; 7.2)	Rautenberg, 2022[276]
Stage I	0.85	Equal utility for stage 0, I, and II	
Stage II	0.85	Utility stage III = x_1 - 0.02	
Stage III	0.83	Utility stage IV = x_1 - 0.12	
Stage IV	0.73		
		Scenario 'Utilities': 0.789 (x_1 = Beta (145.2; 38.8) for stage 0, I, and II)**, x_1 - 0.015 for stage III and x_1 - 0.103 for stage IV BC.	Wang, 2018[277]
Disutilities:			
Screening participation	No disutility		Rijnsburger, 2004[216]
False-positive findings	Disutility of 0.07 (95% CI 0.05-0.09)	Beta (43.7; 580.5)	Pillay et al.[278]
		Scenario 'Disutility FP': no disutility	

Variable	Value	Uncertainty	Source
		Scenario 'Disutility FP': disutility also for overdiagnosed true positive findings	
Duration of disutilities FP (& overdiagnosed TP findings)	1 month*	Scenario 'Disutility FP': no disutility	BCR (2022)
Disutility related to treatment of overdiagnosed cancers	No disutility (No values available in literature)	Narrative inclusion in discussion	Expert opinion

Abbreviations: *BC: breast cancer.*
Remark: ** Based on information from true-positive findings (see text for explanation). ** No values were provided for stage 0 in the study of Wang et al.[277] In line with results from other studies, an equal utility was assumed for stage 0 and I. The mean utility value for stage II was 0.004 higher than for stage I (0.793 versus 0.789). To avoid illogical higher values for stage II in comparison with stage I breast cancer, the same utility as for stage I is assumed.*

Breast cancer screening can have positive and negative effects on quality of life. On the positive side, quality of life improves if breast cancer is diagnosed at an earlier stage. Assigning lower utilities to more advanced breast cancers (i.e., higher stages) combined with the modeled stage shift (see section 6.2.7.10) and stage-dependent mortality (see section 6.2.7.7) results in an estimate of the number of QALYs for both the intervention and comparator groups. On the negative side, there is the potential impact of e.g. screening participation, false-positive findings or overdiagnosis.

A non-systematic search strategy was used to identify relevant systematic reviews and meta-analyses on HRQoL (see section 6.1.2.2). Table 108 in Appendix 22 provides an overview of the included and excluded studies and the reasons for exclusion. Seven systematic reviews and meta-analyses were selected. The results of these studies, as well as the underlying primary studies, were used to identify HRQoL scores. In what follows, we present the results of our targeted literature search on HRQoL for stage-specific utilities and screening-related (dis)utilities.

HRQOL – STAGE-SPECIFIC UTILITIES

Table 9 shows the results for stage-specific health utilities. Note that most of the studies, marked with *** in the overview table, were not included in the economic evaluations and therefore do not appear in the summary of the systematic economic literature review (see section 4.2.3).

Table 26 – Overview of studies reporting stage-specific utilities in breast cancer

Study (First author, year)	Population & sample	Methods	Results	Limitations
Pillay, 2024***[278]	Meta-analysis: 2 studies	EQ-5D (3L & 5L)	• Stage II-III vs I: -0.02 95% CI [-0.03, -0.01] • Stage III vs I-II: -0.03 95% CI [-0.05, -0.02]	Stage 0 and IV not included Low certainty of evidence
Rautenberg, 2022***[276]	Meta-analysis: 3 QoL studies in Chinese BC patients	EQ-5D (3L & 5L)	• Stage 0 & I: 0.85 95% CI [0.750, 0.950] • Stage II: 0.85 95% CI [0.780, 0.930] • Stage III: 0.83 95% CI [0.770, 0.900] • Stage IV: 0.73 95% CI [0.630, 0.820]	High heterogeneity Treatment settings highly variable Potential ceiling effect of EQ-5D-3L
Criscitiello, 2021***[279] *Included in Pillay, 2024[278]*	1083 patients diagnosed with stage I to III HR+/HER2– breast cancer from USA, Japan, France, Germany, Italy, Spain, and UK	EQ-5D-5L	• Stage I: 0.879 SD: 0.137 • Stage II: 0.875 SD: 0.129 • Stage III: 0.841 SD: 0.133 • Stage IV: not included	Potential for selection bias All patients were disease free at the moment the questionnaire was completed Subgroup of all breast cancer patients
Gong, 2020***[280]	Meta-analysis: 35 QoL studies (patients & non-patients)** Patients only (16 studies)	EQ-5D, VAS, SG, TTO EQ-5D only (3L & 5L)	• Early stage: 0.742 95% CI [0.624, 0.859] • Late stage: 0.526 95% CI [0.460, 0.592] • Early stage: 0.819 95% CI [~0.676, ~0.967] • Late stage: 0.603 95% CI [~0.476, ~0.750]	Due to variations in defining health states, only early vs late stage was possible Sample size and response rate of the included studies varied greatly, and little information was available regarding missing data, loss-to-follow up, or evaluating the appropriateness of the measure
Porciello, 2020***[281] *Included in Pillay, 2024[278]*	309 women enrolled within 12 months of breast cancer diagnosis without metastasis in Italy	EQ-5D-3L	• Stage I: 0.866 SD: 0.121 • Stage II: 0.851 SD: 0.116 • Stage III: 0.855 SD: 0.105	Sample size Part of randomized controlled trial of the effectiveness of a treatment programme including dietary modification, physical activity and vitamin D supplementation
Guerra, 2019***[282] *Included in Pourrahmat, 2021[283]*	196 diagnosed cancer patients before routine therapy at a Brazilian public cancer centre	EQ-5D-3L	• Stage 0 & I & II: 0.729 95% CI [0.698, 0.760] • Stage III: 0.689 95% CI [0.648, 0.730] • Stage IV: 0.689 95% CI [0.648, 0.730]	Metastatic patients poorly represented in convenience sample Pre-treatment
Li, 2019***[284]	608 breast cancer patients at a tertiary	EQ-5D-5L	• Stage 0 & I: 0.83 SD: 0.17 • Stage II: 0.86 SD: 0.18	One hospital may not be representative

Study (First author, year)	Population & sample	Methods	Results	Limitations
Included in Rautenberg, 2022[276] & Gong, 2020[280]	hospital in eastern China		• Stage III: 0.82 SD: 0.18 • Stage IV: 0.78 SD: 0.22 *VAS & TTO not reported here*	
Ou, 2019***[285] *Included in Rautenberg, 2022[276]*	193 patients with hormone receptor-positive/human epidermal growth factor receptor-2-negative breast cancer at Taiwanese hospital	EQ-5D-5L	• Stage 0: 0.93 SD: 0.06 • Stage I: 0.94 SD: 0.06 • Stage II: 0.91 SD: 0.10 • Stage III: 0.91 SD: 0.12 • Stage IV: 0.91 SD: 0.03	Limited size of sample (e.g. only 3 respondents in stage IV) No random sampling Participants mainly comprised women with HR-positive/HER2-negative breast cancer leading to lack of generalizability to all breast cancers
Wang, 2018***[277] *Included in Rautenberg, 2022[276] and Pourrahmat, 2021[283]*	2626 breast cancer patients from Chinese screening programme	EQ-5D-3L VAS	• Stage 0: not examined • Stage I: ○ EQ-5D: 0.789 95% CI [0.774, 0.805] ○ VAS: 0.736 95% CI [0.724, 0.747] • Stage II: ○ EQ-5D: 0.793 95% CI [0.783, 0.802] ○ VAS: 0.741 95% CI [0.734, 0.748] • Stage III: ○ EQ-5D: 0.774 95% CI [0.759, 0.788] ○ VAS: 0.725 95% CI [0.714, 0.736] • Stage IV: ○ EQ-5D: 0.686 95% CI [0.654, 0.717] ○ VAS: 0.653 95% CI [0.625, 0.682]	Potential ceiling effect of EQ-5D-3L No DCIS and LCIS included
Kim, 2017[213] *Included in Pourrahmat, 2021,[283] Gong, 2020[280] and Kaur, 2022[217]*	509 randomly selected individuals from the Korean general population: male & female >19 years old	VAS (8 scenarios per respondent) SG (1 scenario per respondent)	1) non-invasive BC with mastectomy only • VAS: mean of 0.669 (SD: 0.199) and median of 0.694 • SG: mean of 0.790 (SD: 0.265) and median of 0.900 2) non-invasive BC with mastectomy and followed by reconstruction • VAS: mean of 0.681 (SD: 0.199) and median of 0.700 • SG: mean of 0.804 (SD: 0.260) and median of 0.900 3) non-invasive BC with breast-conserving surgery & radiation therapy • VAS: mean of 0.662 (SD: 0.201) and median of 0.684	Hypothetical bias

Study (First author, year)	Population & sample	Methods	Results	Limitations
			• SG: mean of 0.779 (SD: 0.261) and median of 0.900 4) invasive BC with surgery (mastectomy or breast-conserving surgery), radiation therapy, and/or chemotherapy → Pourrahmat: Stage I & II† • VAS: mean of 0.579 (SD: 0.202) • SG: mean of 0.731 (SD: 0.255) 5) locally advanced BC with radical mastectomy & radiation therapy → Pourrahmat: stage III A&B† • VAS: mean of 0.435 (SD: 0.178) • SG: mean of 0.610 (SD: 0.261) 6) inoperable, locally advanced BC → Pourrahmat: Stage III C† • VAS: mean of 0.415 (SD: 0.173) • SG: mean of 0.587 (SD: 0.259) 7) loco-regional recurrent BC • VAS: mean of 0.333 (SD: 0.176) • SG: mean of 0.496 (SD: 0.260) 8) metastatic BC → Pourrahmat: Stage IV† • VAS: mean of 0.170 (SD: 0.220) • SG: mean of 0.352 (SD: 0.275)	
Kim, 2015***[286] *Included in Gong, 2020*[280]	827 patients visiting the ambulatory cancer centre of a tertiary hospital after breast cancer surgery in Korea	EQ-5D-3L	• Stage 0: 0.925 SD: 0.081 • Stage I: 0.918 SD: 0.093 • Stage II: 0.909 SD: 0.083 • Stage III: 0.895 SD: 0.088 • Stage IV: not reported	Only one hospital, ambulatory setting No stage IV patients included
Lidgren, 2007[204]	361 consecutive breast cancer patients attending outpatient	EQ-5D-3L	First year after primary BC	No random selection of patients

Study (First author, year)	Population & sample	Methods	Results	Limitations
Included in Pourrahmat, 2021,[283] Gong, 2020[280], Kaur, 2022[217] and Peasgood, 2010[287]	clinic at Karolinska University hospital (Sweden)	(TTO not reported here)	• EQ-5D: 0.696 95% CI [0.634, 0.747] → Pourrahmat: Stage I† First year after recurrence • EQ-5D: 0.779 95% CI [0.700, 0.849] → Pourrahmat: Stage II† 2nd & following years after primary BC or recurrence • EQ-5D: 0.779 95% CI [0.700, 0.849] → Pourrahmat: Stage III† Metastatic disease • EQ-5D: 0.685 95% CI [0.620, 0.735] → Pourrahmat: Stage IV†	Selection bias: non-metastatic patients with more severe symptoms more likely treated in outpatient clinic Selection bias: metastatic patients with less severe symptoms more likely treated in outpatient clinic
Schleinitz, 2006[203] *Included in Pourrahmat, 2021,[283] Gong, 2020[280], Kaur, 2022[217] and Peasgood, 2010[287]*	Age- and race-stratified sample of 156 women not undergoing BC treatment above 25 years old in the USA	SG (for BC stages) TTO (for therapeutic modalities)	• Stage I: 0.68 SD: 0.06 • Stage II: 0.61 SD: 0.06 • Stage III: 0.56 SD: 0.06 • Stage IV (oestrogen receptor positive): 0.41 SD: 0.06 • Stage IV (oestrogen receptor negative): 0.42 SD: 0.06	Hypothetical bias

Abbreviations: *BC: breast cancer; BCS: breast cancer screening; BSE: breast self-examination; CBE: clinical breast examination; QoL: Quality of Life; RT: radiotherapy; SD: Standard Deviation; SG: Standard Gamble; TTO: Time Trade-Off; USA: United States of America; VAS: Visual Analogue Scale; y: year.*

Notes: *† These studies do not present stage-specific utility values. However, Pourrahmat's systematic review[283] reclassified these health states into stage-specific utility values.*

*** The list of underlying studies was checked. No additional studies were identified (beyond those already included in this overview) that report stage-specific utilities for all invasive stages.*

**** These studies were not mentioned in the HRQoL overview of the identified economic literature review.*

Both the Belgian guidelines for economic evaluations[3] and the EUnetHTA guidelines for measuring HRQoL[288] recommend that patients assess their own health states. Of the studies mentioned above, several conduct their analysis on individuals without breast cancer. The meta-analysis by Gong et al.,[280] which includes 35 HRQoL studies, combines results from patients and non-patients. Kim et al.[213] included 509 randomly selected individuals from the general Korean population. Schleinitz et al.[203] included 156 women who were not undergoing breast cancer treatment. These studies observe the greatest differences in utilities: a 0.41 difference between stages I and II versus stage IV breast cancer, based on the utilities in the study of Kim et al.,[213] which Pourrahmat et al.[283] classified into breast cancer stages; and a 0.27 difference between stage I and estrogen receptor-positive stage IV breast cancer.[203] Pourrahmat et al.[283] also reclassified the results of Lidgren et al.[204] into breast cancer stages. This approach involves additional uncertainty and is therefore not preferred.

We focus on utilities based on research involving women with breast cancer from studies presenting results for multiple stages of breast cancer (0, I, II, III, and IV). We examine the differences in utilities across these stages, preferably using a generic utility instrument, such as the EQ-5D. We observe the following:

- No or only minimal differences are observed between stages 0/I and II:
  - St. 0/I vs II: 0.00[276]
  - St. I vs II: 0.00[279]
  - St. I vs II: 0.02[281]
  - St. 0/I vs II: -0.03[284]
  - St. 0/I vs II: 0.02-0.03[285]
  - St. I vs II: 0.00[277]
  - St. 0/I vs II: 0.01-0.02[286]
- Minor differences of 0.01–0.04 between stages (0)/I/II and stage III:
  - St. I/II vs III: 0.03[278]
  - St. 0/I/II vs III: 0.02[276]
  - St. I/II vs III: 0.03-0.04[279]
  - St. I vs III: 0.01[281]
  - St. 0/I/II vs III: 0.04[282]
  - St. 0/I and II vs III: 0.01 and 0.04, respectively[284] (and an unlogic lower value for stage I versus stage II)
  - St. 0/I vs III: 0.02-0.03[285]
  - St. I/II vs III: 0.02[277]
  - St. 0/I vs III: 0.03[286]
- Larger differences are observed comparing stages (0)/I/II with stage IV:
  - St. 0/I/IIvs IV: 0.12[276]
  - St. 0/I/II vs IV: 0.04[282]
  - St. 0/I and II vs IV: 0.05 and 0.08, respectively[284] (and an unlogic lower value for stage I versus stage II)
  - St. 0/I vs IV: 0.02-0.03[285]
  - St. I/II vs IV: 0.10[277]

Although the studies were conducted in a Chinese population, the studies by Rautenberg et al.[276] and Wang et al.[277] included the largest sample of breast cancer patients. Rautenberg et al. also covered all relevant stages of breast cancer. This study identified the greatest differences in utilities between stages 0/I/II and stage IV breast cancer, based on results obtained from breast cancer patients. In the base case, the utilities from Rautenberg et al. are included. Taking utilities from other studies involving breast cancer patients would have a negative impact on the intervention's cost-effectiveness. We will address this point in the discussion.

Utilities are assigned based on the stage of breast cancer at the time of diagnosis. There is no information on how breast cancer progresses over time. In all identified HRQoL studies, utilities are not differentiated according to age. Due to a lack of such information, and to avoid more complexity, we assume an equal utility for an aging population in our study. This is also a choice in the advantage of breast cancer screening. This will be explored in greater detail later in the discussion.

Consequently, the following average utility values are assigned to the different stages in our hypothetical economic evaluation: 0.85 for stages 0, I and II, 0.83 for stage III and 0.73 for stage IV breast cancer. In a scenario analysis (scenario 'Utilities'), we include the values published by Wang et al.[277]

HRQOL – BREAST CANCER SCREENING-RELATED (DIS)UTILITIES

Table 27 provides an overview of the information identified relating to screening-specific (dis)utilities.

Table 27 – Overview of studies reporting QoL for screening-specific events or health states

Study (First author, year)	Population & sample	Methods	Results	Limitations
Pillay, 2024***[278]	/	Meta-analysis Screening mammography, biopsy, false positive: 3 studies using EQ-5D-3L True positive: EQ-5D-3L (5 studies), EQ-5D-5L (3 studies), 15D (1 study)	Disutility of positive **screening mammography** before diagnostic testing: 0.07 95% CI [0.05, 0.09] (moderate certainty of evidence) Disutility after **biopsy**: 0.15 95% CI [0.11, 0.19] (very low certainty of evidence) Disutility **false positive** requiring imaging only or imaging plus biopsy: 0.04 95% CI [0.03, 0.05] (low certainty of evidence) Disutility of **true positive** (screen-detected BC): 0.08 95% CI [0.07, 0.09] (moderate certainty of evidence) *All disutilities are based on a reference utility of 0.94 95% CI [0.93, 0.94]*	Due to machine learning in screening articles, potentially some articles were missed *Heterogeneity not assessed*
Timmers, 2014***[289] *Included in Pillay, 2024[278] & Li, 2019[205]*	366 Dutch women who were screened and recalled with BI-RADS 0	EQ-5D-3L	Directly after the women were told that their screening examination was positive but before the assessment in the hospital: 0.89 After completion of the diagnostic process: 0.91	*Only BI-RADS 0 women*
Tosteson, 2014[218] *Included in Pillay, 2024[278]*	1028 randomly selected screening trial participants (positive & negative results) without BC	EQ-5D-3L with US scoring	*Difference between true-negative (TN) and false-positive (FP) groups:* ○ Before work-up showed negative result in FP group (woman unaware of not having BC): 0 *Results here only reported for part of the sample interviewed before results of the diagnosis. Women with diagnosed breast cancer in trial excluded*	Trial participants might not be representative for full eligible population
Domeyer, 2010***[290] *Included in Pillay, 2024[278]*	102 patients undergoing vacuum-assisted breast biopsy (VABB) and having benign lesions	EQ-5D-3L	Morning before VABB: 0.729 (SD: 0.224) 4 days after VABB: 0.787 (SD: 0.208) 18 months after VABB: 0.769 (SD: 0.225)	VABB was exclusively performed under stereotactic guidance Volume of tissue removed in comparison to other settings is high *Only BI-RADS 3 & 4 women with non-palpable lesions*
Grann, 2010***[291]	160 women without a personal or family	TTO	Screening mammography: • Non-BRCA carriers: 0.97 (SD: 0.11)	*Scenario-based approach (hypothetical bias)*

Study (First author, year)	Population & sample	Methods	Results	Limitations
Included in Kaur, 2022[217]	history of BC or ovarian cancer & 83 women who carry BRCA gene in USA		• BRCA carriers: 1.00 (SD: 0.004)	
Bonomi, 2008[215] *Included in Peasgood, 2010*[287] *& Li, 2019*[205]	131 randomly selected women between 50 & 79 years participating in BCS in the USA	VAS	Screening mammography: 0.804 (SD: 0.143) for 3 weeks False positive after diagnostic work-up: 0.810 (SD: 0.147) False negative after diagnostic work-up: 0.485 (SD: 0.207) for 2 weeks True positive after diagnostic work-up: 0.457 (SD: 0.205) for 2 weeks or until surgery True negative after diagnostic work-up: 0.891 (SD:0.098) Diagnostic mammography after uncertain screening result: 0.553 (SD: 0.197) for 2 weeks	None of the scenarios included prognostic or life-expectancy information *Scenario-based approach (hypothetical bias)*
Rijnsburger, 2004[216] *Included in Li, 2019*[205]	329 Dutch participants of magnetic resonance imaging screening (MRISC) study (high-risk population) without screen-detected or interval cancer	EQ-5D-3L (UK value set) *(SF-36 & VAS not reported here)*	Negative screening result: 0.88 (25^{th} – 75^{th} percentile: 0.80-1.00) *QoL 2 months prior to screening: 0.88 (25th – 75th percentile: 0.80-1.00)*	*High-risk population* *No random selection* *Intervention = biennial CBE + annual MRI + annual mammography + monthly BSE* *No analysis for (false-)positive result*
Gerard, 1999*[219] *Included in Peasgood, 2010*[287] *& Li, 2019*[205]	440 women eligible for BCS selected from registers of 4 local UK BCS centres	EQ-5D-3L (only 'usual activity', 'pain/discomfort', 'anxiety/depression' dimensions with other assumed non-impacted by BCS; UK value set) *(TTO not reported here)*	True negative: 0.94 (SD: 0.14) for 12 months (no significant difference with full health) False positive: 0.79 (SD: 0.21) for 12 months (significant difference with full health, but significance level not reported) True positive: 0.48 (SD: 0.30) for remaining life expectancy (significant difference with full health, but significance level not reported) False negative: 0.45 (SD: 0.30) for remaining life expectancy (significant difference with full health, but significance level not reported)	Not all EQ-5D dimensions were scored *Scenario-based approach (hypothetical bias)*

Study (First author, year)	Population & sample	Methods	Results	Limitations
de Haes, 1991[206] *Included in Peasgood, 2010[287] & Li, 2019[205]*	15 employees of Department of Public Health and Social Medicine and 12 experts in breast cancer treatment & epidemiology in the Netherlands	VAS (horizontal line without subdivision) *Converted VAS scores to TTO (function & original source unretrievable)*	Screening participation: 0.92 (SD: 0.069) converted to 0.994 for 1 week Diagnostic phase: 0.71 (SD: 0.155) converted to 0.867 for 5 weeks	Difficult questionnaire for non-experts Experts & patients have different preferences *Non-transparent conversion to TTO values* *Scenario-based approach (hypothetical bias)*

***Abbreviations**: BC: breast cancer; BCS: breast cancer screening; BSE: breast self-examination; CBE: clinical breast examination; QoL: Quality of Life; SD: Standard Deviation; SG: Standard Gamble; TTO: Time Trade-Off; USA: United States of America; VABB: vacuum-assisted breast biopsy; VAS: Visual Analogue Scale; y: year.*
***Notes**: Only relevant health states and events were reported. *Johnston et al. (1998)[220] was excluded since they used the same respondents as Gerard et al. (1999) with another instrument (VAS). As Gerard et al. (1999) reported the EQ-5D-3L results, the choice was made to only report this study.*
**** These studies were not mentioned in the HRQoL overview of the identified economic literature review.*

The systematic review of Pillay et al.[278] examined the relative importance placed by patients on the potential benefits and harms of mammography-based breast cancer screening. Their search was performed up to June 2023 and also searched for studies reporting disutilities related to screening. "*Disutilities from outcomes during the breast cancer screening process were calculated in comparison with (subtracted from) the pooled utility (0.94) of a healthy screen-eligible population including individuals scheduled for genetic counseling (n = 33; results of genetic testing unknown), with a known negative screening result (n = 531), and healthy age and education matched comparators (n = 7992).*"[278] As this method is indirect, it involves a high degree of uncertainty and should therefore be interpreted with caution.

Amongst others, the following relevant information was presented in their review:[278]

- Disutility of screening test process: no evidence identified
- Disutility of positive screening mammography (before diagnostic testing): "*The disutility value for a positive screening mammography is probably 0.07.*" (based on 3 studies[218, 289, 290] with 565 participants, moderate certainty of evidence).
- Disutility of knowledge of a false positive requiring imaging only or imaging plus biopsy: *"The disutility value for having knowledge of a false positive requiring imaging only or imaging plus biopsy may be 0.03 to 0.04."* (based on 2 studies[218, 289] with 696 participants, low certainty of evidence)
- Disutility of false-positive result requiring imaging plus biopsy: *"We are very uncertain about the disutility of a false-positive result after invasive testing."*
- Disutility of a true-positive result, before treatment: *"The disutility of a screen-detected cancer is probably on average 0.08, but may be higher for older ages and advanced stage operable cancer."* (based on 9 studies[ddddd] with 6657 participants, moderate certainty of evidence)

The study by Pillay et al.[278] used the results of the three studies of Timmers et al.,[289] Tosteson et al.,[218] and Domeyer et al.[290] Although all these individual studies have their limitations (see Table 27), which therefore also apply to the meta-analysis, we have chosen to use the results of the systematic review by Pillay et al. According to the analysis by Pillay et al.[278] the disutility of a positive screening mammography prior to diagnostic testing was estimated to be 0.07 (95% CI: 0.05-0.09) (moderate certainty of evidence).

Pillay et al. also report separate results for disutilities following biopsy (0.15), false positives requiring imaging alone or imaging plus biopsy (0.04), and true-positive findings (0.08). However, we prefer to apply the general results of a positive test, since the screening participant cannot distinguish between the status of the test at that point (i.e. whether it is a true- or a false-positive finding). To assess the impact of including this disutility, a scenario was run in which it was not included (scenario 'Disutility PF').

Studies including a disutility for false-positive results only apply it to such findings. However, as this is an incremental analysis, comparing the impact of screening with no screening, this disutility could also be applied to true-positive findings representing overdiagnosis. Those that do not constitute overdiagnosis would also be confronted with a cancer diagnosis later and experience this disutility. Neglecting the difference in timing, which would be impacted by the discount rate, favours screening (i.e. a conservative assumption). Nevertheless, we also include a scenario in which the disutility is also applied to overdiagnosed cancers (scenario 'Disutility PF').

The inconvenience caused by a false-positive test result is only temporary. This is the period between a positive screening test result being identified and the final diagnosis of a false positive being made. However, the Belgian database does not provide an exact date on which it is reported that the result

---

[ddddd] We refer to the study of Pillay et al.[278] for the full references of the underlying studies.

is a false positive. Only the incidence date for true-positive tests is available. This is why our analysis is based on the information available for true-positive test results. We calculated the duration between the positive screening test and the date of breast cancer incidence, which on average is 26.6 days. Based on this, we assume that the above disutility lasts for 1 month.

Rijnsburger et al.[216] analysed HRQoL of 329 Dutch participants in an MRI-based screening programme in which no screen-detected or interval cancers were identified. HRQoL remained unchanged two months prior to participation in the screening programme and after receiving a negative result. However, it is unclear whether there may have been a decline in HRQoL while participants were awaiting the results, due to the potential stress this may have caused. The study of Grann et al.[291] used a TTO approach to estimate, among others, the preference ratings for annual mammography in both mutation and non-mutation carriers. In non-mutation carriers, the mean preference rating for mammography was 0.97, meaning mammography would only have a slight effect on quality of life. Based on these two studies, and in line with previous conservative assumptions in favour of screening, no decline in HRQoL linked to screening participation is assumed.

The other studies included in the non-systematic review of (dis)utilities yield results that are difficult to incorporate into the economic evaluation. The studies by Bonomi et al.[215] and Gerard et al.[219] use a scenario-based approach in which, for example, the impact of a false-negative result is examined. In reality, however, patients will not be able to distinguish between true- and false-negative results, which would therefore yield the same HRQoL. By contrast, the hypothetical scenarios in both studies result in a significant HRQoL impact: 0.89 versus 0.49[215] and 0.94 versus 0.45[219] for true- and false-negative results, respectively. This highlights the problem of using hypothetical scenarios that do not reflect the actual HRQoL impact. The study by de Haes et al.,[206] dating back to 1991, was also not conducted among breast cancer screening participants and employs a scenario-based approach. The use of hypothetical values was avoided wherever possible, with preference being given to values extracted from the appropriate populations.

Finally, we note that none of the studies in the clinical or economic part of this report identified or included a disutility for overdiagnosis, which is linked to overtreatment. We also encountered this issue in the KCE study on lung cancer screening.[264] "*No QoL information was identified related to the loss of QoL due to overdiagnosis. Some argue that people with an overdiagnosed stage I lung cancer might not have a major loss in QoL because these people are 'relieved' lung cancer was detected so early and they thank their lives due to the early screening (expert opinion). On the other hand, there is still the impact on QoL due to the 'unnecessary' treatment and because these people are faced with the (unnecessary) diagnosis of cancer. Due to the lack of evidence, no disutility was modelled.*" In this report, we have chosen not to include a disutility for neither the diagnosis nor the treatment of overdiagnosed cancers in the base case scenario, which again favours screening. In a scenario analysis, only a disutility for the diagnosis of overdiagnosed cancers is applied (see above). We address this element further in the discussion.

6.2.7.9 *Diagnosis, follow-up and treatment costs*

Table 28 refers to the input (tables) related to the costs of diagnosis, follow-up and treatment of patients aged 50-69 years in which breast cancer was detected after organised screening.

Table 28 – Overview of input variables (costs related to diagnosis, follow-up and treatment of breast cancer – per BC stage)

Variable	Value	Uncertainty	Source
Stage 0 Stage I Stage II Stage III Stage IV	Health care payer costs (excl. suppl.) presented in Table 29 to Table 36 and Table 39 to Table 42	Normal distribution (mean & 95% CI +/- 20%) (see section 6.2.7.12) Scenario 'Supplements': inclusive supplements	BCR (2018-2022)* IMA (2022-2023)*
Percentage under MEA (per stage)	See Table 37	/	BCR (2018-2022)* IMA (2022-2023)*
Discount for expenditure under MEA	40%	Scenario 'Price discount': 0, 10, 20, 30, 40, 50, 60, 70, 80, and 90% (fixed percentages)	KCE report 400[3] and MORSE report[292]

BC: breast cancer; excl. suppl.: exclusive supplements; MEA: managed entry agreement.
***Remark**: * Cost information is based on IMA data (from January 2022 to December 2023) for breast cancers diagnosed in 2022 (costs 1st year), 2021 (costs 2nd year), 2020 (costs 3rd year), 2019 (costs 4th year) and 2018 (costs 5th year).*

We refer to Chapter 5 for an overview of the available databases and methodological approaches to extract Belgian costs related to diagnosis and treatment of breast cancer. In this section, the costs are shown for six categories: diagnosis and follow-up; radiotherapy; surgery; systemic therapy; palliative care; and miscellaneous. We refer to Appendix 24.1 for an overview of all included Nomenclature codes per cost category.

We use cost information from patients aged 50–69 years who were diagnosed with breast cancer. Costs are presented per breast cancer stage. Initially, we only selected patients with screen-detected cancers. However, due to the small number of patients with stage IV breast cancer in this sample (less than 10 patients in the third, fourth and fifth year after diagnosis), we decided to extract stage-specific cost information from a larger sample of all patients diagnosed with breast cancer between the ages of 50 and 69.[eeeee] This resulted in more stable cost information with increased face validity. Furthermore, the cost differences between the higher and lower stages of breast cancer were higher (see section 5.2.1.2), favouring breast cancer screening. To avoid too many small cells, we chose not to break down these costs by age group. This does not affect the results, as we are calculating the cost-effectiveness of the 50-69 age category as a whole.

We have used the most recent available cost information. The costs for the first year are based on incidence year 2022 (the most recent year available with tumour registration). The first-year cost data for patients diagnosed with cancer between January and December 2022 includes costs from January

[eeeee] The number of patients present at mid-year in the 2nd, 3rd, 4th, and 5th year for stage IV breast cancer (aged 50-69 years) was 208, 202, 155, and 140, respectively.

2022 (i.e. the first month for patients diagnosed in January 2022) up to December 2023 (i.e. the twelfth month for patients diagnosed in December 2022).

Cost data from IMA is available for up to five years after cancer incidence. Cost information for subsequent years is based on the following incidence years: 2021 for the second year, 2020 for the third year, 2019 for the fourth year, and 2018 for the fifth year. Consequently, the IMA data linked to the BCR data always reflects costs incurred between January 2022 and December 2023. We chose not to adjust the cost data using indexation.

Finally, the costs are expressed per stage at diagnosis. If disease progression occurs during the five-year follow-up period, this is reflected in the presented costs. This means that the impact of disease progression on costs is implicitly incorporated into the underlying cost data. Thus, no separate modelling of disease progression is needed, which would be complex given the lack of supporting data. We will return to this advantage in the discussion.

HALF-CYCLE CORRECTION AND THE DURATION OF THE TIME PERIOD

To calculate the average cost, the total cost must be calculated over a period of time and then divided by the number of patients in the sample. However, the number of patients in the sample decreases over time due to mortality. Therefore, it is incorrect to use the number of patients at the beginning or end of the period as the denominator. One solution is to apply the half-cycle correction. Here, the denominator is the number of patients halfway through the time period. When applying this approach, it is important to check that the period is not unnecessarily long or short, i.e. that the number of patients in the sample at the halfway point is approximately equal to the average number of patients at the beginning and end of the period. We applied the same approach as in the KCE report on lung cancer screening, in which the average costs are calculated for the following ten time periods: the first month, the second month, the third month, months four to six, months seven to nine, months ten to twelve, year two, year three, year four and year five. For costs after the fifth year, we assume they are equal to the costs reported for the last available fifth year.

IN- AND EXCLUDING SUPPLEMENTS

In accordance with the Belgian guidelines for economic evaluations,[3] payments from the government's healthcare budget, as well as patient co-payments, are included. In this base case, supplements are excluded. In a scenario analysis, however, patients' supplements are included (scenario 'Supplements').

The following section presents the costs, both with and without additional supplements, for the six cost categories.

DIAGNOSIS AND FOLLOW-UP

Table 29 and Table 30 present the diagnosis and follow-up costs, excluding and including supplements, respectively. Higher costs are observed in the month before or after the date of diagnosis. In subsequent years, follow-up costs are relatively lower for patients with stage 0 or I breast cancer and substantially higher for patients with stage IV breast cancer.

Table 29 – Diagnosis and follow-up costs, excluding supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
Pre-period	€920	€779	€791	€893	€1066
1st month	€426	€917	€1189	€1350	€1478
2nd month	€236	€343	€381	€405	€476
3rd month	€102	€141	€222	€327	€305
Months 4-6	€90	€229	€455	€763	€909
Months 7-9	€52	€152	€302	€514	€717
Months 10-12	€47	€134	€167	€277	€631
Total 1st year	€1873	€2694	€3507	€4529	€5582
2nd year	€193	€444	€610	€879	€2374
3rd year	€146	€399	€520	€714	€2271
4th year	€144	€367	€457	€707	€2083
5th year	€129	€308	€389	€607	€2049

Abbreviations: *BC: breast cancer;*
Notes: *Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

Table 30 – Diagnosis and follow-up costs, including supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
Pre-period	€953	€804	€817	€922	€1131
1st month	€440	€936	€1212	€1370	€1536
2nd month	€241	€352	€390	€415	€490
3rd month	€103	€143	€227	€334	€317
Months 4-6	€94	€237	€467	€780	€931
Months 7-9	€55	€156	€309	€523	€728
Months 10-12	€50	€138	€171	€281	€641
Total 1st year	€1936	€2767	€3592	€4624	€5775
2nd year	€199	€461	€632	€916	€2460
3rd year	€153	€413	€539	€729	€2319
4th year	€152	€381	€474	€725	€2131
5th year	€133	€318	€398	€624	€2087

Abbreviations: *BC: breast cancer.*
Notes: *Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

RADIOTHERAPY

Table 31 and Table 32 present the radiotherapy costs, excluding and including supplements, respectively. In general, the highest costs related to radiotherapy are incurred in the first year after diagnosis. The higher the stage, the later in time during this first year the highest costs of radiotherapy are incurred. The costs after the first year are relatively low compared to the first year.

Table 31 – Radiotherapy costs, excluding supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€16	€45	€13	€9	€377
2nd month	€253	€285	€80	€0	€251
3rd month	€1011	€1497	€550	€55	€162
Months 4-6	€1165	€1360	€1056	€196	€203
Months 7-9	€81	€474	€1587	€2494	€651
Months 10-12	€20	€82	€461	€989	€520
Total 1st year	€2546	€3744	€3747	€3744	€2163
2nd year	€31	€17	€69	€134	€487
3rd year	€16	€16	€66	€154	€535
4th year	€54	€12	€51	€130	€554
5th year	€17	€21	€43	€92	€320

Abbreviations: *BC: breast cancer.*
Notes: *Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

Table 32 – Radiotherapy costs, including supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€16	€54	€15	€9	€379
2nd month	€257	€301	€82	€0	€254
3rd month	€1017	€1526	€555	€55	€162
Months 4-6	€1215	€1378	€1082	€196	€204
Months 7-9	€81	€483	€1606	€2506	€671
Months 10-12	€20	€83	€461	€989	€570
Total 1st year	€2606	€3824	€3801	€3755	€2240
2nd year	€31	€17	€73	€156	€513
3rd year	€16	€16	€66	€154	€535
4th year	€55	€12	€51	€130	€554
5th year	€17	€21	€43	€104	€340

Abbreviations: *BC: breast cancer.*
Notes: *Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

SURGERY

Table 33 and Table 34 present the costs related to surgery, excluding and including supplements, respectively. Costs related to surgery generally increase between stages 0 and III, with the largest share of costs incurred in the first year after diagnosis. In stage IV, the costs related to surgery are substantially lower. While for all other cost categories the inclusion of supplements minimally increases costs, the impact is larger for surgery with an average difference of more than €1000 during the first year for stage 0-III breast cancers. In absolute terms, this increase is much lower in stage IV breast cancer.

Table 33 – Surgery costs, excluding supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€1145	€1673	€1273	€930	€84
2nd month	€1121	€1390	€1329	€1257	€135
3rd month	€473	€251	€210	€221	€21
Months 4-6	€315	€258	€696	€940	€307
Months 7-9	€221	€181	€993	€1821	€691
Months 10-12	€190	€75	€109	€120	€139
Total 1st year	€3463	€3828	€4609	€5288	€1377
2nd year	€370	€243	€690	€802	€120
3rd year	€153	€102	€404	€780	€150
4th year	€174	€92	€249	€456	€120
5th year	€117	€66	€131	€208	€134

***Abbreviations**: BC: breast cancer.*
***Notes**: Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

Table 34 – Surgery costs, including supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€1494	€2162	€1602	€1145	€109
2nd month	€1426	€1806	€1661	€1618	€156
3rd month	€656	€323	€265	€239	€21
Months 4-6	€416	€323	€835	€1074	€397
Months 7-9	€309	€221	€1147	€2096	€839
Months 10-12	€261	€107	€142	€172	€206
Total 1st year	€4562	€4942	€5652	€6345	€1727
2nd year	€494	€323	€878	€1073	€141
3rd year	€210	€135	€558	€1007	€187
4th year	€227	€120	€312	€653	€137
5th year	€145	€96	€166	€227	€212

***Abbreviations**: BC: breast cancer.*
***Notes**: Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

SYSTEMIC THERAPY

Table 35 and Table 36 present the systemic therapy costs, excluding and including supplements, respectively. A clear trend is visible with substantially increasing costs with increasing stage and remaining high costs over the years in stage IV breast cancer.

Table 35 – Systemic therapy costs, excluding supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€3	€74	€537	€696	€2280
2nd month	€4	€233	€1376	€1744	€3641
3rd month	€3	€394	€1765	€2537	€4068
Months 4-6	€6	€1086	€4980	€6975	€11 136
Months 7-9	€6	€777	€3668	€4444	€9659
Months 10-12	€14	€716	€3886	€6548	€9044
Total 1st year	€35	€3280	€16 214	€22 943	€39 829
2nd year	€56	€596	€2964	€6422	€34 098
3rd year	€61	€286	€1258	€2392	€26 927
4th year	€168	€330	€1156	€2800	€25 611
5th year	€225	€428	€1188	€2906	€24 211

***Abbreviations**: BC: breast cancer.*
***Notes**: Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

Table 36 – Systemic therapy costs, including supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€3	€74	€540	€698	€2287
2nd month	€4	€234	€1382	€1750	€3649
3rd month	€3	€397	€1773	€2546	€4076
Months 4-6	€6	€1094	€5004	€7007	€11 154
Months 7-9	€6	€778	€3679	€4456	€9667
Months 10-12	€14	€717	€3892	€6558	€9047
Total 1st year	€36	€3295	€16 270	€23 015	€39 880
2nd year	€56	€598	€2970	€6431	€34 141
3rd year	€63	€287	€1260	€2396	€26 956
4th year	€168	€330	€1157	€2806	€25 640
5th year	€230	€428	€1189	€2908	€24 230

***Abbreviations**: BC: breast cancer.*
***Notes**: Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

As described in the Belgian economic evaluation guidelines, pharmaceutical products may be reimbursed under a managed entry agreement, which may include confidential price discounts. These discounts are, by definition, unknown. The Belgian guidelines on economic evaluations recommend creating price discount scenarios for pharmaceutical products reimbursed under confidential contracts as part of the reference case.

The IMA data available to us reflect gross expenditures and cannot take the confidential discounts into account. Nevertheless, information is publicly available on which products are reimbursed under such contracts (https://webappsa.riziv-inami.fgov.be/ssp/Contracts). Table 37 presents the percentage of the total gross expenditure on systematic therapy that is linked to a managed entry agreement. This percentage is calculated in relation to the gross expenditure excluding (Table 35) and including supplements (Table 36).

Table 37 – Percentage of expenditure on systemic therapy under managed entry agreement

In relation to	Stage 0	Stage I	Stage II	Stage III	Stage IV
Gross expenditures excl. suppl.	27.4%	31.2%	46.0%	45.6%	64.3%
Gross expenditures incl. suppl.	27.0%	31.1%	45.8%	45.5%	64.2%

***Abbreviations**: Excl. suppl.: excluding supplements; incl. suppl.: including supplements.*
***Notes**: Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

The part that is not reimbursed under contract does not require any adjustment. For the remainder, a correction is included in accordance with the guidelines. The Belgian MORSE (Monitoring of Reimbursement Significant Expenses) report[292] presents the overall discounts received on pharmaceuticals reimbursed under managed entry agreements. These percentages are provided per calendar year, as well as cumulatively, i.e. calculated on the cumulative turnover and cumulative compensation over the previous years (see Table 38). The costs for the first year are based on data from patients diagnosed with breast cancer in 2022 (who therefore incurred costs in 2022 or 2023). The average discount could easily be higher or lower than the (cumulative) discounts shown in Table 38. As the size of this discount is unknown, we work with different scenarios. In the base case, we have decided to apply a 40% discount (i.e. around the cumulative discount for the year 2022-2023) to that part of the systemic therapy costs that are reimbursed under the managed entry agreement.[fffff] In the scenario analyses, the price discount varies in 10% increments from 0% to 90% (scenario 'Price discount').

Table 38 – Overview of discounts on pharmaceuticals reimbursed under managed entry agreement (per year)

Period	2015	2016	2017	2018	2019	2020	2021	2022	2023
Per year	12.2%	18.6%	25.5%	27.2%	39.3%	39.9%	48.8%	50.8%	53.7%
Cumulative	12.2%	15.9%	21.0%	23.1%	28.0%	31.3%	35.3%	38.7%	41.7%

***Source**: MORSE report[292]*
***Notes**: Information for 2024 is not included in this table, as data for that year is not yet complete.*

[fffff] This results in a reduction in total systemic therapy expenditure of approximately 10.96%, 12.48%, 18.4%, 18.2% and 25.7% for stages 0, I, II, III and IV, respectively.

As an example, combining the information from the above tables results in an average saving of €10 244[ggggg] in the first year for patients with stage IV breast cancer due to this discount.

PALLIATIVE CARE

Table 39 and Table 40 present the palliative care costs, excluding and including supplements, respectively. Almost no codes related to palliative care were identified during the first year in stage 0-III breast cancers. This is also observed during the following years in stage 0 and I. As expected, most costs related to palliative care are incurred in stage IV breast cancers.

Table 39 – Palliative care costs, excluding supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€0	€0	€0	€4	€37
2nd month	€0	€0	€1	€2	€59
3rd month	€0	€0	€1	€0	€53
Months 4-6	€0	€0	€3	€0	€162
Months 7-9	€0	€1	€11	€4	€168
Months 10-12	€0	€3	€5	€0	€160
Total 1st year	€0	€4	€22	€11	€639
2nd year	€0	€9	€14	€154	€483
3rd year	€0	€6	€34	€96	€649
4th year	€0	€7	€40	€99	€770
5th year	€0	€28	€50	€141	€541

Abbreviations: *BC: breast cancer.*
Notes: *Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

Table 40 – Palliative care costs, including supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€0	€0	€0	€4	€37
2nd month	€0	€0	€1	€2	€59
3rd month	€0	€0	€1	€0	€53
Months 4-6	€0	€0	€3	€0	€163
Months 7-9	€0	€1	€11	€4	€170
Months 10-12	€0	€3	€5	€0	€161
Total 1st year	€0	€4	€22	€11	€643
2nd year	€0	€9	€14	€168	€483
3rd year	€0	€6	€34	€96	€649
4th year	€0	€7	€40	€103	€770
5th year	€0	€28	€50	€141	€541

Abbreviations: *BC: breast cancer.*
Notes: *Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

---

ggggg €39 829 (average stage IV systemic therapy expenditure during the 1st year) x 64.3% (part under managed entry agreement) x 40% (assumed discount) (or €39 829 x 25.7%) = ~€10 244 (differences due to rounding).

MISCELLANEOUS

Table 41 and Table 42 present the miscellaneous costs, excluding and including supplements, respectively. A similar trend to the other cost categories is observed: in general, costs increase with increasing stage. While the largest share of costs is predominantly incurred in the first year in stages 0-III, substantial costs are also incurred in subsequent years in stage IV.

Table 41 – Miscellaneous costs, excluding supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€7	€29	€131	€207	€220
2nd month	€15	€82	€265	€429	€266
3rd month	€50	€170	€320	€465	€267
Months 4-6	€51	€280	€750	€1237	€611
Months 7-9	€8	€101	€337	€556	€388
Months 10-12	€5	€63	€188	€335	€338
Total 1st year	€136	€724	€1991	€3230	€2090
2nd year	€29	€74	€223	€371	€1364
3rd year	€20	€23	€86	€183	€1417
4th year	€43	€32	€91	€207	€1187
5th year	€29	€29	€93	€214	€1528

Abbreviations: *BC: breast cancer.*
Notes: *Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

Table 42 – Miscellaneous costs, including supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€7	€31	€149	€239	€245
2nd month	€15	€97	€331	€553	€283
3rd month	€50	€190	€368	€548	€282
Months 4-6	€51	€289	€779	€1254	€642
Months 7-9	€8	€102	€339	€557	€397
Months 10-12	€9	€63	€189	€340	€341
Total 1st year	€140	€773	€2155	€3491	€2189
2nd year	€29	€74	€225	€375	€1388
3rd year	€20	€24	€87	€185	€1442
4th year	€47	€32	€94	€212	€1210
5th year	€30	€30	€97	€216	€1548

Abbreviations: *BC: breast cancer*
Notes: *Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

TOTAL COSTS

In Table 43 and Table 44, the average costs mentioned in the previous tables are summed up, excluding and including supplements, respectively. We note that the diagnostic costs from the pre-period are included in the first month and that these costs do not include the confidential price discounts applied in the context of managed entry agreements. The largest part of these total costs are attributable to the systemic therapy costs for the treatment of stage IV breast cancers.

Table 43 – Total costs, excluding supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€2516	€3516	€3934	€4089	€5543
2nd month	€1629	€2333	€3433	€3837	€4828
3rd month	€1639	€2452	€3068	€3606	€4877
Months 4-6	€1627	€3213	€7940	€10 111	€13 326
Months 7-9	€368	€1687	€6898	€9834	€12 274
Months 10-12	€276	€1073	€4817	€8270	€10 832
Total 1st year	€8055	€14 274	€30 090	€39 746	€51 680
2nd year	€679	€1383	€4570	€8762	€38 927
3rd year	€397	€832	€2368	€4318	€31 948
4th year	€583	€840	€2046	€4400	€30 326
5th year	€518	€880	€1894	€4168	€28 784

Abbreviations: *BC: breast cancer.*
Notes: *Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

Table 44 – Total costs, including supplements (per time period and BC stage)

Period	Stage 0	Stage I	Stage II	Stage III	Stage IV
1st month	€2912	€4062	€4333	€4388	€5723
2nd month	€1944	€2790	€3849	€4338	€4891
3rd month	€1829	€2579	€3190	€3722	€4912
Months 4-6	€1782	€3321	€8170	€10 311	€13 491
Months 7-9	€459	€1742	€7090	€10 141	€12 471
Months 10-12	€354	€1111	€4860	€8340	€10 966
Total 1st year	€9280	€15 605	€31 492	€41 240	€52 454
2nd year	€809	€1482	€4791	€9118	€39 127
3rd year	€462	€881	€2545	€4567	€32 087
4th year	€649	€882	€2129	€4629	€30 442
5th year	€556	€921	€1943	€4219	€28 959

Abbreviations: *BC: breast cancer.*
Notes: *Based on stage 0-IV cancers in women aged 50-69 years (incidence years 2018-2022).*

6.2.7.10 Treatment effect through stage shift

Table 45 presents the input related to the modelling of the treatment effect.

Table 45 – Overview of input variables (treatment effect)

Variable	Value	Uncertainty	Source
Risk ratio stage II, III or IV breast cancer in intervention group versus comparator group.	0.66 (95% CI: 0.61-0.72)	Normal distribution on the log scale Scenario 'Risk ratio': RR = 0.83 (95% CI: 0.68-1.00)	Meta-analysis (see section 3.2.2.5)

***Abbreviations**: RR: risk ratio.*

We cannot rely on a (meta-analysis of) recent RCT(s) to assess the treatment effect. All the evidence presented in the systematic review of the clinical evidence is based on relatively old RCTs, a large proportion of which are at high risk of selection bias. Therefore, we adopt an alternative approach in this hypothetical economic evaluation, the various steps of which are explained below. The Markov section of Figure 35 may help the reader to follow these steps.

First, we provide an overview of the current situation with regard to screening (intervention group), based on the existing population-based screening programme in Belgium. Using the Belgian Nomenclature, we can reliably identify women participating in the breast cancer screening programme. For these women, we can categorise screen-detected (see section 6.2.7.3) and interval cancers (see section 6.2.7.4) by age and stage (see section 6.2.7.5). The sum of screen-detected and interval cancers represents the total number of cancers in the intervention group.

- Step 1: number of screen-detected cancers (according to age & stage) + number of interval cancers (according to age and stage) = total number of cancers in the intervention group (according to age and stage)

In the second step, we exclude the percentage of cancers that are overdiagnosed (see section 6.2.7.6). This percentage is applied to the total number of cancers in the intervention group. The number of overdiagnoses is initially classified as stage 0 cancers. If the actual number of overdiagnoses exceeds the number of identified stage 0 cancers, the remainder are classified as stage I breast cancers. This allows us to calculate the number of cancers in the intervention group that are not overdiagnosed by age and stage.

- Step 2: number of screen-detected cancers (according to age & stage) that are no overdiagnosis = total number of cancers in the intervention group (according to age and stage) – overdiagnosed cancers (according to age & stage (0/I))

In the third step, we establish how many cases of breast cancer in the intervention group are classified as stage II, III or IV. We then use the relative risk (see further) to estimate the number of stage II, III or IV cases that would have occurred in the comparator group if screening had not been carried out. This difference in the occurrence of stage II, III and IV breast cancers between the intervention and comparator groups reflects the stage shift which can be attributed to screening.

- Step 3: number of stage II, III and IV breast cancers in the comparator group = number of stage II, III and IV breast cancers in the intervention group / (1 – relative risk reduction)

In the fourth step, we calculate the total number of stage 0, I, II, III and IV breast cancers in the comparator group. In step 3, we determined the higher number of stage II, III and IV cancers. We also know the total number of cancers in the comparator group by subtracting the number of overdiagnoses from the total number of cancers in the intervention group. The remaining cancers are divided into stage 0 and stage I cancers. First, we look at the small number of stage 0 cancers diagnosed among the interval cancers. The remainder are classified as stage I breast cancers.

- Step 4: Determine number of stage 0, I, II, III or IV breast cancers in the comparator group:
  - number of stage II, III or IV (see step 3);
  - calculate the number of remaining cancers by subtracting the number of stage II, III and IV cancers from the total number of non-overdiagnosed cancers. (see step 2).
  - classify the remaining cancers as stage 0 or stage I breast cancers. To determine the number of stage 0 cancers, consider the number of stage 0 cancers among the interval cancers (3a). The remaining cancers are classified as stage I breast cancer (3b).

In order to apply the above calculation, one further variable needs to be estimated: the relative risk (RR) of stage II, III or IV breast cancer in the intervention group compared to the comparator group. The systematic review of the clinical evidence in this report concludes that "In a general population of adult women at average risk of breast cancer, invitation to screening with mammography results in a statistically significant reduction of the probability of being diagnosed with advanced disease (stage II or higher: RR: 0.66, 95% CI: 0.61-0.72". For stage III or higher, the average is also 0.66 (see section 3.2.2.5). This is why, in the base case, we use an RR of 0.66 (95% CI: 0.61–0.72) for stages II, III and IV. "When the data are restricted to the single trial with a low risk of selection bias, the Malmö I study, no statistically significant result is observed (RR: 0.83, 95% CI: 0.68-1.00, p=0.05)." This is applied in an alternative scenario (scenario 'Risk ratio').

We would like to point out that the evidence relates to women invited to attend mammography screening, rather than to those who actually participate in population-based screening. While participation in the underlying randomised controlled trials (RCTs) is relatively high, it is certainly not 100%.[hhhhh] Applying a correction could result in a more accurate estimation of the relative risk. However, there are also other factors that could negatively impact this relative risk. For example, there may be increased awareness of breast cancer and the importance of self-examination compared to when the trials were performed. Additionally, the meta-analysis identified a high risk of selection bias in four of the five RCTs included. As it is unclear in which direction the RR should be adjusted, we have chosen not to apply any correction to this hypothetical analysis.

Once all the preceding steps have been completed, we know the number of cancer cases in both the intervention and comparator groups, as well as their stage distribution. On this basis, these cancers can be linked to survival (see section 6.2.7.7), quality of life (see section 6.2.7.8), and costs (see section 0) to calculate the total costs and effects (in life years (LY) and quality-adjusted life years (QALYs)) in both groups. These form the basis for calculating incremental costs, effects and cost-effectiveness ratios (€/QALY gained or €/LY gained).

6.2.7.11 Lead time

Due to the lead time, the survival time of people diagnosed through screening will appear longer due to early detection, even if this would have no effect on the course of the disease or the time of death itself (see section 2.2.3). Including the lead time poses challenges for the model. Based on Belgian data, we have survival information for breast cancer patients diagnosed with the disease between the ages of 50 and 69, categorised by stage. These are the survival probabilities for patients diagnosed with breast cancer within a population-based screening programme. This means that the lead time is included in these survival data.

[hhhhh] Attendance rate in the 5 RCTs included in the meta-analyses on the relative risk: 1) Malmö I: ~70%; 74% in first round ranging from 64% in oldest age group to 79% in youngest; 2) CNBSS: 100% in first round, CNBSS1: 89% in second, decreasing to 86% in fifth round; CNBSS2: 90% in second, decreasing to 87% in fifth round. 3) HIP: 65% in total population; 4) Stockholm: ~80%; 5) Swedish Two-County: Kopparberg: age group 40-59 years: 91-94%, age group 60-74 years: 50-80%; Östergötland: ~90% in first round, ~80% in second. See more information in Appendix 7.5.

In theory, the solution would be to isolate the lead time from these survival curves. This would need to be done for each stage of breast cancer, since the lead time can vary between them. It is possible that lead time is longer for earlier stages of breast cancer. However, we do not have the necessary stage-specific information to perform this isolation. This is most likely to the advantage of screening, as a longer lead time in earlier stages leads to an artificially high gain in life years via the stage shift. Lead time would also mean that breast cancer would be diagnosed later in the comparator group and treatment would therefore start later, resulting in a greater discount. Excluding this effect is at the advantage of screening.

6.2.7.12 Uncertainty analysis

Both probabilistic and deterministic sensitivity analyses are performed.

PROBABILISTIC SENSITIVITY ANALYSIS

Uncertainty surrounding the input variables was incorporated into the model using probability distributions. The following probability distributions were chosen for the different variables:

Beta distributions for the following variables:

- Participation rate (see Table 13)
- Screening results (positive/negative screening test results) (see Table 14)
- True- and false-positive findings (see Table 15)
- Frequencies of tests performed in relation to false-positive findings (see Table 15)
- Interval cancers (see Table 16)
- Overdiagnosis (see Table 21)
- Health-related quality of life – stage-specific utilities and disutilities related to false-positive findings (see Table 25)

Conditional beta distributions for the following variables:

- Number of cancers per age category and age-specific cancer stages at diagnosis – true positive cancers (see Table 17)
- Number of cancers per age category and age-specific cancer stages at diagnosis – interval cancers (see Table 18)

Uniform distribution for the following variables:

- Invitation cost, usage eBox, cost result letter, lump sum costs related to screening (see Table 13)
- Outpatient visit to medical specialist, possibly in combination with visit to a GP (see Table 15)

Normal distribution for the following variables:

- Diagnosis, follow-up and treatment costs (stage specific) (see Table 28)

Normal distribution on the log scale:

- Risk ratio for stage II, III or IV breast cancer (see Table 45)

Beta distributions are used for variables that are bounded by 0 and 1. Given this property, they are the ideal distribution for utilities and probabilities. Conditional beta distributions are chosen where the probabilities must sum to 100%. A uniform distribution is chosen for costs that are not based on a sample, for which there is no information regarding the associated uncertainty. In such cases, the minimum and maximum of the uniform distribution is -/+ 20% of the average cost.

When applying probability distributions, it is important to check whether logical values are modelled. For example, it is preferable for utilities to satisfy the following conditions: utility stage 0 ≥ utility stage I ≥ utility stage II ≥ utility stage III ≥ utility stage IV. However, if these utilities are modelled as independent values using a beta distribution, the rationale of better utilities for lower breast cancer stages is not respected in most simulations. This is because the utilities for the lower stages of breast cancer are close to each other (0.85 for stages 0, I and II, and 0.83 for stage III, see Table 13). The occurrence of these illogical values is undesirable for population-level modelling. Following the approach taken in KCE's lung cancer screening report,[264] we first model the utility for stages 0, I and II breast cancers as a beta distribution reflecting the reported mean and standard deviation. Then, for the utilities for stages III and IV, we subtract the difference between the reported means for the different stages: minus 0.02 for stage III (i.e. 0.83 versus 0.85), and minus 0.12 for stage IV (i.e. 0.73 versus 0.85).

We also have to be careful not to model illogical values for costs at the population level. For example, it is possible that an individual incurs no costs for radiotherapy or surgery. However, it is not possible for the average cost for our total population to be zero. Therefore, a probability distribution that yields correct values for individuals is not necessarily suitable for modelling at the population level. This would be the case if we applied a gamma distribution, which is nevertheless a widely used probability distribution for cost data. An alternative approach would be to apply a normal distribution via the Central Limit Theorem. This theorem states that, if sufficiently large samples are taken from a population, the means of these samples will be normally distributed, even if the original sample is not. When applying the Central Limit Theorem to the original cost data, which had a wide standard deviation, the resulting normal distribution also had a wide confidence interval. This meant that, when 1000 simulations were run, negative values were drawn from this distribution. This is clearly an illogical value for cost data. Therefore, similar to the approach taken in KCE's lung cancer screening report,[264] we opted to apply a normal distribution where the standard deviation was determined so that the 95% confidence interval (CI) equals ±20% of the mean (formula: ((mean x 1.2) – (mean x 0.8)) / (1.96 x 2)). This enabled us to reflect the correct mean costs extracted from the Belgian IMA–AIM data and to simulate only positive values for all our cost input variables.

Uncertainty surrounding the input variables translates into uncertainty about the incremental costs, incremental effects (per LY gained or QALY gained) and ICERs (€/LY gained or €/QALY gained). This is transparently reported by presenting the 95% confidence interval for all these output variables in table format. The results of the base case analysis are also presented on the cost-effectiveness plane and cost-effectiveness acceptability curve.

DETERMINISTIC SENSITIVITY ANALYSIS

In previous sections, we provided an overview of the different variables and a description of the scenario analyses. Table 46 provides a supplementary overview, clarifying which values were chosen for the base case and how these were varied in the scenario analyses. The results of these analyses are presented in table and figure format, as well as in the form of a tornado graph.

Table 46 – Overview variables in the base case analysis and scenario analyses

Variable	Base case*	Scenario analysis	Source
Time horizon	Lifetime	10, 20, 30, and 40 years	Section 6.2.5
Discount rate	Costs: 3%; Effects: 1.5%	Scenario 'Discount rate': • Costs and effects: 0% • Costs and effects: 3% • Costs and effects: 5%	Section 6.2.5
Participation rate	40.8%	Scenario 'Participation': • 10% - Beta (54 454; 490 082)	Section 6.2.7.1

Variable	Base case*	Scenario analysis	Source
		• 20% - Beta (108 907; 435 629) • 30% - Beta (163 361; 381 175) • 40% - Beta (217 814; 326 722) • 50% - Beta (272 268; 272 268) • 60% - Beta (326 722; 217 814) • 70% - Beta (381 175; 163 361)	
Invitation cost	See section 6.2.7.1	Scenario 'Invitation cost letter': €0.76 Scenario 'Invitation cost excluded': €0 Scenario 'Invitation cost alternative': total cost of €5.52, €5.99, or €3.69	Section 6.2.7.1
Diagnosis	TP: 19.9% FP: 80.1%	Scenario 'True positives': TP: 21.1%; FP: 78.9%	Section 6.2.7.3
Overdiagnosis	20.7%	Scenarios 'Overdiagnosis': 27.1% (Beta(1059; 2847) 0, 10, 20, and 30%	Section 6.2.7.6
HRQoL - utilities	Stage 0, I & II: 0.85; stage II: 0.83 & stage IV: 0.73	Scenario 'Utilities': Stage 0, I & II: 0.789; stage II: 0.774 & stage IV: 0.686	Section 6.2.7.8
HRQoL - disutilities	0.07 (1 month)	Scenario 'Disutility FP': no disutility; disutility also for overdiagnosed true positive findings	Section 6.2.7.8
Costs	Exclusive supplements	Scenario 'Supplements': inclusive supplements	Section 0
Price discount (~MEAs)	40%	Scenario 'Price discount': 0, 10, 20, 30, 40, 50, 60, 70, 80, and 90%	Section 0
Risk ratio stage II, III & IV BC	0.66	Scenario 'Risk ratio': 0.83	Section 6.2.7.10

***Abbreviations**: BC: breast cancer; FP: false positive; HRQoL: health-related quality of life; MEA: managed entry agreement; TP: true positive.*
***Notes**: * given the poor quality of the evidence for several of these variables, any reference to a base case should be nuanced (e.g. with regard to the relative risk reduction of stages II, III and IV breast cancers and the percentage of overdiagnosis).*

We would like to stress that, for specific variables, there is no preference for the chosen base case variable. For example, in the case of the percentage of overdiagnosis and the relative risk reduction related to the stage shift, a balance must be found between the most comprehensive meta-analysis, which includes RCTs at high risk of bias, and a more selective selection of RCTs at lower risk of bias. Given the limitations in the underlying evidence, both options are still subject to considerable uncertainty. To reflect this uncertainty, several results will be presented alongside each other.

MONTE CARLO SIMULATIONS

The uncertainty surrounding the input variables is translated to the uncertainty in the outcomes by applying Monte Carlo simulations. In total, for every scenario, 1000 simulations are performed, in which a value for each input parameter is randomly drawn simultaneously from their respective probability distributions (see above). In this way, the model generates 1000 estimates of the outputs (IC, IE, ICERs). For this purpose, we use ModelRisk, an Excel add-in developed by Vose Software.

Scenario analyses are conducted using the VoseSimTable function. With this function, the model retains the draws from all probability distributions generated in previous scenarios and performs new draws only for the single variable under investigation in the scenario analysis. Consequently, any difference in results can be fully attributed to variation in that specific variable (i.e., differences due to random variation are eliminated). This is reflected in the results: for example, a scenario in which only costs are varied will affect only the incremental costs and will not alter the incremental effects (neither the mean nor the confidence interval – see results in section 6.3).

6.2.7.13 Validation

When constructing an economic model, it is essential to verify that the model has been implemented correctly and does not produce illogical or erroneous results. In addition, it is important to assess whether the underlying evidence is accurately represented within the model. Therefore, several validation steps were undertaken:

- The model structure and formulas were reviewed by a second health economist.
- A check was performed to ensure that changes in input parameters resulted in logically consistent changes in the model outputs.
- Intermediate results were extracted as model outputs (e.g., screening test results and different cost categories).
- A visual inspection of the survival curves was conducted after applying different extrapolation scenarios, along with a comparison to age-specific general population survival curves for women. Based on this assessment, the most valid extrapolation assumption was selected and two scenarios were excluded (see section 6.2.7.7).
- The model output for the difference in number of individuals alive in the intervention and comparator groups at 13 years were compared with the corresponding results from the clinical part of this HTA.
- A check was performed to ensure that applying the inverse risk ratio to the number of stage II, III, and IV breast cancers in the intervention group does not result in a total number of stage II, III or IV breast cancers in the comparator group exceeding the total expected number of breast cancer cases in this group (defined as the total number of breast cancers in the intervention group minus the number of overdiagnosed cases).[iiiii]

6.3 Results of de novo hypothetical economic evaluation

The following sections present the results of all modelled scenarios. First, we present the results for one of the most influential variables: the relative risk of stages II, III and IV breast cancer. This variable is used to model the stage shift. The results for this scenario are presented in tabular form, on the cost-effectiveness (CE) plane and on the cost-effectiveness acceptability curve (CEAC) (see section 6.3.1). For the other scenarios, all results are presented in tabular form, and a figure shows how the averages of these extra scenarios compare to the base case for both stage shift scenarios.

[iiiii] This was only the case in the scenarios combining a risk ratio for the diagnosis of stage II, III and IV breast cancer of 0.66 (95% CI: 0.61 - 0.72) in combination with an overdiagnosis of 27.1% (in 6 of the 1000 simulations) or 30% (in 22 of the 1000 simulations). Excluding these simulations did not impact the ICERs (change of <1%). This problem did not occur in any of the simulations applying a risk ratio of 0.83 (95% CI: 0.68 - 1.00).

6.3.1 Risk ratio (incl. cost-effectiveness plane and cost-effectiveness acceptability curves)

Based on 1000 simulations and an initial invitation to 633 471 people, an average of 258 289 women (40.8%) aged 50–69 participated in the breast cancer screening programme. Of these, 250 056 (96.8%) received a negative result and 8233 (3.2%) received a positive result. Of the positive test results, on average 6595 (80.1%) were false positives and 1638 (19.9%) were true positives. Of these true positives, 501 were categorised as overdiagnosis. Combined with an average of 793 interval cancers (0.31%), this equates to an average overdiagnosis rate of 20.6%[jjjjj] (501 / (1638+793)) when expressed in relation to all cancers diagnosed in the intervention group. If expressed in relation to the number of screen-detected cancers, this figure would be 30.6% (501 / 1638).

Calculating the costs for the intervention group over a lifetime results in a total cost of €120 737 386. This consists of the following components:

- €2 681 481 is linked to invitations and communication.
- €21 192 619 for screening.
- €1 173 586 for false-positive tests
- €85 571 056 for treating patients not categorised as overdiagnosis (stages 0-IV breast cancer).
- €10 118 644 for treating patients categorised as overdiagnosis (all stage 0 or I breast cancers).

The total cost in the comparator group depends on the applied risk ratio. For a risk ratio of 0.66, the total cost is €102 003 370. This results in an incremental cost of €18 734 015 (see Table 47). If a less favourable risk ratio of 0.83 is applied, the cost in the comparator group is €92 226 144 – an incremental cost of €28 511 242.

Based on the stage shift, a relative risk for stages II, III and IV breast cancers of 0.66 results in an average gain of 1657 discounted life-years for those participating in screening. When adjusted for quality of life, this equates to an average gain of 1421 QALYs. This means that, if expressed in relation to the number of women diagnosed with breast cancer who are not categorized as an overdiagnosis, there is an average gain of approximately 15.0 months adjusted for QoL (95% CI: 10.6 – 20.0). However, if the gain is expressed relative to all participants in breast cancer screening, it amounts to an average of 2.0 days (95% CI: 1.4 – 2.6). Combining the above incremental costs and effects gives an average incremental cost-effectiveness ratio (ICER) of around €13 700 per QALY gained (see Table 47).

Assuming a risk ratio of 0.83 results in an average gain of 666 life years or 549 QALYs. This equates to around 5.8 quality-adjusted months (95% CI: -0.8 – 13.6) of benefit when measured against all women diagnosed with breast cancer who have not been overdiagnosed, or an average of 0.8 days (95% CI: -0.1 – 1.8) when measured against all participants in breast cancer screening programmes. In this case, the average ICER is around €51 900 per QALY gained (see Table 47).

[jjjjj] All reported figures are derived from the model outputs. The input parameters are closely reflected in the outputs (within a precision of 0.1% in absolute values. E.g. after 1000 simulations, the mean overdiagnosis rate was 20.6% compared to the expected 20.7%).

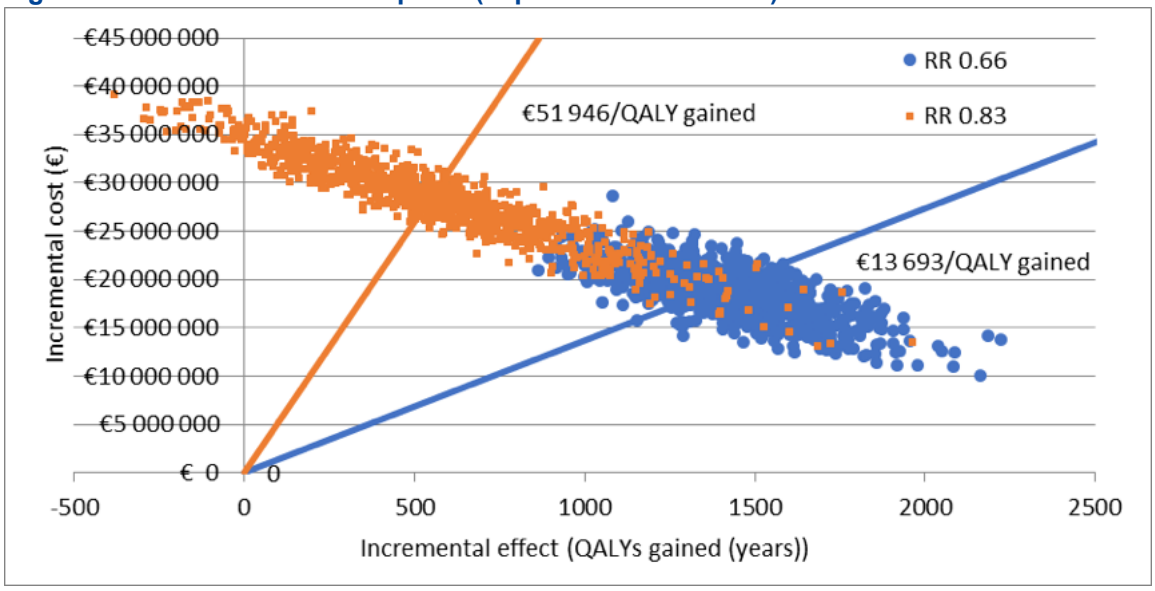
Table 47 – Results of the economic evaluation for different risk ratio scenarios

	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG)* ICER (€/QALY gained)*
Stage shift	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
RR = 0.66	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078)	€11.684 (6.643 - 18.412)
		1421 (1.019 - 1.854)	€13.693 (7.538 - 22.232)
RR = 0.83	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518)	*€42.800*
		549 (-72 - 1.303)	*€51.946*

***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*
***Notes**: * The calculation of the average ICER based on the 1000 simulations is not reliable if the outcomes are situated in different quadrants of the cost-effectiveness plane (see Figure 45). In these cases, the presented ICERs are calculated by dividing the mean incremental cost by the mean incremental benefit.*

The cost-effectiveness plane (Figure 45) illustrates the significant impact of the chosen risk ratio. As expected, when applying the RR of 0.83 (95% CI: 0.68 – 1.00), in combination with the QALYs lost due to false-positive findings, some of the simulations fall within the fourth quadrant of the cost-effectiveness plane.

Figure 45 – Cost-effectiveness plane (impact of the risk ratio)



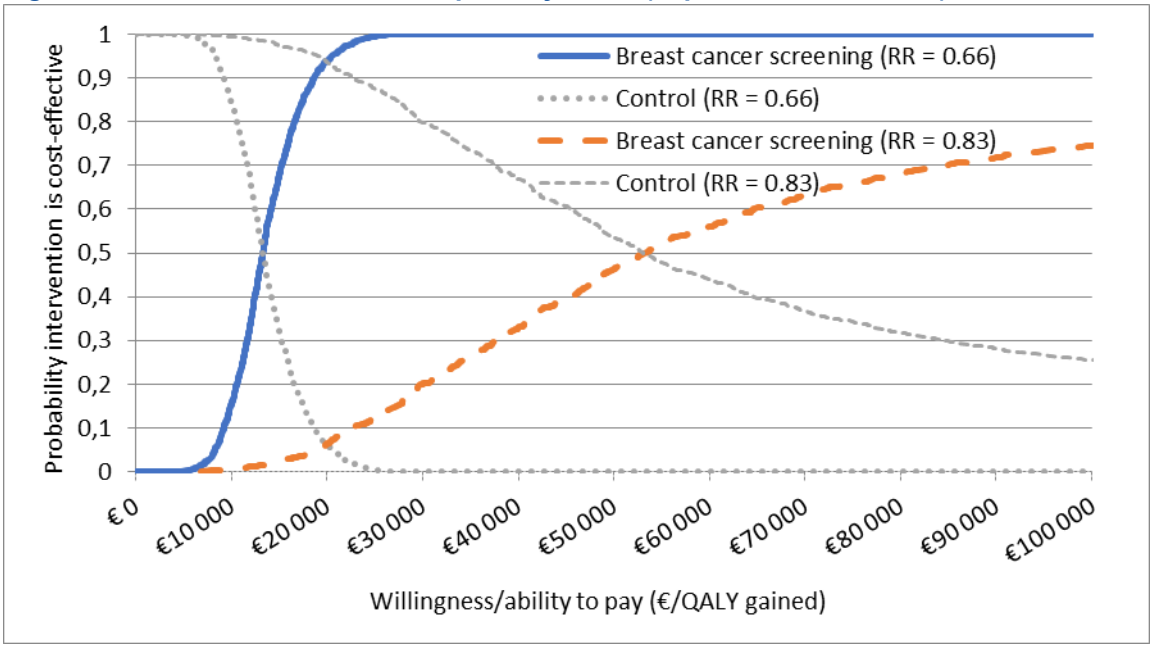
***Abbreviations**: QALY: quality-adjusted life year; RR: risk ratio.*

Figure 46 shows the cost-effectiveness acceptability curve, which expresses the probability that an intervention or comparator will be considered cost-effective as a function of the willingness to pay (WTP) per QALY gained. In the scenario with an RR of 0.66 and a WTP of €10 000 per QALY gained,

there is a 16% probability that breast cancer screening would be considered cost-effective. This probability increases to 94% or 100% at a WTP of €20 000 or €30 000 per QALY gained, respectively.

In the scenario with an RR of 0.83, these figures are 0.5%, 6% and 20% at a WTP of €10 000, €20 000 and €30 000 per QALY gained, respectively. If the WTP rises to €50 000 per QALY gained, the probability is 46%. A WTP of about €53 000 per QALY gained is required to achieve a 50% probability.

Figure 46 – Cost-effectiveness acceptability curve (impact of the risk ratio)



Abbreviations: *QALY: quality-adjusted life year.*

6.3.2 Time horizon

Table 48 and Table 49 illustrate the impact of changing the time horizon. The impact on incremental costs is minimal, since the majority of additional costs are incurred during the screening process. The impact of the stage shift on costs is also felt immediately in the early years, since, as with all previous economic evaluations, no account is taken of lead-time bias.

The impact of a shorter time horizon on years of life gained is clear, as the survival curves for the different stages of breast cancer diverge significantly over time (see section 6.2.7.7). Consequently, the incremental effect over 10 years, for example, is only 291 QALYs gained, resulting in an ICER of almost €69 000 per QALY gained in the 0.66 RR scenario (Table 48). With time horizons of 20 and 30 years, however, the average ICER falls to below €24 000 and €16 000 per QALY gained, respectively. With an RR of 0.83, the average ICER is approximately €277 000, €88 000, and €59 000 per QALY gained over time horizons of 10, 20, and 30 years, respectively (see Table 49). With a time horizon of 40 years or more, the results hardly vary (see Figure 47), given the advanced age of the population by that point and the significant impact of discounting costs and benefits.

Table 48 – Results of the economic evaluation for time horizon scenarios (risk ratio 0.66)

RR: 0.66	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Time horizon	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
10 years	€19.226.246 (15.182.895 - 23.028.759)	354 (266 - 450) 291 (198 - 395)	€55.931 (35.054 - 84.409) €68.804 (40.185 - 111.473)
20 years	€18.659.878 (13.788.481 - 23.057.855)	970 (729 - 1.227) 828 (594 - 1.089)	€19.847 (11.682 - 30.711) €23.400 (13.200 - 37.741)
30 years	€18.689.884 (13.408.415 - 23.424.366)	1452 (1.090 - 1.827) 1244 (894 - 1.627)	€13.296 (7.673 - 20.720) €15.597 (8.668 - 25.457)
40 years	€18.730.467 (13.351.142 - 23.547.831)	1635 (1.224 - 2.052) 1402 (1.005 - 1.830)	€11.840 (6.747 - 18.657) €13.877 (7.647 - 22.562)
Lifelong	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078) 1421 (1.019 - 1.854)	€11.684 (6.643 - 18.412) €13.693 (7.538 - 22.232)

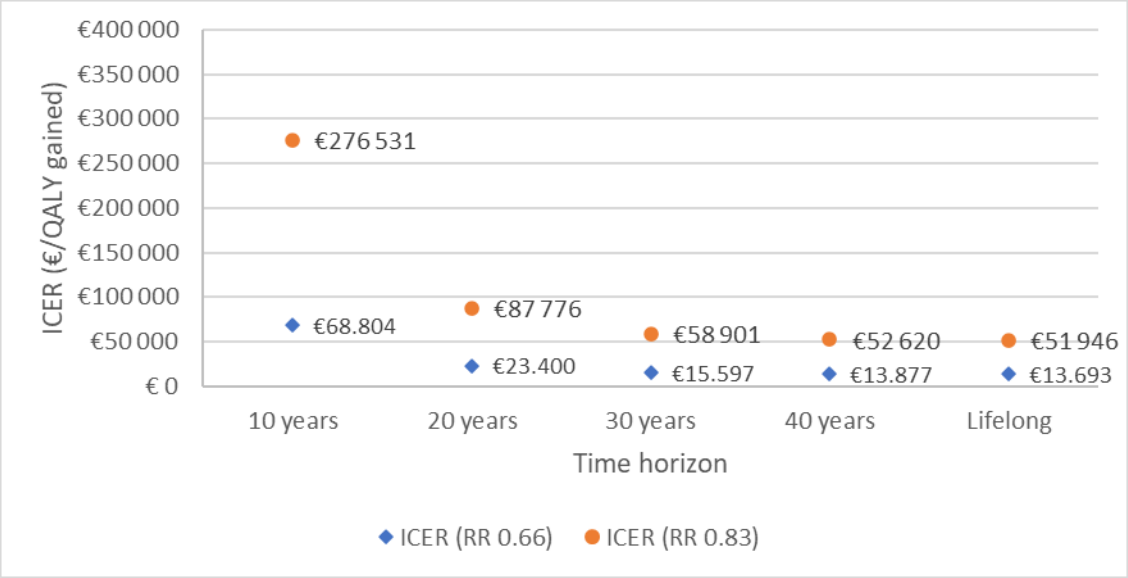
***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Table 49 – Results of the economic evaluation for time horizon scenarios (risk ratio 0.83)

RR: 0.83	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Time horizon	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
10 years	€27.065.077 (20.201.095 - 33.513.413)	146 (-4 - 325) 98 (-41 - 268)	*€185.204* *€276.531*
20 years	€27.812.017 (19.890.866 - 35.401.791)	397 (-16 - 887) 317 (-47 - 760)	*€69.973* *€87.776*
30 years	€28.330.087 (19.959.711 - 36.306.675)	587 (-36 - 1.328) 481 (-63 - 1.139)	*€48.228* *€58.901*
40 years	€28.494.490 (20.043.259 - 36.553.412)	658 (-45 - 1.497) 542 (-71 - 1.285)	*€43.329* *€52.620*
Lifelong	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€42.800* *€51.946*

***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Figure 47 – Results of the economic evaluation for time horizon scenarios



***Abbreviations**: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RR: risk ratio.*

6.3.3 Discount rate

The discount rate scenarios clearly demonstrate the impact of the Belgian guideline, which applies a lower discount rate of 1.5% to benefits compared with the rate of 3% applied to costs. Applying an equal discount rate of 3% or 5%, the average ICER rises to almost €18 000 or about €25 000 per QALY gained, respectively, in the RR 0.66 scenario (see Figure 48). In the RR 0.83 scenario, this increases to over €68 000 and €94 000, respectively.

As expected, the scenario in which no discounting is applied results in the lowest ICERs. This scenario also illustrates the undiscounted gain in (quality-adjusted) life years: 2202 life years or 1893 QALYs (RR 0.66 scenario, Table 50). This equates to approximately 23.3 months (95% CI: 17.0 – 30.1) of benefit when expressed against all women diagnosed with breast cancer who have not been categorized as overdiagnosis, or an average of 3.1 days (95% CI: 2.3 – 3.9) when expressed against all individuals participating in breast cancer screening. When adjusted for QoL, these figures become 20 months and 2.7 days, respectively. In the RR 0.83 scenario (Table 51), these numbers decrease to 882 life-years gained (8.2 months per non-overdiagnosed breast cancer or 1.1 days per participant) and 736 QALYs gained (7.8 months per non-overdiagnosed breast cancer or 1.0 days per participant).

Table 50 – Results of the economic evaluation for discount rate scenarios (risk ratio 0.66)

RR: 0.66	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Discount rate	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
C: 3% E: 1.5%	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078) 1421 (1.019 - 1.854)	€11.684 (6.643 - 18.412) €13.693 (7.538 - 22.232)
C&E: 0%	€18.028.778 (11.240.265 - 24.129.526)	2202 (1.647 - 2.762) 1893 (1.363 - 2.468)	€8.496 (4.126 - 13.918) €9.923 (4.696 - 16.704)
C&E: 3%	€18.734.015 (13.345.328 - 23.563.309)	1276 (956 - 1.604) 1090 (780 - 1.426)	€15.167 (8.631 - 23.855) €17.854 (9.809 - 29.182)
C&E: 5%	€19.146.640 (14.381.721 - 23.496.827)	931 (697 - 1.173) 790 (564 - 1.038)	€21.212 (12.592 - 32.636) €25.167 (14.401 - 40.495)

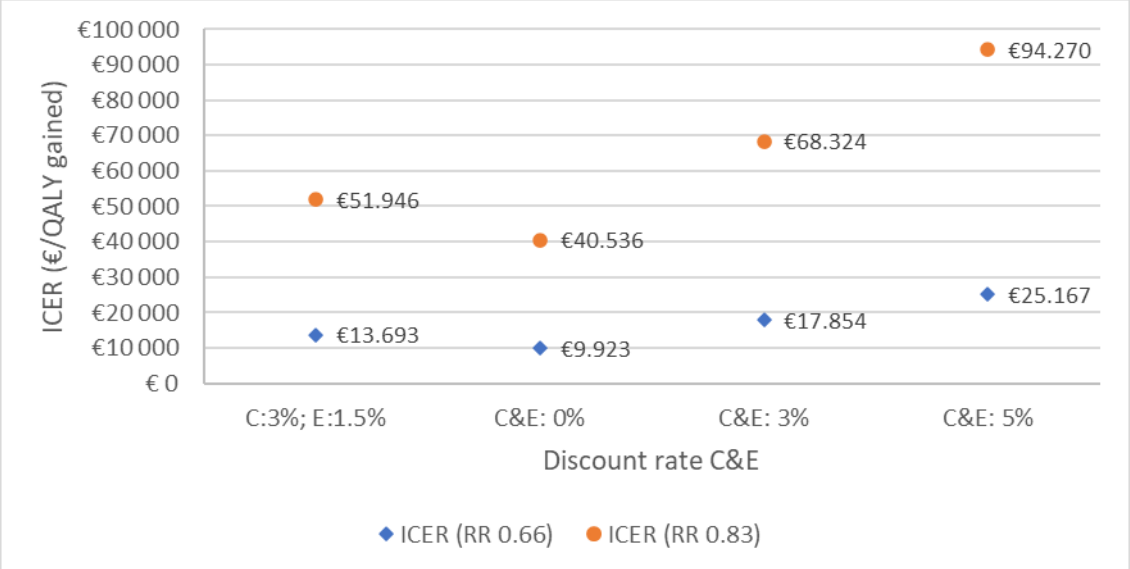
Abbreviations: *ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Table 51 – Results of the economic evaluation for discount rate scenarios (risk ratio 0.83)

RR: 0.83	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Discount rate	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
C: 3% E: 1.5%	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€42.800* *€51.946*
C&E: 0%	€29.841.823 (19.567.059 - 39.524.994)	882 (-66 - 2.018) 736 (-89 - 1.739)	*€33.836* *€40.536*
C&E: 3%	€28.511.242 (20.052.379 - 36.579.645)	515 (-33 - 1.169) 417 (-62 - 998)	*€55.385* *€68.324*
C&E: 5%	€28.029.101 (20.305.291 - 35.418.283)	377 (-22 - 852) 297 (-53 - 721)	*€74.350* *€94.270*

Abbreviations: *ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Figure 48 – Results of the economic evaluation for discount rate scenarios



Abbreviations: *ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RR: risk ratio.*

6.3.4 Invitation and communication cost

Table 52, Table 53, and Figure 49 show how changing the costs associated with invitations and communication affects results. The unjustified complete exclusion of these costs causes the ICER to decrease. Once these costs are included, changes in their level have a relatively smaller impact on the cost-effectiveness of the intervention.

Table 52 – Results of the economic evaluation for invitation cost scenarios (risk ratio 0.66)

RR: 0.66	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Invitation cost	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
Base case	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078)	€11.684 (6.643 - 18.412)
		1421 (1.019 - 1.854)	€13.693 (7.538 - 22.232)
Cost letter (€0.76)	€18.796.521 (13.416.872 - 23.626.390)	1657 (1.240 - 2.078)	€11.722 (6.683 - 18.452)
		1421 (1.019 - 1.854)	€13.738 (7.569 - 22.302)
Invitation cost excl.	€16.052.535 (10.714.747 - 20.829.995)	1657 (1.240 - 2.078)	€10.036 (5.380 - 16.341)
		1421 (1.019 - 1.854)	€11.762 (5.906 - 19.461)
Alternative cost (€5.52)	€19.543.537 (14.144.989 - 24.439.903)	1657 (1.240 - 2.078)	€12.181 (7.013 - 18.994)
		1421 (1.019 - 1.854)	€14.276 (8.016 - 23.143)
Alternative cost (€5.99)	€19.840.778 (14.409.172 - 24.764.832)	1657 (1.240 - 2.078)	€12.364 (7.146 - 19.227)
		1421 (1.019 - 1.854)	€14.490 (8.152 - 23.432)
Alternative cost (€3.69)	€18.386.194 (12.952.220 - 23.203.586)	1657 (1.240 - 2.078)	€11.470 (6.442 - 18.178)
		1421 (1.019 - 1.854)	€13.442 (7.368 - 21.891)

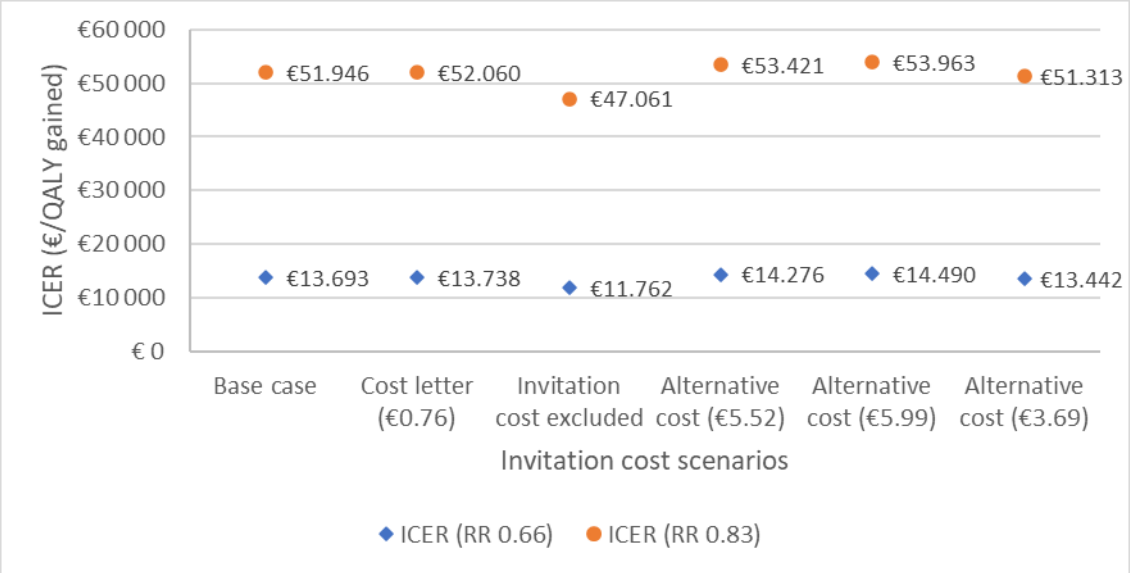
***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Table 53 – Results of the economic evaluation for invitation cost scenarios (risk ratio 0.83)

RR: 0.83	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Invitation cost	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
Base case	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€42.800* *€51.946*
Cost letter (€0.76)	€28.573.748 (20.116.390 - 36.656.381)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€42.893* *€52.060*
Invitation cost excl.	€25.829.761 (17.246.552 - 33.757.406)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€38.774* *€47.061*
Alternative cost (€5.52)	€29.320.763 (20.925.058 - 37.492.099)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€44.015* *€53.421*
Alternative cost (€5.99)	€29.618.004 (21.234.761 - 37.786.394)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€44.461* *€53.963*
Alternative cost (€3.69)	€28.163.420 (19.694.292 - 36.160.448)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€42.277* *€51.313*

***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Figure 49 – Results of the economic evaluation for invitation cost scenarios



***Abbreviations**: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RR: risk ratio.*

6.3.5 Participation

The following tables and Figure 50 show how a change in participation rates affects cost-effectiveness. A participation rate below 40% has a greater impact on cost-effectiveness than a further increase in participation. Costs associated with invitations and communication (see section 6.3.1) are distributed among a larger number of participants. As a result of the decreasing average invitation and communication cost, cost-effectiveness improves relatively more if participation rises from 10% to 20%, 30% or 40%. After this, the costs associated with invitations and communication account for a smaller proportion of the total incremental costs in percentage terms, while the improvement in (quality-adjusted) life years increases proportionally (see Table 54 and Table 55). As a result, the improvement in the ICER becomes less pronounced (see Figure 50).

Table 54 – Results of the economic evaluation for participation scenarios (risk ratio 0.66)

RR: 0.66	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Participation	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
10%	€6.424.581 (5.061.585 - 7.709.496)	406 (305 - 511)	€16.272 (10.185 - 24.487)
		348 (250 - 454)	€19.069 (11.449 - 29.278)
20%	€10.424.705 (7.775.662 - 12.896.901)	813 (608 - 1.018)	€13.232 (7.909 - 20.298)
		697 (500 - 906)	€15.507 (8.810 - 24.519)
30%	€14.424.121 (10.439.107 - 18.037.931)	1219 (909 - 1.533)	€12.219 (7.036 - 19.043)
		1045 (749 - 1.360)	€14.320 (8.056 - 23.212)
40%	€18.423.910 (13.135.322 - 23.151.167)	1626 (1.217 - 2.039)	€11.713 (6.669 - 18.452)
		1394 (1.001 - 1.817)	€13.727 (7.563 - 22.285)
50%	€22.424.443 (15.747.622 - 28.307.847)	2032 (1.522 - 2.550)	€11.409 (6.397 - 18.089)
		1742 (1.251 - 2.267)	€13.370 (7.310 - 21.811)
60%	€26.422.319 (18.413.927 - 33.463.231)	2438 (1.825 - 3.060)	€11.206 (6.254 - 17.811)
		2091 (1.498 - 2.724)	€13.133 (7.078 - 21.492)
70%	€30.422.095 (21.138.394 - 38.639.149)	2845 (2.125 - 3.570)	€11.061 (6.153 - 17.614)
		2439 (1.749 - 3.177)	€12.964 (6.912 - 21.245)

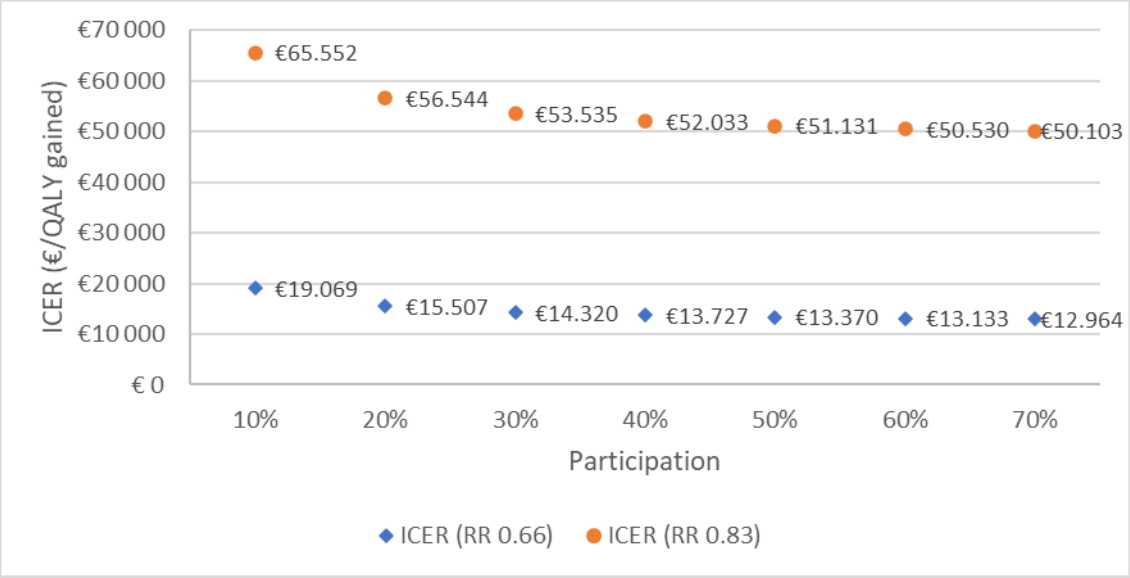
***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Table 55 – Results of the economic evaluation for participation scenarios (risk ratio 0.83)

RR: 0.83	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Participation	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
10%	€8.822.339 (6.628.511 - 10.870.847)	163 (-11 - 373) 135 (-18 - 318)	*€54.010* *€65.552*
20%	€15.220.727 (11.057.499 - 19.184.464)	327 (-23 - 745) 269 (-35 - 641)	*€46.587* *€56.544*
30%	€21.617.719 (15.443.042 - 27.630.106)	490 (-34 - 1.116) 404 (-53 - 956)	*€44.108* *€53.535*
40%	€28.015.855 (19.767.307 - 35.921.291)	653 (-45 - 1.488) 538 (-71 - 1.277)	*€42.871* *€52.033*
50%	€34.414.345 (24.048.743 - 44.266.844)	817 (-57 - 1.859) 673 (-89 - 1.596)	*€42.128* *€51.131*
60%	€40.809.181 (28.339.072 - 52.596.271)	980 (-68 - 2.234) 808 (-106 - 1.915)	*€41.633* *€50.530*
70%	€47.207.626 (32.609.842 - 60.956.157)	1144 (-79 - 2.606) 942 (-124 - 2.232)	*€41.281* *€50.103*

***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Figure 50 – Results of the economic evaluation for participation scenarios



***Abbreviations**: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RR: risk ratio.*

6.3.6 True positives

Classifying a number of positive tests as false or true positives has virtually no impact on cost-effectiveness (see Table 56, Table 57 and Figure 51), given that the number of cases in question (n = 81) is relatively small compared to the total number of positive tests (6717).

Table 56 – Results of the economic evaluation for true-positives scenarios (risk ratio 0.66)

RR: 0.66	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
True positives	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
Base case (19.9%)	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078)	€11.684 (6.643 - 18.412)
		1421 (1.019 - 1.854)	€13.693 (7.538 - 22.232)
Alternative (21.1%)	€18.617.482 (13.112.341 - 23.596.670)	1702 (1.276 - 2.126)	€11.308 (6.380 - 17.941)
		1461 (1.046 - 1.902)	€13.239 (7.178 - 21.596)

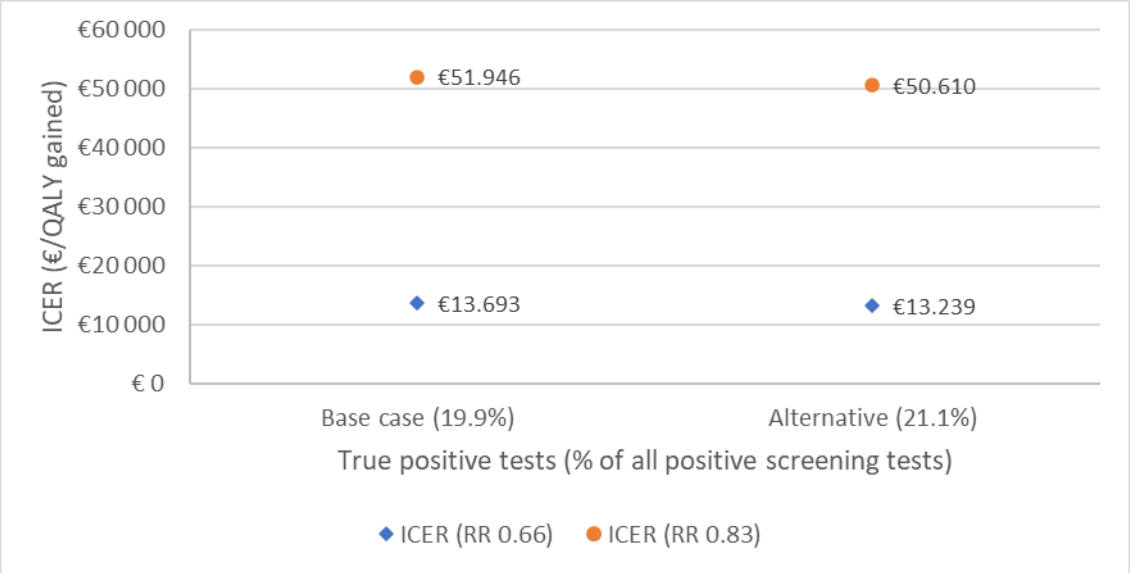
***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Table 57 – Results of the economic evaluation for true-positives scenarios (risk ratio 0.83)

RR: 0.83	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
True positives	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
Base case (19.9%)	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518)	*€42.800*
		549 (-72 - 1.303)	*€51.946*
Alternative (21.1%)	€28.662.256 (20.016.902 - 36.869.486)	686 (-46 - 1.556)	*€41.809*
		566 (-72 - 1.333)	*€50.610*

Abbreviations: *ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Figure 51 – Results of the economic evaluation for true-positives scenarios



Abbreviations: *ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RR: risk ratio.*

6.3.7 Overdiagnosis

In addition to the RR assumption, overdiagnosis is one of the variables that most strongly affects the cost-effectiveness of screening. The unjustified assumption that there is no overdiagnosis significantly improves the ICER to about €3000 per QALY gained (RR 0.66, Table 58) or almost €17 000 per QALY gained (RR 0.83, Table 59). With an assumed overdiagnosis rate of 10%, the ICER is about €8 000 and €34 000 per QALY gained, respectively. However, if the rates are based on evidence from all RCTs and all ages combined, with an overdiagnosis rate of 27.1%, the ICER rises to approximately €17 000 (RR 0.66) or €60 000 (RR 0.83) per QALY gained (Figure 52).

Face validity was checked by adding additional scenarios for all simulations. The impact of the overdiagnosis variable on incremental costs is proportional. Incremental costs rise by approximately

€6.6 million each time overdiagnosis increases by 10% (see Table 58 and Table 59). At first glance, one might not expect there to be any impact on incremental effects, given that overdiagnosed cancers in the model only generate additional costs and no health benefits. Patients having stage II, III or IV breast cancer in the comparator group (without screening), who would have a stage 0 or I breast cancer as a result of screening (i.e. the stage shift), would benefit from the screening.

In the economic model, breast cancer stages are initially modelled for the screening group based on how they are actually diagnosed (see section 6.2.7.5 for details on the stages of screen-detected and interval cancers). Before the stage shift is applied, the overdiagnosed cancers are isolated. As the literature indicates that overdiagnosis mainly concerns stage 0 and I breast cancers, the overdiagnosed cancers were removed from these stages. For simplicity, these cancers were first removed from stage 0. We have ensured that, in doing so, the total number of overdiagnosed stage 0 cancers does not exceed the total number of stage 0 cancers actually diagnosed in the screening group. As the percentage of overdiagnoses increases, a larger proportion of stage 0 cancers are classified as overdiagnosed cancers. As a consequence, after the stage shift fewer cancers can be categorized as stage 0 (since these cancers have already been placed in the overdiagnosis category), and more are categorized as stage I. Consequently, the incremental effect decreases with higher rates of overdiagnosis until all stage 0 breast cancers are classified as overdiagnosis. After that, the incremental effect remains stable (i.e. in the scenarios with an overdiagnosis >20%). This explains the results presented in Table 58 and Table 59.

Table 58 – Results of the economic evaluation for overdiagnosis scenarios (risk ratio 0.66)

RR: 0.66	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Overdiagnosis	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
Base case (20.7%)	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078) 1421 (1.019 - 1.854)	€11.684 (6.643 - 18.412) €13.693 (7.538 - 22.232)
Alternative (27.1%)	€23.083.271 (18.135.707 - 27.440.218)	1657 (1.240 - 2.081) 1421 (1.019 - 1.854)	€14.355 (8.930 - 21.753) €16.827 (10.170 - 26.480)
Alternative (0%)	€5.054.724 (536.064 - 9.132.338)	2051 (1.638 - 2.481) 1757 (1.346 - 2.213)	€2.608 (224 - 5.565) €3.062 (247 - 6.691)
Alternative (10%)	€11.697.946 (7.222.513 - 15.883.307)	1757 (1.345 - 2.184) 1506 (1.109 - 1.943)	€6.917 (3.349 - 11.684) €8.112 (3.876 - 14.074)
Alternative (20%)	€18.341.167 (13.788.376 - 22.481.564)	1657 (1.240 - 2.078) 1421 (1.019 - 1.854)	€11.442 (6.705 - 18.105) €13.414 (7.728 - 21.794)
Alternative (30%)	€24.976.179 (20.257.133 - 29.317.044)	1658 (1.240 - 2.086) 1421 (1.019 - 1.855)	€15.514 (9.821 - 23.338) €18.185 (11.314 - 28.174)

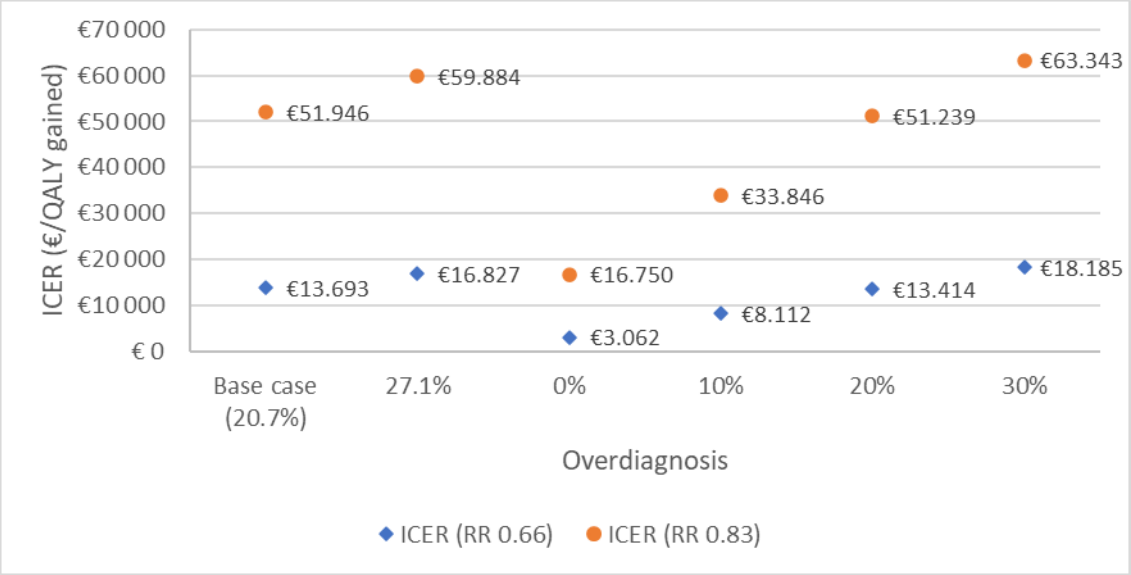
Abbreviations: *ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Table 59 – Results of the economic evaluation for overdiagnosis scenarios (risk ratio 0.83)

RR: 0.83	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Overdiagnosis	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
Base case (20.7%)	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€42.800* *€51.946*
Alternative (27.1%)	€32.862.067 (24.493.448 - 40.195.965)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€49.340* *€59.884*
Alternative (0%)	€14.831.950 (6.418.522 - 22.123.399)	1060 (341 - 1.912) 885 (263 - 1.649)	*€13.989* *€16.750*
Alternative (10%)	€21.475.172 (13.173.250 - 28.853.498)	766 (47 - 1.615) 635 (12 - 1.404)	*€28.020* *€33.846*
Alternative (20%)	€28.118.394 (19.494.464 - 35.641.750)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€42.217* *€51.239*
Alternative (30%)	€34.761.368 (26.261.373 - 42.148.372)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€52.189* *€63.343*

***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Figure 52 – Results of the economic evaluation for overdiagnosis scenarios



***Abbreviations**: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RR: risk ratio.*

6.3.8 HRQoL

The study by Rautenberg et al. reported the greatest difference in HRQoL, measured in breast cancer patients, between stage IV and other stages of breast cancer. If a smaller difference in utilities is modelled, as in the study by Wang et al., the ICER rises from approximately €13 700 to €15 000 (RR 0.66, Table 60) or from €52 000 to €57 000 (RR 0.83, Table 61) per QALY gained. Excluding a disutility attributed to false-positive tests, which is not justified, improves the ICER slightly. However, the impact on total QALY gain is minimal: a loss of 38 QALYs out of a total of over 1400 (RR 0.66 scenario) or almost 550 (RR 0.83 scenario). This is due to the short duration for which this disutility is assigned (only one month). If this short-term disutility were also assigned to overdiagnosis, the impact would be negligible (Figure 53).

Table 60 – Results of the economic evaluation for utility scenarios (risk ratio 0.66)

RR: 0.66	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Utilities	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
Base case Rautenberg et al. (2022)	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078) 1421 (1.019 - 1.854)	€11.684 (6.643 - 18.412) €13.693 (7.538 - 22.232)
Alternative Wang et al. (2018)	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078) 1303 (965 - 1.675)	€11.684 (6.643 - 18.412) €14.898 (8.285 - 23.599)
Alternative No disutility FPs	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078) 1459 (1.059 - 1.898)	€11.684 (6.643 - 18.412) €13.310 (7.411 - 21.306)
Alternative Disutility for overdiagnosed cancers*	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078) 1418 (1.016 - 1.851)	€11.684 (6.643 - 18.412) €13.723 (7.549 - 22.310)

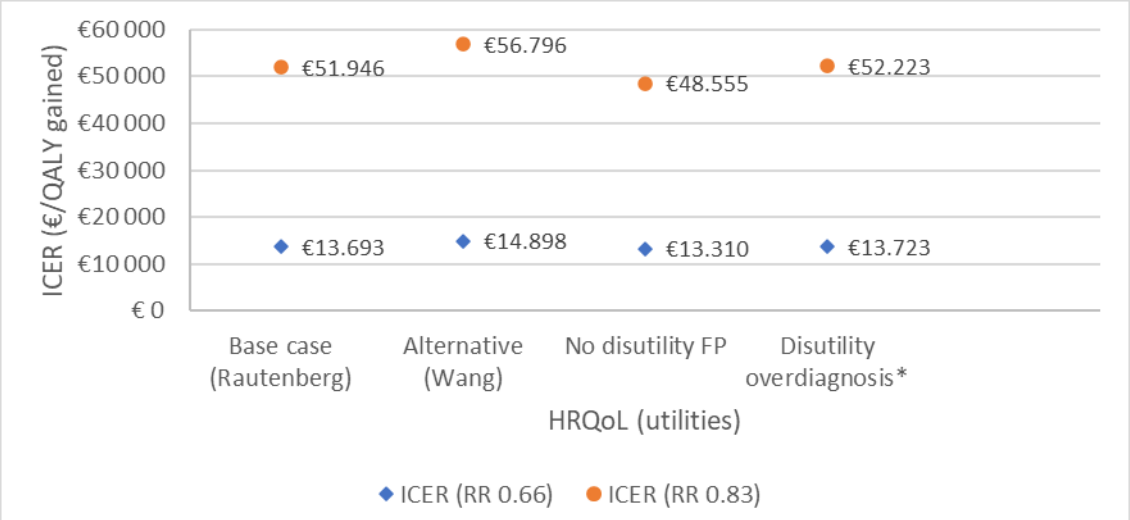
***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Table 61 – Results of the economic evaluation for utility scenarios (risk ratio 0.83)

RR: 0.83	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Utilities	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
Base case Rautenberg et al. (2022)	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€42.800* *€51.946*
Alternative Wang et al. (2018)	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518) 502 (-70 - 1.201)	*€42.800* *€56.796*
Alternative No disutility FPs	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518) 587 (-39 - 1.340)	*€42.800* *€48.555*
Alternative Disutility for overdiagnosed cancers*	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518) 546 (-75 - 1.300)	*€42.800* *€52.223*

***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Figure 53 – Results of the economic evaluation for utility scenarios



***Abbreviations**: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RR: risk ratio.*
** This scenario only includes a one-month disutility and does not include a disutility related to overtreatment.*

6.3.9 Price discount

As expected, the impact of a price reduction is that smaller reductions lead to higher treatment costs, primarily for stage IV breast cancers. Consequently, the incremental cost decreases and the cost-effectiveness of the intervention improves after applying the stage shift (see Table 62 and Table 63). Conversely, larger discounts lead to lower treatment costs, primarily for stage IV breast cancers. This results in higher incremental costs after applying the stage shift and reduces the cost-effectiveness of the intervention. However, the effect is smaller than expected because the discount is only applied to the proportion of expenditure on systemic therapy that falls under an MEA (see Table 37). As shown in Figure 54, the applied RR has a greater impact on the final result. With an RR of 0.66, the average ICER varies from €11 000 to €17 000 per QALY gained with discounts ranging from 0% to 90%. With an RR of 0.83, the respective range is approximately €50 000 to €55 000 per QALY gained.

Table 62 – Results of the economic evaluation for price discount scenarios (risk ratio 0.66)

RR: 0.66	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Price discount	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
0%	€15.472.249 (9.028.709 - 21.099.309)	1657 (1.240 - 2.078)	€9.718 (4.482 - 16.501)
		1421 (1.019 - 1.854)	€11.391 (5.166 - 19.864)
10%	€16.287.691 (10.125.110 - 21.714.442)	1657 (1.240 - 2.078)	€10.210 (5.010 - 16.979)
		1421 (1.019 - 1.854)	€11.966 (5.754 - 20.449)
20%	€17.103.132 (11.177.357 - 22.325.167)	1657 (1.240 - 2.078)	€10.701 (5.545 - 17.457)
		1421 (1.019 - 1.854)	€12.542 (6.343 - 21.034)
30%	€17.918.574 (12.227.217 - 22.948.213)	1657 (1.240 - 2.078)	€11.192 (6.094 - 17.935)
		1421 (1.019 - 1.854)	€13.117 (6.936 - 21.625)
40%	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078)	€11.684 (6.643 - 18.412)
		1421 (1.019 - 1.854)	€13.693 (7.538 - 22.232)
50%	€19.549.457 (14.457.522 - 24.209.159)	1657 (1.240 - 2.078)	€12.175 (7.179 - 18.882)
		1421 (1.019 - 1.854)	€14.268 (8.141 - 22.839)
60%	€20.364.898 (15.488.638 - 24.846.648)	1657 (1.240 - 2.078)	€12.666 (7.649 - 19.317)
		1421 (1.019 - 1.854)	€14.844 (8.696 - 23.446)
70%	€21.180.340 (16.554.488 - 25.429.288)	1657 (1.240 - 2.078)	€13.158 (8.150 - 19.752)
		1421 (1.019 - 1.854)	€15.419 (9.281 - 24.053)
80%	€21.995.782 (17.556.350 - 26.038.163)	1657 (1.240 - 2.078)	€13.649 (8.639 - 20.211)
		1421 (1.019 - 1.854)	€15.995 (9.867 - 24.660)
90%	€22.811.223 (18.577.463 - 26.654.623)	1657 (1.240 - 2.078)	€14.140 (9.169 - 20.720)
		1421 (1.019 - 1.854)	€16.571 (10.407 - 25.274)

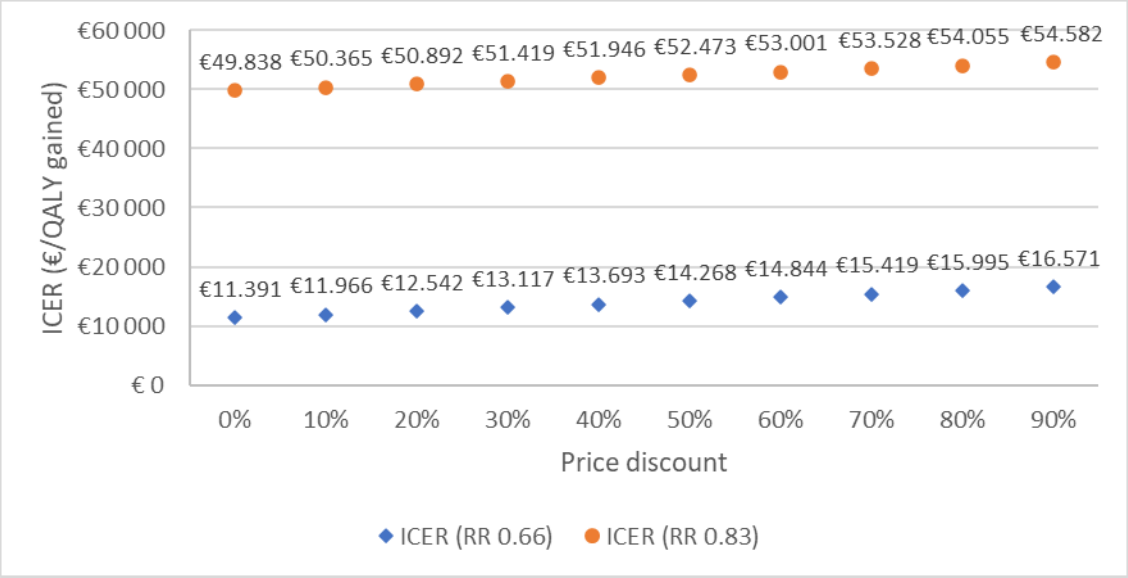
Abbreviations: *ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Table 63 – Results of the economic evaluation for price discount scenarios (risk ratio 0.83)

RR: 0.83	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Price discount	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
0%	€27.353.851 (17.195.043 - 36.964.845)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€41.062* *€49.838*
10%	€27.643.199 (17.908.675 - 36.940.018)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€41.496* *€50.365*
20%	€27.932.546 (18.597.152 - 36.852.832)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€41.931* *€50.892*
30%	€28.221.894 (19.285.812 - 36.737.025)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€42.365* *€51.419*
40%	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€42.800* *€51.946*
50%	€28.800.589 (20.845.041 - 36.444.316)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€43.234* *€52.473*
60%	€29.089.937 (21.574.911 - 36.388.844)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€43.668* *€53.001*
70%	€29.379.285 (22.219.687 - 36.272.447)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€44.103* *€53.528*
80%	€29.668.632 (22.828.208 - 36.164.448)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€44.537* *€54.055*
90%	€29.957.980 (23.484.600 - 36.069.467)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€44.971* *€54.582*

***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Figure 54 – Results of the economic evaluation for price discount scenarios



***Abbreviations**: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RR: risk ratio.*
Note: the price discount is only applied on the part of the systemic therapy cost that is reimbursed under a managed entry agreement.

6.3.10 Supplements

The following Table 64, Table 65 and Figure 55 demonstrate that the impact of adding supplements is limited. This is mainly because the impact on total treatment costs for all cancer stages, and therefore incremental costs, was also limited.

Table 64 – Results of the economic evaluation for supplements scenarios (risk ratio 0.66)

RR: 0.66	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Supplements	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
Base case (exclusive)	€18.734.015 (13.345.328 - 23.563.309)	1657 (1.240 - 2.078)	€11.684 (6.643 - 18.412)
		1421 (1.019 - 1.854)	€13.693 (7.538 - 22.232)
Alternative (inclusive)	€19.560.921 (14.319.690 - 24.474.188)	1657 (1.240 - 2.078)	€12.191 (7.095 - 19.113)
		1421 (1.019 - 1.854)	€14.289 (8.168 - 23.035)

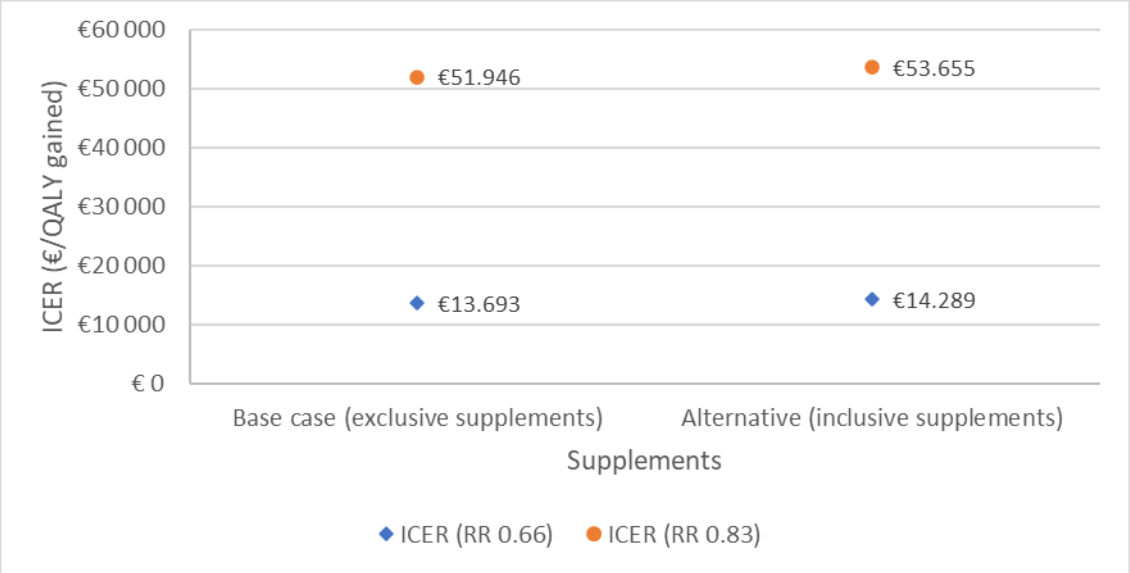
***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Table 65 – Results of the economic evaluation for supplements scenarios (risk ratio 0.83)

RR: 0.83	Incremental cost	IE (LYG) IE (QALYs gained)	ICER (€/LYG) ICER (€/QALY gained)
Supplements	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)	mean (2.5% - 97.5%)
Base case (exclusive)	€28.511.242 (20.052.379 - 36.579.645)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€42.800* *€51.946*
Alternative (inclusive)	€29.449.378 (20.868.848 - 37.940.589)	666 (-46 - 1.518) 549 (-72 - 1.303)	*€44.208* *€53.655*

***Abbreviations**: ICER: incremental cost-effectiveness ratio; IE: incremental effect; LYG: life-year gained; QALY: quality-adjusted life year; RR: risk ratio.*

Figure 55 – Results of the economic evaluation for supplements scenarios

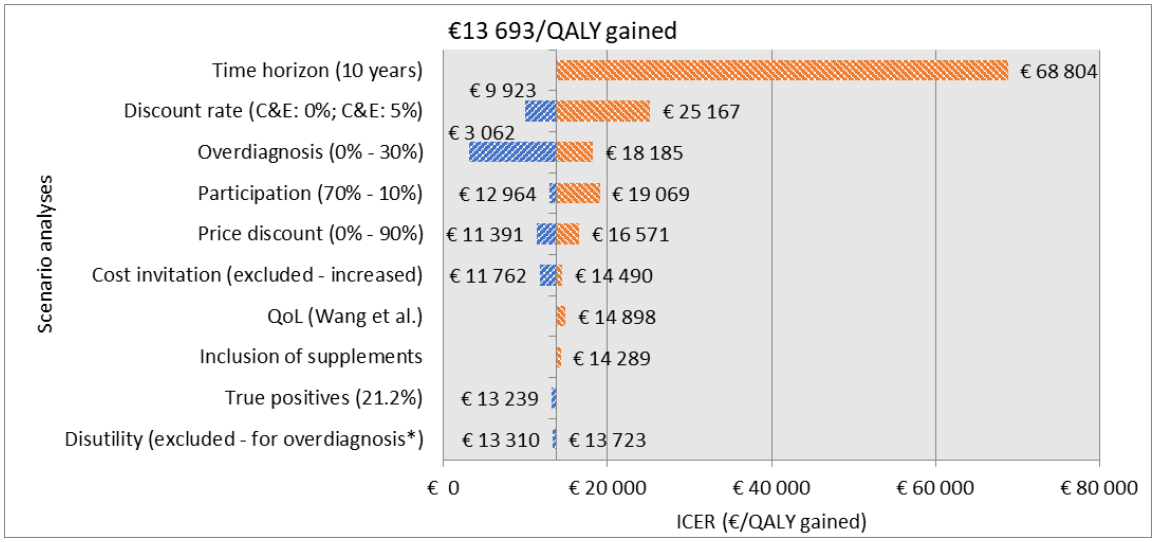


***Abbreviations**: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RR: risk ratio.*

6.3.11 Tornado graphs

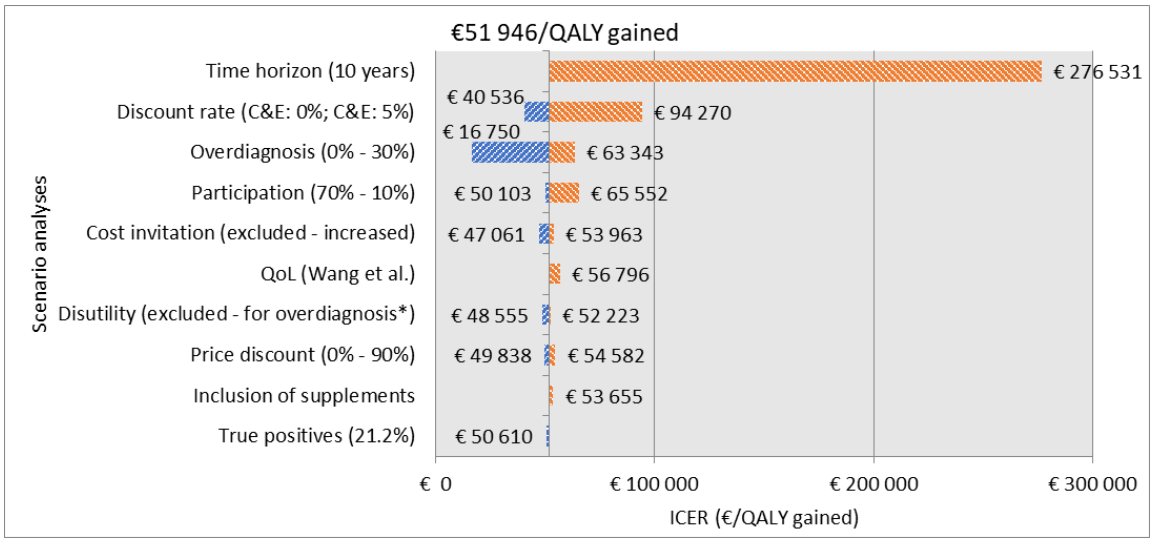
The tornado graphs (Figure 56 and Figure 57) demonstrate that the chosen time horizon and discount rate are two key factors that determine the cost-effectiveness of screening. However, these variables are independent of the evidence regarding screening and reflect the requirements set out in the Belgian guidelines for economic evaluations. When we examine the clinical input variables, it becomes clear that overdiagnosis, along with the risk ratio, has the most significant influence on the cost-effectiveness of screening.

Figure 56 – Results of the economic evaluation – tornado graph (risk ratio 0.66)



***Abbreviations**: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RR: risk ratio.*

Figure 57 – Results of the economic evaluation – tornado graph (risk ratio 0.83)



***Abbreviations**: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RR: risk ratio.*

6.3.12 Validation

We successfully modelled all input variables correctly after 1000 simulations. This applies not only to the outputs presented in section 6.3.1, but also to other variables, such as the presented stage distribution and treatment costs.

One output deserves special attention: the difference in survival. Based on our model, after 13 years and with 258 456 women participating, in the scenario with a risk ratio of 0.66, there are 76 more survivors in the screening group than in the comparator group, representing a difference of 29.3 additional survivors (95% CI: 22.2 - 36.8) per 100 000 participants. The peak is reached after 17 years, with a difference of 31.9 per 100 000 participants. Applying a risk ratio of 0.83, this is only a difference of 12.0 survivors per (95% CI: -0.4 - 26.8) 100 000 participants after 13 years, with a peak of 13 extra survivors after 17 years. Both results fall within the confidence interval presented in the clinical section of this report, but are lower than the average of 61 persons (95% CI: -61 - 182, see section 3.2.2.1). It makes sense that our results do not reflect the average perfectly, given that we did not base our cost-effectiveness calculation on a single RCT or meta-analysis, but instead used an alternative 'what

if no screening was performed' analysis. One possible explanation for the difference is that the RCTs in the clinical section are much older studies, where survival was influenced by the available treatments at the time. Since then, survival rates for women with breast cancer have improved. Applying more recent age- and stage-specific survival curves results in a smaller incremental difference in survival rates between higher and lower stages of breast cancer, limiting the incremental impact. We will come back to this in the discussion.

6.4 Discussion

6.4.1 Why performing a 'hypothetical' economic evaluation

The strength of an economic evaluation depends on the strength and reliability of the underlying evidence. The RCTs on which the clinical section of this report is based are performed in the period 1963 to 1997 and most trials carry a high risk of selection bias. More recent observational data are not a suitable alternative to determine evidence related to the relative effect of screening compared to no screening, as they are even more susceptible to various forms of bias. Therefore, we know in advance that it will be difficult to carry out a robust, evidence-based economic evaluation. We want to highlight this issues with the underlying evidence by explicitly referring to **a 'hypothetical' economic evaluation**, where the results depend heavily on the underlying assumptions.

The decision to conduct an economic evaluation was motivated by the observation of shortcomings in existing evaluations and the difficulty of applying their results to the Belgian context. These relate to a combination of factors, such as the failure to include the costs associated with organising a screening campaign, the use of a high participation rate (up to 100% in the base case), a lack of transparency regarding how overdiagnosis was included, the inclusion of utilities lacking face validity, or uncertainty regarding how the effect of screening was modelled. Because these variables are highly influential and difficult to adjust in previous models, we decided to conduct our own context-specific economic evaluation, taking these elements into account wherever possible.

For this evaluation, we focused on **evaluating Belgium's current screening programmes**, which invite asymptomatic women aged 50–69 for a screening mammogram every two years. This is primarily because available Belgian data allows us to determine the current screening status for this group, as a specific invoicing code is used for mammograms performed as part of the organised screening programmes (the intervention). By performing a hypothetical 'what if' analysis, it is possible to model the counterfactual, i.e. the situation as if no screening had been carried out (the comparator) and calculate the cost-effectiveness of the intervention.

However, for the younger (40–49 years) and older (70–74 years) age groups, this approach cannot be applied. The available data for these age groups cannot distinguish between mammograms performed due to the onset of symptoms and those performed as opportunistic screening. Therefore, it is not possible to form a clear picture of the comparator group without screening and apply the inverse 'what if' approach. Furthermore, the supporting evidence for carrying out screening in these age groups was also very limited. For these groups, we begin with the results obtained in women aged 50–69 and indicate the likely impact of extending screening to younger or older women.

6.4.2 Strengths and limitations of the economic evaluation

6.4.2.1 Strengths

Our access to Belgian real-world data gives us the advantage of being able to form an accurate picture of screening within the context of population-based screening programmes. This enables us to create the most accurate possible picture of the intervention group. The **BCR** provides us with context-specific information on the number of invitations, participants, positive screening tests, true and false positives, interval cancers and the stages of screen-detected and interval cancers.

Another advantage is that we have access to vital status data (up to 2025), obtained via linkage of the BCR data with the **Crossroads Bank for Social Security**, which allows us to include long-term age- and stage-specific survival probabilities in the model. To accurately reflect the target population invited for screening, we excluded women with an invasive breast tumour in the preceding 10 years or another invasive tumour in the preceding 5 years. Finally, we utilised data from 98 025 women aged 50–69 years with (combined[yyy]) stage 0–IV breast cancer diagnosed between 2004 and 2022.[d]

When modelling mortality, we deliberately chose to model **overall mortality**. If we were to use disease-specific mortality, we would need to explicitly include non-breast cancer-related mortality and verify that the total mortality estimate remained realistic. However, this is not necessary, as total mortality data are already available by age and breast cancer stage. The advantage of this approach is that no statements about the cause of death need to be made. It implicitly considers all expected and unexpected positive and negative impacts on mortality associated with the diagnosis and/or treatment of breast cancer. Consequently, there is less potential for bias.

Another advantage of using real-world, age- and stage-specific survival data is that it eliminates the need to make assumptions about breast cancer progression. Not only would this add complexity to the model, but a lack of data on progression would also make the modelling challenging. We only have information on the stage of breast cancer at diagnosis. Nevertheless, the **impact of progression on mortality is implicitly taken into account in the available stage-specific survival figures**, which is an additional strength of these data. Similarly, the **impact of progression is implicitly incorporated in the underlying cost data**, available per stage at diagnosis, and no complex assumptions need to be made.

Another strength of the evaluation is our access to the **Health insurance data (i.e. IMA-AIM database)**, which allows us to analyse costs reimbursed by the national health insurance and the patient's share for diagnostic, therapeutic and follow-up procedures. These costs were categorised transparently into six groups (diagnostic and follow-up procedures; surgery; radiotherapy; systemic therapy; palliative care; and a miscellaneous group) For each group, the selected billing codes and Anatomical Therapeutic Chemical (ATC) product codes were submitted to clinical experts for amendment, correction and validation. This enabled us to calculate relevant real-world costs for the treatment of the various (combined) stages (0 to IV) of breast cancer up to five years after cancer incidence (i.e. the maximum duration for which IMA–AIM data are linked). In order to map costs relating to the diagnosis, follow-up and treatment of breast cancer as accurately as possible, women who had experienced any other invasive tumour within five years prior to or following the diagnosis of breast cancer were excluded from the cost analysis.

As false positives are not included in the BCR-IMA-AIM database (as they were finally not diagnosed with breast cancer), we relied on the DD database (the medical acts in screening population database; see chapter 5, section 5.1.3). Based on the frequencies of procedures performed in women with a false-positive screening result, combined with the most recent tariffs and a limited number of assumptions, it was possible to estimate the cost of the false-positive tests. In general, every effort was made to **exhaustively cover all costs while taking into account the complexity of this analysis**, and reporting on the costs included in the model was made as **transparent** as possible.

In addition, the **most recent and complete cost data available was used**. As the available IMA-AIM dataset extended to 30 June 2024, 2023 is the most recent year for which there are almost complete records of healthcare interventions. Consequently, 2022 was selected as the incidence year to identify costs incurred during the first year after diagnosis of breast cancer. For each subsequent year (after diagnosis of breast cancer), an earlier incidence year was chosen. This approach was favoured over longitudinal follow-up of the same population. More specifically, it allowed us to avoid using outdated cost data, given that more expensive systemic therapies have been introduced and reimbursed in recent years. Inclusion of more expensive treatments used in later stages of breast cancer is in favour of screening. By shifting the stage, the costs associated with these higher stages can be avoided, reducing the incremental costs of screening and having a positive impact on the ICER.

We were also able to identify **which systemic therapies are reimbursed under a managed entry agreement (MEAs or confidential contracts)** in order to account for the impact of a potential confidential discount, as required by the Belgian guidelines for economic evaluations.[3]

Finally, it has been observed that for HRQoL, various economic evaluations make use of values based on expert opinion or non-patients (see section 4.2.3.2) These evaluations lead to overly optimistic results and should be interpreted with great caution. As both the Belgian guidelines for economic evaluations and the EUnetHTA guidelines for measuring HRQoL recommend that patients assess their own health states,[3, 288] we conducted a pragmatic, non-systematic search for systematic reviews and meta-analyses to identify stage-specific health utilities and screening-specific (dis)utilities measured in breast cancer patients (see chapter 6, section 6.2.7.8). The data confirm that, based on hypothetical assumptions, non-patients estimate a greater impact of breast cancer on quality of life. In our model, we only retained **HRQoL values derived from breast cancer patients**.

6.4.2.2 Limitations

As previously mentioned, the main weakness of the economic evaluation is the **lack of recent, unbiased RCTs**. However, this limitation is shared by all existing economic evaluations of breast cancer screening. We will take a closer look at the evidence for two of the most critical variables in section 6.4.4.

We also encountered several limitations with regard to costs. Firstly, the most recent year for which quasi-**complete invoicing data were available was 2023**. We chose not to apply expert-based adjustments to include cost information from more recent years because it is unclear for which stages of breast cancer such adjustments should be made. The scenarios with a higher or lower discount for systemic therapy might already demonstrate the potential magnitude of the impact of such a change. Nevertheless, newer treatments may be more expensive, but both their impact on survival and the associated additional costs should be considered. Higher incremental costs for treating stage IV cancers would improve the cost-effectiveness of screening. Conversely, improved survival outcomes for stage IV breast cancer would reduce the cost-effectiveness of screening. Both effects need to be taken into account. Paradoxically, the greater the use of therapies in stage IV disease that are not cost-effective – namely those associated with high additional costs and little or no survival and/or QoL benefit – the more favourable the cost-effectiveness of screening becomes (and vice versa). To support the efficient use of resources, reimbursement of treatments with such characteristics should be avoided.

We intentionally chose to work with a **smaller sample size for our cost analysis**, as we only used data from the most recent available year. We opted for this approach to ensure the validity of the obtained results. Using older samples would result in lower costs for systemic therapy, for example, and these would no longer accurately reflect the current situation regarding reimbursed treatments. As a result of the smaller sample size, we were unable to break down the costs according to the patient's age, only according to the stage of breast cancer. However, this should not impact the incremental costs calculated in the model, since the same stage shift was modelled for all 5-year age categories between 50 and 69 years, and the results for the 50–69 age group were presented as a single outcome.

In the case of palliative care costs, there may be an **underestimation of costs**. Costs in this category were identified using predefined codes. However, it is possible that palliative care is provided during a hospital stay. Only surveillance fees were included for in-hospital palliative care. No complex reconstruction of hospital stays was performed, unlike for surgical hospitalisation costs (see Appendix 24). Nevertheless, the presence of many empty cells (i.e. no such costs were observed for most patients) suggests that the non-exhaustive inclusion of hospital costs is unlikely to substantially bias the results.

With regard to HRQoL, we have selected data from studies in which breast cancer patients were questioned. One limitation is that the **selected HRQoL studies primarily involved populations of**

Asian origin. It is unclear to what extent the results from this population are applicable to the Belgian population. Unfortunately, none of the non-Asian studies by Perciello et al. (2020) or Criscitiello et al. (2021) was suited for use in the model as none included stage IV patients.[279, 281] A better understanding of the HRQoL of Belgian breast cancer patients would help to address this limitation.

Furthermore, we have been **unable to model changes in HRQoL either as the breast cancer population ages or as the disease progresses**. While this is implicitly reflected in the survival and cost data, the same principle does not apply to the utilities used in the model. To simplify the model, a single utility was assumed based on the initial cancer diagnosis for an ageing population. This assumption favours the screening intervention (see below).

For methodological reasons, we note that various data points are based on the 2020 screening year. This applies to the number of positive and negative screening results, the percentage of true and false positives, and the number of interval cancers. For the latter group, more recent data cannot be used, as two-year follow-up data is required to accurately estimate this figure. However, 2020 was also the year of **the COVID-19 pandemic** and this may have influenced the recorded values. Regarding specificity, we note that it corresponds to the value reported for Flanders in 2021. For other data, such as the number of invitees (2023), participation rate (2022) and cancer stage at diagnosis for true-positive and interval cancers (2020–2022), we selected more recent data or data covering a longer period to obtain more 'stable' results. It is unclear to what extent the values have been impacted by the pandemic. However, what is more important is that the variables that are most decisive for cost-effectiveness, i.e. the risk ratios for stages II, III and IV breast cancer used in the stage shift and the percentage of overdiagnosis, are based on available evidence and applied as relative percentages. Therefore, there is no impact of the pandemic on these input variables used in the hypothetical economic evaluation.

6.4.3 Optimistic assumptions

In situations of high uncertainty, it can be helpful to develop worst- and **best-case scenarios**. If only pessimistic assumptions were applied, the screening intervention involving an additional initial cost would never be cost-effective. This scenario was therefore not developed because it would not be considered acceptable. Conversely, modelling a scenario incorporating optimistic assumptions could be of interest. If the result were to yield relatively high ICERs, the cost-effectiveness of the intervention could be questioned. In such a situation, less optimistic assumptions only worsen the cost-effectiveness. While the optimistic assumptions are prone to justified criticism, the outcome regarding the cost-effectiveness would nevertheless be clear. This is why we initially carried out a hypothetical calculation in which **optimistic assumptions** were made for various variables. These include the following elements:

- **No negative impact on quality of life (QoL) was modelled for overdiagnosis and overtreatment**. Screening is undeniably associated with overdiagnosis, leading to adverse consequences and psychological distress. This may negatively impact patients' QoL. Based on an analysis of screen-detected stage 0 breast cancers (incidence year 2022), 11.3% received endocrine therapy, 15.0% radiotherapy, 0.2% systemic therapy and 21.8% surgery (or 22.0% one of these interventions) during the first year after diagnosis. In stage I, this is already 69.5%, 69.9%, 12.1% and 77.0% (combined 77.1%), respectively. It is realistic to assume that undergoing these treatments negatively impacts women's quality of life. Only Morton et al. corrected for all screening-related harms by multiplying the QALY gain by 0.253.[194] However, the evidence base for this correction is weak. None of the other identified economic evaluations applied a specific disutility for overdiagnosis. The main reason for this omission is the lack of evidence. Neither the systematic review by Li et al.,[205] nor our updated pragmatic review identified any studies examining the impact of overdiagnosis on quality of life. From an individual perspective, it is impossible to determine which patients have been overdiagnosed. This is not surprising, since the concept overdiagnosis is only to be applied at a population level, not at an individual level. At an individual level one cannot distinguish between a patient being "truly" diagnosed with breast

cancer and someone “over-”diagnosed with breast cancer. Patients may therefore perceive a cancer diagnosis following screening as life-saving and may not necessarily view the harms resulting from overdiagnosis and overtreatment negatively. However, if we were to model the potential negative impact on HRQoL from overdiagnosis, the consequences could be substantial. This is because the population affected is relatively large, and the impact could endure over an extended period.

- **No decline in HRQoL was foreseen with ageing or disease progression**. In fifteen of the seventeen identified economic evaluations, utilities were not differentiated based on age. For the sake of simplicity, our model also assumed a constant utility for an ageing population. Given the longer survival in the intervention group, driven by the modelled stage shift, including a decline in HRQoL would reduce the intervention's incremental effects and thus worsen its cost-effectiveness.
- **The study showing the greatest difference in quality of life between earlier stages and stage IV, measured in breast cancer patients, was selected**.[276] If a study with a smaller incremental difference in utilities between lower and higher stages of breast cancer had been chosen, this would have had a negative impact on the gain in QALYs and, consequently, on the ICER.
- **No disutility was modelled for participation in screening**. Of the identified studies that considered a disutility for participants due to the physical and psychological burden of mammography screening, two-thirds used a disutility value of 0.006 over a period of one week (see section 4.2.3), which was rounded to 0.01 by Blankenburg et al. and van Luijt et al.[179, 192] Rafia et al. assumed a 20% reduction in QoL for two hours as a result of participating in screening.[195] However, given the limited impact in terms of both magnitude and duration, applying this disutility will only have a minor influence on the result. A back-of-the-envelope calculation indicates that applying a decrease of 0.006 for one week would result in an additional loss of 29.7 QALYs if applied to 258 289 participants, which would only have a minor negative effect on the cost-effectiveness of screening.
- **The model did not include a lead time**. In principle, a lead time implies that the cancer, and thus the necessary treatment, would be diagnosed and treated at a later point in time. Unfortunately, the duration of this lead time is still debated in the literature without a clear best estimate. Even if the lead time were limited to a few years, an annual discount rate of 3% would reduce the costs in the comparator group and have a negative impact on the cost-effectiveness of screening.
- **The long-term survival curves were calculated for a population over the incidence years 2004–2022**. Due to developments in various areas, survival has improved since the screening RCTs were performed (1963-1997). This improvement could explain the smaller absolute treatment effect observed in the model compared to the evidence in the clinical section of this report (see section 6.2.7.13). We also note that the survival curves for the years 2016–2022 are more favourable (see Figure 34 in Section 5.3.4). In absolute terms, the greatest effect is seen for stages III and IV breast cancer, meaning that the incremental benefit due to the stage shift would decrease if more recent survival curves were used. While the additional costs of more recent, expensive treatments have already been taken into account as much as possible, this is not the case for potentially improved survival. We chose to use the older survival curves, which have a longer follow-up period, and to explicitly acknowledge this as an optimistic assumption. Adjusting for improved survival would have a negative impact on the cost-effectiveness of screening.

A number of smaller elements were also modelled relatively optimistically. For example, the disutility assigned to false positives for a period of one month was set at 0.07. Most studies in the economic literature have assigned a disutility of 0.105 for five weeks to participants with a false-positive result

requiring diagnostic work-up.[180, 181, 183, 184, 189, 192, 200] However, as previously mentioned, the incremental impact on the loss of QALYs is minimal.

Finally, the relatively high specificity of 97.5%, derived from Belgian real-world data and included in the model, is worth mentioning. The periodic reports on the regional screening programmes report specificity and sensitivity for each of the regions separately.[80, 81, 85] For Flanders (2022), a specificity of 97.5% and a sensitivity of 67.5% were reported. For Wallonia (2022), these figures were 89.4% and 71.4%. For Brussels (2022), these figures were 90.1% and 72.5%, respectively. The review by Shi et al.[293] reported that the sensitivity of mammography techniques ranged from 55% to 91%, and the specificity from 84% to 97%. However, it is not possible to simply adjust the specificity in the model and assess the impact on the ICER. Lower rates of specificity are associated with higher numbers of false-positive cases. If the specificity in the model is reduced to the values reported for Wallonia or Brussels, the number of false positives would exceed the total number of positives, which is not possible. In other words, if the specificity is adjusted, the sensitivity must also be adjusted. However, we start from the current situation in Belgium regarding the number of positive tests, as well as reflecting the Belgian real-world stage distribution of screen-detected breast cancers, without explicitly including sensitivity as a variable. Since we have data that accurately reflect the current situation of population-based screening in Belgium, we have chosen to present only the results derived directly from the available real-world data.

6.4.4 The most determining variables

Two important variables for the cost-effectiveness of screening are the **time horizon** over which costs and benefits are calculated and the applied **discount rate**. However, there is no issue of choice, as the Belgian guidelines for conducting economic evaluations are followed. Using a longer time horizon (as was done in our model) ensures that any increase in survival is taken into account. The lower discount rate for effects (1.5% instead of 3%), favours interventions that require short-term investment and have long-term effects.

In our hypothetical approach, we have chosen to start from the current situation and represent it as accurately as possible using the available data. To mimic a situation without breast cancer screening, various risk ratios and rates of overdiagnosis were used. The resulting QALYs are calculated by applying a stage shift after excluding overdiagnosis, which only incurs costs and has no beneficial effect. Therefore, the **percentage of overdiagnosis and the applied risk ratio for being diagnosed with stages II, III and IV breast cancer** are two key variables that determine the cost-effectiveness of screening.

Several approaches are applied in the identified economic evaluations for the **stage shift**: extract the stage distribution from data before the introduction of population screening (i.e. data dating back to the 70s,[180] 80s,[193] and 90s[196] respectively) to simulate the situation without screening; using the stage distribution of the closest screened age group;[178] or simulating tumour growth or size at various ages, followed by simulating tumour detection by screening based on screening sensitivity and specificity parameters or a detection probability function.[183, 187, 189-191, 195, 200] All these approaches are associated with considerable uncertainty. The latter approach also often lacks transparency regarding the parameters, techniques and/or underlying sources used.

As indicated by Pillay et al.,[278] **overdiagnosis** is often not fully understood by participants: "*It is well documented that uninformed people overestimate the beneficial effects of screening[294] and often do not have any awareness or conceptual understanding of overdiagnosis,[295, 296] such as when non-lifesaving early-stage breast cancer detection is not viewed as a harm but rather a benefit when discussed among social networks.[297]*" Evidence on overdiagnosis is also ambiguous, as many studies have a high risk of bias. It cannot be ruled out that individuals in the comparator group underwent screening during or after the study ended. This contamination masks the overdiagnosis. Limiting the analysis to studies that did not offer screening to the control group at the end of the study, after a follow-up of 7-9 years, provides rates above 20%. This percentage is expressed relative to all cancers

diagnosed in the intervention group (i.e. screen-detected and interval cancers, and cancers diagnosed in non-participants).

An overdiagnosis rate of 20.7% was used in the base case. However, if we base our analysis on all studies with a follow-up of 7-9 years, including those subject to bias, we arrive at an overdiagnosis rate of 27.1%. When extending the FU to ten to fifteen years , 9.2% of all cancers detected in the screening group were overdiagnosed. However, it is unclear to what extent participants in the control (and intervention) group underwent screening after the end of the study and during the subsequent follow-up period. Some degree of contamination cannot be excluded, which would result in overdiagnosis being underestimated.

Of the seventeen studies, only fourteen explicitly mention overdiagnosis. Input values ranged between 2%[192] and 31%[191] and varied considerably depending on age eligibility criteria and screening interval. For example, Tina Shih et al. applied overdiagnosis rates of 12.4% and 10.5% to biennial and triennial screenings, respectively, for women aged 50-75.[183] Rafia et al. reported age-specific overdiagnosis rates, estimating overdiagnosis at 6.2% at age 72 and 37.9% at age 90.[195] By definition, overdiagnosis does not lead to improved survival, resulting in overtreatment which is associated with additional costs and a potential loss of quality of life. Therefore, an underestimation of the number of overdiagnoses inevitably leads to an overestimation of the cost-effectiveness of screening. As previously mentioned, including the loss of HRQoL due to overtreatment would have an even more negative impact on the results.

When it comes to the stage shift in diagnosed breast cancers, we are faced with weak evidence. However, we can draw on the results of a meta-analysis based on older RCTs, most of which are subject to a high risk of bias. The meta-analysis indicates that mammography screening results in a statistically significant reduction of 34% of breast cancers diagnosed at stage II or higher compared to usual care (RR: 0.66, 95% CI: 0.61-0.72, p<0.001). If restricted to the single trial with a low risk of selection bias, the Malmö I study, the respective RR is 0.83 (95% CI: 0.68-1.00, p=0.05). Using the first estimate in combination with the optimistic assumptions mentioned above results in an average ICER of around €13 700 per QALY gained. Using the second estimate results in an ICER of around €51 900 per QALY gained. This demonstrates how sensitive the result is to this stage shift variable, which is discussed further in section 6.4.5.

Lastly, the evidence from the systematic review on clinical effectiveness was expressed in relation to the number of people invited for screening rather than the actual number of participants, which might impact the reported probabilities. Yet, in the breast cancer screening trials the participation rate is relatively high in most RCTs (see footnote[hhhhh]). Reporting results in relation to the actual number of participants (i.e. reducing the denominator) could have an impact. For both overdiagnosis and the risk ratio for the detection of stage II, III or IV breast cancer, this would increase the percentage. A higher percentage of overdiagnosis increases the ICER, while a better (i.e lower) risk ratio decreases the ICER. No further corrections were applied to either variable since the necessary information was not readily available in the underlying RCTs. Rather, preference was given to demonstrating transparently the significant impact of these variables on the intervention's cost-effectiveness.

6.4.5 A reflection based on real-world data

The evidence from RCTs is limited to older trials, many of which have a high risk of selection bias. Drawing firm conclusions from real-world data is also not possible, as a comparative group is missing. Therefore, it is unclear what the situation would be if no screening were carried out. Nevertheless, we will briefly examine a selection of real-world data to see whether the expected and modelled stage shift (i.e. a decline in stages II, III and IV breast cancers at diagnosis) is also observed under real-world conditions.

6.4.5.1 European population-based study (Cardoso et al.)

The recently published study of Cardoso et al. (2026) aimed to analyse and compare changes over time in breast cancer incidence, by stage at diagnosis, and breast cancer mortality across countries in relation to the timing of screening implementation and age at diagnosis.[17] This population-based study was conducted in 21 European countries. As mentioned by the authors, their study does not allow quantification of the screening's specific contribution to the observed trends or direct comparison with randomised trial estimates.[17]

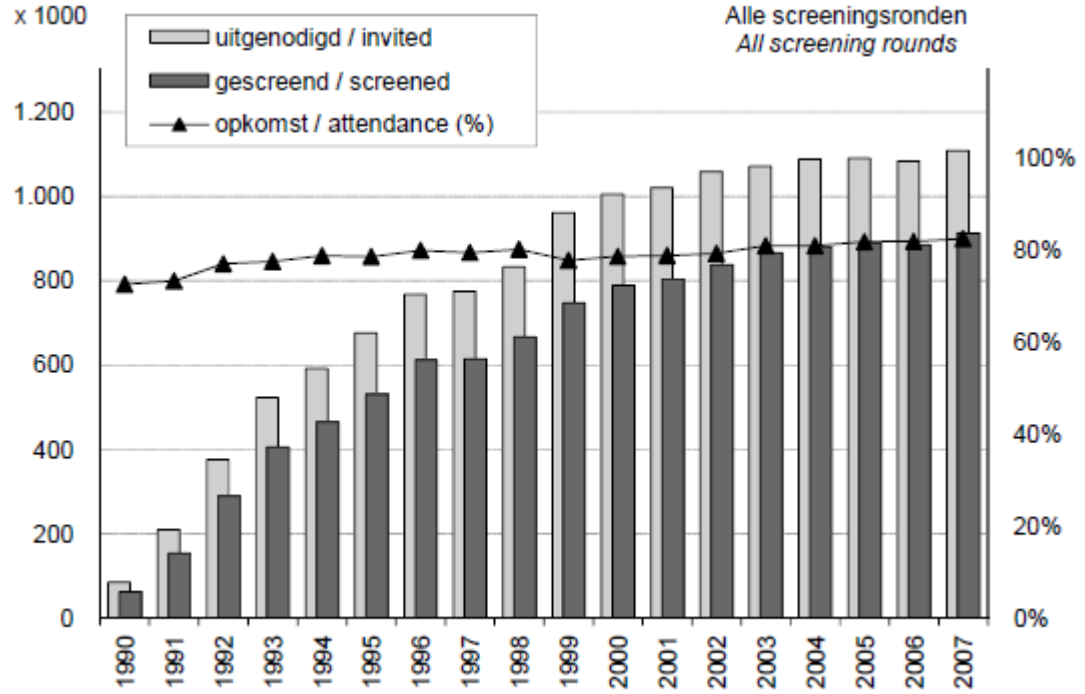
The results of this study show "*the largest increases in incidence were seen for in situ and stage I cancers ([Average annual percent changes] AAPCs ranging from non-significant to 12.03 (95% CI, 7.40–16.86) following screening implementation). Incidence of stage IV cancer declined or remained stable in most countries, with AAPCs down to −6.16 (95% CI, −8.28 to −4.00) after screening introduction. These trends were particularly pronounced among age groups targeted by screening (mostly 50–69 years).*"[17] The impact on stage IV breast cancer is thus very heterogeneous across European countries. Autier (2026) remarks that, for the most optimistic AAPC, the average annual percent changes of stage IV breast cancers was based on results of one country (Portugal), where 20% of stage were unclassified or unknown. Breast cancer patients with missing stage data have a poorer survival,[298] most probably because they may be diagnosed with more advanced-stage cancer.[18] The results of this observational study may thus have been affected by missing stage data. Given the diverse impact across countries, it is not possible to extract whether or not there is a decrease, or the size of this decrease, in stage IV breast cancer due to the introduction of screening.

Concerning overdiagnosis, the "*data also suggests that mammography screening has led to overdiagnosis, as the increases in incidence of in situ and stage I cancers were much larger than the decreases in incidence of advanced-stage cancers, and overall, no marked changes in incidence were observed among the post-screening age groups.*"[17]

6.4.5.2 The Netherlands Cancer Registry (NCR)

The Netherlands Cancer Registry (NCR), maintained by the Netherlands Comprehensive Cancer Organisation (IKNL – Integraal Kankercentrum Nederland), provides statistics on cancer in the Netherlands. It provides cancer statistics since 1989, also available through an interactive web-based portal (https://nkr-cijfers.iknl.nl/viewer). The following figures present the number of individuals invited and screened in the Netherlands (Figure 58) and incidence figures (Figure 59).

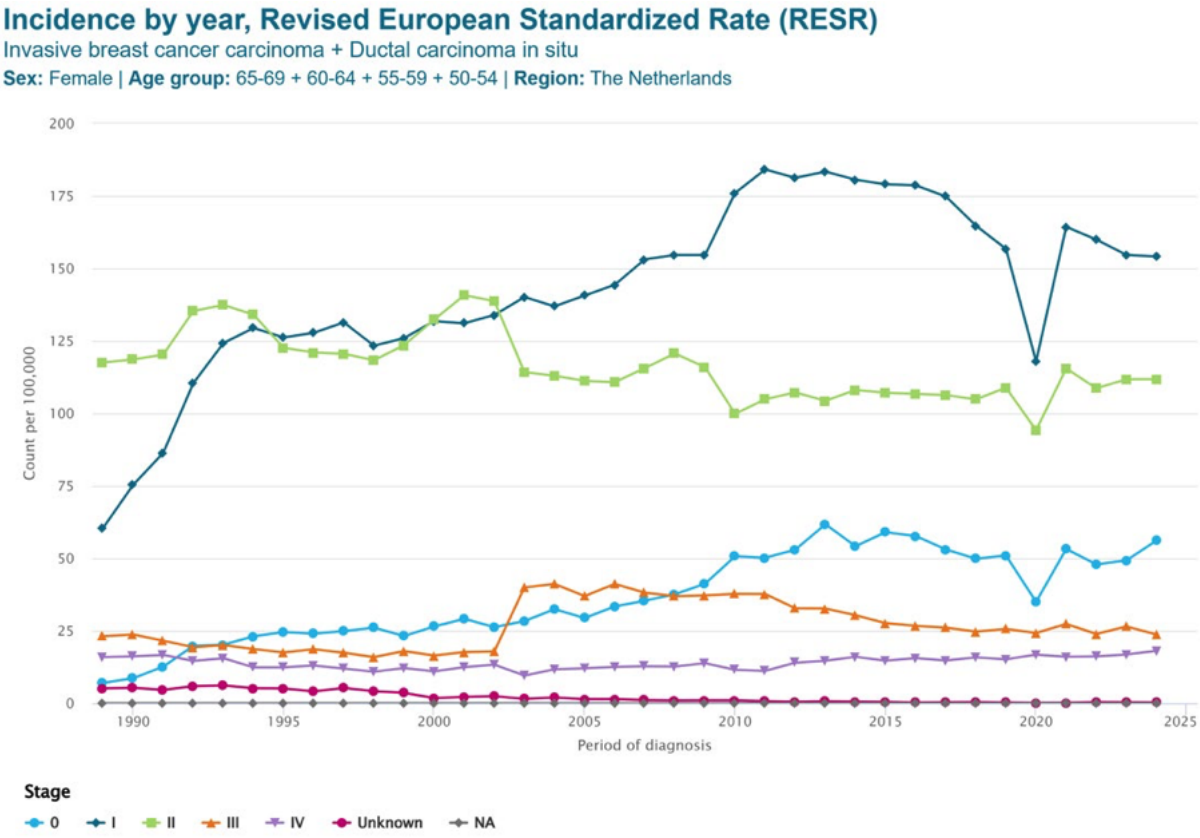
Figure 58 – Number of invited and screened women and attendance rate in the Netherlands (1990 – 2007)



***Source**: NETB (2009)[243]. 1990-1997: women aged 50–69; 1998–2007: women aged 50–75.*

The figure shows a sharp increase in the number of women invited for screening in the early years. In 1998, the screening programme was extended to include women aged 75. While fewer than 100 000 women were screened in 1990, this figure had risen to over 600 000 in 1996, increasing further to around 900 000 by 2007.

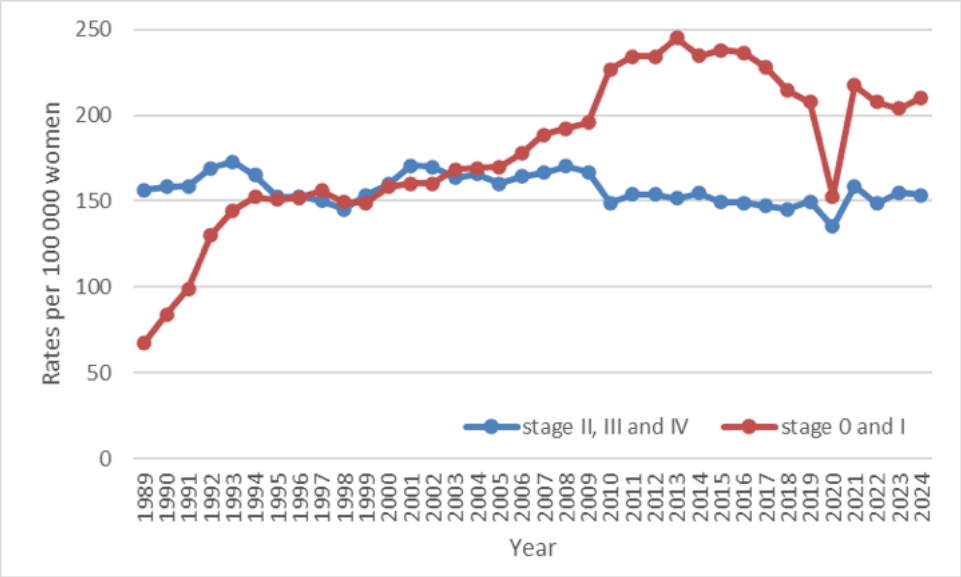
Figure 59 – Breast cancer incidence by year and stage (The Netherlands)



***Source**: Netherlands Cancer Registry (2026, https://nkr-cijfers.iknl.nl/)*
Revised European Standardized Rate (RESR): Number of events per 100 000 persons per year corrected for the age structure of the European population in 2010. Data for 2024 is provisional.

Figure 59 shows the breast cancer incidence by year and stage for the age group 50-69 years. As mentioned by Cardoso et al. (2026), "*a likely stage migration (from stage II to III) might have occurred over time, particularly around 2003, due to reclassification of tumours with more than three positive lymph nodes as stage III instead of stage II.*"[17] The impact of this reclassification is observed in the above figure. According to Autier (2026), a better approach would be to consider stage II, III and IV breast cancers together.[18] Figure 60 provides the combined incidence of stage 0 and I versus II, III and IV breast cancers in women 50-69 years over the same time period.

Figure 60 – Stage 0 and I versus II, III and IV breast cancer incidence by year (age group 50-69 years, The Netherlands)



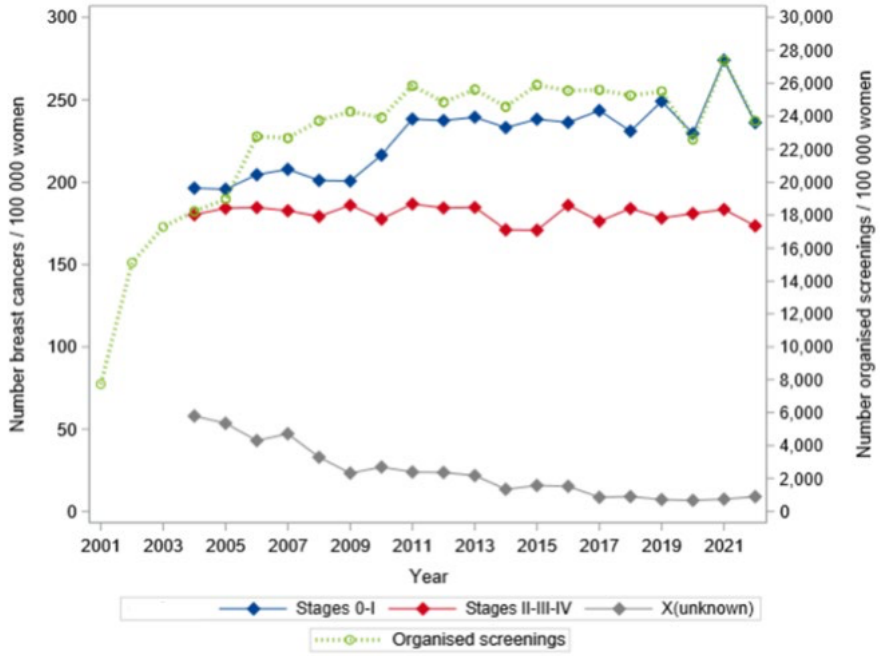
***Source**: Derived from data presented in Figure 59 (Netherlands Cancer Registry, 2026, https://nkr-cijfers.iknl.nl/). Data for 2024 is provisional. Expressed as revised European Standardized Rates (RESR)*

As mentioned by Mali et al. (2026), epidemiological data show a steady increase in stage I cancers, while the number of stage IV cancers has remained stable, which is a sign of overdiagnosis.[299] While there is a considerable increase in the number of participants (Figure 58), especially during the first years screening was introduced, no substantial decrease in stage II, III or IV breast cancers is observed (Figure 60). This raises questions on whether the optimistic scenario including a risk ratio of 0.66 for the occurrence of stage II, III and IV breast cancers, can be defended. As a result, the ICER of this most optimistic scenario can also be questioned. Once a less optimistic scenario with a risk ratio of 0.83 is modelled (still modelled in combination with other optimistic assumptions – see section 6.4.3), the average ICER already increases above €50 000 per QALY gained.

6.4.5.3 The Belgian Cancer Registry (BCR)

For Belgium, we are unable to present the data in the same way as for the Netherlands. Regarding participation, BCR only has data for Brussels and Wallonia from 2005 and 2007 respectively, despite the fact that screening in these regions was launched in June 2002. Therefore, we present the data for Flanders, for which participation figures are available from 2001 onwards, i.e. the year the organized breast cancer screening started. The number of diagnosed cancers is available from 2004 onwards. These are presented in groups (stages 0 and I versus II–IV) in Figure 61. Due to a higher number of cancers for which the stage is not defined, these data are more difficult to interpret. Nevertheless, even based on these data, it is challenging to justify the optimistic risk ratio of 0.66.

Figure 61 – Stage 0 and I versus II, III and IV breast cancer incidence by year (age group 50-69 years, Flanders)



***Source**: BCR data.*

6.4.6 Extension to other age groups, screening frequency, combination with ultrasound or other changes

The original scope of the research questions was broader. Unfortunately, it was not possible to apply the 'what if' approach to the younger (40–49 years) or older (70–74 years) populations, nor to implement other screening strategies in an evidence-based manner. However, the calculations for the current screening programmes can be used as a starting point to deduce the likely impact of a number of changes, which we do in a narrative format.

- **An extension of the current screening programmes to asymptomatic women aged 40–49 with no increased risk of breast cancer**: As shown in Figure 1 in Chapter 2, the incidence of breast cancer among younger women is approximately four times lower than in the 50–69 age group. This means that the screening will be less cost-effective compared to the current target group. The effort required to detect cancer will be much greater, while the harm caused by false positives and overdiagnosis will remain. The greater gain in life years due to higher life expectancy among younger women carries less weight in terms of preventing mortality because of the significant impact of discounting over the longer term. Furthermore, a higher probability of faster growing cancers would also negatively impact the balance between screen-detected and interval cancers. Therefore, it is very likely that extending screening to this target group will lead to less favorable ICERs. The uncertainty surrounding the cost-effectiveness of screening 50–69-year-olds (see above), combined with the deterioration in cost-effectiveness when extending screening to younger women, leads us to conclude that, based on the currently available evidence, serious questions can be raised about studies concluding that this extension would be cost-effective.
- **An extension of the current screening programmes to asymptomatic women aged 70–74 who are not at an increased risk of breast cancer**: Figure 1 in Chapter 2 shows that the incidence of breast cancer among women in the 70–74 age group is slightly higher than among those in the 50–69 age group. However, the benefit of averting a breast cancer death will be much lower in this age group than in the 50–69 age group, as mortality from other causes is much

higher. This has an impact on the number of life years gained, whether or not quality of life is taken into account. Furthermore, this will lead to more overdiagnosis, more precisely to the detection of a tumour at a time when the patient is more likely to die from other diseases or old age. This has a further negative impact on the cost-effectiveness of breast cancer screening.

Based on current evidence and economic argumentation, extending screening to younger and older women cannot be defended as efficient use of limited resources. As mentioned by Mali et al. (2026), "*Before we can begin screening, we must demonstrate that it reduces cancer-specific mortality, thereby reducing overall mortality. Opponents argue that this would require studies on a scale that is too large. This is remarkable when you consider that we would be screening millions of people using an unproven method that costs far more than an RCT.*"[299]

- **Reducing the screening interval from two to one year**: This will result in a doubling of the workload associated with the screening programmes. While certain costs will double (e.g. the number of screening mammograms to be performed), economies of scale will apply to others (e.g. lump-sum costs associated with the screening campaign). Conducting a screening mammogram and a second reading for around 258 000 participating women would result in additional expenditure of over €21 million. Therefore, if the screening interval is halved (i.e. from biennial to annual screening), this incremental cost will be incurred. Based on the results of the model using biennial screening, there were 6595 false-positive results, 1638 true-positive results, and 793 interval cancers (0.31%) based on approximately 258 000 participating women. With an overdiagnosis rate of 20.7%, 501 cancers would fall into this category. This means there are 1137 non-overdiagnosed screen-detected cancers, as well as almost 800 interval cancers. Doubling the screening effort will never prevent all interval cancers and will result in a smaller benefit, negatively impacting the cost-effectiveness of screening.
  Conversely, extending the screening interval would very probably lead to a smaller proportional reduction in benefits compared to the reduction in costs, improving cost-effectiveness. However, there is no information available on the increased number of interval cancers this would cause.
- **The combination of mammography and ultrasonography**: Based on one RCT, no improved breast cancer or all-cause mortality was found, in combination with a higher risk of false-positive results.[167] This makes an economic evaluation redundant.

Several other topics were out of scope of this HTA. There are for example frequent referrals to personalised, risk-based screening. As mentioned by Cardoso et al. [17] the "*development and implementation of personalised, risk-based, screening offers on the one hand, and improvements in our understanding of the biological behaviours of breast cancers on the other hand have, however, the potential to considerably reduce overdiagnosis and overtreatment and revolutionise breast cancer screening and outcomes in the future.*[300, 301]" Without evidence regarding the impact on false-positive findings, overdiagnosis, the reduction in the number of invasive breast cancers at an advanced stage, etc. – both in groups where more intensive screening would be carried out and in those where less intensive screening would be carried out – it is difficult to assess the extent to which this will influence the benefits, harms and cost-effectiveness. Similarly, reference is also made to "*the use of artificial intelligence to support, for example, screen reading might improve detection and also allow application in countries and settings where the need for numerous qualified human resources has made screening implementation difficult so far.*[302] *However, such improvements in detection should also be assessed in relation to potential additional harms, namely overdiagnosis and overtreatment.*"[17]

6.4.7 Other elements

Opportunistic screening was not included in the reference case. However, the impact of altered participation rates in the official population-based screening programme was examined through scenario analyses, exploring participation levels ranging from 10% to 70%. The results indicate that the largest improvement in cost-effectiveness occurs during the initial increase in participation, particularly from 10% to 20% and 30%. Beyond a participation rate of 40%, the marginal impact of further increases diminishes substantially. Consequently, the inclusion of higher participation rates

above the 40.8% included in the reference case does not lead to further substantial improvements in the ICER. Moreover, it cannot be assumed that opportunistic screening would affect costs and health outcomes to the same extent. The cost of the initial opportunistic screening may be higher if complementary imaging techniques are applied, and sensitivity and/or specificity may be lower, for instance due to the absence of systematic second reading. A Danish study reported a markedly lower sensitivity with comparable specificity for opportunistic screening compared with organised screening.[303] This would negatively affect the cost-effectiveness.

Organised breast cancer screening programmes are in place in Flanders, Brussels and Wallonia. No separate analysis was carried out for each of the three regional programmes. The economic evaluation was based on available data for Belgium for all participants in the organised programmes. With participation rates of 54%, 8% and 4% in 2024 for the three regions respectively (see Table 3), the results are most applicable to Flanders. In Flanders, the specificity was 97.6% in 2021 (see Table 3). This is consistent with the specificity of 97.5% derived from Belgian data applied in the model (see Section 6.2.7.3). Calculating the sensitivity as the number of screen-detected cancers divided by the sum of screen- and interval-detected cancers gives a percentage of 67.4%. This was similar to the 67.5% reported for Flanders in 2022[s] and slightly lower than the 70.3% reported for Flanders in 2021 (see Table 3). Therefore, the results are most applicable to Flanders, where scenario analyses suggest that the impact of higher participation on the ICERs is limited (see section 6.3.5). In the other regions, participation is much lower, which negatively impacts the ICERs.

The hypothetical economic evaluation was conducted from the perspective of a healthcare payer. If the evaluation were to be extended to a **societal perspective**, the impact on productivity would also be taken into account. However, we note that this extension should consider not only the potential positive consequences, but also the negative impact on productivity. The positive consequences would apply to a relatively small group: those diagnosed with breast cancer at an earlier stage. The negative consequences of participating in screening are limited in time; i.e. potential absence from work due to participation in screening during the working week. However, these consequences would affect a much larger group of participants. Therefore, it is incorrect to assume that applying a societal perspective by including the impact on productivity would lead to an improvement in cost-effectiveness since both positive and negative consequences need to be taken into account.

Finally, we would like to refer to the Belgian guidelines[3] for economic evaluations, which state that **there is no explicit ICER threshold in Belgium**. "*Referring to such thresholds from other countries or international organizations should be avoided, such thresholds should not be adjusted to the Belgian context or indexed for inflation.*" "*It is not sustainable to systematically use very high ICER threshold values if this creates chronic deficits in the budget and/or if more efficient interventions are replaced as a consequence. The budget constraint affects the optimal cost-effectiveness threshold. In that respect, it is more about what we can pay for within our budgets (ability to pay) rather than what we want to pay (willingness to pay).*[304]" "*To allow policymakers to take cost-effectiveness into account without imposing an explicit ICER threshold, it is recommended that the reference case presents the results on the so-called cost-effectiveness acceptability curve (i.e. expressing the probability of the cost-effectiveness of an intervention as a function of varying ICER thresholds […]. The implicit threshold that policymakers use might be influenced by other factors, such as whether there is a high unmet need or severity of the disease, the strength of the evidence for both safety, effectiveness and cost-effectiveness, equity considerations, the pressure on healthcare budgets, etc. While consideration of the cost-effectiveness of an intervention is necessary, it is clearly not the sole basis for decision-making.*[305]"[3]

6.4.8 Conclusion

If various optimistic assumptions are applied in combination with a risk ratio of 0.66 for the occurrence of stage II, III and IV breast cancer, the current population-based screening programmes for asymptomatic women aged 50-69 with no increased risk of breast cancer may have an ICER of approximately €13 700 per QALY gained. However, if the probability to be diagnosed with an

advanced stage breast cancer is adjusted to 0.83 (i.e. limited to studies with a low risk of selection bias), the ICER increases to around €51 900 per QALY gained. In addition, not only the probability to be diagnosed with an advanced stage breast cancer is a determining factor, as overdiagnosis also impacts the results substantially. However, due to the limitations of the available evidence, which consists mainly of older RCTs, it is unclear which scenario is the most realistic and applicable to the current Belgian situation. These limitations therefore translate into uncertainty regarding the cost-effectiveness of screening, as reflected in this hypothetical economic evaluation. Unfortunately, real-world data seem not to support the scenario with a risk ratio of 0.66 for the occurrence of stage II, III or IV breast cancer, which was mainly based on RCTs with a high risk of selection bias. As a result, less optimistic scenarios might better reflect current screening's cost-effectiveness, which already need a much higher willingness-to-pay to be considered cost-effective.

7 ETHICAL CONSIDERATIONS

This chapter addresses the ethical considerations related to breast cancer screening. Beyond the four classical principles of biomedical ethics, it also examines key ethical issues specific to this type of screening.

KEY MESSAGES

- In breast cancer screening, the balance between benefits and harms is complex and may be impacted by the perspective (collective versus individual) from which they are considered.
- Individuals may find it difficult to comprehend the benefits and harms of screening and to weigh them against each other, which raises important challenges for informed decision-making.
- Effective informed consent requires clear, balanced, and accessible communication, taking into account framing effects and variations in individuals' ability to understand and use information.
- Participation in screening may be influenced by organisational features and social norms, raising concerns about the extent to which decisions are fully autonomous.

7.1 Methods

A pragmatic literature search was conducted to address the main ethical principles and concerns related to (extending) breast cancer screening. To ensure comprehensive and targeted identification of relevant literature, references were identified from three complementary sources: (i) relevant articles identified during the literature searches for clinical evidence, (ii) a targeted web-based search and (iii) a structured database search. Articles were selected based on predefined inclusion and exclusion criteria (Table 155 in Appendix 26).

First, ten articles, identified during the systematic search for literature on clinical effectiveness and safety, were screened at full-text level. Three were retained based on their relevance to the research question.[306-308]

Second, a targeted search was conducted using Google in 2024, based on the keywords "ethical", "considerations", "issues" and "breast cancer screening". Seventeen unique references were identified and screened at full-text level, of which five were retained for inclusion.[309-313]

Third, a structured literature search was conducted in MEDLINE and the Cochrane Database of Systematic Reviews, covering the period 2015–2025. The search strategy combined the text words and controlled vocabulary used in the clinical chapter with ethics-related keywords. This search yielded 224 records, of which 17 duplicates were removed using EndNote.

The selection process was subsequently managed using the Covidence® tool. A total of 207 records were imported into Covidence (191 from MEDLINE and 16 from Cochrane). After removal of an additional 10 duplicates, 197 records were screened based on title and abstract, leading to the exclusion of 129 records. 68 articles were assessed at full-text level, of which 56 were excluded. 12 articles were ultimately retained from this database search.[314-325]

The detailed search strategy and inclusion and exclusion criteria are provided in Appendix 26.

Overall, this pragmatic search thus resulted in the inclusion of 20 references, covering both general ethical frameworks applicable to cancer screening and ethical issues specific to breast cancer screening.

7.2 Normative principles and key ethical issues related to breast cancer screening

7.2.1 The four principles of Beauchamps and Childress

In biomedical ethics, the four principles formulated by Beauchamp and Childress are commonly used to structure normative analysis.[326] These principles provide a framework for examining the expected benefits and harms associated with screening, as well as the fundamental values that should guide public health decisions.

The four principles are defined as follows:

- Beneficence: the moral obligation to act for the benefit of others, promoting their welfare and seeking to achieve a favourable balance of benefits over harms.
- Non-maleficence: the obligation to avoid causing harm and to minimise unnecessary injury or suffering.
- Autonomy: respect for individuals' capacity for self-determination, including their right to make informed and voluntary decisions regarding their own health.
- Justice: fairness in the distribution of benefits, risks, and costs, as well as equitable treatment of individuals within healthcare systems.

When applied to breast cancer screening, these principles highlight a number of ethical dilemmas.

Beneficence is commonly invoked to justify screening programmes, as they aim to detect cancer at an earlier stage and thereby attempt to reduce morbidity and mortality at the population level.[327]

Importantly, the ethical justification of screening programmes also depends on the availability of population-level evidence demonstrating that the expected benefits outweigh the potential harms.

However, these justifications are based on population-level outcomes and do not necessarily translate into benefit for each individual.[312]

Non-maleficence draws attention to the harms associated with screening, including false-positive results leading to unnecessary procedures and potential psychological distress, as well as overdiagnosis resulting in unnecessary diagnosis and treatment.[309] Empirical evidence indicates that false-positive results may generate sustained anxiety.[306, 309] As Elton et al. (2020) argue, screening presents a particular ethical challenge because it may expose healthy individuals to harm even when an overall population benefit is anticipated.[310] This challenges the assumption that a favourable benefit–risk balance at the population level is sufficient to justify screening practices.

Autonomy requires that individuals are able to make informed and voluntary decisions regarding participation in screening. In the context of breast cancer screening, this implies that women should receive clear, balanced, and accessible information about potential benefits, risks, and uncertainties, including overdiagnosis and false results.[307, 309] Participation should therefore be based on informed consent rather than on implicit pressure to meet participation targets. However, the literature suggests

that individuals' understanding of benefits and harms is often limited, raising concerns about the extent to which consent can be considered fully informed (see below 7.2.2).

Justice, in the context of breast cancer screening, primarily concerns equitable access to screening programmes. However, the literature indicates that access and participation are not evenly distributed across the population. In particular, women with disabilities are consistently less likely to participate in screening programmes due to a range of barriers, including physical accessibility, communication difficulties, and inadequate healthcare support specifically recognised. In addition, structural and social barriers may further influence access to and participation in screening programmes, contributing to persistent inequalities in uptake.[308, 312, 316, 324]

These principles illustrate that breast cancer screening raises ethical questions that cannot be fully resolved by appealing to a single principle. Rather, they highlight inherent tensions between collective benefits and individual risks, between population screening approaches and respect for individual autonomy, and between encouraging uptake and preserving voluntary choice based on complete information. These tensions are further explored in the following section, which examines ethical issues specific to breast cancer screening.

7.2.2 Key ethical issues in breast cancer screening

BALANCING BENEFITS AND HARMS

Across the literature, the ethical justification of breast cancer screening is primarily grounded in the assumption that the benefits of mammography outweigh its potential harms at the population level.[307, 309] Mammography screening aims to detect cancers at an earlier stage, in an attempt to reduce breast cancer mortality and to improve clinical outcomes.[327]

In addition, the balance between benefits and harms is complex in the context of breast cancer screening. Mammography is associated with a number of well-documented harms, including false-positive results and overdiagnosis, both of which lead to unnecessary procedures and treatments.[306, 312] Overdiagnosis is of particular concern, as it involves the detection of cancers that would not have become clinically relevant during a woman's lifetime, thereby exposing individuals to treatments without any benefit.[309, 325]

Despite these risks, the justification of screening relies on population-level evidence showing a significant reduction in breast cancer-specific mortality. However, a significant effect on all-cause mortality is lacking, which raises questions on the effectiveness of breast cancer screening.[kkkkk] In addition, this population-based benefit does not translate into benefit for each individual participant. As highlighted in the literature, some women will experience harms without deriving any personal benefit from screening.[310, 312]

Moreover, individuals differ in how they perceive and weigh these benefits and harms. Grimault (2022) emphasises that each woman may attribute different importance to avoiding breast cancer death, the risk of overdiagnosis, or the consequences of false-positive results, depending on her personal values, experiences, and level of risk perception.[307] Carter et al. (2016) similarly note that there is no consensus on which outcomes should be prioritised and how benefits and harms should be weighed, as policymakers, clinicians, and individuals may have different perspectives on what matters most,[312] which are often imperfectly aligned with available evidence (see below).[320, 321]

Furthermore, the experience of screening-related events may shape future attitudes towards screening. As discussed by Health Quality Ontario, false-positive results can generate significant anxiety, which may persist even after cancer has been ruled out, yet many women remain willing to continue participating in screening programmes.[306] In addition, evidence suggests that such

---

[kkkkk] For more information on the clinical effectiveness of breast cancer screening, we refer the reader to Chapter 3.

experiences may also have longer-term psychosocial consequences, extending beyond the immediate screening context.[314] This suggests that emotional responses and prior experiences may play a significant role in shaping individual preferences, beyond purely rational assessments of risks and benefits.

While public health approaches focus on population-level outcomes, individual decisions are shaped by personal values, risk perception, and tolerance for uncertainty. These differing perspectives may lead to divergent interpretations of the benefits and harms of screening, reflecting ethical tensions between population-based and individual-centred approaches.[315] These findings imply that the ethical evaluation of breast cancer screening cannot rely solely on population-level evidence. It must also take into account how benefits and harms are perceived, understood, and experienced by individuals, and how these perceptions influence decision-making.

INFORMED CONSENT IN MAMMOGRAPHY SCREENING: THE ROLE OF COMMUNICATION AND HEALTH LITERACY

Elton et al. (2020) contrast screening with therapeutic care, noting that in therapeutic contexts patients typically seek medical attention for a specific condition they suffer from, whereas screening programmes invite asymptomatic individuals to undergo testing.[310] This difference is not only structural but also affects the conditions under which consent is obtained. In traditional clinical settings, valid consent is generally understood to require adequate information, comprehension, voluntariness, and a patient-initiated request for care. In contrast, breast cancer screening, like every cancer screening programme, involves inviting individuals who have not expressed a prior health concern to participate in a medical intervention. This contrast raises questions about whether consent in this context can meet the same ethical standards, particularly in terms of voluntariness and understanding, as in a therapeutic setting.

Respect for autonomy in breast cancer screening requires that women receive clear, balanced, and comprehensible information about both the benefits and the harms of mammography screening, including false-positive results, overdiagnosis, and the uncertainty of individual benefit.[307, 309, 311] However, the literature consistently shows that this requirement is difficult to fulfil in practice.

Empirical studies suggest that many women have an incomplete or inaccurate understanding of the outcomes of mammography screening. In a systematic review, Seaman et al. (2018) concluded that a substantial proportion of women overestimate the effectiveness of screening in reducing mortality and underestimate its potential harms, including false-positive results and overdiagnosis. Some participants even believe that mammography prevents breast cancer altogether, reflecting a fundamental misunderstanding of the purpose of screening.[320] These findings raise concerns about the extent to which participation decisions can be considered genuinely informed.

Beyond the content of information, the way it is presented plays a crucial role in shaping perceptions. Carter et al. (2016) highlight that screening information often relies on relative risk reductions rather than absolute risks, a practice that may “inflate perceived benefits and discount harms.”[312] Similarly, Elton et al. (2020) demonstrate that individuals may make different decisions depending on whether outcomes are framed in terms of survival or mortality, despite the equivalence of these measures. This so-called ‘framing effect’ is particularly relevant in mammography screening, where the magnitude of benefit is modest and highly dependent on how it is communicated.[310] This suggests that ethical considerations extend beyond the accuracy of information to include how it is presented and interpreted.

In this context, the concept of health literacy is particularly important. Breast cancer screening requires individuals to engage with complex probabilistic information, such as the likelihood of benefit, the risk of false-positive results, and the concept of overdiagnosis, which are inherently difficult to understand. Health literacy encompasses not only the ability to access information, but also to interpret and use it appropriately. As noted by several authors, individuals with limited health literacy or numeracy may be particularly disadvantaged in this regard, raising concerns about the equity of informed decision-

making.[312, 318, 322] Beyond these cognitive dimensions, broader socio-cultural and belief-related factors may also shape individuals' understanding of screening information.[316]

These challenges are further illustrated in specific aspects of mammography screening, such as breast density notification and supplemental screening. Schifferdecker et al. (2020) show that women's knowledge of breast density and its implications for cancer risk and screening is often limited and heterogeneous. Even when information is provided, it does not necessarily translate into meaningful understanding or informed decision-making.[319] This highlights the gap between information provision and actual comprehension.

More broadly, several authors emphasise that communication in breast cancer screening may not always be neutral. Moutel et al. (2019) stress the importance of distinguishing between information intended to support informed choice and communication strategies aimed at increasing participation. In practice, invitations to screening and public health messages may implicitly convey normative expectations, suggesting that participation is desirable or even expected.[309] Grimault (2022) similarly notes that some approaches resemble social marketing strategies, potentially encouraging participation rather than supporting autonomous decision-making.[307]

Taken together, these findings indicate that informed consent in mammography screening cannot be reduced to the provision of standardised information materials. It depends on whether women are genuinely able to understand complex information, recognise uncertainties, and make decisions aligned with their own values and preferences. This requires not only improving the content of information, but also addressing issues of communication, framing, and health literacy that shape how information is interpreted and used.

CHALLENGES TO AUTONOMY IN BREAST CANCER SCREENING

Breast cancer screening programmes occupy an intermediate position between individual clinical care and public health intervention. As Carter et al. (2016) explain, unlike vaccination programmes, the justification for screening programmes does not rest on collective protection mechanisms but rather on the aggregation of outcomes at the individual level.[312]

This situation creates a persistent tension between respect for individual autonomy and the pursuit of public health objectives. Women are typically invited through systematic mailing programmes, reminders, and targeted communication campaigns. While these strategies aim to improve public health outcomes, they may also influence individual decision-making in ways that are not fully neutral.[309, 312]

Several authors have highlighted that communication strategies used in breast cancer screening may resemble forms of soft paternalism or social marketing. Grimault (2022) notes that invitations to mammography screening can implicitly convey the message that participation is the expected or "correct" behaviour, thereby framing screening as a normative choice rather than a personal decision.[307] In this context, the boundary between encouraging participation and respecting autonomy becomes blurred. Empirical evidence further suggests that participation decisions are not determined by information alone, but are also influenced by cognitive and social factors. For example, Ritchie et al. (2022) found that social norms and perceived behavioural control were significant predictors of intention to participate in screening, whereas the effect of health literacy was limited. [318]

Qualitative research suggests that these implicit mechanisms operate within a broader social context that shapes how women perceive their role in screening programmes. Seaman et al. (2021) describe how participation in mammography screening may be associated with being a "responsible" woman, whereas non-participation may be perceived negatively.[321] In line with this, Moutel et al. (2019) emphasise that non-participation should be recognised as a legitimate expression of individual autonomy, although the need some women feel to justify their decision suggests that these social norms are not neutral and may shape choices in practice.[309] This indicates that autonomy may be influenced by social and institutional factors that extend beyond the provision of information.

In addition, the organisational features of mammography screening programmes may contribute to shaping behaviour. Regular invitations, age-based eligibility criteria, and standardised pathways may encourage repeated participation over time, potentially leading to a more routinised form of attendance rather than a fully deliberative choice. This raises questions about the extent to which repeated participation reflects ongoing informed consent or, at least in part, adherence to established routines.[307, 309, 312]

ETHICAL CHALLENGES RELATED TO EXTENDING BREAST CANCER SCREENING

The potential extension of breast cancer screening to additional age groups introduces further ethical complexity. While the present analysis did not identify a substantial body of literature specifically addressing the ethical implications of extending screening to women under 50 or over 70 years of age, some studies provide indirect insights into these issues. In particular, research focusing on vulnerable populations, such as individuals with disabilities or reduced capacity for decision-making, highlights additional ethical challenges related to informed consent, vulnerability, and the balance between benefits and harms.[308] These issues are especially salient in older populations, where evidence suggests that decisions regarding screening should take into account not only age, but also factors such as comorbidities, functional status, life expectancy, and the potential burden of interventions.[323] Moreover, ethical dilemmas arise in situations where decision-making capacity may be impaired, such as in patients with cognitive decline, requiring careful consideration of autonomy, best interests, and proportionality of care.[323] These findings suggest that extending eligibility criteria may amplify existing ethical tensions rather than simply reproduce them.

7.3 Conclusion

Breast cancer screening raises complex ethical challenges that require balancing population-level objectives with individual-level considerations. Ensuring that participation is based on clear, balanced, and understandable information, and that individuals are able to make decisions aligned with their own values, remains essential. In this context, ethical decision-making in breast cancer screening requires careful attention to both effectiveness and respect for autonomy.

8 DISCUSSION

Breast cancer screening programmes have been widely introduced across European countries over the past decades.[17] In Belgium the population-based mammography screening programme was implemented about 25 years ago, first in the Flemish region, and soon Brussels and the Walloon region followed.

The rationale for cancer screening is mainly based on the "**early detection paradigm**": when cancer is detected at an early stage through screening, a large proportion of cases can be prevented from being diagnosed at a later stage, when treatment is more cumbersome and prognosis is overall less favourable, and ultimately breast cancer mortality can be reduced. This concept is based on the linear growth model, for which the German pathologist Virchow (1821–1902) laid the basis for breast cancer.[299] While the linear growth model assumes that all cancers follow the same path of progression, starting from the primary tumour to locoregional and then to distant metastases, ultimately resulting in death, the prevailing model is more complex and heterogeneous.[299, 328, 329] Insights gained over the last decades, in part based on knowledge gained from cancer screening, emphasize that cancers can grow at very different rates. Some cancers metastasize from the outset, some are fatal through locally invasive growth while they do not metastasize, and other cancers may never cause symptoms because their growth is so slow that those affected die from other causes before symptoms appear, or because they stop growing or even regress.[14, 299] Cancer screening programmes are more likely to detect slower-growing, less aggressive cancers, as for these cancers the preclinical period, also known as the sojourn time, is longer. A phenomenon also known as length time bias.

Population screening programmes must adhere to the principles of beneficence and non-maleficence. They should maximise benefits and minimise harms, costs, and waste,[330] even more so since they target asymptomatic people.

Does invitation to breast cancer screening reduce **breast cancer mortality**? Most likely, it does. Our meta-analysis points to a reduction of 19% in breast cancer related mortality among women aged 40 years and older invited to mammography screening. This result is in line with previous SRs,[13, 14, 86, 87] which are based on the same RCTs. This reduction is somewhat less pronounced in women younger than fifty, while in women aged seventy or older invitation to breast cancer screening does not affect breast cancer mortality.

The results for breast cancer mortality should not be viewed in isolation from the results for **all-cause mortality**.[40, 43, 44] While power calculation for the RCTs was indeed based on the former outcome, the latter is less prone to bias and does take deaths resulting from the harms related to the screening programme into account. For all age groups combined, invitation to screening most likely does not reduce all-cause mortality. It does not affect all-cause mortality in the youngest or oldest age groups either.

For screening programmes to be considered effective, an essential premise is that the **incidence of advanced-stage cancers** drops following their implementation, and this decline should precede (or accelerate) reductions in cancer-specific mortality rates.[157, 331, 332] A major benefit of this metric is that it is not affected by improvements in cancer treatments, while this is indeed the case for cancer mortality.[157] Overall, invitation to screening with mammography results in a statistically significant reduction of the incidence of advanced-stage breast cancers. This impact was less pronounced in the youngest age group (i.e. 40-49 years), while for the older age group (i.e. 70 years and older) no relevant RCT data were retrieved. These results are in line with a recent multi-country analysis (including Belgian data), which indicated that following the implementation of mammography screening programmes, the incidence of advanced-stage cancer, especially of stage IV and among the age groups above fifty years, decreased.[17] However, large intercountry variations were observed. In addition, the same research highlighted a significant rise in early-stage (in situ and stage I) breast cancer cases, mostly among women aged 50–69 years, after mammography screening programmes were introduced.[17]

This observation brings us to the most important harm related to (breast) cancer screening: **overdiagnosis**, i.e. diagnosing a person with a screen-detected problem that would never have manifested clinically. Patients overdiagnosed with breast cancer suffer from unnecessary anxiety and costs, alongside the risk of developing complications or experiencing fatal side effects from treatments they never needed.[63] Although the definition of overdiagnosis is clear, variability in reported estimates is compounded by the choice of the denominator.[129] Notably, de Gelder and colleagues described seven distinct methodologies.[136] Consequently, the magnitude of overdiagnosis as a result of breast cancer screening remains uncertain.[14]

Our meta-analysis indicated that seven to nine years after randomization, across all age groups, 27% of all cancers (invasive and non-invasive) detected in women invited to screening were overdiagnosed. This proportion was lower when the analyses were restricted to women younger than fifty years old (23%) or to women between fifty and sixty-nine years (21%), while for the older age group again no relevant RCT data were retrieved. Ten to fifteen years after the end of the screening period, the respective proportion dropped to 9%. However, this result should be interpreted with caution because it is based on only two trials. Additionally, it is difficult to avoid, or even control for, screening contamination - in the intervention as well as in the control groups - over such a long follow-up period.

The clinical results presented in this report are based on the **meta-analyses of seven RCTs**. Following the publication of these trials, their methodological rigor has been subject to extensive debate.[14, 95, 333-335] Only two trials were considered having a low risk of selection bias. For some outcomes like all-cause mortality and overdiagnosis, the result restricted to the trials with a low risk of selection bias were comparable to the overall result, yet for other outcomes like mortality ascribed to

breast cancer, the point estimate was comparable, but the 95% CI was situated much closer to the null value (of no significant effect).

These trials were performed **between 1963 and 1997**. Given the (very) old age of the trials, the level of evidence for all outcomes was downgraded due to indirectness. In essence, we considered the certainty of the evidence for all outcomes - benefits as well as harms - as low or very low.

Since these trials were run, **substantial advances in mammography screening technology** have been made.[336-340] Most breast cancer screening programmes in high-income countries have transitioned from screen-film mammography to full-field digital mammography, as the latter provides (among others) better image quality and dose optimisation, easier and more efficient archiving and image transfer capabilities.[341, 342] However, not all advances come (solely) with a benefit. It has been argued that the shift from screen-film to digital mammography may have increased early detection and recall rates as well as false-positive rates and overdiagnosis.[14, 343, 344] The To-Be trial in Norway illustrated that digital breast tomosynthesis including synthetic 2D mammograms as a screening tool did not offer an added value compared to standard digital mammography.[345]

In addition, since the mammography screening trials were performed, **major advances** were also observed in **breast cancer care**. For instance, the introduction of systemic therapy has reduced breast cancer mortality substantially at all stages,[346] with the largest decreases observed in young women, women younger than the standard screening age.[347] According to a US modeling study, treatment (63%) played a larger role than screening (37%) in reducing breast cancer mortality as of 2012, compared with a situation with no screening and no treatment.[346] Yet, the associations varied by breast cancer molecular subtype. With improved treatments, the current impact of screening on mortality is a subject of ongoing debate.[14, 348]

Considering the above concerns, the **relevance** of these old trials to current screening programmes has been questioned.[14, 129] **Are contemporary cohort and case-control studies a valuable alternative** for evaluating the effectiveness of (breast) cancer screening programmes? While they may indeed better reflect current incidence rates and care standards, self-selection poses a serious risk of inflating the perceived benefits of screening.[157, 349] Self-selection bias arises from inherent prognostic differences between women who attend mammography screening and those who do not. Non-attenders typically have lower education, reduced health awareness, higher deprivation, and more comorbidities, all of which correlate with a higher risk of aggressive cancers and increased mortality (both breast cancer-specific and all-cause). These elevated risks exist regardless of the effect of screening.[157, 350-352] This was clearly demonstrated in a SR of cohort studies that evaluated the risk of breast cancer mortality and of all-cause (or of non-breast cancer) mortality associated with breast cancer screening attendance.[49] Interestingly, women who do not attend mammography screening, did not only have a twofold higher risk of breast cancer death compared to their peers who do attend, they also have a similarly increased risk to die from other causes than breast cancer, while attending a screening programme does not impact any other causes of death than breast cancer.[49] Hence, observational studies have been considered not suitable for assessing screening effectiveness,[49, 353] as they are subject to healthy user bias and unmeasured confounding factors, leading to overestimations of screening benefits.[49] Therefore, in this report, the results obtained in observational studies were deliberately not included in the meta-analyses, nor in the conclusions.

What do we learn from the **cost-effectiveness analysis**? The significant uncertainty regarding key outcomes like the probability of being diagnosed with advanced stage breast cancer (i.e. stage II-IV) and overdiagnosis translates directly into uncertainty regarding cost-effectiveness. In the Belgian hypothetical economic evaluation the current situation in which women aged 50-69 years are invited for breast cancer screening was compared with the counterfactual (i.e. 'what if there was no screening'). The model assessed the potential impact of breast cancer screening on costs and health outcomes. Under a set of optimistic assumptions, combined with an overdiagnosis rate of 21% and a risk ratio for being diagnosed with advanced disease of 0.66 (based on the meta-analysis), we obtained an **ICER of about €13 700 per QALY gained**. However, when for the probability of being

diagnosed with advanced disease the result was restricted to trials with a low risk of selection bias (i.e. a risk ratio of 0.83), the ICER increases to approximately **€51 900 per QALY gained**.

Which value should we select? The real-world data seem not to support the most optimistic scenario applying a risk ratio for being diagnosed with advanced disease of 0.66. As a result, the plausibility of this scenario, and thus the cost-effectiveness of current screening, can be questioned. Furthermore, deviations from several optimistic assumptions (such as the use of relatively old survival curves with longer follow-up or the absence of a QoL loss associated with overdiagnosis) further increase the value of the ICER.

Similarly, extending screening to younger or older age groups would also negatively affect cost-effectiveness: in the former group due to lower incidence rates and in the latter due to higher levels of overdiagnosis.

Last, **no relevant RCT data** were retrieved to answer the research question on the **frequency** of mammography screening. Due to a proportionally smaller impact on benefits relative to the additional costs, annual screening will likely have an unfavourable effect on the cost-effectiveness of breast cancer screening.

Also for the third research question the currently available RCT evidence is limited to one large trial in Japanese women younger than fifty years.[167] Long-term FU data reveal that population screening with the **combination of mammography and ultrasonography** does not result in better breast cancer or all-cause mortality compared to mammography alone, yet it leads to a significantly higher rate of false-positive results. Currently, there is insufficient rigorous evidence for routinely adding ultrasound to mammography for average-risk breast cancer screening. Therefore, the cost-effectiveness of this approach was not evaluated in this report.

In conclusion, the evidence underpinning breast cancer screening is based on old trials which have many (methodological) shortcomings. Observational studies do not provide a valuable alternative as they are prone to healthy user bias and unmeasured confounding factors, leading to overestimations of screening benefits.

Translating this evidence into an economic model carries over the same inherent flaws. Given a balanced evaluation, the cost-effectiveness of these screening programs is still up for debate. Methodological choices, such as including all RCTs versus limiting the analysis to studies with a low-risk of selection bias, heavily influence the final conclusions. We therefore opted to present the results as transparently as possible, in order to inform potential participants, healthcare professionals, and policymakers in a clear and balanced manner.

REFERENCES

1. (ECIS) ECIS. Breast cancer burden in EU-27. 2022. 2023. Available from: https://ecis.jrc.ec.europa.eu/sites/default/files/2024-01/jrc_Breast_cancer_2022_Oct_2023.pdf
2. European Commission Initiative on Breast Cancer (ECIBC). European guidelines on breast cancer screening and diagnosis [Web page].2025 [cited 06/06/2025]. Available from: https://cancer-screening-and-care.jrc.ec.europa.eu/en/ecibc/european-breast-cancer-guidelines?topic=63&usertype=60&updatef2=0
3. Neyt M, Thiry N, Cleemput I. Belgian guidelines for economic evaluations and budget impact analyses: third edition. Method. 2025;KCE Reports 400C.
4. Pesapane F, Cassano E. Screening for Breast Cancer. JAMA. 2026.
5. Roux A, Cholerton R, Sicsic J, Moumjid N, French DP, Giorgi Rossi P, et al. Study protocol comparing the ethical, psychological and socio-economic impact of personalised breast cancer screening to that of standard screening in the "My Personal Breast Screening" (MyPeBS) randomised clinical trial. BMC Cancer. 2022;22(1):507.
6. Walker MJ, Neely A, Antoniou AC, Broeders MJM, Brooks JD, Carver T, et al. Barriers and Facilitators to Delivering Multifactorial Risk Assessment and Communication for Personalized Breast Cancer Screening: A Qualitative Study Exploring Implementation in Canada. Current Oncology. 2025;32(3):155.
7. Abu Abeelh E, Abuabeileh Z. Screening Mammography and Artificial Intelligence: A Comprehensive Systematic Review. Cureus. 2025;17(2):e79353.
8. Gommers J, Hernström V, Josefsson V, Sartor H, Schmidt D, Hjelmgren A, et al. Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study: a randomised, controlled, non-inferiority, single-blinded, population-based, screening-accuracy trial. The Lancet. 2026;407(10527):505-14.
9. Paulus D, Mambourg F, Bonneux L. Breast cancer screening. Good Clinical Practice (GCP). Brussels: Belgian Health Care Knowledge Centre (KCE); 2005 02/05/2005. KCE Reports 11 Available from: https://kce.fgov.be/publication/report/breast-cancer-screening
10. Mambourg F, Robays J, Camberlin C, Vlayen J, Gailly J. Breast cancer screening with mammography for women in the agegroup of 40-49 years. Good Clinical Practice (GCP). Brussels: Belgian Health Care Knowledge Centre (KCE); 2010 07/07/2010. KCE Reports 129 Available from: https://kce.fgov.be/publication/report/breast-cancer-screening-with-mammography-for-women-in-the-agegroup-of-40-49-years
11. Mambourg F, Robays J, Gerkens S. Breast cancer screening amongst women aged 70-74 years of age. Good Clinical Practice (GCP). Brussels: Belgian Health Care Knowledge Centre (KCE); 2012. KCE Reports 176 (D/2012/10.273/20) Available from: https://kce.fgov.be/publication/report/breast-cancer-screening-amongst-women-aged-70-74-years-of-age
12. Kohn L, Mambourg F, Robays J, Albertijn M, Janssens S, Hoefnagels K, et al. Informed choice on breast cancer screening: messages to support informed decision. Good Clinical Practice (GCP). Brussels: Belgian Health Care Knowledge Centre (KCE); 2014 08/01/20214. KCE Reports 216 (D/2014/10.273/04) Available from: https://doi.org/10.57598/R216C
13. Bennett A, Shaver N, Vyas N, Almoli F, Pap R, Douglas A, et al. Screening for breast cancer: A systematic review update to inform the Canadian Task Force on Preventive Health Care guideline. 2024. Available from: https://www.medrxiv.org/content/10.1101/2024.05.29.24308154v1.full.pdf
14. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. [Preprint]. 2024]
15. Ma T, Semsarian CR, Barratt A, Parker L, Pathmanathan N, Nickel B, et al. Should low-risk DCIS lose the cancer label? An evidence review. Breast Cancer Res Treat. 2023;199(3):415-33.
16. Belgian Cancer Registry (BCR). Cancer Fact Sheets, Breast cancer (ICD-10 C50). Incidence year 2023. Brussels: 2025. Available from: https://kankerregister.org/nl/publicaties/cancer-fact-sheet-borstkanker-2023
17. Cardoso R, Ola I, Jansen L, Hackl M, Ihle P, Francart J, et al. Breast cancer incidence, by stage at diagnosis, and mortality in 21 European countries in the era of mammography screening: an international population-based study. Lancet Reg Health Eur. 2026;62:101574.
18. Autier P. Questions regarding the reduction of the incidence of stage 4 breast cancer in Europe. Lancet Reg Health Eur. 2026;63:101645.

19. Belgian Cancer Registry (BCR). Cancer Fact Sheets, Breast cancer (ICD-10 C50). Incidence year 2023. Appendix EN. Brussels: 2025. Available from: https://kankerregister.org/nl/publicaties/cancer-fact-sheet-borstkanker-2023

20. Brierley JD, Gospodarowicz MK, Wittekind C. TNM Classification of Malignant Tumours, 8th edition. 2017.

21. Sciensano. Sterfte en Doodsoorzaken: Doodsoorzaken, Health Status Report [Web page].Brussel, België;2025. Available from: https://www.gezondbelgie.be/nl/gezondheidstoestand/sterfte-en-doodsoorzaken/algemene-sterfte-naar-doodsoorzaak

22. Perme MP, Stare J, Estève J. On estimation in relative survival. Biometrics. 2012;68(1):113-20.

23. Carneiro I. Introduction to Epidemiology. Second ed. Plowman R, Thorgood N, editor. London (England): Open University Press, McGraw-Hill Education; 2017.

24. Herman CR, Gill HK, Eng J, Fajardo LL. Screening for preclinical disease: test and disease characteristics. AJR Am J Roentgenol. 2002;179(4):825-31.

25. Barth RJ, Jr., Gibson GR, Carney PA, Mott LA, Becher RD, Poplack SP. Detection of breast cancer on screening mammography allows patients to be treated with less-toxic therapy. AJR Am J Roentgenol. 2005;184(1):324-9.

26. Ahn S, Wooster M, Valente C, Moshier E, Meng R, Pisapati K, et al. Impact of Screening Mammography on Treatment in Women Diagnosed with Breast Cancer. Ann Surg Oncol. 2018;25(10):2979-86.

27. Elder K, Nickson C, Pattanasri M, Cooke S, Machalek D, Rose A, et al. Treatment Intensity Differences After Early-Stage Breast Cancer (ESBC) Diagnosis Depending on Participation in a Screening Program. Ann Surg Oncol. 2018;25(9):2563-72.

28. Wilkinson AN, Seely JM, Rushton M, Williams P, Cordeiro E, Allard-Coutu A, et al. Capturing the True Cost of Breast Cancer Treatment: Molecular Subtype and Stage-Specific per-Case Activity-Based Costing. Curr Oncol. 2023;30(9):7860-73.

29. Freedman GM, Anderson PR, Goldstein LJ, Hanlon AL, Cianfrocca ME, Millenson MM, et al. Routine mammography is associated with earlier stage disease and greater eligibility for breast conservation in breast carcinoma patients age 40 years and older. Cancer. 2003;98(5):918-25.

30. Coldman AJ, Phillips N, Speers C. A retrospective study of the effect of participation in screening mammography on the use of chemotherapy and breast conserving surgery. Int J Cancer. 2007;120(10):2185-90.

31. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. N Engl J Med. 2018;379(2):111-21.

32. Whelan TJ, Smith S, Parpia S, Fyles AW, Bane A, Liu F-F, et al. Omitting Radiotherapy after Breast-Conserving Surgery in Luminal A Breast Cancer. N Engl J Med. 2023;389(7):612-9.

33. Leung JWT. Screening mammography reduces morbidity of breast cancer treatment. AJR Am J Roentgenol. 2005;184(5):1508-9.

34. World Health Organisation (WHO). Screening programmes: a short guide. Increase effectiveness, maximize benefits and minimize harm. Copenhagen: 2020. Available from: https://iris.who.int/server/api/core/bitstreams/b3922afd-3ff9-4bf3-9254-fa6b3f2b6023/content

35. Wilson JM, Jungner YG. Principles and Practice of Screening for Disease. WHO papers No 34 1968.

36. Dobrow MJ, Hagens V, Chafe R, Sullivan T, Rabeneck L. Consolidated principles for screening based on a systematic review and consensus process. CMAJ. 2018;190(14):E422-E9.

37. Morrison AS. Intermediate determinants of mortality in the evaluation of screening. Int J Epidemiol. 1991;20(3):642-50.

38. Autier P, Boniol M, Koechlin A, Pizot C, Boniol M. Effectiveness of and overdiagnosis from mammography screening in the Netherlands: population based study. BMJ. 2017;359:j5224.

39. Steele RJ, Brewster DH. Should we use total mortality rather than cancer specific mortality to judge cancer screening programmes? No. BMJ. 2011;343:d6397.

40. Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. J Natl Cancer Inst. 2002;94(3):167-73.

41. Fang F, Fall K, Mittleman MA, Sparén P, Ye W, Adami H-O, et al. Suicide and cardiovascular death after a cancer diagnosis. N Engl J Med. 2012;366(14):1310-8.

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

43. Dobbin KK, Ebell M. Should we expect all-cause mortality reductions in large screening studies? Br J Gen Pract. 2018;68(671):290-1.

44. Penston J. Should we use total mortality rather than cancer specific mortality to judge cancer screening programmes? Yes. BMJ. 2011;343:d6395.

45. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. Ann Fam Med. 2013;11(2):106-15.

46. DeFrank JT, Barclay C, Sheridan S, Brewer NT, Gilliam M, Moon AM, et al. The psychological harms of screening: the evidence we have versus the evidence we need. J Gen Intern Med. 2015;30(2):242-8.

47. To T, Wall C, Baines CJ, Miller AB. Is carcinoma in situ a precursor lesion of invasive breast cancer? Int J Cancer. 2014;135(7):1646-52.

48. Shrank WH, Patrick AR, Brookhart MA. Healthy user and related biases in observational studies of preventive interventions: a primer for physicians. J Gen Intern Med. 2011;26(5):546-50.

49. Autier P, Jørgensen KJ, Smans M, Støvring H. Effect of screening mammography on the risk of breast cancer deaths and of all-cause deaths: a systematic review with meta-analysis of cohort studies. J Clin Epidemiol. 2024;172:111426.

50. Cox B, Sneyd MJ. Bias in breast cancer research in the screening era. Breast. 2013;22(6):1041-5.

51. Welch HG. Should I be tested for cancer? : University of California Press; 2004.

52. Lee CI, Elmore JG. Beyond survival: a closer look at lead-time bias and disease-free intervals in mammography screening. J Natl Cancer Inst. 2024;116(3):343-4.

53. MeSH Descriptor Data. Overdiagnosis. 2022. Available from: https://meshb-prev.nlm.nih.gov/record/ui?ui=D000088522

54. Woloshin S, Kramer B. Overdiagnosis: it's official. BMJ. 2021;375:n2854.

55. Akin-Odanye EO, Husman AJ. Impact of stigma and stigma-focused interventions on screening and treatment outcomes in cancer patients. Ecancermedicalscience. 2021;15:1308.

56. Autier P, Boniol M. Caution needed for country-specific cancer survival. Lancet. 2011;377(9760):99-101.

57. Autier P, Boniol M, Héry C, Masuyer E, Ferlay J. Cancer survival statistics should be viewed with caution. Lancet Oncol. 2007;8(12):1050-2.

58. Poelhekken K, Greuter MJW, Dorrius MD, de Bock GH. Influence of follow-up, screening age, interval, and compliance on overdiagnosis of ductal carcinoma in situ (DCIS): A modelling study. PLoS One. 2026;21(1):e0331821.

59. Beaulac SM, McNair LA, Scott TE, LaMorte WW, Kavanah MT. Lymphedema and quality of life in survivors of early-stage breast cancer. Arch Surg. 2002;137(11):1253-7.

60. Brownlee Z, Garg R, Listo M, Zavitsanos P, Wazer DE, Huber KE. Late complications of radiation therapy for breast cancer: evolution in techniques and risk over time. Gland Surg. 2018;7(4):371-8.

61. Tommasi C, Balsano R, Corianò M, Pellegrino B, Saba G, Bardanzellu F, et al. Long-Term Effects of Breast Cancer Therapy and Care: Calm after the Storm? J Clin Med. 2022;11(23).

62. Piessens V, Van den Bruel A, Piessens A, Van Hecke A, Brodersen JB, Lauwerier E, et al. Do health professionals know about overdiagnosis in screening, and how are they dealing with it? A mixed-methods systematic scoping review. PLoS One. 2025;20(2):e0315247.

63. Jatoi I, Pinsky PF. Breast Cancer Screening Trials: Endpoints and Overdiagnosis. J Natl Cancer Inst. 2021;113(9):1131-5.

64. Ripping TM, Ten Haaf K, Verbeek ALM, van Ravesteyn NT, Broeders MJM. Quantifying Overdiagnosis in Cancer Screening: A Systematic Review to Evaluate the Methodology. J Natl Cancer Inst. 2017;109(10).

65. Brodersen J, Schwartz LM, Heneghan C, O'Sullivan JW, Aronson JK, Woloshin S. Overdiagnosis: what it is and what it isn't. BMJ Evid Based Med. 2018;23(1):1-3.

66. Baker SG, Kramer BS, Prorok PC. Statistical issues in randomized trials of cancer screening. BMC Med Res Methodol. 2002;2:11.

67. National Cancer Institute at the National Institutes of Health. Mammograms [Web page].2023 [cited 10 July 2024]. Available from: https://www.cancer.gov/types/breast/mammograms-fact-sheet

68. Chen I, Chalfant J. 2D vs. 3D Tomosynthesis [Web page].2024 [cited 10 July 2024]. Available from: https://www.uclahealth.org/departments/radiology/education/breast-imaging-teaching-resources/screening-mammogram/2d-vs-3d-tomosynthesis

69. Chikarmane SA, Offit LR, Giess CS. Synthetic Mammography: Benefits, Drawbacks, and Pitfalls. Radiographics. 2023;43(10):e230018.

70. Marcon M, Fuchsjäger MH, Clauser P, Mann RM. ESR Essentials: screening for breast cancer - general recommendations by EUSOBI. Eur Radiol. 2024.

71. Cohen SL, Margolies LR, Schwager SJ, Zuckerman S, Patel N, Szabo J, et al. Early discussion of breast density and supplemental breast cancer screening: is it possible? Breast J. 2014;20(3):229-34.

72. Boca Bene I, Ciurea AI, Ciortea CA, Dudea SM. Pros and Cons for Automated Breast Ultrasound (ABUS): A Narrative Review. J Pers Med. 2021;11(8).

73. De Jesus C, Moseley T, Diaz V, Vishwanath V, Jean S, Elhatw A, et al. Supplemental Screening for Breast Cancer. Curr Breast Cancer Rep. 2023;15:97-102.

74. Expert Panel on Breast I, Weinstein SP, Slanetz PJ, Lewin AA, Battaglia T, Chagpar AB, et al. ACR Appropriateness Criteria® Supplemental Breast Cancer Screening Based on Breast Density. J Am Coll Radiol. 2021;18(11S):S456-S73.

75. D'Orsi CJ, Mendelson EB, Ikeda DM, al. e. Breast imaging reporting and data system: ACR BI-RADS – breast imaging atlas. Reston (VA): American College of Radiology; 2003.

76. European Commission - Joint Research Centre. European Commission Initiative on Breast Cancer - European Quality Assurance Scheme for Breast Cancer Services. Luxembourg: 2025. Available from: https://data.europa.eu/doi/10.2760/7067385,JRC140425

77. Muratov S, Canelo-Aybar C, Tarride JE, Alonso-Coello P, Dimitrova N, Borisch B, et al. Monitoring and evaluation of breast cancer screening programmes: selecting candidate performance indicators. BMC Cancer. 2020;20(1):795.

78. Gerkens S, Merkur S. Belgium - Health system review. 2020. Health Systems in Transition 22(5) Available from: https://eurohealthobservatory.who.int/publications/i/belgium-health-system-review-2020

79. Goossens M, De Brabander I, De Grève J, Van Ongeval C, Martens P, Van Limbergen E, et al. Flemish breast cancer screening programme: 15 years of key performance indicators (2002-2016). BMC Cancer. 2019;19(1):1012.

80. Belgian Cancer Registry (BCR). Indicateurs – évaluation du dépistage du cancer du sein 2021-2024 – Wallonie. 2026

81. Belgian Cancer Registry (BCR) B, Vivalis. Evaluatie van het screeningsprogramma voor borstkanker in Brussel, screeningsjaren 2021-2024. 2026

82. Centrum voor Kankeropsporing (CvKO), Belgian Cancer Registry (BCR), Departement Zorg. Infosheet belangrijkste indicatoren Bevolkingsonderzoek Borstkanker - screeningsjaar 2023. 2024

83. BruPrev. Activiteitenverslag 2024. 2024

84. Centrum voor Kankeropsporing (CvKO), Belgian Cancer Registry (BCR). Jaarrapport CvKO. 2022

85. Centrum voor Kankeropsporing (CvKO), Belgian Cancer Registry (BCR), Departement Zorg. Infosheet belangrijkste indicatoren Bevolkingsonderzoek Borstkanker - screeningsjaar 2024. 2025

86. Canelo-Aybar C, Ferreira DS, Ballesteros M, Posso M, Montero N, Solà I, 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;28(4):389-404.

87. Henderson JT, Webber EM, Weyrich MS, Miller M, Melnikow J. Screening for Breast Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force. Jama. 2024.

88. Institute for Quality and Efficiency in Health Care (IQWiG). Review of age limits in the mammography screening programme. 2022. S21-01 Available from: https://www.iqwig.de/download/s21-01_age-limits-in-the-mammography-screening-programme_extract-of-final-report_v1-1.pdf

89. Harada-Shoji N, Suzuki A, Ishida T, Yamamoto S, Kanemura S, Yamaguchi T, et al. Cumulative incidence of advanced breast cancer in women aged 40-49 years in the Japan Strategic Anti-cancer

Randomised Trial (J-START) of adjunctive ultrasonography: a prespecified secondary analysis. Lancet. 2026;407(10530):784-93.

90. Glasziou PP. Meta-analysis adjusting for compliance: the example of screening for breast cancer. J Clin Epidemiol. 1992;45(11):1251-6.

91. Duffy SW, Cuzick J, Tabár L, Vitak B, Hsiu-Hsi Chen T, Yen AM-F, 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(2):236-43.

92. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008.(doi):10.1136/bmj.j4008.

93. Higgins JPT, Altman DG, on behalf of the Cochrane Statistical Methods Group and the Cochrane Bias Methods Group. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions: John Wiley & Sons; 2012. p. 187-241.

94. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.

95. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2013;2013(6):CD001877.

96. Tarone RE. The excess of patients with advanced breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. Cancer. 1995;75(4):997-1003.

97. Kirkwood BR, Sterne JAC. Medical statistics. Second edition ed. Carlton, Victoria, Australia: Wiley-Blackwell; 2005.

98. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383-94. doi: 10.1016/j.jclinepi.2010.04.026. Epub Dec 31.

99. Glechner A, Wagner G, Mitus JW, Teufer B, Klerings I, Böck N, et al. Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk. Cochrane Database Syst Rev. 2023;3(3):CD009632.

100. Canelo-Aybar C, Posso M, Montero N, Solà I, Saz-Parkinson Z, Duffy SW, et al. Benefits and harms of annual, biennial, or triennial breast cancer mammography screening for women at average risk of breast cancer: a systematic review for the European Commission Initiative on Breast Cancer (ECIBC). Br J Cancer. 2022;126(4):673-88.

101. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64(4):401-6. doi: 10.1016/j.jclinepi.2010.07.015. Epub 1 Jan 5.

102. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, et al. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. J Clin Epidemiol. 2020;119:126-35.

103. Morgano GP, Wiercioch W, Piovani D, Neumann I, Nieuwlaat R, Piggott T, et al. Defining decision thresholds for judgments on health benefits and harms using the grading of recommendations assessment, development, and evaluation (GRADE) evidence to decision (EtD) frameworks: a randomized methodological study (GRADE-THRESHOLD). J Clin Epidemiol. 2025;179:111639.

104. Mathieu E, Noguchi N, Li T, Barratt AL, Hersch JK, De Bock GH, et al. Health benefits and harms of mammography screening in older women (75+ years)-a systematic review. Br J Cancer. 2024;130(2):275-96.

105. Zielonke N, Gini A, Jansen EEL, Anttila A, Segnan N, Ponti A, 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.

106. Nicholson WK, Silverstein M, Wong JB, Barry MJ, Chelmow D, Coker TR, et al. Screening for Breast Cancer: US Preventive Services Task Force Recommendation Statement. Effect of screening mammography on the risk of breast cancer deaths and of all-cause deaths: a systematic review with meta-analysis of cohort studies. JAMA. 2024;331(22):1918-30.

107. Olsen O, Gøtzsche PC. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2001(4):CD001877.

108. Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2006(4):CD001877.

109. Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2009(4):CD001877.

110. Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2011(1):CD001877.

111. Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R, et al. Chapter 7: Systematic reviews of etiology and risk In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis: JBI; 2020. Available from: https://synthesismanual.jbi.global

112. Kopans DB. The Canadian National Breast Screening Studies are compromised and their results are unreliable. They should not factor into decisions about breast cancer screening. Breast Cancer Res Treat. 2017;165(1):9-15.

113. Seely JM, Eby PR, Gordon PB, Appavoo S, Yaffe MJ. Errors in Conduct of the CNBSS Trials of Breast Cancer Screening Observed by Research Personnel. J Breast Imaging. 2022;4(2):135-43.

114. Yaffe MJ, Seely JM, Gordon PB, Appavoo S, Kopans DB. The randomized trial of mammography screening that was not-A cautionary tale. J Med Screen. 2022;29(1):7-11.

115. Shapiro S, Venet W, Strax P. Current results of the breast cancer screening randomized trial: the health insurance plan (HIP) of greater New York study. Hans Huber. 1988:3–15.

116. Clinical Trials. Extending the National Health Service (NHS) Breast Screening Age Range (AgeX) [Web page].2023 [cited 18 December]. Available from: https://clinicaltrials.gov/study/NCT01081288

117. Moser K, Sellars S, Wheaton M, Cooke J, Duncan A, Maxwell A, et al. Extending the age range for breast screening in England: pilot study to assess the feasibility and acceptability of randomization. J Med Screen. 2011;18(2):96-102.

118. Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE, et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. Lancet. 1999;353(9168):1903-8.

119. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64.

120. García-Albéniz X, Hernán MA, Logan RW, Price M, Armstrong K, Hsu J. Continuation of Annual Screening Mammography and Breast Cancer Mortality in Women Older Than 70 Years. Ann Intern Med. 2020;172(6):381-9.

121. Yang L, Wang S, Huang Y. An exploration for quantification of overdiagnosis and its effect for breast cancer screening. Chin J Cancer Res. 2020;32(1):26-35.

122. Richman IB, Long JB, Soulos PR, Wang SY, Gross CP. Estimating Breast Cancer Overdiagnosis After Screening Mammography Among Older Women in the United States. Ann Intern Med. 2023;176(9):1172-80.

123. Puliti D, Bucchi L, Mancini S, Paci E, Baracco S, Campari C, et al. Advanced breast cancer rates in the epoch of service screening: The 400,000 women cohort study from Italy. Eur J Cancer. 2017;75:109-16.

124. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. CMAJ. 1992;147(10):1477-88.

125. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. CMAJ. 1992;147(10):1459-76.

126. de Glas NA, de Craen AJM, Bastiaannet E, Op 't Land EG, Kiderlen M, van de Water W, et al. Effect of implementation of the mass breast cancer screening programme in older women in the Netherlands: population based study. BMJ. 2014;349:g5410.

127. McCarthy EP, Burns RB, Freund KM, Ash AS, Shwartz M, Marwill SL, et al. Mammography use, breast cancer stage at diagnosis, and survival among older women. J Am Geriatr Soc. 2000;48(10):1226-33.

128. Malmgren JA, Parikh J, Atwood MK, Kaplan HG. Improved prognosis of women aged 75 and older with mammography-detected breast cancer. Radiology. 2014;273(3):686-94.

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

130. 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. Lancet Oncol. 2007;8(12):1129-38.

131. Puliti D, Miccinesi G, Paci E. Overdiagnosis in breast cancer: design and methods of estimation in observational studies. Prev Med. 2011;53(3):131-3.

132. Miller AB. The costs and benefits of breast cancer screening. Am J Prev Med. 1993;9(3):175-80.

133. Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP. Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. BMJ. 2006;332(7543):689-92.

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

135. Lund E, Nakamura A, Thalabard JC. No overdiagnosis in the Norwegian Breast Cancer Screening Program estimated by combining record linkage and questionnaire information in the Norwegian Women and Cancer study. Eur J Cancer. 2018;89:102-12.

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

137. Blichert-Toft M, Rose C, Andersen JA, Overgaard M, Axelsson CK, Andersen KW, et al. Danish randomized trial comparing breast conservation therapy with mastectomy: six years of life-table analysis. Danish Breast Cancer Cooperative Group. J Natl Cancer Inst Monogr. 1992(11):19-25.

138. Fisher B, Anderson S. Conservative surgery for the management of invasive and noninvasive carcinoma of the breast: NSABP trials. National Surgical Adjuvant Breast and Bowel Project. World J Surg. 1994;18(1):63-9.

139. Jacobson JA, Danforth DN, Cowan KH, d'Angelo T, Steinberg SM, Pierce L, et al. Ten-year results of a comparison of conservation with mastectomy in the treatment of stage I and II breast cancer. N Engl J Med. 1995;332(14):907-11.

140. Lichter AS, Lippman ME, Danforth DN, Jr., d'Angelo T, Steinberg SM, deMoss E, et al. Mastectomy versus breast-conserving therapy in the treatment of stage I and II carcinoma of the breast: a randomized trial at the National Cancer Institute. J Clin Oncol. 1992;10(6):976-83.

141. Veronesi U, Salvadori B, Luini A, Greco M, Saccozzi R, del Vecchio M, et al. Breast conservation is a safe method in patients with small cancer of the breast. Long-term results of three randomised trials on 1,973 patients. Eur J Cancer. 1995;31A(10):1574-9.

142. Blichert-Toft M, Nielsen M, Düring M, Møller S, Rank F, Overgaard M, et al. Long-term results of breast conserving surgery vs. mastectomy for early stage invasive breast cancer: 20-year follow-up of the Danish randomized DBCG-82TM protocol. Acta Oncol. 2008;47(4):672-81.

143. Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med. 2002;347(16):1233-41.

144. Litière S, Werutsky G, Fentiman IS, Rutgers E, Christiaens M-R, Van Limbergen E, et al. Breast conserving therapy versus mastectomy for stage I-II breast cancer: 20 year follow-up of the EORTC 10801 phase 3 randomised trial. Lancet Oncol. 2012;13(4):412-9.

145. Simone NL, Dan T, Shih J, Smith SL, Sciuto L, Lita E, et al. Twenty-five year results of the national cancer institute randomized breast conservation trial. Breast Cancer Res Treat. 2012;132(1):197-203.

146. Veronesi U, Cascinelli N, Mariani L, Greco M, Saccozzi R, Luini A, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. N Engl J Med. 2002;347(16):1227-32.

147. Jatoi I, Proschan MA. Randomized trials of breast-conserving therapy versus mastectomy for primary breast cancer: a pooled analysis of updated results. Am J Clin Oncol. 2005;28(3):289-94.

148. Yang SH, Yang KH, Li YP, Zhang YC, He XD, Song AL, et al. Breast conservation therapy for stage I or stage II breast cancer: a meta-analysis of randomized controlled trials. Ann Oncol. 2008;19(6):1039-44.

149. Vila J, Gandini S, Gentilini O. Overall survival according to type of surgery in young (≤40 years) early breast cancer patients: A systematic meta-analysis comparing breast-conserving surgery versus mastectomy. Breast. 2015;24(3):175-81.

150. Chen Y, Jiang L, Gao B, Cheng Z-Y, Jin J, Yang K-H. Survival and disease-free benefits with mastectomy versus breast conservation therapy for early breast cancer: a meta-analysis. Breast Cancer Res Treat. 2016;157(3):517-25.

151. McCready D, Holloway C, Shelley W, Down N, Robinson P, Sinclair S, et al. Surgical management of early stage invasive breast cancer: a practice guideline. Can J Surg. 2005;48(3):185-94.

152. Kreienberg R, Kopp I, Lorenz W, Budach W, Dunst J, Lebeau A, et al. Interdisciplinary S3 Guidelines for the Diagnosis and Treatment of Breast Cancer in Women. 2006. Available from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=bcfa82c008361bdcee6875d1699a197d0a941e2a

153. Christiaens M-R, Vlayen J, Gailly J, Neven P, Carly B, Schobbens J-C, et al. Scientific support of the College of Oncology: a national clinical practice guideline for breast cancer. Good Clinical Practice (GCP). Brussels: Belgian Health Care Knowledge Centre (KCE); 2007 26/07/2007. KCE Reports 63 (D2007/10.273/37) Available from: https://kce.fgov.be/publication/report/scientific-support-of-the-college-of-oncology-a-national-clinical-practice-guid-0

154. Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. Psychooncology. 2005;14(11):917-38.

155. Bond M, Pavey T, Welch K, Cooper C, Garside R, Dean S, et al. Psychological consequences of false-positive screening mammograms in the UK. Evid Based Med. 2013;18(2):54-61.

156. Salz T, Richman AR, Brewer NT. Meta-analyses of the effect of false-positive mammograms on generic and specific psychosocial outcomes. Psychooncology. 2010;19(10):1026-34.

157. Autier P, Jørgensen KJ, Støvring H. Evaluation of screening mammography effectiveness: The IARC recommendations of 2015 need revision. Eur J Cancer. 2025;227:115657.

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

159. Parvinen I, Chiu S, Pylkkänen L, Klemi P, Immonen-Räihä P, Kauhava L, et al. Effects of annual vs triennial mammography interval on breast cancer incidence and mortality in ages 40-49 in Finland. Br J Cancer. 2011;105(9):1388-91.

160. Coldman AJ, Phillips N, Olivotto IA, Gordon P, Warren L, Kan L. Impact of changing from annual to biennial mammographic screening on breast cancer outcomes in women aged 50-79 in British Columbia. J Med Screen. 2008;15(4):182-7.

161. Miglioretti DL, Zhu W, Kerlikowske K, Sprague BL, Onega T, Buist DS, et al. Breast Tumor Prognostic Characteristics and Biennial vs Annual Mammography, Age, and Menopausal Status. JAMA Oncol. 2015;1(8):1069-77.

162. O'Meara ES, Zhu W, Hubbard RA, Braithwaite D, Kerlikowske K, Dittus KL, et al. Mammographic screening interval in relation to tumor characteristics and false-positive risk by race/ethnicity and age. Cancer. 2013;119(22):3959-67.

163. Klemi PJ, Toikkanen S, Räsänen O, Parvinen I, Joensuu H. Mammography screening interval and the frequency of interval cancers in a population-based screening. Br J Cancer. 1997;75(5):762-6.

164. Kerlikowske K, Zhu W, Hubbard RA, Geller B, Dittus K, Braithwaite D, et al. Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy. JAMA Intern Med. 2013;173(9):807-16.

165. Braithwaite D, Zhu W, Hubbard RA, O'Meara ES, Miglioretti DL, Geller B, et al. Screening outcomes in older US women undergoing multiple mammograms in community practice: does interval, age, or comorbidity score affect tumor characteristics or false positive rates? J Natl Cancer Inst. 2013;105(5):334-41.

166. Dittus K, Geller B, Weaver DL, Kerlikowske K, Zhu W, Hubbard R, et al. Impact of mammography screening interval on breast cancer diagnosis by menopausal status and BMI. J Gen Intern Med. 2013;28(11):1454-62.

167. Ohuchi N, Suzuki A, Sobue T, Kawai M, Yamamoto S, Zheng Y-F, et al. Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): a randomised controlled trial. Lancet. 2016;387(10016):341-8.

168. Harada-Shoji N, Suzuki A, Ishida T, Zheng Y-F, Narikawa-Shiono Y, Sato-Tadano A, et al. Evaluation of Adjunctive Ultrasonography for Breast Cancer Detection Among Women Aged 40-49 Years With

Varying Breast Density Undergoing Screening Mammography: A Secondary Analysis of a Randomized Clinical Trial. JAMA Netw Open. 2021;4(8):e2121505.

169. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.

170. Buchberger W, Geiger-Gritsch S, Knapp R, Gautsch K, Oberaigner W. Combined screening with mammography and ultrasound in a population-based screening program. Eur J Radiol. 2018;101:24-9.

171. Chae EY, Kim HH, Cha JH, Shin HJ, Kim H. Evaluation of screening whole-breast sonography as a supplemental tool in conjunction with mammography in women with dense breasts. J Ultrasound Med. 2013;32(9):1573-8.

172. Giuliano V, Giuliano C. Improved breast cancer detection in asymptomatic women using 3D-automated breast ultrasound in mammographically dense breasts. Clin Imaging. 2013;37(3):480-6.

173. Starikov A, Drotman M, Hentel K, Katzen J, Min RJ, Arleo EK. 2D mammography, digital breast tomosynthesis, and ultrasound: which should be used for the different breast densities in breast cancer screening? Clin Imaging. 2016;40(1):68-71.

174. Chough DM, Berg WA, Bandos AI, Rathfon GY, Hakim CM, Lu AH, et al. A Prospective Study of Automated Breast Ultrasound Screening of Women with Dense Breasts in a Digital Breast Tomosynthesis-based Practice. J Breast Imaging. 2020;2(2):125-33.

175. Lee JM, Arao RF, Sprague BL, Kerlikowske K, Lehman CD, Smith RA, et al. Performance of Screening Ultrasonography as an Adjunct to Screening Mammography in Women Across the Spectrum of Breast Cancer Risk. JAMA Intern Med. 2019;179(5):658-67.

176. Schiller-Frühwirth 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(3):333-51.

177. IMS Health Pty Ltd Australia. BreastScreen Australia Evaluation. Economic evaluation and modelling study. 2009.

178. Shafik N, Ilmonen P, Viitasaari L, Sarkeala T, Heinävaara S. Flexible transition probability model for assessing cost-effectiveness of breast cancer screening extension to include women aged 45-49 and 70-74. PLoS ONE. 2023;18(6 June).

179. Blankenburg M, Sánchez-Collado I, Soyemi BO, Åkerborg Ö, Caleyachetty A, Harris J, et al. Economic evaluation of supplemental breast cancer screening modalities to mammography or digital breast tomosynthesis in women with heterogeneously and extremely dense breasts and average or intermediate breast cancer risk in US healthcare. Journal of Medical Economics. 2023;26(1):850-61.

180. Schousboe JT, Sprague BL, Abraham L, O'Meara ES, Onega T, Advani S, et al. Cost-Effectiveness of Screening Mammography Beyond Age 75 Years : A Cost-Effectiveness Analysis. Annals of Internal Medicine. 2022;175(1):11-9.

181. Kregting LM, Sankatsing VDV, Heijnsdijk EAM, de Koning HJ, van Ravesteyn NT. Finding the optimal mammography screening strategy: A cost-effectiveness analysis of 920 modelled strategies. International Journal of Cancer. 2022;151(2):287-96.

182. Muhlberger N, Sroczynski G, Gogollari A, Jahn B, Pashayan N, Steyerberg E. Cost Effectiveness of Breast Cancer Screening and Prevention: A Systematic Review with a Focus on Risk-Adapted Strategies. European Journal of Health Economics. 2021;22(8):1311-44.

183. Tina Shih YC, Dong W, Xu Y, Etzioni R, Shen Y. Incorporating baseline breast density when screening women at average risk for breast cancer: A cost-effectiveness analysis. Annals of Internal Medicine. 2021;174(5):602-12.

184. Chootipongchaivat S, Wong XY, Ten Haaf K, Hartman M, Tan KB, Van Ravesteyn NT, et al. Cost-effectiveness analysis of breast cancer screening using mammography in singapore: A modeling study. Cancer Epidemiology Biomarkers and Prevention. 2021;30(4):653-60.

185. Yuniar P, Robinson S, Moorin R, Norman R. Economic Evaluation of Breast Cancer Early Detection Strategies in Asia: A Systematic Review. Value in Health Regional Issues. 2020;21((Yuniar P., popy.yuniar@postgrad.curtin.edu.au; Robinson S.; Moorin R.; Norman R.) School of Public Health, Curtin University, Perth, Western Australia, Australia):252-63.

186. 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.

187. Tina Shih YC, Dong W, Xu Y, Shen Y. Assessing the Cost-Effectiveness of Updated Breast Cancer Screening Guidelines for Average-Risk Women. Value in Health. 2019;22(2):185-93.

188. Mandrik O, Ekwunife OI, Meheus F, Severens JL, Lhachimi S, Uyl-de Groot CA, et al. Systematic reviews as a “lens of evidence”: Determinants of cost-effectiveness of breast cancer screening. Cancer Medicine. 2019;8(18):7846-58.

189. 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. CMAJ open. 2018;6(1):E77-E86.

190. Koleva-Kolarova RG, Daszczuk AM, de Jonge C, Abu Hantash MK, Zhan ZZ, Postema EJ, et al. A modelling study to evaluate the costs and effects of lowering the starting age of population breast cancer screening. Maturitas. 2018;109:81-8.

191. Carter KJ, Castro F, Morcos RN. Insights into breast cancer screening: A computer simulation of two contemporary screening strategies. American Journal of Roentgenology. 2018;210(3):564-71.

192. van Luijt PA, Heijnsdijk EAM, de Koning HJ. Cost-effectiveness of the Norwegian breast cancer screening program. International Journal of Cancer. 2017;140(4):833-40.

193. Schiller-Frühwirth I, Jahn B, Einzinger P, Zauner G, Urach C, Siebert U. The Long-Term Effectiveness and Cost Effectiveness of Organized versus Opportunistic Screening for Breast Cancer in Austria. Value in Health. 2017;20(8):1048-57.

194. Morton R, Sayma M, Sura MS. Economic analysis of the breast cancer screening program used by the UK NHS: Should the program be maintained? Breast Cancer: Targets and Therapy. 2017;9((Morton R.; Sayma M., meeladsayma@yahoo.co.uk; Sura M.S.) Imperial College Business School, Imperial College London, London, United Kingdom):217-25.

195. Rafia R, Brennan A, Madan J, Collins K, Reed MWR, Lawrence G, et al. Modeling the Cost-Effectiveness of Alternative Upper Age Limits for Breast Cancer Screening in England and Wales. Value in Health. 2016;19(4):404-12.

196. Arrospide A, Rue M, van Ravesteyn NT, Comas M, Soto-Gordoa M, Sarriugarte G, et al. Economic evaluation of the breast cancer screening programme in the Basque Country: Retrospective cost-effectiveness and budget impact analysis. BMC Cancer. 2016;16(1).

197. Sankatsing VDV, Heijnsdijk EAM, Van Luijt PA, Van Ravesteyn NT, Fracheboud J, De Koning HJ. Cost-effectiveness of digital mammography screening before the age of 50 in the Netherlands. International Journal of Cancer. 2015;137(8):1990-9.

198. 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(4):354-63.

199. Health Quality Ontario (HQO). Ultrasound as an Adjunct to Mammography for Breast Cancer Screening: A Health Technology Assessment. 2016

200. Lelie M, Vandenbulcke B, Verhaeghe N, Annemans L, Simoens S, Putman K. De kosteneffectiviteit van het bevolkingsonderzoek borstkanker in Vlaanderen. Interuniversity Centre for Health Economics Research (I-CHER); 2024.

201. !!! INVALID CITATION !!! 173-175, 177, 178, 181, 183, 186, 188-191, 194.

202. Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG. Retrospective Cost-effectiveness Analysis of Screening Mammography. JNCI: Journal of the National Cancer Institute. 2006;98(11):774-82.

203. Schleinitz MD, DePalo D, Blume J, Stein M. Can differences in breast cancer utilities explain disparities in breast cancer care? Journal of General Internal Medicine. 2006;21(12):1253-60.

204. Lidgren M, Wilking N, Jönsson B, Rehnberg C. Health related quality of life in different states of breast cancer. Quality of Life Research. 2007;16(6):1073-81.

205. Li L, Severens JL, Mandrik O. Disutility associated with cancer screening programs: A systematic review. PLOS ONE. 2019;14(7):e0220148.

206. de Haes JCJM, de Koning HJ, van Oortmarssen GJ, van Agt HME, de Bruyn AE, van der Maas PJ. The impact of a breast cancer screening programme on quality-adjusted life-years. International Journal of Cancer. 1991;49(4):538-44.

207. Tengs TO, Wallace A. One Thousand Health-Related Quality-of-Life Estimates. Medical Care. 2000;38(6).

208. de Koning HJ, Martin van Ineveld B, van Oortmarssen GJ, de Haes JCJM, Collette HJA, Hendriks JHCL, et al. Breast cancer screening and cost-effectiveness; Policy alternatives, quality of life considerations and the possible impact of uncertain factors. International Journal of Cancer. 1991;49(4):531-7.

209. Jansen SJT, Stiggelbout AM, Wakker PP, Vliet Vlieland TPM, Leer J-WH, Nooy MA, et al. Patients' Utilities for Cancer Treatments:A Study of the Chained Procedure for the Standard Gamble and Time Tradeoff. Medical Decision Making. 1998;18(4):391-9.

210. Carter KJ, Ritchey NP, Castro F, Caccamo LP, Kessler E, Erickson BA, et al. Treatment of Early-Stage Breast Cancer in the Elderly:A Health-outcome-Based Approach. Medical Decision Making. 1998;18(2):213-9.

211. Ahern CH, Shih YC, Dong W, Parmigiani G, Shen Y. Cost-effectiveness of alternative strategies for integrating MRI into breast cancer screening for women at high risk. Br J Cancer. 2014;111(8):1542-51.

212. Stout NK, Lee SJ, Schechter CB, Kerlikowske K, Alagoz O, Berry D, et al. Benefits, Harms, and Costs for Breast Cancer Screening After US Implementation of Digital Mammography. JNCI: Journal of the National Cancer Institute. 2014;106(6):dju092.

213. Kim SH, Jo MW, Ock M, Lee HJ, Lee JW. Estimation of health state utilities in breast cancer. Patient Prefer Adherence. 2017;11:531-6.

214. Raftery J, Chorozoglou M. Possible net harms of breast cancer screening: updated modelling of Forrest report. BMJ. 2011;343:d7627.

215. Bonomi AE, Boudreau DM, Fishman PA, Ludman E, Mohelnitzky A, Cannon EA, et al. Quality of life valuations of mammography screening. Quality of Life Research. 2008;17(5):801-14.

216. Rijnsburger AJ, Essink-Bot ML, van Dooren S, Borsboom GJJM, Seynaeve C, Bartels CCM, et al. Impact of screening for breast cancer in high-risk women on health-related quality of life. British Journal of Cancer. 2004;91(1):69-76.

217. Kaur MN, Yan J, Klassen AF, David JP, Pieris D, Sharma M, et al. A Systematic Literature Review of Health Utility Values in Breast Cancer. Medical Decision Making. 2022;42(5):704-19.

218. Tosteson ANA, Fryback DG, Hammond CS, Hanna LG, Grove MR, Brown M, et al. Consequences of False-Positive Screening Mammograms. JAMA Internal Medicine. 2014;174(6):954-61.

219. Gerard K, Johnston K, Brown J. The role of a pre-scored multi-attribute health classification measure in validating condition-specific health state descriptions. Health Econ. 1999;8(8):685-99.

220. Johnston K, Brown J, Gerard K, O'Hanlon M, Morton A. Valuing temporary and chronic health states associated with breast screening. Social Science & Medicine. 1998;47(2):213-22.

221. Gatta G, Cappabianca S, La Forgia D, Massafra R, Fanizzi A, Cuccurullo V, et al. Second-Generation 3D Automated Breast Ultrasonography (Prone ABUS) for Dense Breast Cancer Screening Integrated to Mammography: Effectiveness, Performance and Detection Rates. Journal of Personalized Medicine; 2021.

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

223. Domingo L, Hofvind S, Hubbard RA, Román M, Benkeser D, Sala M, et al. Cross-national comparison of screening mammography accuracy measures in U.S., Norway, and Spain. European Radiology. 2016;26(8):2520-8.

224. Perry NM, Patani N, Milner SE, Pinker K, Mokbel K, Allgood PC, et al. The impact of digital mammography on screening a young cohort of women for breast cancer in an urban specialist breast unit. European Radiology. 2011;21(4):676-82.

225. Skaane P, Skjennald A. Screen-Film Mammography versus Full-Field Digital Mammography with Soft-Copy Reading: Randomized Trial in a Population-based Screening Program—The Oslo II Study. Radiology. 2004;232(1):197-204.

226. van Luijt PA, Fracheboud J, Heijnsdijk EAM, den Heeten GJ, de Koning HJ. Nation-wide data on screening performance during the transition to digital mammography: Observations in 6 million screens. European Journal of Cancer. 2013;49(16):3517-25.

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

228. Souza FH, Wendland EM, Rosa MI, Polanczyk CA. Is full-field digital mammography more accurate than screen-film mammography in overall population screening? A systematic review and meta-analysis. The Breast. 2013;22(3):217-24.

229. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, et al. Individual and Combined Effects of Age, Breast Density, and Hormone Replacement Therapy Use on the Accuracy of Screening Mammography. Annals of Internal Medicine. 2003;138(3):168-75.

230. Kavanagh AM, Cawson J, Byrnes GB, Giles GG, Marr G, Tong B, et al. Hormone Replacement Therapy, Percent Mammographic Density, and Sensitivity of Mammography. Cancer Epidemiology, Biomarkers & Prevention. 2005;14(5):1060-4.

231. Kerlikowske K, Carney PA, Geller B, Mandelson MT, Taplin SH, Malvin K, et al. Performance of Screening Mammography among Women with and without a First-Degree Relative with Breast Cancer. Annals of Internal Medicine. 2000;133(11):855-63.

232. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of Age, Breast Density, and Family History on the Sensitivity of First Screening Mammography. JAMA. 1996;276(1):33-8.

233. Kerlikowske K, Hubbard RA, Miglioretti DL, Geller BM, Yankaskas BC, Lehman CD, et al. Comparative Effectiveness of Digital Versus Film-Screen Mammography in Community Practice in the United States. Annals of Internal Medicine. 2011;155(8):493-502.

234. Kolb TM, Lichy J, Newhouse JH. Comparison of the Performance of Screening Mammography, Physical Examination, and Breast US and Evaluation of Factors that Influence Them: An Analysis of 27,825 Patient Evaluations. Radiology. 2002;225(1):165-75.

235. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, et al. Breast Density as a Predictor of Mammographic Detection: Comparison of Interval- and Screen-Detected Cancers. JNCI: Journal of the National Cancer Institute. 2000;92(13):1081-7.

236. Rosenberg RD, Hunt WC, Williamson MR, Gilliland FD, Wiest PW, Kelsey CA, et al. Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: review of 183,134 screening mammograms in Albuquerque, New Mexico. Radiology. 1998;209(2):511-8.

237. Sinclair N, Littenberg B, Geller B, Muss H. Accuracy of Screening Mammography in Older Women. American Journal of Roentgenology. 2011;197(5):1268-73.

238. Kemp Jacobsen K, O'Meara ES, Key D, S.M. Buist D, Kerlikowske K, Vejborg I, et al. Comparing sensitivity and specificity of screening mammography in the United States and Denmark. International Journal of Cancer. 2015;137(9):2198-207.

239. Hofvind S, Geller B, Vacek PM, Thoresen S, Skaane P. Using the European guidelines to evaluate the Norwegian Breast Cancer Screening Program. European Journal of Epidemiology. 2007;22(7):447-55.

240. Kerlikowske K, Barclay J. Outcomes of Modern Screening Mammography. JNCI Monographs. 1997;1997(22):105-11.

241. Hofvind S, Geller BM, Skelly J, Vacek PM. Sensitivity and specificity of mammographic screening as practised in Vermont and Norway. British Journal of Radiology. 2012;85(1020):e1226-e32.

242. NCI-funded Breast Cancer Surveillance Consortium co-operative agreement [Web page]. [cited 1 April 2017]. Available from: http://breastscreening.cancer.gov/data/performance/screening/2008/perf_age.html

243. National Evaluation Team for Breast cancer screening (NETB). National evaluation of breast cancer screening in the Netherlands 1990-2007. 2009

244. Nederlandse Kankerregistratie. [cited December, 2014]. Available from: http://www.cijfersoverkanker.nl/

245. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, et al. Diagnostic Accuracy of Digital versus Film Mammography: Exploratory Analysis of Selected Population Subgroups in DMIST. Radiology. 2008;246(2):376-83.

246. Centrum voor Kankeropsporing (CvKO), Belgian Cancer Registry (BCR). Jaarrapport CvKO. 2021

247. Centrum voor Kankeropsporing (CvKO), Belgian Cancer Registry (BCR). Jaarrapport CvKO. 2020

248. Centrum voor Kankeropsporing (CvKO), Belgian Cancer Registry (BCR). Jaarrapport CvKO. 2019

249. Centrum voor Kankeropsporing (CvKO), Belgian Cancer Registry (BCR). Jaarrapport CvKO. 2018

250. Centrum voor Kankeropsporing (CvKO), Belgian Cancer Registry (BCR). Jaarrapport CvKO. 2017

251. van Ravesteyn NT, van den Broek JJ, Li X, Weedon-Fekjær H, Schechter CB, Alagoz O, et al. Modeling Ductal Carcinoma In Situ (DCIS): An Overview of CISNET Model Approaches. Medical Decision Making. 2018;38(1_suppl):126S-39S.

252. Fobelets M, Pil L, Putman K, Annemans L. De kosteneffectiviteit van het bevolkingsonderzoek naar borstkanker in Vlaanderen: Gezondheidseconomische evaluatie. 2015

253. Informatieveiligheidscomité Kamer van Sociale Zekerheid en Gezondheid - Comité de sécurité de l'information -Chambre sécurité sociale et santé. Beraadslaging nr. 24/220 van 3 december 2024 met betrekking tot de mededeling van gepseudonimiseerde persoonsgegevens die de gezondheid betreffen door de Stichting Kankerregister aan het Federaal Kenniscentrum (KCE) in het kader vanhet project «2024-11: Breast cancers screening»/ Délibération n° 21/220 du 3 décembre 2024 portant sur la communication de données à caractère personnel pseudonymisées relatives à la santé par la Fondation Registre du cancer au Centre fédéral d'expertise (KCE) dans le cadre du projet «2024-11: Breast cancers screening». 2024. IVC/KSZG/24/448 - CSI/CSSS/24/448

254. Henau K, Van Eycken E, Silversmit G, Pukkala E. Regional variation in incidence for smoking and alcohol related cancers in Belgium. Cancer Epidemiol. 2015;39(1):55-65.

255. Koninklijk besluit van 21 maart 2003 houdende vaststelling van de normen waaraan het zorgprogramma voor oncologische basiszorg en het zorgprogramma voor oncologie moeten voldoen om te worden erkend. Artikel 11, § 1, B.S. 2003.

256. Wet houdende diverse bepalingen betreffende gezondheid van 13 december 2006, Hoofdstuk VI (Kankerregister), Artikel 39, B.S. 2006.

257. Commissie voor de bescherming van de persoonlijke levenssfeer (CBPL) – Commission de la protection de la vie privée (CPVP). Beraadslaging nr 09/071 van 15 september 2009, laatst gewijzigd op 18 februari 2014, met betrekking tot de mededeling van persoonsgegevens door de verzekeringsinstellingen aan de Stichting Kankerregister in het kader van artikel 45 quinquies van het KB nr. 78 van 10 november 1967 betreffende de uitoefening van de gezondheidsberoepen / Délibération n°09/071 du 15 septembre 2009, modifiée le 18 février 2014, relative à la communication de données à caractère personnel par les organismes assureurs à la Fondation Registre du Cancer dans le cadre de l'article 45quinquies de l'AR n° 78 du 10 novembre 1967 relatif è l'exercice des professions des soins de santé.2014. 2009. Available from: https://www.ehealth.fgov.be/ehealthplatform/file/view/467ba5d6fbccb4d3c29cbd21f77e5032737beddf?filename=sector_committee_09-071-n033.pd

258. Comité de sécurité de l'information Csses. Délibération n° 16/020 du 15 mars 2016, modifiée en dernier lieu le 4 juin 2019, relative à l'échange de données à caractère personnel relatives à la santé entre le Centre Communautaire de Référence pour le dépistage des cancers, la Fondation Registre du Cancer, les organismes assureurs et l'agence intermutualiste dans le cadre du dépistage du cancer du sein organisé par la Communauté Française. 2019 Available from: https://ehealth.fgov.be/ehealthplatform/file/cc73d96153bbd5448a56f19d925d05b1379c7f21/129e7c1553258e354c54e178b470089c4effd8ab/16-020-f154-depistage-du-cancer-du-sein-communaute-francaise-modifiee-le-4-juin-2019.pdf

259. Comité de sécurité de l'information Csses. Délibération n° 16/039 du 19 avril 2016, modifiée le 24 novembre 2020, relative à l'échange de données à caractère personnel relatives à la santé entre le Centre Bruxellois de Coordination pour le Dépistage et la Drévention des Dancers (bruprev), la Fondation Registre du Cancer, les organismes assureurs et l'agence intermutualiste dans le cadre du dépistage du cancer du sein dans la région de Bruxelles-Capitale. Comité de sécurité de l'information; 2020 Available from: https://ehealth.fgov.be/ehealthplatform/file/cc73d96153bbd5448a56f19d925d05b1379c7f21/22df184e0b057a54af5ee30abb4721a67adcb975/16-039-f350-depistage-du-cancer-du-sein-bruprev-modifiee-le-24-novembre-2020.pdf

260. Informatieveiligheidscomité Kszeg. Beraadslaging nr. 14/115 van 16 december 2014, laatst gewijzigd op 2 juli 2019, betreffende de uitwisseling van persoonsgegevens die de gezondheid betreffen door het centrum voor kankeropsporing, de stichting kankerregister en de verzekeringsinstellingen in het kader van het bevolkingsonderzoek naar borstkanker. Informatieveiligheidscomité; 2019 Available from: https://ehealth.fgov.be/ehealthplatform/file/cc73d96153bbd5448a56f19d925d05b1379c7f21/d53b0329f903b54deeb4fdb6e5b015003adb771f/14-115-n234-bevolkingsonderzoek-borstkanker-gewijzigd-op-2-juli-2019.pdf

261. Brierley JD, Wittekind C. TNM Classification of Malignant Tumours. 6th edition ed.; 2002.

262. Sobin lH, Gospodarowicz MK, Wittekind C. TNM Classification of Malignant Tumours. UICC International Union Against Cancer. 7th edition ed. Sons JW, editor.; 2009.

263. Brierley J, Wittekind C. TNM Classification of Malignant Tumours. 8th Edition ed. Wiley-Blackwell, editor.; 2017.

264. Desimpel F, Luyten J, Camberlin C, Primus-de Jong C, Verleye L, Neyt M. Lung cancer screening in a high-risk population. Health Technology Assessment (HTA). Brussels: Belgian Health Care Knowledge Center (KCE); 2024 02/2024. KCE Reports 379 (379C) Available from: https://doi.org/10.57598/R398C

265. Leroy R, De Gendt C, Bourgeois J, Verbeeck J, Savoye I, Silversmit G, et al. Survival and Quality of care offered in Belgian hospitals with and without recognition for breast cancer. Health Services Research (HSR). Brussels: Belgian Health Care Knowledge Centre (KCE); 2023 03/2023. KCE Reports 365 (365C) Available from: https://doi.org/10.57598/R365C

266. SAS Institute Inc. SAS 9.4 BASE & STAT/GRAPH. Cary, NC: {SAS Institute Inc.}; 2017.

267. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.

268. Peacock HM, van Walle L, Silversmit G, Neven P, Han SN, Van Damme N. Breast cancer incidence, stage distribution, and treatment shifts during the 2020 COVID-19 pandemic: a nationwide population-level study. Arch Public Health. 2024;82(1):66.

269. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) Statement: Updated Reporting Guidance for Health Economic Evaluations. Value Health. 2022;25(1):3-9.

270. Gerkens S, Lefèvre M, Bouckaert N, Levy M, Maertens de Noordhout C, Obyn C, et al. Performance of the Belgian health system: Report 2024. Health Services Research (HSR). Brussels: Belgian Health Care Knowledge Center (KCE); 2024 02/2024. KCE Reports 376C (376) Available from: https://doi.org/10.57598/R376C

271. Centrum voor Kankeropsporing (CvKO). Jaarverslag CvKO 2023. 2023

272. Centrum voor Kankeropsporing (CvKO), Belgian Cancer Registry (BCR). Bevolkingsonderzoek borstkanker: definities en berekeningswijzen. 2024

273. Welch HG, Schwartz LM, Woloshin S. Ramifications of screening for breast cancer: 1 in 4 cancers detected by mammography are pseudocancers. BMJ. 2006;332(7543):727.

274. Harris R, Yeatts J, Kinsinger L. Breast cancer screening for women ages 50 to 69years a systematic review of observational evidence. Preventive Medicine. 2011;53(3):108-14.

275. Neyt M, Devos C, Thiry N, Silversmit G, De Gendt C, Van Damme N, et al. Benefits and costs of innovative oncology drugs in Belgium (2004-2017). Health Technology Assessment (HTA). Brussel: Belgian Health Care Knowledge Centre (KCE). 2021. KCE Reports 343 Available from: https://doi.org/10.57598/R343C

276. Rautenberg T, Hodgkinson B, Zerwes U, Downes M. Meta-analysis of health state utility values measured by EuroQol 5-dimensions (EQ5D) questionnaire in Chinese women with breast cancer. BMC cancer. 2022;22(1):52.

277. Wang L, Shi JF, Zhu J, Huang HY, Bai YN, Liu GX, et al. Health-related quality of life and utility scores of patients with breast neoplasms in China: A multicenter cross-sectional survey. Breast. 2018;39:53-62.

278. Pillay J, Guitard S, Rahman S, Saba S, Rahman A, Bialy L, et al. Patient preferences for breast cancer screening: a systematic review update to inform recommendations by the Canadian Task Force on Preventive Health Care. Systematic Reviews. 2024;13(1):140.

279. Criscitiello C, Spurden D, Piercy J, Rider A, Williams R, Mitra D, et al. Health-Related Quality of Life Among Patients With HR+/HER2- Early Breast Cancer. Clin Ther. 2021;43(7):1228-44.e4.

280. Gong JR, Han J, Lee D, Bae S. A Meta-Regression Analysis of Utility Weights for Breast Cancer: The Power of Patients' Experience. International journal of environmental research and public health. 2020;17(24).

281. Porciello G, Montagnese C, Crispo A, Grimaldi M, Libra M, Vitale S, et al. Mediterranean diet and quality of life in women treated for breast cancer: A baseline analysis of DEDiCa multicentre trial. PLoS One. 2020;15(10):e0239803.

282. Guerra RL, Dos Reis NB, Corrêa FM, Fernandes MM, Ribeiro Alves Fernandes R, Cancela MC, et al. Breast Cancer Quality of Life and Health-state Utility at a Brazilian Reference Public Cancer Center. Expert Rev Pharmacoecon Outcomes Res. 2020;20(2):185-91.

283. Pourrahmat MM, Kim A, Kansal AR, Hux M, Pushkarna D, Fazeli MS, et al. Health state utility values by cancer stage: a systematic literature review. Eur J Health Econ. 2021;22(8):1275-88.

284. Li S, Wang M, Liu L, Chen G. Which approach is better in eliciting health state utilities from breast cancer patients? Evidence from mainland China. Eur J Cancer Care (Engl). 2019;28(2):e12965.

285. Ou HT, Chung WP, Su PF, Lin TH, Lin JY, Wen YC, et al. Health-related quality of life associated with different cancer treatments in Chinese breast cancer survivors in Taiwan. Eur J Cancer Care (Engl). 2019;28(4):e13069.

286. Kim SH, Jo MW, Lee JW, Lee HJ, Kim JK. Validity and reliability of EQ-5D-3L for breast cancer patients in Korea. Health Qual Life Outcomes. 2015;13:203.

287. Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. Expert Rev Pharmacoecon Outcomes Res. 2010;10(5):553-66.

288. EUnetHTA. Guideline Health-related quality of life and utility measures. 2015

289. Timmers JM, Damen JA, Pijnappel RM, Verbeek AL, den Heeten GJ, Adang EM, et al. Cost-effectiveness of non-invasive assessment in the Dutch breast cancer screening program versus usual care: A randomized controlled trial. Canadian Journal of Public Health. 2014;105(5):e342-e7.

290. Domeyer PJ, Sergentanis TN, Zagouri F, Zografos GC. Health-related Quality of Life in Vacuum-Assisted Breast Biopsy: short-term effects, long-term effects and predictors. Health and Quality of Life Outcomes. 2010;8(1):11.

291. Grann VR, Patel P, Bharthuar A, Jacobson JS, Warner E, Anderson K, et al. Breast cancer-related preferences among women with and without BRCA mutations. Breast Cancer Res Treat. 2010;119(1):177-84.

292. National Institute for Health and Disability Insurance (NIHDI). Monitoring Of Reimbursement Significant Expenses –M.O.R.S.E., 2024 data. 2026

293. Shi J, Li J, Gao Y, Chen W, Zhao L, Li N, et al. The screening value of mammography for breast cancer: an overview of 28 systematic reviews with evidence mapping. J Cancer Res Clin Oncol. 2025;151(3):102.

294. Hoffmann TC, Del Mar C. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA Intern Med. 2015;175(2):274-86.

295. Ghanouni A, Meisel SF, Renzi C, Wardle J, Waller J. Survey of public definitions of the term 'overdiagnosis' in the UK. BMJ Open. 2016;6(4):e010723.

296. Moynihan R, Nickel B, Hersch J, Beller E, Doust J, Compton S, et al. Public Opinions about Overdiagnosis: A National Community Survey. PLoS One. 2015;10(5):e0125165.

297. Nowak SA, Parker AM. Social network effects of nonlifesaving early-stage breast cancer detection on mammography rates. Am J Public Health. 2014;104(12):2439-44.

298. Plichta JK, Rushing CN, Lewis HC, Rooney MM, Blazer DG, Thomas SM, et al. Implications of missing data on reported breast cancer mortality. Breast Cancer Res Treat. 2023;197(1):177-87.

299. Mali W, van der Graaf Y. [Cancer screening: blessing or curse?]. Ned Tijdschr Geneeskd. 2026;170.

300. Esserman LJ, Study W, Athena I. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. NPJ Breast Cancer. 2017;3:34.

301. Brawley OW, Ramalingam R. Understanding the Varying Biological Behaviors of Breast and Other Types of Cancer to Avoid Overdiagnosis. Ann Intern Med. 2023;176(9):1273-4.

302. Lang K, Josefsson V, Larsson AM, Larsson S, Hogberg C, Sartor H, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol. 2023;24(8):936-44.

303. Bihrmann K, Jensen A, Olsen AH, Njor S, Schwartz W, Vejborg I, et al. Performance of systematic and non-systematic ('opportunistic') screening mammography: a comparative study from Denmark. J Med Screen. 2008;15(1):23-6.

304. Neyt M. Value-Based Pricing: Do Not Throw Away the Baby with the Bath Water. Pharmacoeconomics. 2018;36(1):1-3.

305. Health Information and Quality Authority (HIQA – Ireland). Draft National Guidelines for the Economic Evaluation of Health Technologies in Ireland 21 October 2024.

306. Health Quality Ontario (HQO). Women's Experiences of Inaccurate Breast Cancer Screening Results: A Systematic Review and Qualitative Meta-synthesis. Ont Health Technol Assess Ser. 2016;16(16):1-22.

307. Grimault H. Dépistage du cancer du sein chez la femme : état des lieux récent sur le dépistage standard et discussion sur un dépistage basé sur le risque individuel et ses limites éthiques; 2022. Available from: https://dumas.ccsd.cnrs.fr/dumas-03652329

308. Almohammed HI. A Systematic Review to Evaluate the Barriers to Breast Cancer Screening in Women with Disability. J Clin Med. 2024;13(11).

309. Moutel G, Darquy S, Jullian O, Duchange N. [Ethics of organized cancer screening in France]. Sante Publique. 2019;S2(HS2):67-74.

310. Elton L. Non-maleficence and the ethics of consent to cancer screening. J Med Ethics. 2020;47(7):510-3.

311. Hofmann B. Informing about mammographic screening: Ethical challenges and suggested solutions. Bioethics. 2020;34(5):483-92.

312. Carter SM. Ethical aspects of cancer screening. In: Proceedings of Cancer Forum; 2016; p. 105-9.

313. Ferretti G. Ethical issues in mammography screening programs and women's decision-making. Milan: European School of Molecular Medicine (SEMM) and University of Milan; 2018.

314. Gram EG, Siersma V, Brodersen JB. Long-term psychosocial consequences of false-positive screening mammography: a cohort study with follow-up of 12-14 years in Denmark. BMJ Open. 2023;13(4):e072188.

315. Ingman WV, Richards B, Street JM, Carter D, Rickard M, Stone J, et al. Breast Density Notification: An Australian Perspective. J Clin Med. 2020;9(3).

316. Moey SF, Sowtali SN, Mohamad Ismail MF, Hashi AA, Mohd Azharuddin NS, Che Mohamed N. Cultural, Religious and Socio-Ethical Misconceptions among Muslim Women towards Breast Cancer Screening: A Systematic Review. Asian Pac J Cancer Prev. 2022;23(12):3971-82.

317. O'Brien E, Ludwigson A, Vemuru S, Higgins M, Hampanda K, Adams M, et al. Interpretation of immediately released health information: Informing patient medical education in breast oncology. Am J Surg. 2025;239:115853.

318. Ritchie D, Van Hal G, Van den Broucke S. Factors affecting intention to screen after being informed of benefits and harms of breast cancer screening: a study in 5 European countries in 2021. Arch Public Health. 2022;80(1):143.

319. Schifferdecker KE, Tosteson ANA, Kaplan C, Kerlikowske K, Buist DSM, Henderson LM, et al. Knowledge and Perception of Breast Density, Screening Mammography, and Supplemental Screening: in Search of "Informed". J Gen Intern Med. 2020;35(6):1654-60.

320. Seaman K, Dzidic PL, Castell E, Saunders C, Breen LJ. A Systematic Review of Women's Knowledge of Screening Mammography. Breast. 2018;42:81-93.

321. Seaman K, Dzidic PL, Castell E, Saunders C, Breen LJ. Subject positions in screening mammography and implications for informed choice. Psychol Health. 2021;36(4):478-95.

322. Sheehy EM, Lehane E, Quinn E, Livingstone V, Redmond HP, Corrigan MA. Information Needs of Patients With Breast Cancer at Years One, Three, and Five After Diagnosis. Clin Breast Cancer. 2018;18(6):e1269-e75.

323. Varghese F, Wong J. Breast Cancer in the Elderly. Surg Clin North Am. 2018;98(4):819-33.

324. Bartholomew T, Colleoni M, Schmidt H. Financial incentives for breast cancer screening undermine informed choice. BMJ. 2022;376:e065726.

325. Rogers WA, Entwistle VA, Carter SM. Risk, Overdiagnosis and Ethical Justifications. Health Care Anal. 2019;27(4):231-48.

326. Beauchamp T, Childress J. Principles of Biomedical Ethics: Marking Its Fortieth Anniversary. Am J Bioeth. 2019;19(11):9-12.

327. Australian Government - Department of Health. Population based screening framework. 2018 Available from: https://www.health.gov.au/sites/default/files/documents/2019/09/population-based-screening-framework_0.pdf

328. Shieh Y, Eklund M, Sawaya GF, Black WC, Kramer BS, Esserman LJ. Population-based screening for cancer: hope and hype. Nat Rev Clin Oncol. 2016;13(9):550-65.

329. Oke JL, Brown SJ, Senger C, Welch HG. Deceptive measures of progress in the NHS long-term plan for cancer: case-based vs. population-based measures. Br J Cancer. 2023;129(1):3-7.

330. Ropers FG, Barratt A, Wilt TJ, Nicholls SG, Taylor-Phillips S, Kramer BS, et al. Health screening needs independent regular re-evaluation. Bmj. 2021;374:n2049.

331. Day NE, Williams DR, Khaw KT. Breast cancer screening programmes: the development of a monitoring and evaluation system. Br J Cancer. 1989;59(6):954-8.

332. Bull A, Mountney L, Sanderson H. Stage distribution of breast cancer: a basis for the evaluation of breast screening programmes. Br J Radiol. 1991;64(762):516-9.

333. Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet. 2000;355(9198):129-34.

334. Jørgensen KJ, Keen JD, Zahl PH, Gøtzsche PC. The Two-County breast screening trial cannot provide a reliable estimate of the effect of breast cancer screening. Radiology. 2012;262(2):729-30; author reply 30-1.

335. Autier P, Boniol M. Mammography screening: A major issue in medicine. Eur J Cancer. 2018;90:34-62.

336. Linder JM, Schiska AD. Progress in diagnosis of breast cancer: Advances in radiology technology. Asia Pac J Oncol Nurs. 2015;2(3):186-91.

337. Bakker MF, de Lange SV, Pijnappel RM, Mann RM, Peeters PHM, Monninkhof EM, et al. Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. N Engl J Med. 2019;381(22):2091-102.

338. Skaane P, Bandos AI, Niklason LT, Sebuødegård S, Østerås BH, Gullien R, et al. Digital Mammography versus Digital Mammography Plus Tomosynthesis in Breast Cancer Screening: The Oslo Tomosynthesis Screening Trial. Radiology. 2019;291(1):23-30.

339. Coffey K, Jochelson MS. Contrast-enhanced mammography in breast cancer screening. Eur J Radiol. 2022;156:110513.

340. Salim M, Liu Y, Sorkhei M, Ntoula D, Foukakis T, Fredriksson I, et al. AI-based selection of individuals for supplemental MRI in population-based breast cancer screening: the randomized ScreenTrustMRI trial. Nat Med. 2024;30(9):2623-30.

341. Obenauer S, Luftner-Nagel S, von Heyden D, Munzel U, Baum F, Grabbe E. Screen film vs full-field digital mammography: image quality, detectability and characterization of lesions. Eur Radiol. 2002;12(7):1697-702.

342. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005;353(17):1773-83.

343. de Munck L, de Bock GH, Otter R, Reiding D, Broeders MJ, Willemse PH, et al. Digital vs screen-film mammography in population-based breast cancer screening: performance indicators and tumour characteristics of screen-detected and interval cancers. Br J Cancer. 2016;115(5):517-24.

344. Farber R, Marinovich ML, Pinna A, Houssami N, McGeechan K, Barratt A, et al. Systematic review and meta-analysis of prognostic characteristics for breast cancers in populations with digital vs film mammography indicate the transition may have increased both early detection and overdiagnosis. J Clin Epidemiol. 2024;171:111339.

345. Hofvind S, Holen Å S, Aase HS, Houssami N, Sebuødegård S, Moger TA, et al. Two-view digital breast tomosynthesis versus digital mammography in a population-based breast cancer screening programme (To-Be): a randomised, controlled trial. Lancet Oncol. 2019;20(6):795-805.

346. Plevritis SK, Munoz D, Kurian AW, Stout NK, Alagoz O, Near AM, et al. Association of Screening and Treatment With Breast Cancer Mortality by Molecular Subtype in US Women, 2000-2012. Jama. 2018;319(2):154-64.

347. Autier P, Boniol M, La Vecchia C, Vatten L, Gavin A, Héry C, et al. Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. Bmj. 2010;341:c3620.

348. de Munck L, Eijkelboom AH, Otten JDM, Broeders MJM, Siesling S. Method of primary breast cancer detection and the disease-free interval, adjusting for lead time. J Natl Cancer Inst. 2024;116(3):370-8.

349. Connor RJ, Prorok PC, Weed DL. The case-control design and the assessment of the efficacy of cancer screening. J Clin Epidemiol. 1991;44(11):1215-21.

350. Diaz A, Kang J, Moore SP, Baade P, Langbecker D, Condon JR, et al. Association between comorbidity and participation in breast and cervical cancer screening: A systematic review and meta-analysis. Cancer Epidemiol. 2017;47:7-19.

351. McWilliams L, Groves S, Howell SJ, French DP. The Impact of Morbidity and Disability on Attendance at Organized Breast Cancer-Screening Programs: A Systematic Review and Meta-Analysis. Cancer Epidemiol Biomarkers Prev. 2022;31(7):1275-83.

352. Mottram R, Knerr WL, Gallacher D, Fraser H, Al-Khudairy L, Ayorinde A, et al. Factors associated with attendance at screening for breast cancer: a systematic review and meta-analysis. BMJ Open. 2021;11(11):e046660.

353. International Agency for Research on Cancer. Breast Cancer Screening. Lyon: 2002.

354. Bakker MF, de Lange SV, Pijnappel RM, Mann RM, Peeters PHM, Monninkhof EM, et al. Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. N Engl J Med. 2019;381(22):2091-102.

355. Contrast-enhanced Mammography and Abbreviated MRI as Supplemental Screening Modalities for Women Aged 50-75 With Extremely Dense Breast Tissue: the DENSE-2 Trial. In: University Medical Center G, Ziekenhuisgroep T, Radboud University Medical C, Rijnstate H, Zuyderland Medical C, Maastricht University Medical C, et al., editors.; 2024.

356. Esserman LJ, Fiscalini AS, Naeim A, van't Veer LJ, Kaster A, Scheuner MT, et al. Risk-Based vs Annual Breast Cancer Screening: The WISDOM Randomized Clinical Trial. JAMA. 2025.

357. Baxter NN, Phillips KA. Unlocking the Potential of Risk-Based Screening for Breast Cancer: From Detection to Prevention. Jama. 2026;335(9):758-60.

358. Zahl PH, Jørgensen KJ, Gøtzsche PC. Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. Br J Cancer. 2013;109(7):2014-9.

359. Zahl PH, Jørgensen KJ, Gøtzsche PC. Lead-time models should not be used to estimate overdiagnosis in cancer screening. J Gen Intern Med. 2014;29(9):1283-6.

360. Lee CI, Etzioni R. Missteps in Current Estimates of Cancer Overdiagnosis. Acad Radiol. 2017;24(2):226-9.

361. GLOBOCAN. Cancer Today: Incidence & Mortality. In: International Agency for Research on Cancer, editor.; 2022.

362. The Cancer Intervention and Surveillance Modeling Network (CISNET) Breast Cancer Working Group. Breast Cancer Screening With Mammography: An Updated Decision Analysis for the U.S. Preventive Services Task Force. 2023. (AHRQ Publication No. 23-05303-EF-2) Available from: https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://uspreventiveservicestaskforce.org/home/getfilebytoken/uRwAnYAnc4HCNY3j3h5v_z&ved=2ahUKEwjntuiBqdiGAxXiV6QEHYgQAhkQFnoECBgQAQ&usg=AOvVaw0HHyn4KwxUkx-jaSsRXLUB

363. Coldman A, Phillips N, Wilson C, Decker K, Chiarelli AM, Brisson J, et al. Pan-Canadian study of mammography screening and mortality from breast cancer. J Natl Cancer Inst. 2014;106(11).

364. Engmann NJ, Golmakani MK, Miglioretti DL, Sprague BL, Kerlikowske K, Breast Cancer Surveillance C. Population-Attributable Risk Proportion of Clinical Risk Factors for Breast Cancer. JAMA Oncol. 2017;3(9):1228-36.

365. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1[updated September 2008] ed. The Cochrane Collaboration, editor.; 2008.

366. Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmö Mammographic Screening Program. J Natl Cancer Inst Monogr. 1997(22):63-7.

367. Warren RML, Young JR, McLean L, Lyons K, Wilson ARM, Evans A, et al. Radiology review of the UKCCCR Breast Screening Frequency Trial: potential improvements in sensitivity and lead time of radiological signs. Clin Radiol. 2003;58(2):128-32.

368. Ishida T, Suzuki A, Kawai M, Narikawa Y, Saito H, Yamamoto S, et al. A randomized controlled trial to verify the efficacy of the use of ultrasonography in breast cancer screening aged 40-49 (J-START): 76 196 women registered. Jpn J Clin Oncol. 2014;44(2):134-40.

369. Moss S, Waller M, Anderson TJ, Cuckle H, Trial Management G. Randomised controlled trial of mammographic screening in women from age 40: predicted mortality based on surrogate outcome measures. Br J Cancer. 2005;92(5):955-60.

370. Bjurstam N, Björneld L, Duffy SW, Smith TC, Cahlin E, Eriksson O, et al. The Gothenburg breast screening trial: first results on mortality, incidence, and mode of detection for women ages 39-49 years at randomization. Cancer. 1997;80(11):2091-9.