Tampere University

DEEPENDRA SINGH

Symptoms and Risk of Breast Cancer

A Population-Based Cohort Study within the Finnish Mammography Screening Programme

Tampere University Dissertations 58

Tampere University Dissertations 58

DEEPENDRA SINGH

Symptoms and Risk of Breast Cancer

A Population-Based Cohort Study within the Finnish Mammography Screening Programme

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty Council of the Faculty of Social Sciences
of Tampere University,
for public discussion in the Yellow Hall
of the Arvo building, Arvo Ylpönkatu 34, Tampere,
on May 31, 2019 at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Social Sciences
Cancer Society of Finland, Finnish Cancer Registry
Finland

Responsible supervisor and Custos	Professor Janne Pitkäniemi Tampere University Finland	
Supervisors	Docent Ahti Anttila Tampere University Finland	Docent Nea Malila University of Helsinki Finland
Pre-examiners	Docent Tarja Rissanen University of Oulu Finland	Professor Esa Läärä University of Oulu Finland
Opponent	Professor Solveig Hofvind Oslo Metropolitan University Norway	

Cover design: Roihu Inc.

ISBN 978-952-03-1081-3 (print)
ISBN 978-952-03-1082-0 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-1082-0

PunaMusta Oy – Yliopistopaino
Tampere 2019

This thesis is dedicated to my wife and our soon-to-be-born daughter, and to my father, my mother and my mentor Ahti

ACKNOWLEDGEMENTS

This study was carried out in the Mass Screening Registry of the Finnish Cancer Registry and in the Epidemiology group, Department of Health Sciences, Tampere University, between 2014 and 2019. The work would not have been possible without the available data collected by several registries in Finland, so a big thank you to all of them.

Writing my thesis would not have been possible without the support and insight provided by a large number of people. Among other, I am indebted to:

Ahti Anttila, primary supervisor, for the introduction to and for helping me gradually integrate into the field of breast cancer screening. He has been an excellent guide in research work and also supported me as a mentor. I first met him during my master's studies while working as an intern at the International Agency for Research on Cancer. Since then, he has been guiding me untiringly and selflessly. He has been generous with his time, trust and broad experience and understanding of the complexities of screening. He supported me exceptionally well when applying for Cancer Foundation Grants and for other grants and scholarships. I will never be able to pay him back for all the support he has offered.

Nea Malila, co-supervisor, for always being there to support all my work and to devote her time and presence to the question at hand. Her sharp intellect and problem-solving skills are truly inspirational. I thank her for all the discussions that deepened my understanding of both epidemiology and mammography screening programme as a whole. She has ensured an excellent working environment at the Finnish Cancer Registry.

I am grateful to Janne Pitkäniemi for his statistical expertise, encouragement and constant optimism, which have carried me through the study. He has been an important scientific collaborator throughout. I am thankful to Joonas Miettinen for all the formal and informal discussions that have deepened my understanding of statistics and R programming. I would also like to thank Stephen Duffy of Queen Mary University of London for providing me the opportunity of a scientific visit for two weeks at his institute and for his valuable comments on the third article of my PhD thesis. I am thankful to one of the co-authors, Arun Pokhrel for his valuable

contribution in the first study. My warm thanks go to the entire staff of the Finnish Cancer Registry for providing a supportive working environment.

I would like to acknowledge my follow-up group, Anssi Auvinen and Rengaswamy Sankaranarayanan, for their valuable comments and suggestions. Also, Christopher Ryan, who revised the language of this thesis. His fast and thorough work is greatly appreciated.

I would like to thank the reviewers of this thesis, Tarja Rissanen and Esa Läärä, for their knowledgeable and constructive criticism and also for a swift review. I learned a lot on the practical aspect of mammography screening process from Tarja. Esa's comments on almost every part of the thesis summary, including study aims, Nordic surnames, the statistical word 'significant' and study results really helped me refine the thesis.

Of course, I owe a lot to the Department of Health Sciences. In particular, I wish to thank:
Anssi Auvinen, Clas-Håkan Nygård, and Pekka Nuorti for many excellent and inspiring lessons in epidemiology and public health. I would like to acknowledge other teaching and administrative staffs members, especially (in alphabetical order) Catarina Ståhle-Nieminen, Kirsi Lumme Sandt, Leena Nikkari, Susanna Lehtinen-Jacks, Tarja Kinnunen, Tiina Kangasluoma and Sinikka Määttä for their continuous support in my studies. I would like to again thank the department for providing me good research facilities. I am also thankful to the entire IPPE cohort of 2013 and my PhD colleagues at the university for offering invaluable moral support and fun during the course work.

Family and friends
First, the love of my life, Indira Adhikari 'Kalu': Your love means everything, and I am so, so grateful for having you in my life. *You are my everything.* I cannot believe how lucky I am to have you. You encouraged me during every moment of my studies and, especially during projects such as this.

My heartfelt thanks go to my family: my mom, my dad, my younger brother and his family, my uncle and his family, my mother-in-law, my father-in-law, my sister-in-law and, my brothers-in-law for their unfailing love and support. My dad always had a dream to see me complete a PhD, so thank you again for your encouragement and support.

I am thankful to Subas Neupane for his guidance and encouragement during my studies and in informal gatherings. I would like to thank my near and dear Nepalese friends living in Tampere and abroad for your support inside and outside of the university. This group includes at least (in alphabetical order) Binita Aryal Neupane,

Biswa Upreti, Durga Neupane Upreti, Kalpana Bastola, Kishor Hadkhale, Kumud Koirala, Luna Rayamajhi, Manita Aryal, Nirmala Shrestha, Pabitra Basnyat, Prakash K.C., Raju Shrestha, Subash Jung Basnet, and Suresh Gnawali.

The financial support of the Cancer Foundation Grant of Finland, the Finnish Cancer Registry, the Department of Health Sciences of Tampere University, Tampere University Foundation Grant, and the city of Tampere is gratefully acknowledged.

Tampere, April 2019

Deependra Singh

ABSTRACT

The breast is the leading cancer site and the first or second cause of malignancy-associated death in women in Western societies. The reason for rapid growth of this disease is complex but reflects the socio-economic development, ageing and growth of the population. Substantial evidence on the potential risk factors has been developed, and mammography screening has been identified as an important public health measure for early detection of the disease. An important risk factor, 'breast symptoms' reported by women at screening visits are not always evaluated adequately, or information regarding the symptoms are left out by many screening programmes. This thesis seek to generate novel evidence on possible breast cancer outcomes in these women.

This thesis is composed of four sub-studies that used individual level data available from various registries in Finland. In the first part of the thesis, we discuss how we measured the mammography-screening programme performance indicators in relation to breast symptoms. We found a higher recall proportion and higher cancer detection rate in women with symptoms as compared to those without symptoms. The sensitivity and positive predictive value (PPV) of the lump was higher than that of retraction and nipple discharge. Mammography test sensitivity and PPVs were higher for visits with a lump or retraction but not for visits with nipple discharge.

The second cohort study estimated the cumulative probability of false-positive test and false-positive referrals and compared visits with and without symptoms. The cumulative risk of a false-positive test after 10 screening visits was 45% for a lump, retraction 25% and nipple discharge 35% as compared to about 18% in visits without symptoms. Likewise, we found higher cumulative risk of a false-positive referral in all symptom types as compared to respective asymptomatic visits.

In the third phase, we assessed the risk of screen-detected and interval breast cancers in relation to breast symptoms. The age-adjusted risk of screen-detected invasive breast cancer in visits with a lump was 8 times higher, retraction 2.4 times higher and nipple discharge 1.7 times higher than the respective visits without symptoms. The risk was 22 times higher if both a lump and retraction were reported in the same visit. Interval cancer risk was assessed by matching symptomatic visits

with asymptomatic visits by background variables. We found a strong positive association between symptoms and the risk of interval breast cancer. The cumulative incidence of interval cancers increased rapidly after a screening visit with a lump.

Finally, we assessed the risk of breast cancer mortality and all-cause mortality by symptom types and compared them with asymptomatic visits. The breast cancer mortality rate was very high within the first five years after screening visits with a lump or retraction. The risk of breast cancer mortality was elevated for women with a lump or retraction. We found an increased all-cause mortality for all symptom types as compared to the respective asymptomatic visit.

Our findings can be applied to various steps in the monitoring of programme performance up to the evaluation of outcomes and health equities. Our study showed the importance of collecting and analysing information on breast symptoms. It is important to continuously improve the awareness of breast symptoms in women, with the advice being to seek care even prior to invitations. High-quality and clinically appropriate services are important for women presenting with symptoms at screening mammography. Our findings indicate the need to improve the guidelines for screening and clinical services for women with symptoms. Further assessment is needed more frequently, including appropriate biopsy. Improved systematic surveillance or follow-up of symptomatic women who did not undergo further assessment is crucial.

TIIVISTELMÄ

Rintasyöpä on naisten tavallisin syöpätyyppi ja yleisin tai toiseksi yleisin syövästä johtuva kuolinsyy länsimaissa. Taudin nopeaan kasvuun on monia syitä, jotka heijastavat sosioekonomista kehitystä, ikääntymistä ja väestönkasvua. Mahdollisista riskitekijöistä on saatavilla paljon tutkimustietoa, ja väestöpohjaisen mammografiaseulonnan on todettu olevan tärkeä keino taudin havaitsemiseksi varhaisessa vaiheessa. Tärkeitä riskitekijöitä ovat naisten seulontakäynneillä ilmoittamat ”rintaoireet”, joita ei kuitenkaan aina arvioida riittävän hyvin tai joita koskevia tietoja ei laisinkaan kerätä seulontaohjelmassa. Tämä väitöstutkimus pyrkii esittämään uutta näyttöä näiden naisten rintasyöpiä koskevista tuloksista.

Työ koostuu neljästä osatutkimuksesta, joissa on hyödynnetty erilaisia rekisteriperäisiä yksilötasoisia tietoja. Tutkimuksen ensimmäisessä osassa selvitimme mammografiaohjelman toimivuutta kuvaavien mittareiden suhdetta rintaoireisiin. Totesimme, että naiset, joilla oli oireita, heidät kutsuttiin useammin jatkotutkimuksiin ja heillä havaittiin useammin syöpä kuin oireettomilla naisilla. Kyhmyjen herkkyys ja positiivinen ennustearvo (PPV) olivat suurempia kuin vetäymien ja nännieritteiden. Mammografiatutkimuksen herkkyys ja PPV-arvot olivat suurempia, kun käyntiin liittyi kyhmy tai vetäymä, mutta ei silloin kun kyseessä oli nännierite.

Toisessa kohorttitutkimuksessa arvioitiin väärän positiivisen tuloksen ja väärän positiivisen lähetteen kumulatiivista todennäköisyyttä ja verrattiin oireisia ja oireettomia käyntejä. Naisilla, joilla oli todettu kyhmy vähintään kerran, väärän positiivisen testituloksen kumulatiivinen riski kymmenen seulontakäynnin jälkeen oli 45 %, ja vastaava arvio oli 25 % vetäymälle ja 35 % nännieritteelle. Vastaava luku oireettomille käynneille oli noin 18 %. Tämän perusteella totesimme, että väärän positiivisen lähetteen kumulatiivinen riski oli suurempi kaikissa oiretyypeissä verrattuna vastaavaan oireettomaan käyntiin.

Kolmannessa vaiheessa arvioimme seulonnassa havaittujen ja seulontakertojen välillä havaittujen rintasyöpien riskiä suhteessa rintaoireisiin. Seulonnassa havaitun rintasyövän ikävakioitu riski käynneillä, joihin liittyi kyhmy, vetäymä tai nännierite oli vastaavasti 8 kertaa, 2,4 kertaa ja 1,7 kertaa suurempi kuin käynneillä, joihin ei liittynyt oireita. Riski oli 22 kertaa suurempi, jos samalla käynnillä todettiin sekä kyhmy että vetäymä. Seulontakertojen välillä ilmenevien syöpien riskiä arvioitiin vertaamalla

oireisia ja oireettomia käyntejä taustamuuttujien avulla. Oireiden ja seulontakertojen välisen rintasyövän riskin välillä oli vahva positiivinen yhteys. Välisyöpien kumulatiivinen ilmaantuvuus lisääntyi seulonnan jälkeen nopeasti, kun seulontakäynnnilä oli todettu kyhmy.

Lopuksi arvioimme rintasyöpäkuolleisuuden ja kaikista syistä johtuvan kuolleisuuden riskiä oiretyyppien mukaan ja vertasimme niitä oireettomiin käynteihin. Rintasyövästä johtuva kuolleisuusaste oli hyvin korkea viiden vuoden sisällä seulontakäynnistä, johon liittyi kyhmy tai vetäymä. Rintasyövästä aiheutuvan kuoleman riski oli korkeampi, kun käyntiin liittyi kyhmy tai vetäymä, mutta ei silloin kun käyntiin liittyi erite nännistä. Kaikista syistä aiheutuva kuolleisuus pysyi suurempana kaikissa oiretyypeissä verrattuna vastaavaan oireettomaan käyntiin.

Havaintojamme voidaan soveltaa eri vaiheisiin ohjelman toimivuuden seurannasta aina tulosten ja terveyden samanarvoisuuden arviointiin saakka. Tutkimuksemme osoitti, kuinka tärkeää on kerätä ja analysoida tietoa rintaoireista. On tärkeää lisätä jatkuvasti tietoisuutta naisten rintaoireista ja neuvoa hakeutumaan hoitoon jo ennen kutsuja. Laadukkaat ja kliinisesti asianmukaiset palvelut ovat tärkeitä naisille, joiden mammografiaseulontaan liittyy oireita. Tuloksemme osoittavat, että on tarpeen kehittää seulontaa ja kliinisiä palveluita koskevia ohjeita oireisille naisille. Jatkotutkimuksia, mm. soveltuvia kudosnäytteitä, tarvitaan nykyistä useammin. On myös erittäin tärkeää seurata seulontaohjelman yhteydessä systemaattisesti rintaoireisia naisia, mikäli heille ei tehty jatkotutkimuksia.

CONTENTS

List of Figures

List of Tables

ABBREVIATIONS

ASR	Age-standardized rate
BMI	Body mass index
BTW	Breast Test Wales
CPP	Cancer Patient Pathway
CI	Confidence interval
CPR	Central Population Registry
DCIS	Ductal Carcinoma In Situ
EU	European Union
FCR	Finnish Cancer Registry
FNBCSP	Finnish National Breast Cancer Screening Programme
GLM	Generalized Linear Model
GP	General Practitioner
HR	Hazard ratio
HRT	Hormone Replacement Therapy
IARC	International Agency for Research on Cancer
IBM	Incidence-based mortality
IC	Interval cancer
MIR	Mortality to incidence ratio
MSR	Mass Screening Registry
MST	Mean Sojourn Time
NHS	National Health Services
NHSBSP	National Health Service Breast Screening Programme
NPV	Negative Predictive Value
OR	Odds ratio
PPV	Positive Predictive Value
UICC	Union for International Cancer Control
UK	United Kingdom
USA	United States of America
THL	National Institute for Health and Welfare
TNM	Tumour Node Metastasis

WHO World Health Organization

ORIGINAL PUBLICATIONS

I. Singh D, Malila N, Pokhrel A, Anttila A (2015): Association of symptoms and breast cancer in population-based mammography screening in Finland. Int J Cancer, 136, E630-E637. https://doi.org/10.1002/ijc.29170

II. Singh D, Pitkäniemi J, Malila N, Anttila A (2016): Cumulative risk of false positive test in relation to breast symptoms in mammography screening: a historical prospective cohort study. Breast Cancer Res Treat 159 (2), 305-313. https://doi.org/10.1007/s10549-016-3931-8

III. Singh D, Miettinen J, Duffy SW, Malila N, Pitkäniemi J, Anttila A (2018): Association of symptoms and interval breast cancers in the mammography-screening programme: population-based matched cohort study. Br J Cancer 119 (11), 1428-1435. https://doi.org/10.1038/s41416-018-0308-2

IV. Singh D, Malila N, Pitkäniemi J, Anttila A (2019): Cancer incidence and mortality patterns in women with breast symptoms in the mammography screening programme: a matched cohort analysis. Int J Cancer 144, 2928-2935. https://doi.org/10.1002/ijc.32035

1 INTRODUCTION

In the 21st-century, cancer is expected to rank as the leading cause of death and the single most important barrier for increasing life expectancy in every country of the world. Breast cancer is the most frequently diagnosed cancer (154 out of 185 countries) and the most common cause of malignancy-associated death in women in over 100 countries (Bray et al., 2018). In Finland, the incidence of this disease continued to rise over decades, while mortality has shown a slight decrease in recent years. Considering the high prevalence of this disease, efforts to reduce deaths are globally an urgent public health priority.

A substantial amount of evidence has been accumulated on the natural history and burden of the disease. In recent decades, efforts have been put into both primary and secondary preventive measures by identifying the potential risk factors and with organized breast cancer screening programmes. Evidence from randomized-controlled trials and observation studies have shown to reduce breast cancer mortality by about 40% in those who underwent mammography screening (Lauby-Secretan et al., 2015). On the other hand, advancement in modern chemotherapy and adjuvant treatment have improved the survival of women with breast cancer. A lot of these improvements are not uniform across countries or geographical areas, even within the same country. Existing socio-economic inequalities, several individual risk factors, evolvement of new risk factors due to so-called westernization of lifestyle, and the ageing and growth of the population have all led to the prominent differences in the rates as observed in different countries.

Countries with existing screening programmes that act as a part of routine health policy invite all women in the targeted age, irrespective of individual history. However, based on the evidence generated in recent decades through epidemiological and biomedical research, several studies have proposed personalized screening modalities for women who are at a higher risk of a breast cancer diagnosis.

A noticeable proportion (about 3% in the Finnish programme) of women present with clinically significant symptoms at the screening visit. A majority of, if not all service screening programmes that offer mammography services to the whole target population refer women presenting with symptoms either directly to breast clinics

or to general practitioners for further consultation. A proportion of women are left out by the programme and, hence, information about breast symptoms is not collected or analysed. One important reason is that many programmes lack proper guidelines on screening and clinical services for women presenting with breast symptoms. As well, EU guidelines on breast cancer screening do not have specific recommendations for referral procedures for these women (Perry et al., 2006). Most of the women in the Finnish programme presenting with symptoms such as a lump, retraction or nipple discharge (also identified as 'urgent referrals' in the UK screening guidelines) (Mansel et al., 1999) likely do not undergo further assessment and are systematically followed-up in the programme up to cancer diagnosis or death. This presents the opportunity for research to generate novel evidence on possible cancer outcomes in these women.

The objective of this study was to identify the importance of analysing breast symptoms in the Finnish mammography screening programme, measured in terms of programme performance indicators and breast cancer outcomes up to mortality outcomes in those women. This helps create an evidence-based guideline and develop a potential screening and clinical service strategies that might improve the potential screening outcomes for women presenting with breast symptoms at the screening visit.

2 LITERATURE REVIEW

2.1 Epidemiology of breast cancer

2.1.1 Breast cancer burden

Breast cancer is the most frequently diagnosed tumour disease in women in both developed and less-developed regions with an age-standardized rate (ASR) of 75.4 per 100000 and 31.3 per 100000 women respectively (Globocan, 2012; Key et al., 2001). Incidence rates are higher in Europe and Northern America than in other regions but are increasing. According to the American Cancer Society, one in eight women will develop breast cancer in her lifetime (DeSantis et al., 2014). More than one million new breast cancer cases are diagnosed each year (Ferlay et al., 2015; Forman et al., 2014) with greater variation in the incidence between different regions (Figure 1, below) (Globocan, 2012). In the Nordic countries from 2000 – 2015, the incidence ASR has increased from 79 to 86.8 per 100,000 women, while the mortality ASR has decreased from 18.3 to 12.2 per 100,000 women respectively (Engholm et al., 2018). Figure 2 shows the time-trend of breast cancer incidence and mortality in Finland from 1953-2016. In Finland, the age-standardized incidence increased from 75 to 94 per 100,000 women during 1998 – 2015, while the mortality rates declined from 15 to 13 per 100,000 women in the same period (Finnish Cancer Registry, 2018c). In 2016, 4961 breast cancers incidents and 888 deaths were registered at the Finnish Cancer Registry. The five-year relative survival of breast cancer patients was 91% (95% CI 90.5-92.0).

The worldwide increase in life-expectancy and continuing demographic and epidemiological transition, particularly in less developed regions, signal the ever-increasing burden of the disease over the next decades (Ferlay et al., 2015). While the incidence of disease is high in Western countries, the mortality-to-incidence ratio (MIR) is high in low-income countries; the MIR ratio of 0.69 in Africa as compared to 0.19 in North America, for example (Forouzanfar et al., 2011). The westernization of the developing world is the main force behind the increase in incidence rates, which encompasses adoption of less desirable habits including lifestyle and

behavioural factors (P. Porter, 2008), whereas in the developed countries, enhancement in early detection and treatment of breast cancer have resulted in a decrease in the mortality rates and continual increase in the five-year survival of the disease. However, facilities are uncommon in low-and middle-income countries, leading to late-stage presentation of disease, and resulting in higher case fatality rates and a lower rate of survival.

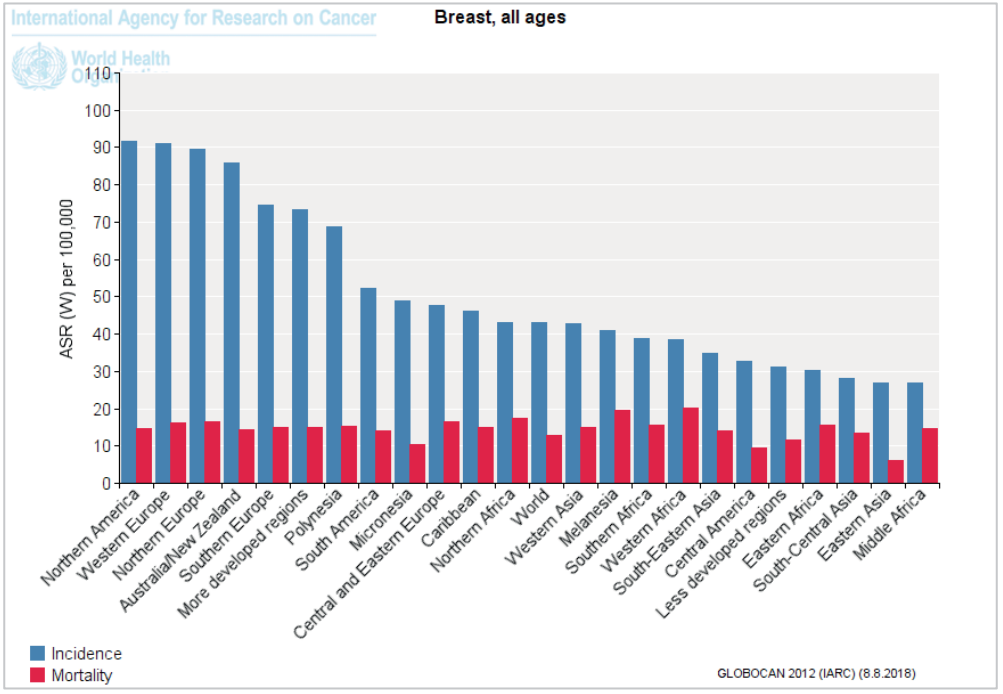

Figure 1. Estimated age-standardized rates (ASR, World) rates of breast cancer per 100,000 women

Source: GLOBOCAN 2012 (IARC), Section of Cancer Surveillance (Accessed 08.08.2018) http://globocan.iarc.fr/Pages/bar_site_sel.aspx

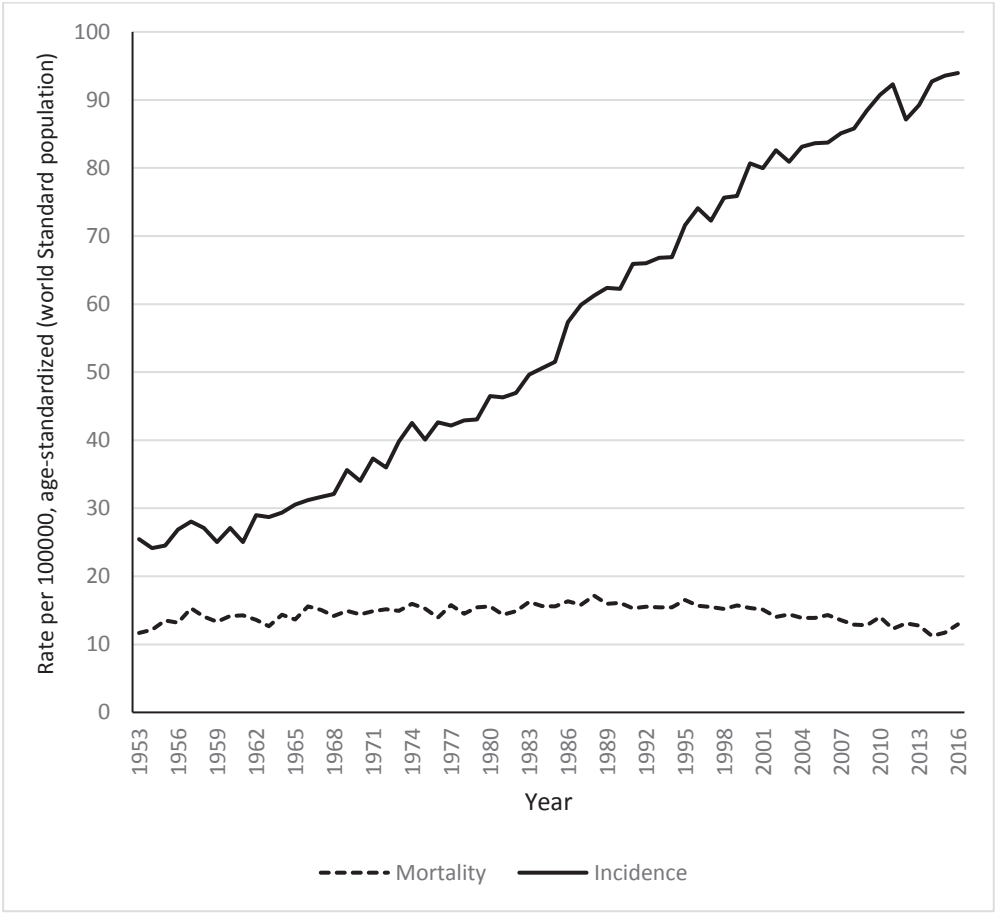

Figure 2. Time trends of breast cancer incidence and mortality in Finland in 1953-2016, age-standardized (World) rates per 100000 women

Source: Finnish Cancer Registry, Cancer Statistics (Accessed 04.01.2018) https://cancerregistry.fi/statistics/cancer-statistics/

2.1.2 Aetiology of breast cancer

Breast cancer is a complex, highly heterogeneous disease at both the molecular and clinical level with multifactorial causations and multiple sub-types (Perou et al., 2000; Sorlie et al., 2001). The disease has more than one natural history, thus cancer progression stages from atypia to carcinoma in situ, invasive cancer and metastasis - may not hold true for all cases (Buerger et al., 1999; Buerger et al., 2001). A common and acceptable understanding is that the epithelial proliferation, both ductal and lobular is associated with benign disease, while atypia confers an increased risk of developing breast cancer (IARC, 2002). Substantial evidence has been developed on the experimental, clinical and epidemiological aspects of breast cancer, providing better understanding of its aetiology. As well, the introduction of screening mammography has led to a significant improvement in the overall prognosis, a shift in tumour stage and survival in breast cancer patients. In parallel, many strong risk factors have been recognized.

2.1.2.1 Risk factors

Several causal factors of breast cancer have been established in the pathways of disease progression, mainly based on tumour heterogeneity, which in turn is mainly related to morphology, gene expression and clinical outcomes (Patel, 2018; Perou et al., 2000; Polyak, 2007; Polyak, 2011). The interaction of these factors together with established risk factors causes the risk profile change in women. This had led to a failure in identifying the exact subset of women in whom the majority of breast cancer arises (IARC, 2016). Many of the risk factors prevalent in high-income countries are well established and evident through epidemiological studies. Of these, age is the single most pronounced risk factor for breast cancer. The incidence rises rapidly till 50 years of age after that, one sees about a 5% increase in cumulative incidence (in Europe) till age 75 (IARC, 2016; Key et al., 2001). Beside age, several non-modifiable and modifiable risk factors, including hormonal and reproductive factors, life-style and environmental factors, are described below in Table 1.

Table 1. Selected list of established risk factors of breast cancer

Risk factor	Categories	Direction	References
Hormonal and reproductive factors	Younger age at menarche	Increased risk	(Anderson et al., 2014; Colditz et al., 2000)
	Younger age at first birth	Decreased risk	(IARC, 2016; Anderson et al., 2014; Barnard et al., 2015)
	Older age at menopause	Increased risk	(Colditz et al., 2000; Collaborative Group on Hormonal Factors in Breast Cancer, 2012)
	Postmenopausal HRT	Increased risk	(Barnard et al., 2015; IARC, 2012; Ritte et al., 2012)
	Oral contraceptives	Increased risk	(IARC, 2012)
	Breast feeding	Decreased risk	(Barnard et al., 2015; Collaborative Group on Hormonal Factors in Breast Cancer, 2002; Islami et al., 2015)
	Parity	Decreased risk	(Anderson et al., 2014; Colditz et al., 2000; Collaborative Group on Hormonal Factors in Breast Cancer, 2002)
Lifestyle factors and environmental exposures	Alcohol intake	Increased risk	(Allen et al., 2009; IARC, 2012; National Cancer Institute, 2018; Seitz et al., 2012)
	Adult BMI	Increased risk	(Arnold et al., 2015; WCRF, 2018)
	Post-menopausal, obesity	Increased risk	(Key et al., 2001)
	Physical activity	Decreased risk	(Pizot et al., 2016; WCRF, 2018; Wu et al., 2013)
	Tobacco smoking	Increased risk	(Catsburg et al., 2015; Cui et al., 2006; IARC, 2012; Lauby-Secretan et al., 2009)
Non-modifiable risk factors	Age (>50 years old)	Increased risk	(Anderson et al., 2006)

Breast density	Increased risk	(Chiu et al., 2010; Kerlikowske et al., 2007; Pettersson et al., 2014)
Family history of breast cancer	Increased risk	(Anderson et al., 2000; Barnard et al., 2015; Collaborative Group on Hormonal Factors in Breast Cancer, 2001)
Benign breast disease	Increased risk	(Colditz et al., 2000; Hartmann et al., 2005)
Chest radiation at younger age	Increased risk	(Henderson et al., 2010; Moskowitz et al., 2014)

2.1.2.2 Risk factors by mode of detection

The mode of cancer detection, mainly screen-detected and interval cancers may be an important factor in identifying the subset of women if the risk factors differ between these sub-types. Performance and evaluation indicators of screening programmes are influenced by the interval cancer; thus, understanding the risk factors for interval cancer is crucial for a programme to detect cancer in the preclinical phase (IARC, 2016). Studies have found that women who have high mammography density, and/or current users of hormonal therapy and/or have previously been diagnosed with benign breast disease have higher risk of interval cancers than screen-detected cancers (Blanch et al., 2014; Domingo et al., 2014; Kirsh et al., 2011; Pollan et al., 2013). In postmenopausal women, the breast is usually fattier, which make lesions easier to detect by mammography, and, hence, more cancers are diagnosed during the screening visits (Surakasula et al., 2014). Hormonal therapy prolongs the density of the breast, which may then disguise the findings in the mammography; this may lead to an increase in interval cancers (Collaborative group on Hormonal Factors in Breast Cancer, 1997; Beral et al., 2003). A study in the USA found that a higher proportion of rapidly proliferating and aggressive cancers in younger women were diagnosed as interval cancers (Gilliland et al., 2000). Another study in Finland found higher interval cancer rates in younger women but no difference in the cell proliferation rate between women younger than 50 and older women (Klemi et al., 1997). The parametrisation of family history showed no clear direction across studies. However, there is no existing valid

approach for identifying the subset of women based on established risk factors; hence, age-related invitation is still the single best screening strategy (IARC, 2016).

2.1.3 Natural history of breast cancer

In recent years, an increase in the understanding at the molecular and tissue levels suggests that breast cancer is the result of an imbalance in the complex regulatory cycles to which breast tissue is exposed (Tkaczuk et al., 2017). The up and down regulation of hormones and epidermal growth influences the genetic pathways, leading to cell proliferation and regression (Georgian-Smith et al., 2014). The real action takes place at the interface between perceived cancer cells and the presumably normal tissues adjacent to those cells. This mechanism forms an early genetic predisposition that alters the cellular and physiological phenotypic traits, which explains the apparent paradox of the natural history of cancer (Lakhani et al., 2012). The ductal and lobular epithelial proliferation and hyperplasia causes benign disease, which confers an increased risk of developing breast cancer. More than 95% of breast tumours arise from the epithelial cell lining of the breast, from either the milk-producing glands (lobular carcinomas) or the draining ducts (ductal carcinomas) (Erban et al., 2010; IARC, 2002).

The heterogeneous nature of breast cancer forms more than one natural history; thus, the idea of a natural progression of cancer from atypia to carcinoma in situ, invasive cancer and metastasis may not actually be certainly true (Buerger et al., 1999; Buerger et al., 2001). The disease has an enigmatic time-variable history, with long, disease-free intervals followed by cyclic recurrence and remission. Despite the variability of disease progression, the degree of malignancy of individual tumours can be predicted with reasonable accuracy. The two most important factors for the prognosis of breast cancers are the extent of spread or stage of the tumour and its degree of differentiation or grade. A tumour with better prognosis is confined to the breast itself (stage I) and well differentiated (grade I), whereas one that has spread to the axillary lymph nodes (stage II and III) and is very poorly differentiated (grade III) results in a poor prognosis (Lakhani et al., 2012)

The natural course of breast cancer progression has three main points: benign disease, in situ carcinoma and invasive cancer. Understanding the progression rates is crucial for answering the question about how intensively the abnormalities should be sought and treated (IARC Handbook - Breast Cancer Screening, 2002). The widespread use of mammographic screening typically aimed at early detection during the detectable pre-clinical phase has altered the natural history of the disease (Figure 3). Thus, the time interval between the onset of disease and the diagnosis of disease (also known as detectable pre-clinical phase or sojourn time) has decreased. Correspondingly, the period when a cancer is found by screening and when it would

appear through clinical signs and symptoms, also known as lead time, has increased. However, the increase or decrease in the time interval is affected by the frequency of screening, characteristics of lesions and the screening test. A screening mammography detects many forms of breast cancer, ranging from low-grade DCIS to large, high-grade, invasive cancer; thus, this technique is well recognized to prevent the development of high-grade, invasive cancer (Cowan et al., 1991; Evans et al., 1997; Evans, 2001a; Evans, 2001b; Klemi et al., 1992; Lampejo et al., 1994; Rajakariar et al., 1995; Walker et al., 1994). The rate of progression at each stage of disease defines the mean sojourn time (MST) and permits a range of behaviours (Tan et al., 2013). Duffy et al. (1995) used the two-parameter Markov chain model to estimate the MST using the data from a Swedish two-country study (Duffy et al., 1995). They found that the estimate of MST ranging from 2.1 to 2.5 years. The lead time following the introduction of digital screening tools has become longer than with analogue mammography. The frequency of screening and increase in life expectancy also affects the lead time (IARC, 2002). While the low-grade tumours identified at screening with excellent prognosis may be indolent, some might progress over time to become aggressive. No method or technique has yet been developed for clinical verification of the over diagnosed cancers. The size, nodal involvement and metastases (TNM) of these tumours defines the disease prognosis, and eventually the success of breast cancer screening programmes.

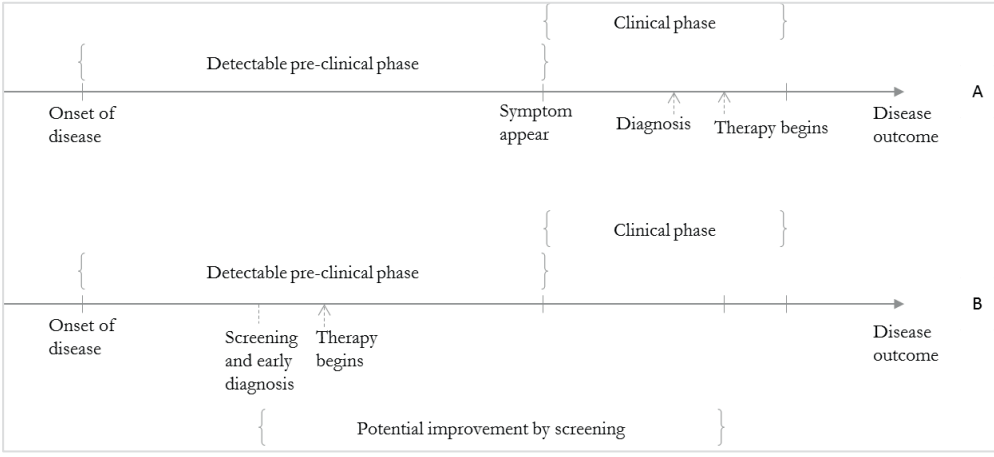

Figure 3. Scheme of progression of cancer, with intervention of screening test for early detection

(A) Natural history progression of breast cancer; (B) Disease progression after intervention by screening

2.2 Screening for breast cancer

While the ultimate goal of breast cancer screening is to reduce mortality from the disease among the screened women, the immediate goal is to detect cancer at the detectable pre-clinical phase or asymptomatic phase (IARC, 2016). Early detection together with early access to effective treatment services, helps achieve the greater effect of screening (Lauby-Secretan et al., 2015). The European Parliament in June 2003 called for the EU member states to develop and implement effective screening strategies for improved preventive health care. Until now, screening by mammography has been identified as the most popular means of screening women for breast cancer. Screening, diagnostic and treatment services are provided at the population level through organized programs in high-income countries. Among the EU member states 25 countries have ongoing, piloted or planned population-based breast cancer screening programmes targeting 50-69 years old (Ponti et al., 2017).

The screening programmes are evaluated based on the set criteria for the process indicators of screening mentioned in the European guidelines for mammography screening (Perry et al., 2006). Some upper middle-income countries offer service through either organized or opportunistic screening, whereas low- and middle-income countries promote breast cancer awareness or clinical breast examination as a means for early detection of the disease (IARC, 2016; Lauby-Secretan et al., 2015). Organized screening programmes are based on firm evidence regarding the effectiveness and appropriate balance between screening benefits and harm (Ponti et al., 2017). The implementation of screening programmes can either be population-based - if the people in the eligible target population in the selected area served by the programme are individually identified and personally invited to attend screening - or performed in a diagnostic or clinical context mainly referred to by general practitioners as so called 'grey', 'wild', or 'opportunistic' testing (Arbyn et al., 2010; Ponti et al., 2017). In several European countries, including Finland, screening by mammography for breast cancer is run as part of public health policy (Hakama et al., 1997). Evidence of screening with other screening modalities or imaging techniques at the population level is limited or inadequate in terms of reduction in mortality from the disease. Some studies have evaluated the effectiveness of other imaging techniques, such as ultrasonography or magnetic resonance imaging together with mammography, in women with an increased risk of breast cancer (Bick et al., 2019; Wilczek et al., 2016).

2.2.1 Impact of screening mammography on breast cancer incidence and mortality

Evidence from the randomised, controlled trials conducted in the 1980s and 1990s (Andersson et al., 1988; Tabár et al., 1992, Bjurstam et al., 2003; Moss et al., 2005; Autier et al., 2009), and the recent cohort and case-controls studies (Moss, 1999; Ascunce et al., 2007; Weedon-Fekjær et al., 2014; Morrison et al., 1988; Hofvind et al., 2013) report a clear reduction in breast cancer mortality in women invited to or who attended organized mammography screening. In 2014, the IARC (International Agency for Research on Cancer) expert working group evaluated the scientific evidence reported by the clinical trials and high-quality observational studies on the effectiveness of mammographic screening programmes (Lauby-Secretan et al., 2015). The working group concluded that a clear reduction in breast cancer mortality was evident in women aged 50-69 years who attended the mammographic screening. The mortality reduction in women aged 50-69 years who were invited for a mammography screening was about 23%, and the mortality risk reduction in women who attended the mammography screening was about 40% (Ascunce et al., 2007; Hakama et al., 1997; Kalager et al., 2010; Morrell et al., 2017; Moss, 1999; Olsen et al., 2005; Olsen et al., 2013; Weedon-Fekjær et al., 2014). In women aged 70-74 years old who attended a mammography screening, a substantial reduction in mortality was reported in several incidence-based mortality (IBM) studies (Coldman et al., 2014; Jonsson et al., 2003; van Dijck et al., 1997). Also, as compared to women screened at ages 50-69, screening mammography performance such as sensitivity, positive predictive value and, specificity, increased in older women aged 70-79 (Sinclair et al., 2011). In contrast, the effectiveness of mammography screening in women aged 40-49 was not well evidenced in terms of a reduction in risk of death (Elmore et al., 2005; Nelson et al., 2016; Norman et al., 2007; Roder et al., 2008; van Schoor et al., 2010).

2.2.1.1 Benefits of screening

The effectiveness of mammography screening can also be measured in terms of a reduction in the incidence of advanced breast cancers and its impact on breast cancer mortality. Results from randomized trials compared the risk of advanced breast cancer between the intervention and control arms and also, the subsequent effect on breast cancer mortality (Andersson et al., 1988; Autier et al., 2009; Bjurstam et al., 2003; IARC, 2016; Miller et al., 1992a; Miller et al., 1992b; Moss et al., 2005; Shapiro, 1997; Tabár et al., 1992; Tabár et al., 2015). A review study by Tabár et al. (2015) showed a substantial and significant effect on breast cancer mortality when the cancer was detected at an early stage in the service screening programme (Autier et al., 2009; Tabár et al., 2015). The reduction in incidence of advanced disease is also associated with a reduction in the risk of greater treatment morbidity and better quality of life. Monitoring advanced breast cancer incidence also provides an accurate and early indication of the subsequent breast cancer mortality (Tabár et al., 2015).

2.2.1.2 Harms of screening: false-positives, overdiagnosis and interval cancers

While evaluating the effectiveness of mammography screening programmes, the associated harms are estimated through performance indicators such as false-positive results, overdiagnosis and interval cancers. The occurrence of false-positive results varies across the screening programmes due to variation in recall or further assessment policies, with and also due to a difference in performance and training of the interpreting radiologists (Perry et al., 2006; Hofvind et al., 2012; Myers et al., 2015; Seely et al., 2018). A pooled estimate of the occurrence of false-positive results in mammography screening in Europe reported an approximately 20% cumulative risk in the screening of middle-aged and older women (between 50 and 70 years of age) after 10 screens (Hofvind et al., 2012). A review study on false-positive results in United States reported 31% cumulative risk after 20 years of biennial screening (Loberg et al., 2015).

The term 'overdiagnosis' is defined as the detection of breast cancer by screening that would never have presented clinically during the woman's lifetime. Overdiagnosed cases of breast cancers are quantified by comparing observed and expected cumulative incidence of breast cancer which extends from the beginning of screening in screened cohort until several years after screening has ended; that is, when the lead time of cancers diagnosed as a result of screening has elapsed (Puliti

et al., 2012a). Several studies conducted in Europe and Canada have tried to quantify overdiagnosis of breast cancer, but estimates vary widely between earlier randomized trial results and recent service screening studies (Gatzsche et al., 2009; Miller et al., 2002; Moss et al., 2005; Paci et al., 2005; Welch et al., 2010; Zackrisson et al., 2006). The estimates from the trials are subject to bias because of shorter follow-up time and chances of or access to mammography outside of the trial. The optimal estimate can be drawn from the service screening programme either following 10 years after the last screen (Duffy et al., 2013) or modelling the lead-time. The estimates range from 1%-10% (Puliti et al., 2012b). This also depends on access to mammography outside of the programme. Also, the technological advancement and continuous development in the practice of screening have modified the risk of overdiagnosis in service screening programmes, and thus the estimates differ from those reported by randomized trials conducted over 25 years ago. A review of the European observational studies (13 primary studies in seven European countries) evaluating the overdiagnosis of breast cancer in service screening programmes analysed by Puliti et al. (2012) suggested a relatively low estimate of overdiagnosis ranging from 1% to 10% (Puliti et al., 2012a). In Finland, less than 10% of breast cancer cases are overdiagnosed (Heinävaara et al., 2014). However, studies might be subject to potential biases that may affect the actual estimates, depending on whether or not the estimates are adjusted for the lead-time bias as well as the difference in the statistical adjustment for lead time. The variation in the overdiagnosis estimates between studies can be explained by the difference in the definition of the population at risk (de Gelder et al., 2011), by the differences in the screening policies and uptake between programmes (Olsen et al., 2006; Waller et al., 2007), the screening age range, the difference in screening interval and recall practice in various programmes and the length of the screening period considered while estimating overdiagnosis (Olsen et al., 2006; Paci et al., 2006; Puliti et al., 2009).

Interval breast cancers inversely affects the effectiveness of the mammography screening programme. The interval cancer rates have been assessed in earlier randomized studies and also in established service screening programmes. Within a single programme, the proportion of interval cancers varies largely between the first and second year (from 15%–60% respectively) after the screening mammography (Bucchi et al., 2008; Fracheboud et al., 1999; Renart-Vicens et al., 2014; Tabár et al., 1992; Weber et al., 2016), and also by age at mammography screening with more cases diagnosed in younger women (Porter et al., 1999). Findings from a meta-analysis showed a higher (27%) proportion of interval cancers in service screening programmes than that of results from randomized trials (19%) (Jacklyn, et al., 2016).

A greater variation in the proportion of interval cancers exists between various screening programmes, mainly because of a difference in the definition, identification and quantification/categorization of interval cancers (Jacklyn et al., 2016; Bulliard et al., 2006; Lekanidi et al., 2017). Further, the completeness of data collection and the use of different inclusion and exclusion criteria may limit the comparability of interval cancer rates in different populations.

2.2.2 Further assessment procedure and histological confirmation

Women invited to attend a screening clinic first undergo mammography. If any abnormality is detected or suspected, women are recalled for further assessment. Imaging procedures during further assessment may include additional mammography or ultrasonography. Additional mammography includes complementary images such as magnified spot compression views, or digital breast tomosynthesis. The primary invasive procedure is ultrasonography guided core needle biopsy. Microcalfications without a tumour are examined with stereotactically guided vacuum assisted biopsy. Both of these provide a histological diagnosis of the breast lesion. Core biopsies are performed in the screening clinics, and if possibility of breast cancer is confirmed, women is informed about the diagnosis and referred to hospital. Few screening units that are part of a hospital breast clinic also perform vacuum biopsy, but in most cases the woman is referred to hospital. Referral for diagnostic surgical biopsy is made in rare cases where core or vacuum biopsy reveals a high-risk lesion, or if there is discrepancy between the clinical or radiological finding and the needle biopsy histology. The duration of further assessment and histological confirmation procedure to confirm the presence or absence of malignancy is about one to two months (in Finland) after the mammography-screening visit (Sarkeala et al., 2014). The proportion of mammography visits recalled for further assessment varies by screening programmes or country, ranging from as low as <2% in the Netherlands (Otten et al., 2013) to >5% in the UK and up to >10% in the USA (Hofvind et al., 2012; Smith-Bindman et al., 2005). The desirable standard in accordance to European Guidelines for recall is <5% at the initial screening round and <3% for subsequent rounds (Perry et al., 2006). In Finland, <3% of mammography visits are recalled and <1% of visits are referred for biopsy examination (Hofvind et al., 2012).

2.3 Breast cancer symptoms

The common symptoms and clinical signs of malignant tumours of the breast ranges from a painless mass to hard lump, nipple retraction, nipple discharge, localized breast skin changes, persistent axillary swelling, to growth of breast (volume) with an inflammatory reddish area and eczematous changes in or around the nipple or areola (Clinical Radiology, 2013; IARC, 2016; Joensuu et al., 2013). For metastatic breast cancer, there can be (in addition various other symptoms), a lump somewhere other than in the breast, pain in the back or hip, or neurological symptoms, depending on where the disease has spread (Joensuu et al., 2013).

A palpable breast lump is the most important symptom of early breast cancer. The majority of lumps are associated with benign breast disease (IARC, 2016; Mahoney et al., 1982; Ohene-Yeboah et al., 2008; Pradhan et al., 2008; Sankaranarayanan et al., 2011). Lumps characterized with a hard consistency and persistent with skin or nipple changes, or unilateral nipple discharge are associated with advanced breast cancer (Chen et al., 2012; Giess et al., 1998; Mahoney et al., 1982). Hospital-based studies reported breast cancer detection between 13% and 25% in women presented with a lump (Mahoney et al., 1982; Ohene-Yeboah et al., 2008; Pradhan et al., 2008).

Changes or asymmetries in the nipple or areola are an important aspect of early detection and breast awareness (IARC, 2002). A tumour located deep in the nipple is associated with extensive nipple retraction towards the tumour. Malignant nipple retraction should be distinguished from the more common inverted nipple that is usually bilateral and lasts for years or decades. Nipple discharge may appear in various colours. Even unilateral bloody discharge is usually caused by benign conditions (Tabár et al., 1983). Breast pain can be unilateral or bilateral, and the distribution can be diffuse or focal. The pain can also be cyclical or non-cyclical depending on the menstrual cycle (Tkaczuk et al., 2017). Pain is uncommon as the only symptom of cancer.

2.4 Breast symptoms reported in the mammography screening programmes

Screening and early diagnosis are important components of comprehensive cancer control. Population-based screening is meant for the unselected target populations (Perry et al., 2006). All eligible women are invited for screening, mainly based on age (mostly by mammography or by other techniques, if existing). The focus of early cancer diagnosis is on people who have symptoms and signs consistent with cancer (WHO, 2017). However, no evidence supports the effectiveness of such a strategy. Early diagnosis of breast cancer can be facilitated by clinical breast examinations by a radiographer or by quering women about breast symptoms at the mammography-screening visit. Different screening or diagnostic protocols for symptomatic and asymptomatic women, depending upon the programmes, may be used.

European guidelines on breast cancer screening recommend recording the symptoms and to make this information availableto the radiologists at the time of the film reading (Perry et al., 2006). The guidelines also mention referral of all symptomatic women to the comprehensive breast units; however, no data is available about the procedure regarding referral to those breast units. The Norwegian screening programme recalls women with clinical symptoms for further assessment (Hofvind et al., 2017). About 0.3% of all screening exams were recalled based on the clinical symptoms. Of those screen-detected cancers, 1.7% were symptomatic. No information is available about the long-term follow-up after a negative screening visit with symptoms. The available online protocols from different countries are mainly targeted at managing symptoms reported to general practitioners (GPs) at the primary health care centres. Thus, these symptoms are handled outside of the screening programme.

The National Health Service Breast Screening Programme (NHSBSP) guidelines in the UK recommend radiologists be alerted about the relevant clinical signs or symptoms at the time of reading the mammograms (NHS Breast Screening Guidance, 2011). In practice, the majority of women with symptoms first consult their general practitioners (GPs). Most of these women are managed by GPs, while some patients with significant breast symptoms are sent to hospital breast units for a specialist opinion or assessments (Clinical Radiology, 2013; Department of Health, 2010). Similarly, the National Cancer Control Plan-II of Denmark recommends that women with a symptom 'see a doctor' with the aim to reduce patient delays, doctor delays and system delay; and hence improve the cancer patient pathways (CPPs) (National Board of Health, 2005). The doctor then decides whether the symptoms

need referral or 'watchful waiting' for unclear symptoms. In Australia, protocol is also based on symptoms reported to the general practitioners (GPs) (Cancer Australia, 2017). We could not locate the protocols of other countries with population-based screening programmes that have published in English. Although not mentioned and/or no publications were found (except from the UK), it is likely that most of the symptoms likely to be malignant are referred to the hospital breast units and are therefore not included as the part of the screening programme. Thus, the symptoms are not subsequently followed-up and, hence not analysed.

Earlier studies in the UK have discussed the referral of breast symptoms reported at mammography-screening visit (Hide et al., 1999; Litherland et al., 2001; Williams et al., 2002). Williams et al (2002) studies the relevance of reported symptoms in a breast-screening programme in Wales (Williams et al., 2002). They found a ten-fold increase in the detection rate in women who reported significant symptoms as compared to the overall screening rate. However, in women with symptoms who had a normal mammography, the detection rate was no higher than that expected of the normal screening population. The Breast Test Wales (BTW) guidelines suggest recording significant symptoms or clinically relevant symptoms such as a new lump, retraction, nipple discharge, persistent localized pain, skin dimpling and red nipple (Breast Test Wales, 1996). Accordingly, Williams' study concluded that recall should be selective based on only the most relevant symptoms when the mammography finding is benign. A similar study in Newcastle, UK, suggested a policy of selectively recalling women with symptoms who had normal mammograms (Hide et al., 1999). The women selected were those who reported significant symptoms developed in the past year prior to the screening visit. Litherland et al. (2001) studied 344 women with symptoms but normal screening mammograms identified in the West of Scotland Breast Screening Programme (Litherland et al., 2001). They found cancers in women complaining of eczematous nipple discharge and dense breast patterns but not lumps. They suggested further assessment in the earlier symptomatic group, irrespective of mammography findings. All of the above-mentioned studies were conducted in a small population size had a short follow-up duration.

In the Finnish mammography screening programme, all women in the target age group are invited irrespective of symptom status (Finnish Cancer Registry, 2018a). Women presenting with symptoms such as a lump, nipple retraction or nipple discharge are also considered within the programme. The complete information is reported electronically to the Mass Screening Registry (Mass Screening Registry, 2018). However, we cannot rule out that some of these symptomatic women are also advised to contact the doctor or specialists to seek further clinical management.

Breast symptoms reported outside of the mammography screening programme, such as in primary health care centres and, outpatient clinics or symptoms reported at opportunistic mammography, are not systematically collected in Finland.

3 AIMS OF THE STUDY

The study was conducted within the ongoing organised population-based mammography-screening programme in Finland. The main research question was to find out whether women with breast symptoms when participating to screening have differential risks of breast cancer diagnosis, breast cancer mortality and false positive results as compared to asymptomatic women. Specific research questions were formulated for each sub-study. How strongly do breast symptoms predict the risks of breast cancer diagnosis and mortality? How do breast symptoms affect screening programme performance indicators? How big is the added risk of false positive results if all women with symptoms are recalled? How great is the added value of information provided by symptoms in the screening context? How can this novel information be best used to improve screening practices?

We then formulated detailed study objectives to obtain trustworthy answers to the above research questions. The specific tasks of the study were therefore to assess:

1. The mammography screening programme performance indicators, and to then compare between visits with and without symptoms.
2. The cumulative probability of false-positive mammography test results and false-positive referrals in screening visits with symptoms as compared to visits without symptoms.
3. The risk of screen-detected and interval breast cancers in women with symptoms as compared to those without symptoms.
4. The mortality outcomes, breast cancer mortality and all-cause mortality, and to then compare between visits with and without breast symptoms

4 MATERIALS AND METHODS

4.1 The organized breast cancer screening programme in Finland

An organized population-based mammography screening programme for breast cancer was introduced in 1987 (Hakama et al., 1997). Initially, the programme invited women aged 50-59, every two years. By 2007, the biennial screening covered the whole country and the upper age of invitation was set at 69. According to the Government Decree on Screenings (1339/2011), the Finnish municipalities are tasked with organizing screening activities. The screening tests and further assessments after recall are performed in organized mammography clinics and are free of charge for the invitee. Referral for more detailed diagnostic examinations and cancer treatment are performed in specialized medical care (central hospitals) at a small outpatient fee. All data gathered in the screening process is reported to the Mass Screening Registry of the Finnish Cancer Registry (FCR), which is then used to monitor and evaluate the quality and effectiveness of the screening programme.

All women in the target population defined by age are invited by personally addressed letters to take part in screening (Finnish Cancer Registry, 2018a). The Population Information System is used to define the target (or eligible) population by using birth year. The Population Information System contains basic information such as personal identity code, date of birth, and address. The date and place of screening is indicated in the invitation with the possibility to change appointment. A reminder letter is sent to those who missed the first invitation (non-attenders). At the screening clinic, breasts are examined by digital mammography taken from two directions, and images are interpreted by two radiologists. Women are notified of the screening results by a personal letter. If the mammography is normal, the women are invited back after two years for the next screening round. Women with abnormal mammography results are recalled for further assessment. The recalled women undergo additional mammography, ultrasound or core needle biopsy if needed. Women are informed about the confirmation of breast cancer diagnosis and are referred to hospital where they are treated (Figure 4). Follow-up after screening is complete for all invited, i.e. until death or emigration.

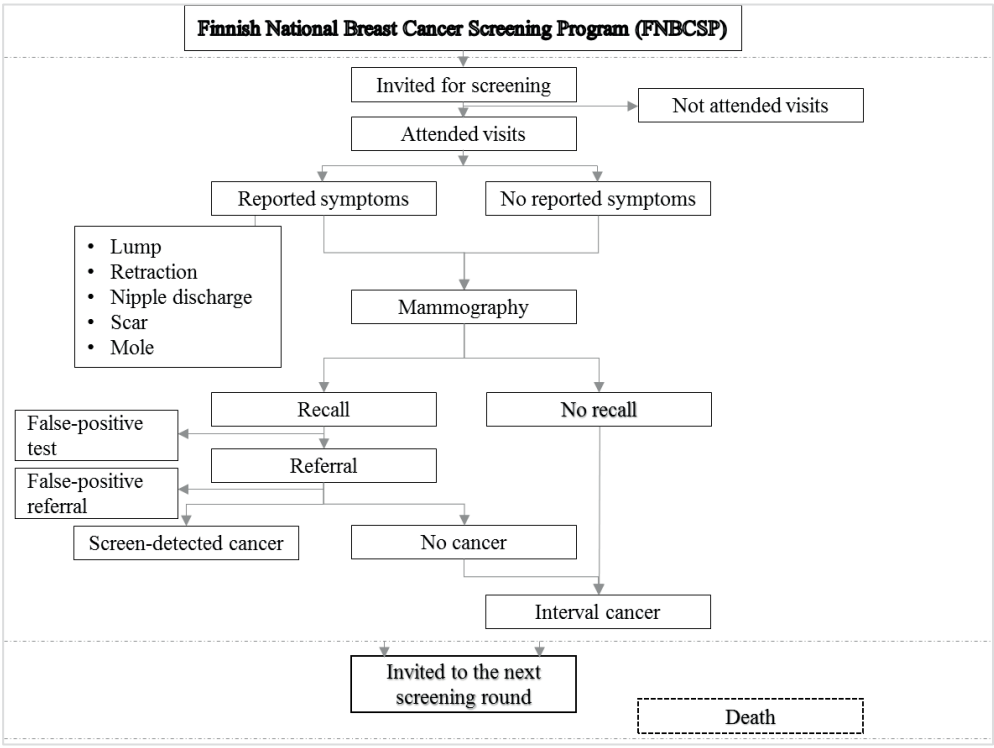

Figure 4. Flow diagram of mammography screening program by symptom status

The statistics from 2016 show that invitational coverage is close to 100% (altogether 381,000 women were invited) with an attendance proportion of 83% (BC screening report, 2017). Of those screened, about 3% were recalled at the screening unit/clinic, 0.8% were referred to hospital and 0.7% (about 2000 women) were diagnosed with invasive breast cancers or carcinomas in situ. About a third of invasive breast cancers and half of in situ carcinomas were diagnosed within the programme. The numbers vary between the health care districts due to differences in both the background risk and diagnostic criteria.

4.2 Mammography screening visits with breast symptoms

Women visit the screening clinic with a pre-set time and place for mammography. Women may present with symptoms or no symptoms at the screening clinic. The radiographer asks if the women have any symptoms during the past two months. If women say yes, the radiographer examines the breasts. (Finnish Cancer Registry, 2018a). Any presence of symptoms such as a lump, retraction, or nipple discharge is recorded and reported to the Mass Screening Registry (Mass Screening Registry, 2018), along with data on screening. Collection of information on lump and nipple discharge started from the early years of the programme, but collection of retraction information started in 1999. In this current thesis, symptoms reported either by the women or the radiographer were considered valid symptoms. After that, breasts are examined by mammography. Confirmation procedures after abnormal mammography are discussed in the previous section. The benign cases with reported symptom status are invited to the next regular screening examination. If a symptom appears after a normal mammography or before the scheduled invitation, women are advised to consult a doctor (Finnish Cancer Registry, 2018b).

4.3 Data source and their linkage

The Mass Screening Register (MSR) of the Finnish Cancer Registry (FCR) receives information about all screened women on a routinely basis through the population files and screening centres (Mass Screening Registry, 2018). Individual-level data using personal identifiers are registered for use on the invitation (date of invitation, municipality) and the screening visit (date of visit, age, symptoms history, mammography test results). The information also includes any further assessment recommendations and results from them, including date and result of histological verification. Histologically confirmed tumours are classified according to the TNM classification of tumours published by the Union for International Cancer Control (UICC) in 2002 (Sobin et al., 2002).

Pathological laboratories and clinicians notify the Finnish Cancer Registry of all cancers (invasive and in situ). Notification of cancers is based on special law and is obligatory. The benign lesions are collected by the Screening Units and sent to the MSR. The FCR uses data to monitor cancer at the individual and population level, produce cancer statistics and reports and provides data for research proposes. Coding of the cancer cases is based on the International Classification of Disease for Oncology - 3rd-edition (Fritz et al., 2000). The coding of breast cancer is based on topography (C50.0 to C50.9) and morphology codes (M8000 to M8999). The registry data covers diagnostic details such as date of diagnosis, cancer site, morphology, stage and, behaviour, as well as type of primary treatment given (FCR, 2018a). In addition, the registry receives/retrieves demographic information from the Population Register Centre. Information on death (date of death, cause of death) is also reviewed and compiled by the cancer registry from Statistics Finland.

To obtain reliable and valid information on screening mammography results, incident breast carcinomas and possible deaths, women were linked individually between different health care registries using the unique personal identifiers available. Information on breast symptoms was extracted from the mammography screening form retrieved from the Mass Screening Registry (MSR). Screen-detected breast cancers (study I) were defined using data from the MSR and the FCR. All breast cancer (screen-detected plus interval cancers) information was available from the Cancer Registry (studies II-IV) data base. The FCR validates breast cancer deaths using information on pathology, clinical notifications and cause of death data. This current thesis utilises breast carcinomas diagnosed before (since 1953; study III, IV) and after the implementation of the mammography screening programme in 1987.

Table 2. Data source, target population and screening participation

	Study I	Study II	Study III	Study IV*
Data source	FNBCSP	FNBCSP	FNBCSP	FNBCSP
Database linkage	MSR (2006–2010), FCR (2006–2010)	MSR (1992–2012), FCR (1992–2012)	MSR (1992–2012), FCR (1953–2014), CPR (1992–2014)	MSR (1992–2012), FCR (1953–2015), CPR (1992–2014), Statistics Finland (1953–2015)
Target population	Women aged 50–69 years old	Women aged 50–69 years old	Women aged 50–69 years old	Women aged 50–69 years old
Study period	2006–2010	1992–2012	1992–2014	1992–2015
Invited visits	1,454,143	2,627,256	4,594,328	4,594,335
Attended visits (%)	1,241,486 (85.4)	2,283,706 (86.9)	3,958,305 (86.1)	3,958,312 (86.1)

* Study IV dataset includes total screening visits of the mammography programme in the given period

FNBCSP: Finnish National Breast Cancer Screening Programme; MSR: Mass Screening Registry; FCR: Finnish Cancer Registry; CPR: Central Population Registry

4.4 Study design and study population by sub-studies

This current thesis is based on information collected by the Finnish National Breast Cancer Screening Programme (FNBCSP). The target population in the sub-studies were women aged 50-69 and invited for screening mammography. Studies I and II used data provided by the Mass Screening Registry (MSR) and the Finnish Cancer Registry (FCR). In addition to these, study III used Central Population Registry (CPR) data and study IV used CPR and Statistics Finland data. (Table 2)

The first sub-study (I) was a cross-sectional study conducted in women who attended screening mammography between 2006 and 2010. Altogether, 1,198,410 screening visits were studied (Table 3). The second study (II) was a historical cohort study conducted in women invited to screening mammography for the first time at ages 50-51 between 1992 and 2004 and who were followed up with until 2012. In total, 2,189,800 visits with and without symptoms were studied.

The matched cohort studies (III and IV) were conducted in women who attended screening mammography between 1992 and 2012. The exposure group (visits in women with breast symptoms at a given screening visit) and reference group (visits in women with no breast symptoms in the screening history before the index visit) were selected independently of the symptom status. Here, the index visit meant any screening visit with any given symptom and a respective asymptomatic visit was selected for every symptomatic visit based on the matching criteria. Altogether, 198,622 visits were analysed in study III and 151,956 visits were analysed in study IV.

Table 3. Study design, main method and variables by the sub-studies

	Study I	Study II	Study III	Study IV
Study design	Cross-sectional study	Historical cohort study	Matched cohort study	Matched cohort study
Main statistical method	Logistic regression	Discrete-time hazard regression	Cox-proportional hazards model	Cox-proportional hazards model
Main independent variables	Lump	Lump	Lump	Lump
	Retraction	Retraction	Retraction	Retraction
	Secretion	Nipple discharge	Nipple discharge	Nipple discharge
	Scar			
	Mole			
Performance indicators	Attendance proportion	False-positive test proportion	First and subsequent round attendance proportion	First and subsequent round attendance proportion
	Recall proportion	False-positive referral proportion		
	Referral proportion	True-positive proportion	Positive predictive value	Duration between symptoms and cancer diagnosis
	Positive predictive value		Test sensitivity	
	Sensitivity		Episode sensitivity	Duration between symptoms and mortality
	Specificity		Specificity	
	Cancer detection rate		Negative predictive value	
Main outcome variables	Screen-detected cancers	False-positive tests	Screen-detected cancers	Screen-detected cancers
	Tumour characteristics (TNM classification)	False-positive referrals	Interval cancers	All cause death
		True-positive referrals	Deaths due to breast cancer	Deaths due to breast cancer

Table 4. Definitions used for the study variables required for the performance indicators

Serial number	Variables	Definition
1.	Test positive	Screening visit with primary mammography positive (recalled for further assessment or diagnostics)
2.	Test negative	Screening visits with negative/normal mammography (i.e. not recalled for further assessment nor diagnostics)
3.	False-positive test	Screening visit with positive mammography but negative further assessment or histological confirmation (i.e. no breast cancer)
4.	False-positive referral	Screening visit with positive mammography and referral for biopsy/surgery, but negative histological confirmation (i.e. no breast cancer diagnosed)
5.	Episode negative	Screening visit with negative mammography (no recall), or positive mammography but negative further assessment or negative histological confirmation
6.	True positive	Screening visit with positive mammography and screen-detected breast cancer diagnosed in the same visit
7.	Screen-detected cancer	Primary breast cancer diagnosed within 6 months following a positive mammography
8.	Interval cancer	Breast cancer diagnosed in a screened woman before the next screening visit or within a period equal to a screening interval after: a negative mammography; a positive mammography but negative further assessment (recall negative); positive further assessment but date of diagnosis more than 6 months after screening mammography
9.	Subsequent round screen-detected cancer	A primary breast cancer diagnosed at the subsequent screening visit in a screen negative woman at the index visit

Table 5. Definitions of the performance and outcome indicators

Serial number	Variables	Definition
1.	Breast cancer detection rate (Study I)	The number of cancer cases detected at screen divided by number of screening visits expressed in 1000 or 10000
2.	Sensitivity of a symptom (Study I)	Number of visits with a symptom with malignant cancers, divided by the total number of visits with breast cancer
3.	Positive predictive value (PPV) (Study I)	Number of visits with cancer diagnosed in those who had symptoms
4.	Test sensitivity (mammography)	Number of visits with screen detected cancers divided by the sum of screen-detected cancers plus interval cancers diagnosed after negative test results (detection method) (Hakama, Pokhrel et al. 2015)
5.	Episode sensitivity	Number of visits with breast cancer detected in the full diagnostic process divided by all cancers detected over a screening round among attenders
6.	Positive predictive value of mammography (PPV)	Number of visits with a positive mammography test and diagnosis of cancer divided by the number of all test positives
7.	Negative predictive value of mammography (NPV)	Number of visits with a negative mammography test and no cancer diagnosed divided by the number of all test negatives
8.	Specificity of mammography	Number of visits with a negative mammography test and no cancer diagnosed divided by the number of all visits with no cancer diagnosed
9.	Breast cancer incidence	New breast cancers diagnosedafter the first invitation by the programme divided by the person-years at risk Incidence of fatal breast cancers (Study III): A subcategory of breast cancer incidence was formed of those breast cancers from which the women died. Follow up closed at the time of cancer incidence
10.	Breast cancer mortality	Deaths due to breast cancer diagnosed after the first invitation by the programme divided by person-years at risk. Follow up closed at the time of death
11.	All-cause mortality	Deaths due to any cause, including breast cancer deaths after the first invitation in the programme divided by person-years at risk

The above tables, 4 and 5 illustrates the definition of variables used in the studies. The variables are defined in the context of the mammography programme performance and the output indicators. In study I, we calculated the clinical validity parameters of symptoms within a screening episode whereas, in study II and study III we analyzed the performance measures of mammography in relation to symptoms in a screening round. Thus, the definition of parameters such as sensitivity, PPV, NPV differ between the studies.

4.5 Matching by background variables (Studies III, IV)

Visits with three symptom types (lump, retraction and nipple discharge) were frequency matched to asymptomatic visits by age at the screening visit (within two years), year of invitation (two-year band), number of past screening visits, and municipality of invitation. At first, visits with symptoms were aggregated based on matching variables. Each symptom stratum was matched to the viable controls (reference group or asymptomatic visits) by random sampling as many times as the number of visits in each stratum of symptoms by replacement sampling method. Hence, a single control had the possibility to be randomly selected to the same stratum more than once. Different symptoms were analysed separately, meaning that if more than one symptom was reported at a single visit then each symptom was analysed separately. The women from the reference group can later be part of the exposed group if symptoms are reported at later screening visits. Thus, symptoms act as time dependent covariates. In study III, symptoms reported more than once in the screening history were aggregated to form the stratum, whereas, in study IV, only the first symptomatic visit with at least one of the symptoms was used for sampling, meaning that symptoms reported in later screening visits were excluded in the exposure group. In both studies, the exposure-to-reference-visits ratio was 1:1 for lump and retraction and 1:2 for nipple discharge.

4.6 Statistical analysis

In study I, we assessed the association between breast symptoms reported at the screening visit and the risk of screen-detected cancer. We also analysed the clinical validity parameters of the symptoms. We utilized logistic regression analysis to estimate the age-adjusted association (odds ratios with 95% confidence interval) of symptoms with the occurrence of breast cancer. Individual and joint exposure to symptoms were analysed in all possible combinations. In addition, breast cancer detection rate by symptoms status (symptomatic versus asymptomatic) using number of cancer cases detected divided by number of screening visits was calculated for all possible combination of symptoms. Breast cancer risks were also compared among different five-year age groups categorized into four groups as '50-54', '55-59', '60-64' and '65-69', and reported using odds ratios with 95% confidence intervals.

In study II, the discrete-time hazard model with $logit\big(P(Y_{ij})\big) = X'_{ij}\beta$ was to estimate the cumulative risk of breast cancer according to symptom status at screen. Here i is the index subjects i=1,…, n and j is the index visits of ith subject j=1,…,Ji. The cumulative risk of first outcome event after k rounds of screening is $q_k = 1 - \prod_{j=1}^{k}\{1 - P(Y_{ij} = 0; Y_{i(j-1)} = 0, \dots, Y_1 = 0)\}$ (Christiansen et al., 2000). The effects of individual symptoms on the false positive and true positive probabilities were estimated using generalized linear model (GLM). Approximate Bayesian inference (INLA) (Rue, 2015) was used to estimate the 95% confidence intervals.

In study III, Cox proportional hazards regression was used to compute the age-adjusted risk of invasive breast cancer, in situ carcinomas, and non-localized cancers in women who reported symptoms as compared to those with no reported symptoms at screening visit. The risk-ratio estimate with 95% confidence interval was used to compute the risk of screen-detected cancers, interval cancers and subsequent round screen-detected cancers. We also estimated the cumulative incidence of invasive interval cancers separately for test negatives (from the index screen to subsequent screening visit) and episode negatives (from the 6-month screening episode to the subsequent screening visit). The follow-up time started from the index visit on 1 January 1992 – 31 December 2012 and ended at the date of emigration or death, diagnosis of cancers or at the end of the follow-up (31 December 2014) – whichever occurred first. In addition, cumulative probability of non-localized interval cancers at 95% confidence interval was also estimated. The programme performance characteristics as defined in section 4.4 were evaluated using basic statistics.

In study IV, the risk of breast cancer incidence and incidence-based mortality (IBM) were computed using Cox proportional hazards regression models. We also separately analysed breast cancer mortality and all-cause mortality. The hazard ratios (HRs) at 95% confidence intervals (CIs) were used to compare the risk between visits with and without symptoms. Likewise, incidence rates and mortality rates were calculated using person-years of follow-up at five-year bands since the date of visit with or without breast symptoms. Incidence rates were calculated as number of breast cancers diagnosed divided by the risk time, whereas mortality rates were calculated as number of deaths divided by the risk time. The rates and absolute difference in breast cancer incidence and mortality were compared between visits with and without symptoms, reported per 10,000 person-years of follow-up at 95% confidence intervals. In addition, the cumulative hazards of breast cancer incidence, breast cancer mortality and all-cause mortality were estimated using years as the underlying time units for risk time.

All statistical analyses in study I were conducted using Stata version 12.0 (STATA statistical software, release 12; Stata Corporation, TX). In study II and study III, all statistical analyses were performed using R-3.4.0 version. All statistical analyses in study IV were conducted using Stata version 14.0 (STATA statistical software, release 14; Stata Corporation, TX).

4.7 Ethical considerations

Observational studies based on registry-based data are certainly at risk of putting our own goals above those of others – especially the study subjects'. Throughout the analysis, we have handled highly sensitive and personal data of mammography-screening participants. When analysing data from the national registries, a personal identification number was necessary for the linkage of information between registries. After linkage, however, identifiers were replaced by running numbers assigned to subjects in all part of the analysis. The researcher has undergone mandatory training at the Finnish Cancer Registry on legislation relating to the handling of sensitive individual data. In addition, the researcher has taken a mandatory university level education in research ethics and proper data management. No contact was made between the researcher and individuals, and the study did not have any effect on providing health care for the invited or screened.

Permission to use the anonymized data was received from the National Institute for Health and Welfare (THL/736/5.05.00/2014; THL/461/5.05.00/2018) and Statistics Finland (TK-53-1258-13).

5 RESULTS

5.1 Symptoms prevalence and performance indicators in the screening programme

Over the years 1992 to 2012, 3, 958, 312 screening visits were made in the Finnish mammography programme, with an attendance proportion of 86.1%. Of these, 51,698 (1.4%) visits were recorded with a lump, 41,326 (1.5%) visits with retraction and 9,131 (0.3%) visits with nipple discharge (see Table 6). The proportion of visits with a lump decreased with an increase in the age group, from 1.6% in the youngest age group of 50-53 to 1.1% in the older age group of 66-69, in visits with retraction, a reverse order was observed from 1.1% in age group 50-53 to 2.1% in age group 66-69. Nipple discharge increased slightly in those of a younger age than of a higher age. The proportion of visits with a lump (2.7% versus 1.2%) were higher in the earlier years (1992-1997) of the programme than in recent years (2008-2012). The opposite was true in the case of retraction and nipple discharge.

Altogether, 14% of visits with a lump, 9% of visits with nipple discharge and 4% of visits with retraction were recalled as compared to less than 3% of visits without symptoms (Table 7). Similarly, a higher proportion of visits with all symptom types were referred for surgery as compared to those visits without symptoms. The cancer detection rate was 30 per 1,000 visits with a lump, 12 per 1,000 visits with a retraction and 9 per 1,000 visits with a nipple discharge as compared to about 5 per 1,000 visits without symptoms.

Table 6. Symptoms prevalence using the whole dataset of attended visits during 1992-2012

		Lump		Retraction#		Nipple discharge	
Characteristics		**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
Number of attended visits*	3958312	51698 (1.41)	3614873 (98.5)	41326 (1.53)	2656405 (98.4)	9131 (0.25)	3676681 (99.7)
Age at visit	50 to 53	18279 (1.60)	1127523 (98.4)	8596 (1.14)	744762 (98.8)	4001 (0.35)	1150694 (99.6)
	54 to 57	15326 (1.39)	1088271 (98.6)	10956 (1.40)	769685 (98.6)	2564 (0.23)	1113826 (99.7)
	58 to 61	10919 (1.29)	836276 (98.7)	11263 (1.72)	644072 (98.3)	1590 (0.19)	849793 (99.8)
	62 to 65	5349 (1.34)	393994 (98.6)	7197 (2.06)	342167 (97.9)	654 (0.17)	393148 (99.8)
	66 to 69	1825 (1.07)	168809 (98.9)	3314 (2.08)	155719 (97.9)	322 (0.19)	168220 (99.8)
Period of visit	1992 to 1997	15502 (2.74)	549551 (97.2)	NA	NA	1136 (0.21)	527757 (99.7)
	1998 to 2002	7562 (1.0)	745440 (99.0)	3570 (1.02)	346196 (98.9)	1681 (0.22)	768628 (99.7)
	2003 to 2007	11936 (1.19)	994557 (98.8)	13083 (1.30)	992935 (98.7)	2880 (0.27)	1050603 (99.7)
	2008 to 2012	16581 (1.24)	1316877 (98.7)	24491 (1.84)	1308967 (98.1)	3412 (0.26)	1321164 (99.7)

*number of attended visits in each symptom category excludes the missing visits; # information collected since 1999

Table 7. Programme performance indicators by symptom status using the whole dataset of attended visits during 1992-2012

		Lump		Retraction#		Nipple discharge	
Characteristics		**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
Recall proportion (%)	Yes (test-positives)	7263 (14.1)	92040 (2.55)	1649 (3.99)	69493 (2.62)	817 (8.95)	98901 (2.69)
	No (test-negatives)	44428 (85.9)	3522466 (97.4)	39677 (96.0)	2586538 (97.4)	8314 (91.0)	3577409 (97.3)
Referral proportion (%)	Yes (episode positives)	1993 (5.21)	21336 (1.0)	604 (2.69)	17104 (1.42)	211 (3.31)	23336 (1.06)
	No (episode negatives)	36247 (94.7)	2111769 (99.0)	21868 (97.3)	1188341 (98.5)	6172 (96.6)	2174073 (98.9)
True positives or cancer detection rate (per 1000 visits)		1546 (29.9)	15560 (4.3)	505 (12.2)	13073 (4.92)	82 (8.98)	17221 (4.68)

#information collected since 1999

5.2 Clinical validity of symptom and screening mammography quality measures (Studies I, III)

The sensitivity to detect cancer in women with a lump (8%) was higher than those with retraction (4%) or nipple discharge (0.7%). Similarly, the PPV was higher for a lump than for retraction and nipple discharge. The specificities of all symptom types were about 99% (Table 8).

The mammography test sensitivity was higher for visits with a lump and retraction but not for visits with a nipple discharge as compared to visits without the respective symptom type (Table 9). The specificity of the mammography test was higher for visits without symptoms as compared to visits with a lump (98% versus 88%), not so different for visits with a retraction (98% versus 97%) and lower for visits with nipple discharge (98% versus 92%). Similarly, the PPVs were higher for visits with a lump and visits with retraction but not for visits with nipple discharge as compared to visits without the respective symptoms.

Table 8. Clinical validity of symptoms with reference to cancer detection in the mammography screening programme

Symptoms	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)
Lump	7.62 (7.04 - 8.32)	98.7 (98.7 - 98.8)	3.19 (2.92 - 3.48)
Retraction	3.71 (3.27 - 4.20)	98.3 (98.2 - 98.3)	1.16 (1.02 - 1.32)
Nipple discharge	0.66 (0.48 - 0.89)	99.7 (99.7 - 99.7)	1.08 (0.79 - 1.47)
Any symptoms	35.5 (34.3 - 36.6)	75.2 (75.1 - 75.3)	0.78 (0.74 - 0.81)

Table 9. Performance quality measures of screening mammography with reference to cancer detection by symptom status

Symptoms		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)
Lump	Yes	81.9 (80.1 to 83.7)	88.3 (88.0 to 88.6)	19.9 (19.4 to 20.4)
	No	64.4 (58.4 to 70.2)	97.5 (97.4 to 97.6)	12.1 (11.1 to 13.3)
Retraction	Yes	76.9 (73.4 to 80.3)	97.1 (96.9 to 97.2)	28.1 (26.7 to 29.6)
	No	66.7 (60.9 to 72.1)	97.8 (97.7 to 98.0)	17.9 (16.4 to 19.5)
Nipple discharge	Yes	62.3 (52.3 to 71.5)	91.7 (91.1 to 92.2)	8.12 (6.98 to 9.42)
	No	70.3 (61.2 to 78.4)	97.5 (97.3 to 97.7)	15.5 (13.7 to 17.6)

5.3 Probabilities of false-positive results by symptom type (Study II)

In total, 12% of visits with a lump, 2.5% of visits with a retraction, and 8% of visits with a nipple discharge had false-positive mammography test results (Table 10). On the other hand, about 2% of visits without symptoms had false-positive mammography. Additionally, 0.8% of visits with a lump, 0.2% of visits with retraction and 1.6% of visits with nipple discharge had false-positive findings during further assessment (referral for surgery) as compared to 0.15% of visits without the respective symptoms.

In the same way, the cumulative risk of false-positive tests after 10 screening visits was higher for visits with a lump (45%), retraction (25%) or nipple discharge (35%) as compared to visits without the respective symptoms. Likewise, the cumulative risk of a false-positive referral after 10 screening visits was higher for visits with a lump, retraction or nipple discharge as compared to those without the respective symptoms.

Table 10. Probabilities of false-positive test and false positive referral after 10 screening visits

Symptoms		Screening visits	False-positive test		False-positive referral	
			Number (%)	Cumulative probability (%)	Number (%)	Cumulative probability (%)
Lump	Yes	26145	3140 (12.0)	45.2	212 (0.81)	3.32
	No	2114103	42527 (2.01)	17.2	3114 (0.15)	1.46
Retraction	Yes	26653	668 (2.51)	24.6	46 (0.17)	2.61
	No	1652257	30460 (1.84)	18.1	2026 (0.12)	2.04
Nipple discharge	Yes	5325	423 (7.94)	34.7	85 (1.60)	7.89
	No	2184475	45873 (2.1)	18.1	3306 (0.15)	1.46

5.4 Risk of screen-detected breast cancers in symptomatic versus asymptomatic women (Studies I, III)

Table 11 shows the rates (per 1,000 visits) and risk (risk ratios at 95% CI) of screen-detected breast carcinomas in women who reported symptoms as compared to those without symptoms at screening visit (study III). The detection rates of invasive breast cancer were elevated in women who reported a lump (28 versus 3 per 1000 visits), a retraction (11 versus 4 per 1,000 visits) or nipple discharge (7 versus 4 per 1000 visits) as compared to respective asymptomatic visits. Similarly, the non-localized invasive cancer detection rates were elevated for all symptom types than that of respective asymptomatic visits. The in situ detection rate was higher for visits with a lump but only slightly higher for visits with retraction or nipple discharge.

In relative terms, the age-adjusted risk of screen-detected invasive breast cancer in women who reported a lump was 8-fold (HR, 7.1 to 9.7) compared to those without a lump (Table 11). The risk was 2.4-fold (HR, 2 to 2.8) for retraction and 1.7-fold (HR, 1.2 to 2.3) for nipple discharge as compared to those without the respective symptoms. Similarly, the risk of non-localized invasive breast cancer increased for all symptom types. In addition, the risk of in situ carcinoma was elevated in those who reported a lump (HR 1.6, 95% CI 1.1 to 2.4) but not for those who reported a retraction or a nipple discharge as compared to the respective asymptomatic visits.

Table 12 illustrates the joint effect of different symptoms on the risk of breast cancer (study I). In screening visits where both lump and retraction were reported, the risk of breast cancer was 22 times higher (95% CI 16.5 to 30.8) than those without symptoms. Similarly, the risk was 5.4 times higher (95% CI 4.3 to 6.7) in screening visits with a lump and scar and 2.3 times higher (95% CI 1.7 to 3.1) in visits with a retraction and scar as compared to those without the respective symptoms. We did not see an additional increase in the risk of breast cancer when scar was reported together with a lump or retraction in a given visit.

We also estimated the risk of screen-detected breast cancer by age groups (five-year interval) as reported in study I. In women who reported a lump or retraction, the risk of breast cancer was higher across all age groups. For nipple discharge, the risk was increased only in the older age groups, i.e., 60-69 years old (Figure 5). We did not find a prominent increase in the risk with increased age in women who reported symptoms.

Table 11. Cases (proportion/1000 visits) and age-adjusted risk (risk ratios) of screen-detected cancers in those who reported symptoms as compared to those without symptoms at screen

Symptom	Cancer outcomes	With symptoms	Without symptoms	Age-adjusted hazards ratio (95% CI)
		Cases per 1000 visits (%)	Cases per 1000 visits (%)	
Lump	Invasive	1440 (28.1)	174 (3.39)	8.26 (7.1 to 9.70)
	In situ	61 (1.19)	38 (0.74)	1.61 (1.08 to 2.43)
	Non-localized	693 (13.5)	57 (1.11)	12.0 (9.22 to 16.0)
Retraction	Invasive	461 (11.3)	192 (4.69)	2.39 (2.02 to 2.84)
	In situ	35 (0.86)	31 (0.76)	1.13 (0.70 to 1.84)
	Non-localized	230 (5.62)	61 (1.49)	3.47 (2.62 to 4.66)
Nipple discharge	Invasive	66 (7.27)	83 (4.57)	1.66 (1.19 to 2.31)
	In situ	15 (1.65)	16 (0.88)	1.87 (0.92 to 3.81)
	Non-localized	25 (2.75)	27 (1.49)	1.85 (1.05 to 3.22)

Table 12. Odds ratios (ORs) of breast cancer with 95% confidence intervals (CI) for joint exposure to symptoms

Symptoms	Screened women	Cancer cases (%)	Age-adjusted OR (95% CI)
Lump | Retraction	36036	695 (1.93)	
No | No	1162374	5829 (0.50)	Reference
No | Yes	20449	198 (0.97)	1.94 (1.68 to 2.24)
Yes | No	15156	453 (2.99)	6.15 (5.55 to 6.74)
Yes | Yes	431	44 (10.2)	22.6 (16.5 to 30.8)
Lump | Scar	152515	1380 (0.90)	
No | No	1045895	5144 (0.49)	Reference
No | Yes	136928	883 (0.64)	1.31 (1.22 to 1.41)
Yes | No	12378	414 (3.34)	7.0 (6.32 to 7.75)
Yes | Yes	3209	83 (2.59)	5.37 (4.31 to 6.69)
Retraction | Scar	157121	1162 (0.74)	
No | No	1041289	5362 (0.51)	Reference
No | Yes	136241	920 (0.68)	1.31 (1.22 to 1.41)
Yes | No	16984	196 (1.15)	2.26 (1.95 to 2.60)
Yes | Yes	3896	46 (1.18)	2.31 (1.72 to 3.09)

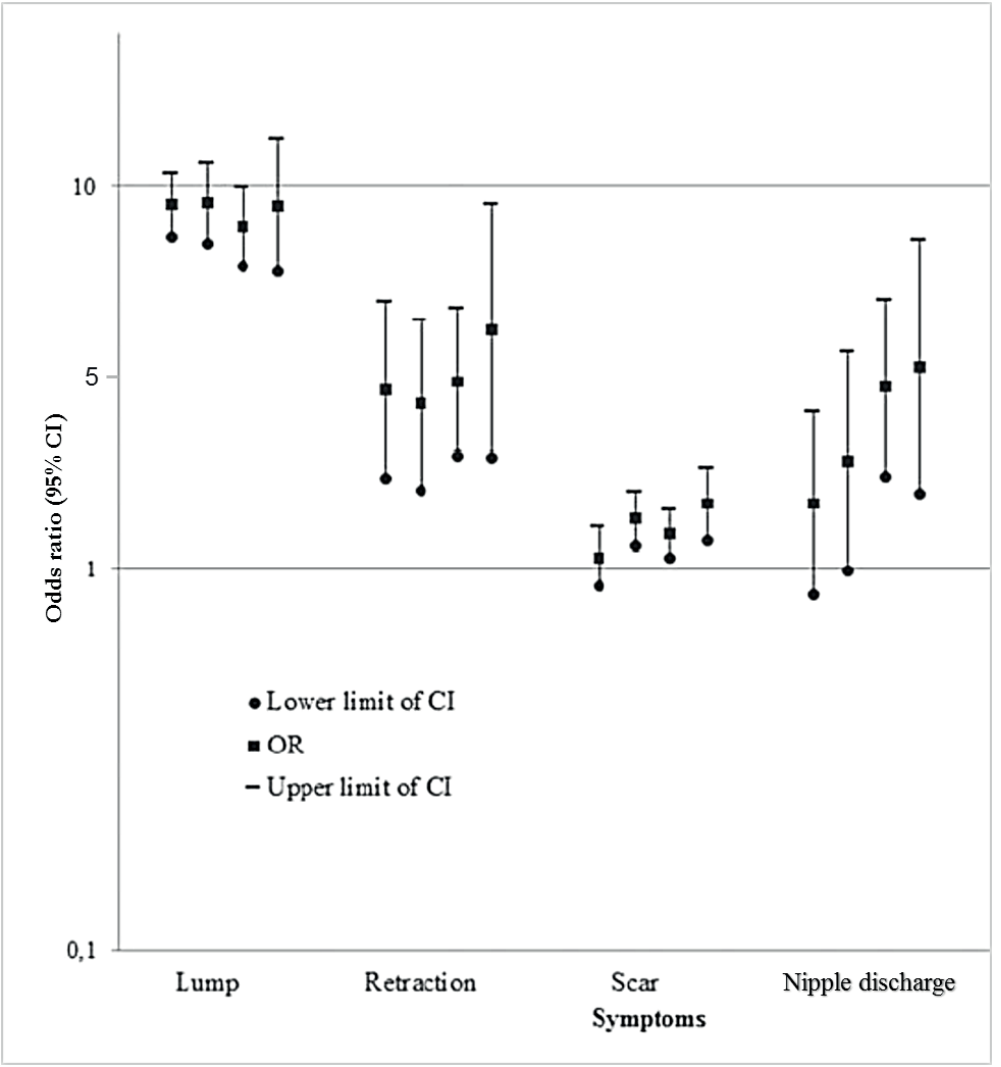

Figure 5. Risk (OR, 95% CI) of breast cancer among women who reported symptoms by age-groups, respectively as '50-54', '55-59', '60-64' and '65-69'

5.5 Risk of interval breast cancers in symptomatic versus asymptomatic women (Study III)

The rates of interval breast cancers (invasive and in situ) were raised after the screening visits with a lump (422 versus 112 cases per 1000 visits), visits with retraction (161 versus 107 cases per 1000 visits) or visits with nipple discharge (61 versus 45 cases per 1000 visits) as compared to visits without respective symptoms (Table 13). For nipple discharge, two asymptomatic visits were selected for each visit (see materials and methods; section 4.6). The age-adjusted risk of invasive interval breast cancer was 4 times (95% CI 3 to 4.7) higher in visits with a lump, 1.5 times (95% CI 1.2 to 2) higher in visits with retraction and 2.5 times (95% CI 1.7 to 3.8) higher for nipple discharge as compared to respective asymptomatic visits. Similarly, the risk of non-localized interval cancers was increased for all-three symptom types, whereas the risk of in situ interval carcinomas was higher for visits with a lump (HR 3.9, 95% CI 1.9 to 8.6) or nipple discharge (HR 4, 95% CI 1.3 to 15) but not for visits with retraction (HR 1.2, 95% CI 0.4 to 3.6).

We also analysed the risk of breast cancer diagnosed at the subsequent screening visit after the index visit with or without symptoms (Table 13). The risk of subsequent round invasive screen-detected cancer was higher in those who reported a lump (HR 1.7, 95% CI 1.4 to 2.1) but not for visits with a retraction (HR 1.1, 95% CI 0.9 to 1.4) or nipple discharge (HR 1, 95% CI 0.6 to 1.5) as compared to visits without the respective symptoms. Notably, the risk of in situ carcinomas in the subsequent round was increased for visits with nipple discharge (two-fold) and for visits with a lump (1.5-fold) but not for visits with a retraction.

The cumulative incidence of interval breast cancers was estimated separately for test positives and negatives with or without symptoms (see Figures 6a-6c). In test negative visits (no recall), the incidence of interval breast cancer in those with a lump increased rather rapidly after the first month as compared to visits without a lump. While considering the cumulative incidence of breast cancer among the asymptomatic visits after 23 months (full follow-up over one round) as a reference, the same cumulative incidence was reached within 6 months after visits with a lump and in 12 months after visits with nipple discharge. In addition, the non-localized breast cancers were diagnosed rather rapidly after the screening visits if a lump was reported as compared to those without a lump. On the other hand, in women who reported a lump and were recalled for further assessment, the difference in the probability of interval breast cancer was clearly visible in the graph soon after the

negative further assessment as compared to those visits without lump but not likely for visits with a retraction or nipple discharge.

Table 13. Cases (proportion/1000 visits) and age-adjusted risk (risk ratios) of interval cancers and subsequent round screen-detected cancers in those who reported symptoms as compared to those without symptoms at screen

Symptom	Cancer outcomes	With symptoms	Without symptoms	Age-adjusted risk of interval cancers (hazards ratio, 95% CI)	With symptoms	Without symptoms	Age-adjusted risk of subsequent round screen-detected cancers (hazards ratio, 95% CI)
		Cases per 1000 visits (%)	Cases per 1000 visits (%)		Cases per 1000 visits (%)	Cases per 1000 visits (%)	
Lump	Invasive	387 (7.54)	103 (2.01)	3.76 (3.04 to 4.69)	264 (5.14)	157 (3.06)	1.68 (1.38 to 2.05)
	In situ	35 (0.68)	9 (0.18)	3.89 (1.95 to 8.61)	159 (3.10)	102 (1.99)	1.54 (1.20 to 1.98)
	Non-localized	178 (3.46)	57 (1.11)	3.36 (2.49 to 4.61)	649 (12.6)	438 (8.53)	1.46 (1.29 to 1.65)
Retraction	Invasive	154 (3.76)	101 (2.47)	1.52 (1.19 to 1.96)	156 (3.81)	158 (3.86)	1.09 (0.87 to 1.37)
	In situ	7 (0.17)	6 (0.15)	1.17 (0.39 to 3.62)	76 (1.86)	66 (1.61)	1.15 (0.83 to 1.60)
	Non-localized	79 (1.93)	48 (1.17)	1.68 (1.18 to 2.43)	244 (5.96)	182 (4.45)	1.35 (1.12 to 1.65)
Nipple discharge	Invasive	52 (5.72)	41 (2.26)	2.49 (1.65 to 3.77)	32 (3.52)	72 (3.96)	1.0 (0.64 to 1.53)
	In situ	9 (0.99)	4 (0.22)	4.0 (1.26 to 14.9)	32 (3.52)	32 (1.76)	2.0 (1.22 to 3.27)
	Non-localized	20 (2.20)	18 (0.99)	2.22 (1.17 to 4.24)	110 (12.1)	181 (9.96)	1.22 (0.96 to 1.55)

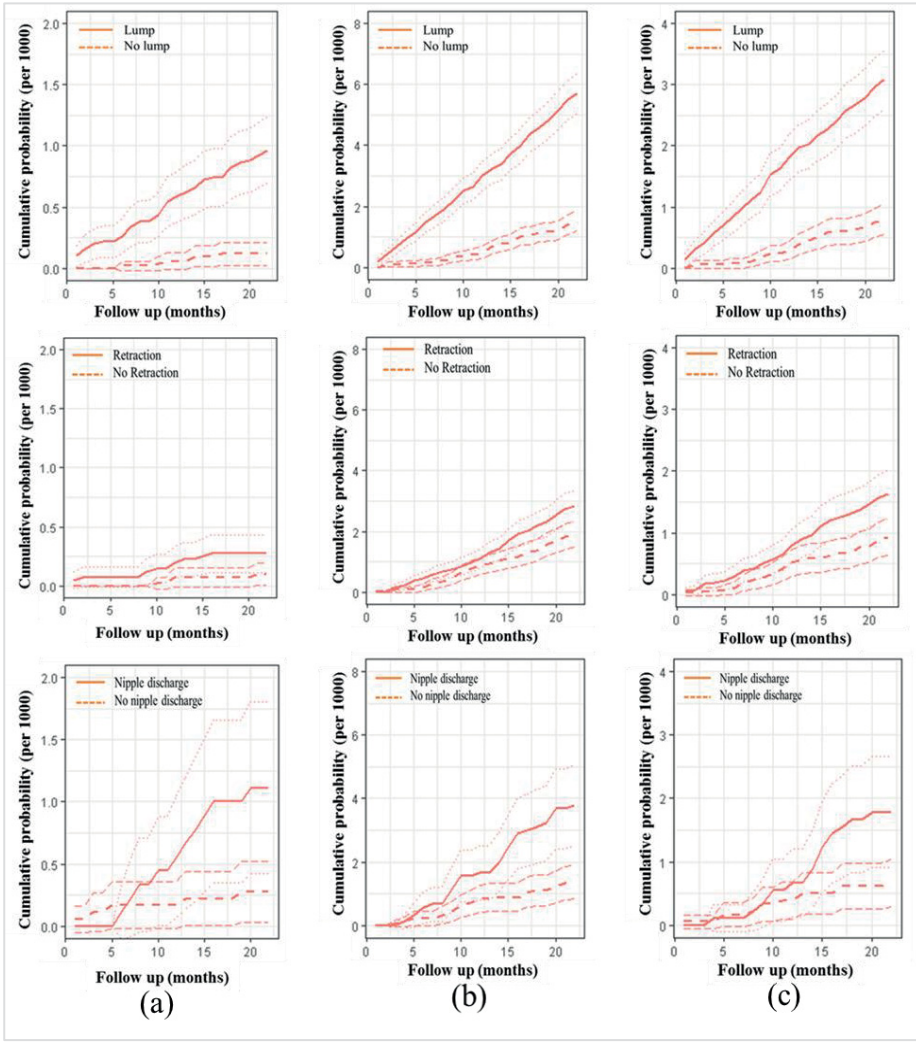

Figure 6. (a-c): Cumulative incidence of invasive (per 1000): a) recalled ICs; b) not recalled ICs; c) non-localized ICs

Note: The confidence intervals lines for cumulative incidence are indicated by light dotted lines in symptomatic and asymptomatic groups

5.6 Breast cancer mortality and all-cause mortality in symptomatic versus asymptomatic women (Study IV)

In Table 14, we report the breast cancer mortality and all-cause mortality rates and risk ratios by symptom types in the screened women after the first invitation by the programme and followed-up with over a period of 24 years. We found an increase in breast cancer mortality rates in women who reported a lump (5.3 per 10,000 person-years, 95% CI 4.5 to 6.2) or retraction (2.9 per 10,000 person-years, 95% CI 2 to 3.8) but not nipple discharge (0.8 per 10000 person-years, 95% CI minus 0.5 to 2) as compared to respective women with asymptomatic visits. The breast cancer mortality rates were very high within the first five-years after a screening visit with a lump as compared to no lump (78 versus 8 deaths per 10,000 person-years) or retraction (19 versus four deaths per 10,000 person-years) (see study IV). Similarly, the difference in all-cause mortality rates were prominent after visits with a lump or retraction but not after visit with a nipple discharge.

In relative terms, breast cancer mortality was increased for visits with a lump (3-fold) or retraction (3.9-fold) but not for visits with nipple discharge. However, all-cause mortality rate ratio was elevated for all symptom types as compared to the respective asymptomatic visits. We also found extra all-cause death cases in women with lower socio-economic class and reported a lump than those in higher class and without lump, respectively. As shown in figure 7 (Figures 7a-7b), the incidence of breast cancer mortality in screening visits with a lump or retraction increased rather rapidly. A similar level of risk was reach in half of the follow-up time in those with symptoms as compared to the risk in the respective asymptomatic visits after 24 years of follow-up (Figure 6). The cumulative hazards curve for visits with a lump or retraction followed a similar pattern of breast cancer mortality and all-cause mortality but no similarity was observed for visits with nipple discharge.

Table 14. Breast cancer mortality and all-cause mortality rates in women with symptom as compared to those without symptom

Symptoms		Person-years of follow-up	Breast cancer mortality				All-cause mortality			
			Deaths	Rate per 10000 person-years	Mortality rate difference (95% CI)	Age-adjusted hazards ratio (95% CI)	Deaths	Rate per 10000 person-years	Mortality rate difference (95% CI)	Age-adjusted hazards ratio (95% CI)
Lump	Yes	560626.7	437	7.8	5.31 (4.47 - 6.15)	3.14 (2.59 - 3.79)	644	11.5	7.27 (6.24 - 8.31)	2.72 (2.35 - 3.17)
	No	562528.5	140	2.5			237	4.2		
Retraction	Yes	206508.7	81	3.9	2.91 (1.95 - 3.87)	3.88 (2.40 - 6.27)	117	5.7	3.93 (2.76 - 5.10)	3.27 (2.25 - 4.75)
	No	207239.4	21	1.0			36	1.7		
Nipple discharge	Yes	86691.9	23	2.7	0.75 (-0.51 - 2.01)	1.40 (0.82 - 2.39)	47	5.4	1.85 (-0.06 - 3.63)	1.52 (1.04 - 2.22)
	No	173437.2	33	1.9			62	3.6		

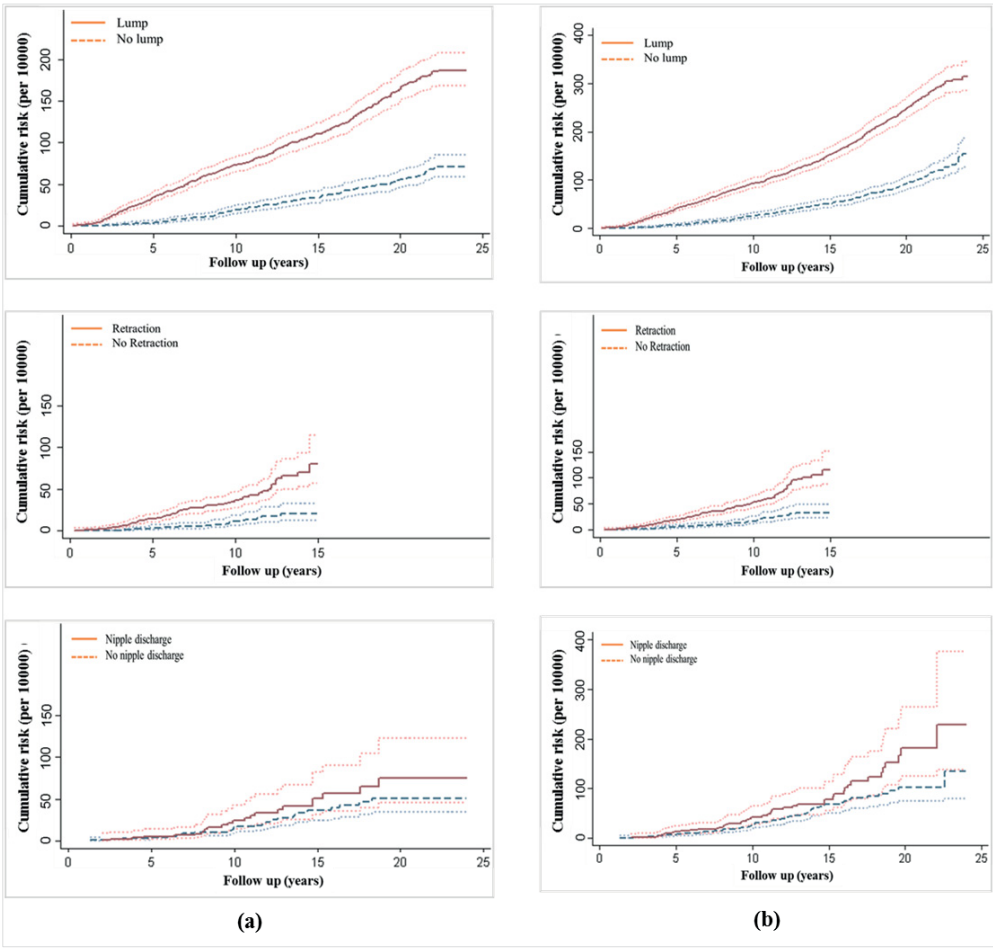

Figure 7 (a-b): Cumulative risk of (a) breast cancer mortality and (b) all-cause mortality per 10000 person-years of follow up

Note: The confidence intervals lines for cumulative risk are indicated by light dotted lines in symptomatic and asymptomatic groups

6 DISCUSSION

6.1 Screening test and performance indicators in relation to symptom status

The results reported in this study (sub-studies II, III, IV) are based on the complete follow-up of the screening aged women from the early years of the implementation of the mammography-screening programme until the most recent year, 2015. The screening attendance proportion (86%) reported in the study is among the highest of the existing mammography screening programmes, and it had near 100% coverage. A recent annual review of the Finnish breast cancer-screening programme reported a slight decline in the attendance proportion to 83% (Finnish Cancer Registry, 2017). We found that about 3% of women reported breast symptoms at the screening visit, which is lower than that reported by other studies. However, most studies are based on the diagnostic mammographic exams (Barlow et al., 2002; Lumachi et al., 2002; Seltzer, 1992; Sterns, 1992) or premenopausal women (Aiello et al., 2004) for whom symptom prevalence is high.

We found that 14% of screening visits with a lump were recalled for further assessment compared with only 3% of those without a lump. In addition, a higher proportion of visits with a retraction or nipple discharge were recalled. In the Norwegian screening programme, only 0.3% of visits based on clinical symptoms, of all those screened were recalled (Hofvind et al., 2017). The usual practice in many screening programmes is to refer women with symptoms to general practitioners or specialized breast clinics irrespective of the mammography findings (Clinical Radiology, 2013; Department of Health, 2010); Cancer Australia, 2017). About 3% of screening visits with a lump and 1% of visits with a retraction or nipple discharge were diagnosed as cancer as compared to about 0.5% in asymptomatic visits. Studies on diagnostic examinations reported higher breast cancer diagnosed among symptomatic women than that reported in our findings (Aiello et al., 2004; Lumachi et al., 2002; Seltzer, 1992; Sterns, 1992). As mentioned earlier, our study involved screening-age women who come for invitation-based mammography screening examination, whereas other studies (Sterns, 1992) were conducted among symptomatic women outside the screening programme. Thus, the variation in the

screening strategies, information collected on symptoms, age of the women or diagnostic examination instead of screening make these findings not directly comparable. A study conducted in the USA based on samples extracted from the medical chart record of screening and clinical breast examinations reported that 10% of women had symptoms; of those, 2% had cancer diagnosed (Ryerson et al., 2015). A study in the UK by Williams et al. (2002) found a higher proportion (45%) of cancers diagnosed in women presenting with symptoms and recalled for further assessment than our study (Williams et al., 2002). However, the recall decision was based on the mammography abnormality findings, and only women labelled as having significant breast symptoms as defined by Breast Test Wales (BSW) guidelines (Breast Test Wales, 1996) were recalled. Similar studies conducted in the UK are based on a small population of screening aged women and, thus the proportion of cancers diagnosed was low (Hide et al., 1999; Litherland et al., 2001; Williams et al., 2002).

In the first study (I), the positive predictive value (PPV) of a lump was 3.2%, which is higher than the PPV of 1%, as reported by other studies (Fenton et al., 2005; Mittra et al., 2010; Sankaranarayanan et al., 2011), mainly based on clinical breast examinations, but few hospital-based studies (Harvey et al., 2003; Kerlikowske et al., 1996; Mahoney et al., 1982; Ohene-Yeboah et al., 2008; Pradhan et al., 2008) reported higher PPV (between 13% and 15%) than our study. In addition, the sensitivity of symptom was low, which can be explained by the difference in the screening population - who normally undergo several rounds of screening in the programme and have access to mammography outside the programme - and the magnitude of diagnostic activities than those reported in the above-mentioned studies. The mammography test sensitivity and PPV were elevated in screening visits with a lump or retraction as compared to those without respective symptoms. Our sensitivity findings were in line with or higher than the screening estimates in other programmes (Perry et al., 2006; Hofvind et al., 2004; Njor et al., 2003; Vitak et al., 1997; Wang et al., 2001) and also higher than those reported in a pooled analysis study in 6 European countries (Tornberg et al., 2010), but we could not find estimates based on the symptom status.

6.2 Symptoms associated with breast cancer outcomes

We found an increased risk of screen-detected breast cancers in screening visits with any given symptom types as compared to respective asymptomatic visits (study I, III). A study by Aiello et al., based on screening or diagnostic exams reported a more than three-fold risk in women who had a lump, but no elevated risk was found for visits with a nipple discharge or breast pain (Aiello et al., 2004). A few other studies (Brouckaert et al., 2013; Gill et al., 2006) have also reported a higher risk of screen-detected breast cancer in those who reported a lump, but a lower hazard ratio than that reported in our study (study III). However, those studies were limited to many factors, such as study design, size, follow-up time and assessment of the possible outcome measures of breast cancer.

The risk of interval breast cancers was also elevated in screening visits with all symptom types as compared to asymptomatic visits. We are not aware of other studies that have reported interval cancer rates in relation to breast symptoms at screening visit. As well, the definition, identification and quantification of interval cancers vary largely by screening programmes (Bulliard et al., 2006) or within the same screening programmes depending on the sensitivity of the mammography test (Day, 1985; Perry et al., 2006; Houssami et al., 2006; Mushlin et al., 1998; Vinnicombe et al., 2009), sensitivity of the further assessment procedures (Perry et al., 2006; Hakama et al., 2007; Taylor et al., 2002; Taylor et al., 2004; Tornberg et al., 2010) or programme sensitivity (Anttila et al., 2002; Zorzi et al., 2010).

We found relatively higher all-cause mortality rates than breast cancer mortality rates between symptomatic and asymptomatic visits and substantial difference in the rates between these groups throughout the follow-up period. In addition, both the all-cause mortality and breast cancer mortality within five years of the screening visit, especially with a lump, were astonishingly higher than those without lump. In absolute terms, in every 10,000 person-years of follow-up with a lump at screen, 180 women died from breast cancer as compared to 70 women without a lump, and 315 cases of all-cause deaths in women with a lump as compared to 160 deaths in women without lump, respectively, after 24 years of follow-up. The difference between the symptomatic and asymptomatic group in mortality was not clearly related to specific causes of death other than breast cancer. We found a surplus number of death cases women of in lower socio-economic status and surplus deaths in those who had a lump. Having a lower socio-economic status might partially affect the findings on symptoms, therefore also affecting mortalities from any cause. Nevertheless, the socio-economic status of the whole cohort was unknown in our study. No other

studies have reported the difference in mortality between these sub-groups. Furthermore, systematic review conducted in Europe has reported that a lower socio-economic status is linked to several factors, such as lower screening attendance, delayed diagnosis and, larger tumour, which all pose an increased risk of mortality in this group (Feinglass et al., 2015; Kaffashian et al., 2003; Lundqvist et al., 2016; Rawshani et al., 2016; Yu, 2009).

6.3 Strengths and limitations

The results represented in this thesis were obtained from the organised Finnish National Breast Cancer Screening Programme (FNBCSP), which covers nearly 100% of the target age population (50-69 years). The unique personal identifier allows accurate linkage of data available at the comprehensive national registers. The collection of information on screening invitations, mammography tests, further assessments by the mass screening registry, and cancer outcomes for the whole population by the cancer registry provides several major advantages to conducting research in the Finnish programme. The collection of information on breast symptoms reported during the mammography screening visits allows for studying the impact of symptoms not yet available or published in any other programme.

No prior studies have estimated the risk of breast cancer, false-positive test findings and cancer mortality in relation to breast symptoms in a prospective manner. Follow-ups were done for a maximum of 24 years and a maximum of 10 screening visits after screening visits with or without symptoms. The study used screening history and symptom history information of all screening-age women in Finland who had biennial screening mammography, and followed up for further assessments until cancer diagnosis and death. The main benefit of using data on breast symptoms reported within the screening programme is that one can study cancer incidence and deaths from breast cancer in symptomatic women as compared to non-symptomatic. Additionally, harms in terms of false-positive findings were assessed for the first time in this study.

One key strength of this study is that follow-ups were performed until the date of the last screening visit or the diagnosis of cancer or death, whichever came earliest. The individual-level data allowed us to extract screening history information of every woman who presented with symptoms or who did not have symptoms. This is unique in the Finnish programme.

In studies III and IV, we made the symptomatic and asymptomatic group more comparable by matching with the background variables at any given period of the screening visits. This minimized the bias in the risk estimates by the confounders. As we used the full screening cohort, the study power was sufficient to compute the difference in risk and rates for symptomatic and asymptomatic visits. Because not all symptoms are equally sensitive, each symptom was analysed independently thus it was possible to analyse the risk of cancer and mortality for every individual symptom.

It is important to ensure that women with symptoms have equal opportunity to participate in the screening programme as those without symptoms. This means that

the screening effectiveness is to be evaluated based on all eligible women invited, regardless of any personal or risk history. Such evaluation reflects the actual scenario in the existing screening-age population. However, this study also highlights the importance of differential follow-up assessment of women who have a differential risk.

The definitions of the programme process and outcome indicators used in these studies adhere to the WHO-IARC and EU guidelines, as well as other international guidelines on breast cancer screening and diagnosis (Perry et al., 2006; IARC, 2016). Thus, future studies conducted in other screening programmes could utilize our study methods and findings. Furthermore, the findings of our studies could be applied to other existing and new programmes to improve the performance and effectiveness of screening mammography.

The studies contained several limitations.

One of the potential limitations is the collection of symptoms information. The collection was based on the women's self-reporting and by radiographers at the screening visits. The radiographers' physical inspection of the breast is likely to be less comprehensive than a full clinical examination. Information on symptoms is mainly collected in order to support the interpretation of the mammograms. However, in most of the cases, the presence of symptoms is confirmed by having the radiographer examine the breasts before the mammography is performed. This supports that the collected symptom information is valid, albeit not perfect.

Women with symptoms were possibly more likely to attend mammography screening than asymptomatic women leading to a self-selection bias. However, about 84% of all invited women participate in mammography screening in the Finnish programme, which is the highest among any existing mammography screening programme. In addition, most of the women who attended were asymptomatic (about 97%), and thus the attendance bias caused by symptoms is likely to be small.

Another issue is the recall bias. We do not know about any possible delay in presentation of symptoms: whether women waited for their first invitation (i.e., at age 50 years) or a subsequent screening invitation. The symptom information is based on the women's reporting of symptoms in the past two to six months and the examination by the radiographer at the screening visit. Women were more likely to remember any recent abnormalities in their breasts, and thus the recall bias is not likely to be of higher importance.

An important limitation is the inability to address the potential confounding effect of important risk factors that do not exist in the database. Thus, effect adjustment by factors such as family history, breast density, hormone use and socio-

economic status, which are known to influence breast cancer risk, was not possible in this study (Anderson et al., 2014; Barnard et al., 2015; Collaborative Group on Hormonal Factors in Breast Cancer, 2001; IARC, 2012a; IARC, 2012b).

The estimate of sensitivity, specificity and positive predictive value (PPV) of diagnosing breast cancer cases with symptoms in study I was limited to those women who attended mammography screening. Cancer cases detected outside of screening were excluded. The low sensitivity of symptoms in this study indicate that a mammography screening programme is still justified.

Study III was sensitive to lead-time bias and overdiagnosis because we used detection method instead of the use of background incidence of breast cancer to estimate the interval cancer rate in the absence of screening. However, it was not possible to find a comparable non-screened group as the Finnish screening programme has a high coverage (almost 100%) and attendance rate (about 84%). As well, there was no possibility to estimate the background incidence of breast cancer in women with symptoms. In addition, the positive predictive value (PPV) of mammography did not differ between the symptomatic and asymptomatic groups between the first and subsequent screening round, and thus the lead time bias due to prevalent screenings is negligible. Furthermore, our estimates of incidence rates and hazard ratios were based on the analysis of invasive breast cancers and on advanced and fatal breast cancers, both of which are less affected by overdiagnosis. Nonetheless, the proportion of in situ carcinomas was only 5% in those with symptoms.

In study IV, we analysed the difference in socio-economic statuses between symptomatic and asymptomatic women who were diagnosed or died of cancer. In women who reported a lump, we observed extra all-cause death cases in lower socio-economic class than in upper classes and also an extra number of deaths as compared to those without a lump. However, the socio-economic status of the whole cohort was unknown; thus, the proportion of breast cancer deaths or all-cause deaths in those with or without symptoms might differ from the study estimates. The higher all-cause mortality rates than breast cancer mortality rates, and also the difference in the rates between symptomatic and asymptomatic group throughout the follow-up period might be explained partially by the difference in socio-economic status.

6.4 Clinical and public health implications

The breast symptoms in this study provide new evidence on the benefits and harms of screening mammography. The study provides novel data on the risk of breast cancer (screen-detected and interval cancers) and mortality from breast cancer in relation to breast symptoms. The symptom status should not be restricted to support the interpretation of the mammography, but needs to be extended to improve programme performance. Our findings can be applied to various steps in the programme, from the initial design to performance assessment and up to the evaluation of outcomes.

The design of the Finnish mammography-screening programme, where invitations include all women of a certain age (including those with symptoms) differs substantially from many other programmes. Women presenting with symptoms are still considered part of the screening programme in Finland, whereas many screening programmes refer symptomatic women directly to hospitals or special breast clinics. However, clinical check-up may still miss cancers and may not offer optimal opportunities for systematic follow-up as the screening programme. An important consequence of excluding women with symptoms is the decreased validity of programme evaluation, since a considerable proportion of women with symptoms are left out. Based on our findings, this means that a prominent number of cancer cases are missed. In addition, assessing symptomatic women outside the programme may demand additional resources (Walker et al., 2014).

The next step is an appropriate inquiry into symptoms information by the radiologists or nurses with additional information, such as duration of symptoms and severity of pain (Walker et al., 2014) or other possible findings based on palpation (Breast Test Wales, 1996; Hide et al., 1999; Litherland et al., 2001). The findings in study III showed a higher screening test and episode sensitivity in women presenting with symptoms compared to asymptomatic women. Particularly, of women who reported a lump, loss occurred in the episode sensitivity in the further assessment, meaning that all women presenting with symptoms were not recalled. On the other hand, further assessment of all women with symptoms possibly leads to additional false-positive cases (study II), lower efficacy of the programme, and anxiety or other psychological distress among the women. In addition, variation in the cancer detection rates or false-positive findings by symptom status implies that not all symptoms are equally sensitive and may in turn lower the positive predictive value of recall. As about 97% of the screening visits were asymptomatic, the number of additional further assessment services would be rather small, as well as would yield

a small improvement in the programme's overall performance and outcome. Still, high quality and clinically appropriate services are important for women presenting with symptoms at screening mammography. The potential of new imaging techniques such as digital breast tomosynthesis to improve diagnostic accuracy of imaging in the cases of symptoms (primarily a lump) could help to decrease false positive recalls if no abnormality is seen on 3-dimensional mammography. Especially in cases of mammograpically dense breast, tomosynthesis has been found to increase cancer detection rate by revealing a tumor that is hidden in the dense tissue (Conant et al., 2019; Skaane et al., 2019). So far, tomosynthesis is used in Finland as part of further assessment but not as part of primary screening technique. A health economic assessment of providing such a comprehensive services to all women presenting with symptoms should be made.

The higher incidence of interval cancers in symptomatic women within 6 months after a negative mammography is a huge concern for radiologists. This indicates that further assessment would be needed if symptoms are present even though this will result in loss of specificity. In addition, because of the high incidence of advanced interval breast cancers before the next screening invitation and higher risk of cancer diagnosis in the subsequent screening round in the symptomatic women, a potential improvement in further assessment would not be clearly sufficient. Furthermore, the programme lacks surveillance or follow-up of symptomatic women who did not undergo further assessment, as these women are more prone to seek out private mammography. The results from study IV showed a substantial difference in the mortality rates within five years after visits with and without symptoms. Women need to be better informed and made aware that if a symptom occurs, it is inadvisable to wait until the next invitation to the programme. Improved guidelines on further assessment and a reduction in the screening interval for women presenting with symptoms as suggested by our study might help to reduce the screening inequities and improve the performance of the mammography screening programme.

6.5 Summary and conclusions

The importance of breast symptoms reported by women during the screening visit was evaluated within the population-based mammography screening programme for breast cancer. The study used the whole cohort of women with or without breast symptoms since the early years of the screening programme (from 1992) until the most recent year available (2014). Thus, the results are directly applicable to routine use. Finland's three health care registries were used to obtain information on symptoms and associated cancer outcomes, including time and cause of death.

Our study showed a strong positive association between breast symptoms and the risk of screen-detected cancers, interval cancers and breast cancer mortality as compared to screening visits without symptoms. The programme performance indicators such as sensitivity and positive predictive value (PPV) were higher in those who reported symptoms as compared to those without symptoms at screening visit. However, we also found a higher risk of false-positive test results in the symptomatic women.

Of the three most common breast symptoms, a lump was a strong predictor of breast cancer and of breast cancer mortality and all-cause mortality. Screening visits with a retraction or nipple discharge were also increasingly associated with the outcomes.

We found that in screening visits with two symptoms, a lump and retraction, the cancer risk was multiplicative. However, we did not find an age-related increase in the cancer risk, except for the symptom of nipple discharge. Given the limited sensitivity and specificity of breast symptoms, prevention programmes based on clinical examination only would not provide a sufficient benefit for breast cancer control. This reinforces the importance of a mammography screening programme in resourceful settings like Finland.

One important harm is the large numbers of false-positive test findings in symptomatic women as compared to asymptomatic women. This creates challenges for radiologists to correctly identify abnormalities in the mammogram that are likely to develop to a malignant tumour and decide whether or not to recall the woman. This should be carefully considered to maintain a balance of benefits and harms of screening for individual women.

Our study showed that the risk of screen-detected and interval breast cancers was higher in those who reported any of the three symptoms. This indicates that all three symptoms are clinically important in detecting cancers. A better diagnostic

workup is needed in the symptomatic women, including detailed indication for biopsies in the further assessment phase.

One of the key findings is that interval breast cancer incidence increased rather rapidly after the negative screening episode in women with symptoms. This is a clear concern for the programme and indicates that these women should have further contact with health care centres due to concern or anxiety. Also, the high incidence of advanced interval breast cancers shortly after a negative episode indicates the need for a follow-up visit shortly after the negative examination, especially in women with a lump or nipple discharge.

The cumulative breast cancer mortality and all-cause mortality pattern remained higher in the symptomatic group throughout the long follow-up time available in our study. The substantial high mortality rates within a period of five years after the screening visit with symptoms reinforces the need for a detailed further assessment and biopsy recommendation. Continuous efforts to increase awareness in women about breast symptoms and encouraging them to seek care even prior to invitations are very important. Improvement in the guidelines on screening and clinical services could reduce inequalities among these women. However, we could not assess the socio-economic differences in the mortality between the symptomatic and asymptomatic groups. Understanding the detailed background information as well as studying or analysing the detailed screening procedure in symptomatic women will help answer most of the questions that this study was not able to.

6.6 Future perspectives

As is often the case, while trying to answer one question we ended up asking more. One important issue to address in the future is how to optimally use information on breast symptoms in the breast cancer-screening programme:

- From the programme perspective, information on breast symptoms needs to be collected more uniformly and symptoms should be asked about during mammography screening visits. An important challenge is the tender process; hence, collecting detailed information on symptoms is an additional administrative burden.
- Future studies should also focus on other important risk factors such as breast density, use of hormones, and family history of breast cancer, all of which might provide insight into the association between symptoms and cancer risk. In addition, studying the difference in socio-economic status between symptomatic and asymptomatic women might help us to understand the association more broadly.
- Also, it would be useful to study the histological classification of tumours, tumour size, grade and lymph node involvement in symptomatic women who were diagnosed with breast cancers and then compare them to asymptomatic women with otherwise similar characteristics.
- More women with breast symptoms should be recalled for further assessment, including appropriate core biopsy. One option would be to have a shorter screening interval, especially for women who report a lump but are not referred for surgery. The impact of such strategies, if implemented, should be closely monitored and assessed.
- National screening guidelines on women presenting with breast symptoms need to be more specific and clearer. The evidence from our study could be used to improve EU guidelines or any international guidelines on screening and further assessment, specifically for those women who present with breast symptoms at screening visits.
- Recent studies have shown that immigrant women have lower mammography screening attendance than native women but similar incidence of breast cancer. Those women might also have higher prevalence of symptoms, thus leading to higher risk of breast cancer. Future studies need to be conducted among those target groups as well.

Much is to be done in the field of breast cancer epidemiology, specifically concerning women with breast symptoms. In general, this extends beyond Finland

to other high-incidence countries that have mammography screening programmes. Our study is an initial example that shows the importance of collecting and analysing information on breast symptoms by the Finnish programme, as well as the importance of not excluding women from invitation to screening. The study findings have clearly shown that other countries could benefit from collecting information on symptoms and from following women thereafter. Our findings are useful to countries that have recently started screening programmes, especially upper middle income countries. Of equal importance is the inclusion of all women in the target age groups in screening and to not exclude those with symptoms.

REFERENCES

Aiello, E. J., Buist, D. S., White, E., Seger, D., & Taplin, S. H. (2004). Rate of breast cancer diagnoses among postmenopausal women with self-reported breast symptoms. J Am Board Fam Med, 17(6), 408-415. doi: https://doi.org/10.3122/jabfm.17.6.408

Allen, N. E., Beral, V., Casabonne, D., Kan, S. W., Reeves, G. K., Brown, A., . . . Million Women Study Collaborators. (2009). Moderate alcohol intake and cancer incidence in women. JNCI, 101(5), 296-305. doi:10.1093/jnci/djn514

Anderson, H., Bladström, A., Olsson, H., & Möller, T. R. (2000). Familial breast and ovarian cancer: A Swedish population-based register study. Am J Epidemiology, 152(12), 1154-1163.

Anderson, K. N., Schwab, R. B., & Martinez, M. E. (2014). Reproductive risk factors and breast cancer subtypes: A review of the literature. Breast Cancer Research and Treatment, 144(1), 1-10. doi:10.1007/s10549-014-2852-7

Anderson, W. F., Pfeiffer, R. M., Dores, G. M., & Sherman, M. E. (2006). Comparison of age distribution patterns for different histopathologic types of breast carcinoma. Cancer Epidemiology, Biomarkers & Prevention, 15(10), 1899-1905. doi:15/10/1899

Andersson, I., Aspegren, K., Janzon, L., Landberg, T., Lindholm, K., Linell, F., . . . Sigfusson, B. (1988). Mammographic screening and mortality from breast cancer: The malmo mammographic screening trial. BMJ (Clinical Research Ed.), 297(6654), 943-948.

Anttila, A., Koskela, J., & Hakama, M. (2002). Programme sensitivity and effectiveness of mammography service screening in Helsinki, Finland. Journal of Medical Screening, 9(4), 153-158. doi:10.1136/jms.9.4.153

Arbyn, M., Anttila, A., Jordan, J., Ronco, G., Schenck, U., Weiner, H., . . . von Karsa, L. (2010). European Guidelines for Quality Assurance in Cervical Cancer Screening. Second Edition- Summary Document. Ann Oncol., 21(3): 448-458. doi: 10.1093/annonc/mdp471

Arnold, M., Pandeya, N., Byrnes, G., Renehan, P. A. G., Stevens, G. A., Ezzati, P. M., . . . Soerjomataram, I. (2015). Global burden of cancer attributable to high body-mass index in 2012: A population-based study. The Lancet Oncology, 16(1), 36-46. doi:10.1016/S1470-2045(14)71123-4

Ascunce, E. N., Moreno-Iribas, C., Barcos Urtiaga, A., Ardanaz, E., Ederra Sanz, M., Castilla, J., & Egues, N. (2007). Changes in breast cancer mortality in navarre (spain) after introduction of a screening programme. Journal of Medical Screening, 14(1), 14-20. doi:10.1258/096914107780154558

Autier, P., Hery, C., Haukka, J., Boniol, M., & Byrnes, G. (2009). Advanced breast cancer and breast cancer mortality in randomized controlled trials on mammography screening. Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology, 27(35), 5919-5923. doi:10.1200/JCO.2009.22.7041

Barlow, W. E., Lehman, C. D., Zheng, Y., Ballard-Barbash, R., Yankaskas, B. C., Cutter, G. R., . . . Taplin, S. H. (2002). Performance of diagnostic mammography for women

with signs or symptoms of breast cancer. Journal of the National Cancer Institute, 94(15), 1151-1159.

Barnard, M. E., Boeke, C. E., & Tamimi, R. M. (2015). Established breast cancer risk factors and risk of intrinsic tumor subtypes. Biochimica et Biophysica Acta, 1856(1), 73-85. doi:10.1016/j.bbcan.2015.06.002

Beral, V., Million Women Study Collaborators. (2003). Breast cancer and hormone-replacement therapy in the Million Women Study. Lancet, 362(9386), 419-27.

Bick, U., Engel, C., Krug, B., Heindel, W., Fallenberg, E.M., Rhiem, K., . . . German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC). (2019). High-risk breast cancer surveillance with MRI: 10-year experience from the German consortium for hereditary breast and ovarian cancer. Breast Cancer Res Treat. doi: 10.1007/s10549-019-05152-9

Bjurstam, N., Björneld, L., Warwick, J., Sala, E., Duffy, S. W., Nyström, L., . . . Wahlin, T. (2003). The Gothenburg breast screening trial. Cancer, 97(10), 2387-2396. doi:10.1002/cncr.11361

Blanch, J., Sala, M., Ibáñez, J., Domingo, L., Fernandez, B., Otegi, A., . . . for the INCA Study Group. (2014). Impact of risk factors on different interval cancer subtypes in a population-based breast cancer screening programme. PLoS ONE, 9(10), e110207. doi:10.1371/journal.pone.0110207

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, doi:10.3322/caac.21492

Breast Test Wales. (1996). Quality manuals. Retrieved from http://www.breasttestwales.wales.nhs.uk/reports-1

Brouckaert, O., Schoneveld, A., Truyers, C., Kellen, E., Van Ongeval, C., Vergote, I., . . . MBC Leuven, B. (2013). Breast cancer phenotype, nodal status and palpability may be useful in the detection of overdiagnosed screening-detected breast cancers. Annals of Oncology, 24(7), 1847-1852. doi:10.1093/annonc/mdt179

Bucchi, L., Ravaioli, A., Foca, F., Colamartini, A., Falcini, F., & Naldoni, C. (2008). Incidence of interval breast cancers after 650,000 negative mammographics in 13 Italian health districts. J Med Screen, 15(1), 30-35. doi:10.1258/jms.2008.007016

Buerger, H., Mommers, E. C., Littmann, R., Simon, R., Diallo, R., Poremba, C., . . . Boecker, W. (2001). Ductal invasive G2 and G3 carcinomas of the breast are the end stages of at least two different lines of genetic evolution. The Journal of Pathology, 194(2), 165-170. doi:10.1002/path.875

Buerger, H., Otterbach, F., Simon, R., Poremba, C., Diallo, R., Decker, T., . . . Boecker, W. (1999). Comparative genomic hybridization of ductal carcinoma in situ of the breast-evidence of multiple genetic pathways. The Journal of Pathology, 187(4), 396-402. doi:10.1002/(SICI)1096-9896(199903)187:4<396::AID-PATH286>3.0.CO;2-L

Bulliard, J. L., Sasieni, P., Klabunde, C., De Landtsheer, J. P., Yankaskas, B. C., & Fracheboud, J. (2006). Methodological issues in international comparison of interval breast cancers. Int J Cancer, 119(5), 1158-1163. doi:10.1002/ijc.21941

Cancer Australia. (2017). Investigation of a new breast symptom - a guide for general practitioners| Cancer Australia. Retrieved from https://canceraustralia.gov.au/publications-and-resources/cancer-australia-publications/investigation-new-breast-symptom-guide-general-practitioners

Catsburg, C., Miller, A. B., & Rohan, T. E. (2015). Active cigarette smoking and risk of breast cancer. Int J Cancer, 136(9), 2204-2209. doi:10.1002/ijc.29266

Chen, L., Zhou, W. B., Zhao, Y., Liu, X. A., Ding, Q., Zha, X. M., & Wang, S. (2012). Bloody nipple discharge is a predictor of breast cancer risk: A meta-analysis. Breast Cancer Research and Treatment, 132(1), 9-14. doi:10.1007/s10549-011-1787-5

Chiu, S. Y., Duffy, S., Yen, A. M., Tabár, L., Smith, R. A., & Chen, H. H. (2010). Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening. Cancer Epidemiology, Biomarkers & Prevention, 19(5), 1219-1228. doi:10.1158/1055-9965.EPI-09-1028

Christiansen, C. L., Wang, F., Barton, M. B., Kreuter, W., Elmore, J. G., Gelfand, A. E., & Fletcher, S. W. (2000). Predicting the cumulative risk of false-positive mammograms. Journal of the National Cancer Institute, 92(20), 1657-1666.

Clinical Radiology. (2013). Guidance on screening and symptomatic breast imaging, third edition. Retrieved from https://www.rcr.ac.uk/publication/guidance-screening-and-symptomatic-breast-imaging-third-edition

Colditz, G. A., Atwood, K. A., Emmons, K., Monson, R. R., Willett, W. C., Trichopoulos, D., & Hunter, D. J. (2000). Harvard report on cancer prevention volume 4: Harvard cancer risk index. risk index working group, Harvard Center for Cancer Prevention. Cancer Causes & Control, 11(6), 477-488.

Coldman, A., Phillips, N., Wilson, C., Decker, K., Chiarelli, A. M., Brisson, J., . . . Ahmad, R. (2014). Pan-canadian study of mammography screening and mortality from breast cancer. Journal of the National Cancer Institute, 106(11). doi:10.1093/jnci/dju261

Collaborative Group on Hormonal Factors in Breast Cancer. (2001). Familial breast cancer: Collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet, 358(9291), 1389-1399. doi:S0140-6736(01)06524-2

Collaborative Group on Hormonal Factors in Breast Cancer. (2002). Breast cancer and breastfeeding: Collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet, 360(9328), 187-195. doi:S0140-6736(02)09454-0

Collaborative Group on Hormonal Factors in Breast Cancer. (2012). Menarche, menopause, and breast cancer risk: Individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. The Lancet Oncology, 13(11), 1141-1151. doi:10.1016/S1470-2045(12)70425-4

Conant, E.F., Barlow, W.E., Herschorn, S.D., Weaver, D.L., Beaber, E.F., Tosteson, A.N.A., . . . The Population-based Research Optimizing Screening Through Personalized Regimen (PROSPR) Consortium. (2019). Association of Digital Breast Tomosynthesis vs Digital Mammography With Cancer Detection and Recall Rates by Age and Breast Density. JAMA Oncol. doi:10.1001/jamaoncol.2018.7078

Cowan, W., Angus, B., Henry, J., Corbett, I., Reid, W., & Horne, C. (1991). Immunohistochemical and other features of breast carcinomas presenting clinically compared with those detected by cancer screening. British Journal of Cancer, 64(4), 780.

Cui, Y., Miller, A. B., & Rohan, T. E. (2006). Cigarette smoking and breast cancer risk: Update of a prospective cohort study. Breast Cancer Research and Treatment, 100(3), 293-299. doi:10.1007/s10549-006-9255-3

Day, N. E. (1985). Estimating the sensitivity of a screening test. Journal of Epidemiology and Community Health, 39(4), 364-366.

de Gelder, R., Heijnsdijk, E. A., van Ravesteyn, N. T., Fracheboud, J., Draisma, G., & de Koning, H. J. (2011). Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiologic Reviews, 33, 111-121. doi:10.1093/epirev/mxr009

Department of Health. (2010). Best practice diagnostic guidelines for patients presenting with breast symptoms. Retrieved from https://www.uhb.nhs.uk/breast-cancer.htm

DeSantis, C., Ma, J., Bryan, L., & Jemal, A. (2014). Breast cancer statistics, 2013. CA: A Cancer Journal for Clinicians, 64(1), 52-62. doi:10.3322/caac.21203

Domingo, L., Salas, D., Zubizarreta, R., Bare, M., Sarriugarte, G., Barata, T., . . . INCA Study Group. (2014). Tumor phenotype and breast density in distinct categories of interval cancer: Results of population-based mammography screening in Spain. Breast Cancer Research, 16(1), R3. doi:10.1186/bcr3595

Duffy, S. W., Chen, H. H., Tabár, L., & Day, N. E. (1995). Estimation of mean sojourn time in breast cancer screening using a markov chain model of both entry to and exit from the preclinical detectable phase. Statistics in Medicine, 14(14), 1531-1543.

Duffy, S. W., & Parmar, D. (2013). Overdiagnosis in breast cancer screening: The importance of length of observation period and lead time. Breast Cancer Research, 15(3), R41. doi:10.1186/bcr3427

Elmore, J. G., Reisch, L. M., Barton, M. B., Barlow, W. E., Rolnick, S., Harris, E. L., . . . Fletcher, S. W. (2005). Efficacy of breast cancer screening in the community according to risk level. Journal of the National Cancer Institute, 97(14), 1035-1043. doi:97/14/1035

Engholm, G., Ferlay, J., Christensen, N., Hansen, H.L., Hertzum-Larsen, R., Johannesen, T.B., . . . Storm, H.H. (2018). Cancer incidence, mortality, prevalence and survival in the Nordic countries. Retrieved from http://www-dep.iarc.fr/NORDCAN/English/frame.asp

Erban, J. K., Smith, B. L., & Taghian, A. G. (2010). Breast cancer : A multidisciplinary approach to diagnosis and management. New York: Demos Medical. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,ip,uid&db=nlebk&AN=320000&site=ehost-live&scope=site&authtype=sso&custid=s4778523

Perry, N., Broeders, M., de Wolf, C., Törnberg, S., Holland, R., von Karsa, L. (2006). European guidelines for quality assurance in breast cancer screening and diagnosis. Luxembuourg: Office of the Official Publications of the European Communities. Retrieved from http://ec.europa.eu/health/ph_projects/2002/cancer/fp_cancer_2002_ext_guid_01.pdf

Evans, A. J. (2001a). Detecting which invasive cancers at mammographic screening saves lives? J Med Screen, 8(2), 86-90. doi:10.1136/jms.8.2.86

Evans, A. J. (2001b). Screen detected ductal carcinoma in situ (DCIS): Overdiagnosis or an obligate precursor of invasive disease? J Med Screen, 8(3), 149-151. doi:10.1136/jms.8.3.149

Evans, A. J., Pinder, S. E., Snead, D. R., Wilson, A. R., Ellis, I. O., & Elston, C. W. (1997). The detection of ductal carcinoma in situ at mammographic screening enables the diagnosis of small, grade 3 invasive tumours. British Journal of Cancer, 75(4), 542-544.

Feinglass, J., Rydzewski, N., & Yang, A. (2015). The socioeconomic gradient in all-cause mortality for women with breast cancer: Findings from the 1998 to 2006 national cancer database with follow-up through 2011. Annals of Epidemiology, 25(8), 549-555. doi:10.1016/j.annepidem.2015.02.006

Fenton, J. J., Barton, M. B., Geiger, A. M., Herrinton, L. J., Rolnick, S. J., Harris, E. L., . . . Elmore, J. G. (2005). Screening clinical breast examination: How often does it miss lethal breast cancer? Journal of the National Cancer Institute. Monographs, (35):67-71. doi(35), 67-71. doi:2005/35/67

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., . . . Bray, F. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer, 136(5), E359-86. doi:10.1002/ijc.29210

Finnish Cancer Registry. (2017). Breast cancer screening programme in Finland. Annual Review. Retrieved from https://syoparekisteri.fi/assets/files/2018/02/Breast-cancer-screening-programme-Annual-review-2017.pdf

Finnish Cancer Registry. (2018a). Breast Cancer Screening. Retrieved from https://cancerregistry.fi/screening/breast-cancer-screening/

Finnish Cancer Registry. (2018b). Breast Cancer Screening; result of breast cancer screening. Retrieved from https://www.freefromcancer.fi/check-your-body/take-part-cancer-screening/

Finnish Cancer Registry. (2018c). Cancer statistics. Retrieved from https://cancerregistry.fi/statistics/cancer-statistics/

Forman, D., Bray, F., Brewster, D., Gombe, C., Kohler B, Piñeros, M., . . . Ferlay, J. (2014). Cancer Incidence in Five Continents Vol. X, Lyon, France. Retrieved from http://ci5.iarc.fr/CI5I-X/Default.aspx

Forouzanfar, M. H., Foreman, K. J., Delossantos, A. M., Lozano, R., Lopez, A. D., Murray, C. J., & Naghavi, M. (2011). Breast and cervical cancer in 187 countries between 1980 and 2010: A systematic analysis. Lancet, 378(9801), 1461-1484. doi:10.1016/S0140-6736(11)61351-2

Fracheboud, J., Koning, H. J., Beemsterboer, P. M. M., Boer, R., Verbeek, A. L. M., Hendriks, J. H. C. L., . . . Maas, P. J. (1999). Interval cancers in the dutch breast cancer screening programme. British Journal of Cancer, 81(5), 912-917. doi:10.1038/sj.bjc.6690786

Fritz A, Percy C, Jack J, Shanmugaratnam K, Sobin L, Parkin MD, Whelan S. (2000). International classification of diseases for oncology (ICD-O-3) - WHO - third edition. Retrieved from http://apps.who.int/iris/bitstream/handle/10665/96612/9789241548496_eng.pdf;jsessionid=A87D543D900372D6DCA85735836EE566?sequence=1

Gøtzsche, P. C., Hartling, O. J., Nielsen, M., Brodersen, J., & Jørgensen, K. J. (2009). Breast screening: The facts or maybe not. BMJ, 338 doi:10.1136/bmj.b86

Georgian-Smith, D., & Lawton, T. J. (2014). Breast imaging and pathologic correlations: A pattern-based approach. First-edition, Wolters Kluwer Health, China.

Giess, C. S., Keating, D. M., Osborne, M. P., Ng, Y. Y., & Rosenblatt, R. (1998). Retroareolar breast carcinoma: Clinical, imaging, and histopathologic features. Radiology, 207(3), 669-673. doi:10.1148/radiology.207.3.9609889

Gill, G., Luke, C., & Roder, D. (2006). Prognostic importance of palpability as a feature of screen-detected breast cancers. Journal of Medical Screening, 13(2), 98-101. doi:10.1258/096914106777589560

Gilliland, F. D., Joste, N., Stauber, P. M., Hunt, W. C., Rosenberg, R., Redlich, G., & Key, C. R. (2000). Biologic characteristics of interval and screen-detected breast cancers. Journal of the National Cancer Institute, 92(9), 743-749.

Globocan. (2012). Globocan - Home. Retrieved from http://globocan.iarc.fr/default.aspx

Hakama, M., Auvinen, A., Day, N. E., & Miller, A. B. (2007). Sensitivity in cancer screening. Journal of Medical Screening, 14(4), 174-177. doi:10.1258/096914107782912077

Hakama, M., Pokhrel, A., Malila, N., & Hakulinen, T. (2015). Sensitivity, effect and overdiagnosis in screening for cancers with detectable pre-invasive phase. Int J Cancer, 136(4), 928-935. doi:10.1002/ijc.29053

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

Hartmann, L. C., Sellers, T. A., Frost, M. H., Lingle, W. L., Degnim, A. C., Ghosh, K., . . . Visscher, D. W. (2005). Benign breast disease and the risk of breast cancer. NEJM, 353(3), 229-237. doi:353/3/229

Harvey, S. C., Geller, B., Oppenheimer, R. G., Pinet, M., Riddell, L., & Garra, B. (2003). Increase in cancer detection and recall rates with independent double interpretation of screening mammography. American Journal of Roentgenology, 180(5), 1461-1467.

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

Henderson, T. O., Amsterdam, A., Bhatia, S., Hudson, M. M., Meadows, A. T., Neglia, J. P., . . . Oeffinger, K. C. (2010). Surveillance for breast cancer in women treated with chest radiation for a childhood, adolescent or young adult cancer: A report from the children's oncology group. Annals of Internal Medicine, 152(7), 444-W154. doi:10.1059/0003-4819-152-7-201004060-00009

Hide, I. G., McLean, L., Simpson, W., & Neilson, F. (1999). Women with normal mammography describing symptoms at screening doi:https://doi.org/10.1054/brst.1999.0043

Hofvind, S., Ursin, G., Tretli, S., Sebuødegård, S., Møller, B. (2013). Breast cancer mortality in participants of the Norwegian Breast Cancer Screening Program. Cancer, 119(17), 3106-3112.

Hofvind, S., Kaitlyn, T., Gunhild, M., Ertzaas, AK., Holen, A.S., Pedersen K. (2017). The Norwegian breast cancer screening program, 1996-2016: Celebrating 20 years of organized mammography screening. in: Cancer in Norway 2016. Retrieved from https://www.kreftregisteret.no/Generelt/Publikasjoner/Mammografiprogrammet/

Hofvind, S., Ponti, A., Patnick, J., Ascunce, N., Njor, S., Broeders, M., . . . Suonio, E. (2012). False-positive results in mammographic screening for breast cancer in Europe: A literature review and survey of service screening programmes. Journal of Medical Screening, 19 Suppl 1, 57-66. doi:19/suppl_1/57 [pii]

Hofvind, S., Wang, H., & Thoresen, S. (2004). Do the results of the process indicators in the Norwegian breast cancer screening program predict future mortality reduction from breast cancer? Acta Oncologica, 43(5), 467-473.

Houssami, N., Irwig, L., & Ciatto, S. (2006). Radiological surveillance of interval breast cancers in screening programmes. The Lancet Oncology, 7(3), 259-265. doi:S1470-2045(06)70617-9

IARC. (2016). IARC Handbook on Breast Cancer Screening, Lyon, France, IARC Press. Retrieved from http://publications.iarc.fr/Book-And-Report-Series/Iarc-Handbooks-Of-Cancer-Prevention/Breast-Cancer-Screening-2016

IARC. (2012a). Pharmaceuticals. Retrieved from https://monographs.iarc.fr/corrigenda/

IARC. (2012b). Personal habits and indoor combustions. Retrieved from https://monographs.iarc.fr/corrigenda/

IARC. (2002). IARC Handbook on Breast Cancer Screening, Lyon, France, IARC Press.

Islami, F., Liu, Y., Jemal, A., Zhou, J., Weiderpass, E., Colditz, G., . . . Weiss, M. (2015). Breastfeeding and breast cancer risk by receptor status--a systematic review and meta-analysis. Annals of Oncology : Official Journal of the European Society for Medical Oncology, 26(12), 2398-2407. doi:10.1093/annonc/mdv379

Jacklyn, G., Glasziou, P., Macaskill, P., Barratt, A. (2016). Meta-analysis of breast cancer mortality benefit and overdiagnosis adjusted for adherence: improving information on the effects of attending screening mammography. British J Cancer, 114, 1269-1276.

Joensuu,H., Roberts, P., Lehtinen, P.K., Jyrkkiö, S., Kouri, M., Teppo, L. (2013). Syöpätaudit. Helsinki: Kustannus Oy Duodecim.

Jonsson, H., Törnberg, S., Nyström, L., & Lenner, P. (2003). Service screening with mammography of women aged 70-74 years in Sweden. Effects on breast cancer mortality. Cancer Detection and Prevention, 27(5), 360-369. doi:S0361090X03001314

Kaffashian, F., Godward, S., Davies, T., Solomon, L., McCann, J., & Duffy, S. W. (2003). Socioeconomic effects on breast cancer survival: Proportion attributable to stage and morphology. British J Cancer, 89(9), 1693-1696. doi:10.1038/sj.bjc.6601339 [doi]

Kalager, M., Zelen, M., Langmark, F., & Adami, H. O. (2010). Effect of screening mammography on breast-cancer mortality in Norway. NEMJ, 363(13), 1203-1210. doi:10.1056/NEJMoa1000727

Kerlikowske, K., Grady, D., Barclay, J., Sickles, E. A., & Ernster, V. (1996). Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA, 276(1), 33-38.

Kerlikowske, K., Ichikawa, L., Miglioretti, D. L., Buist, D. S., Vacek, P. M., Smith-Bindman, R., . . . National Institutes of Health Breast Cancer Surveillance Consortium. (2007). Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk. JNCI, 99(5), 386-395. doi:99/5/386

Key, T. J., Verkasalo, P. K., & Banks, E. (2001). Epidemiology of breast cancer. Lancet Oncol, 2(3), 133-140. doi:S1470-2045(00)00254-0

Kirsh, V. A., Chiarelli, A. M., Edwards, S. A., O'Malley, F. P., Shumak, R. S., Yaffe, M. J., & Boyd, N. F. (2011). Tumor characteristics associated with mammographic detection of breast cancer in the ontario breast screening program. JNCI, 103(12), 942-950. doi:10.1093/jnci/djr138

Klemi, P. J., Joensuu, H., Toikkanen, S., Tuominen, J., Räsänen, O., Tyrkkä, J., & Parvinen, I. (1992). Aggressiveness of breast cancers found with and without screening. BMJ : British Medical Journal, 304(6825), 467-469.

Klemi, P. J., Toikkanen, S., Räsänen, O., Parvinen, I., & Joensuu, H. (1997). Mammography screening interval and the frequency of interval cancers in a population-based screening. British J Cancer, 75(5), 762-766.

Lakhani, S.R., Ellis, I.O., Schnitt, S.J., Tan, P.H., van de Vijver, M.J. (2012). WHO Classification of Tumours of the Breast, 4th Edition Geneva.

Lampejo, O. T., Barnes, D. M., Smith, P., & Millis, R. R. (1994). Evaluation of infiltrating ductal carcinomas with a DCIS component: Correlation of the histologic type of the in situ component with grade of the infiltrating component. Seminars in Diagnostic Pathology, 11(3), 215-222.

Lauby-Secretan, B., Scoccianti, C., Loomis, D., Benbrahim-Tallaa, L., Bouvard, V., Bianchini, F., . . . International Agency for Research on Cancer Handbook Working Group. (2015). Breast-cancer screening--viewpoint of the IARC working group. NEMJ, 372(24), 2353-2358. doi:10.1056/NEJMsr1504363

Lekanidi, K., Dilks, P., Suaris, T., Kennett, S., & Purushothaman, H. (2017). Breast screening: What can the interval cancer review teach us? Are we perhaps being a bit too hard on ourselves? European Journal of Radiology, 94, 13-15. doi:S0720-048X(17)30289-9

Litherland, J. C., & Dobson, H. M. (2001). Symptomatic women with normal screening mammograms: Is assessment by the breast screening programme justifiable? doi:https://doi.org/10.1054/brst.2000.0218

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

Lumachi, F., Ermani, M., Brandes, A. A., Boccagni, P., Polistina, F., Basso, S. M., . . . D'Amico, D. F. (2002). Breast complaints and risk of breast cancer. population-based study of 2,879 self-selected women and long-term follow-up. Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie, 56(2), 88-92. doi:S0753-3322(01)00160-3

Lundqvist, A., Andersson, E., Ahlberg, I., Nilbert, M., & Gerdtham, U. (2016). Socioeconomic inequalities in breast cancer incidence and mortality in Europe- a systematic review and meta-analysis. The European Journal of Public Health, 26(5), 804-813. doi:10.1093/eurpub/ckw070

Mahoney, L., & Csima, A. (1982). Efficiency of palpation in clinical detection of breast cancer. Canadian Medical Association Journal, 127(8), 729-730.

Mansel R, Khonji N, Clarke D. (1999). Updating national guidelines for referral of breast problems. Retrieved from https://www.guidelinesinpractice.co.uk/updating-national-guidelines-for-referral-of-breast-problems/300417.article

Mass Screening Registry. (2018). Organising Cancer Screening - Syöpärekisteri. Retrieved from https://cancerregistry.fi/screening/organising-cancer-screening/

Miller, A. B., Baines, C. J., To, T., & Wall, C. (1992a). Canadian national breast screening study: 1. breast cancer detection and death rates among women aged 40 to 49 years. CMAJ : Canadian Medical Association Journal, 147(10), 1459-1476.

Miller, A. B., Baines, C. J., To, T., & Wall, C. (1992b). Canadian national breast screening study: 2. breast cancer detection and death rates among women aged 50 to 59 years. CMAJ: Canadian Medical Association Journal, 147(10), 1477-1488.

Miller, A. B., To, T., Baines, C. J., & Wall, C. (2002). The canadian national breast screening study-1: Breast cancer mortality after 11 to 16 years of follow-up: A randomized screening trial of mammography in women age 40 to 49 years. Annals of Internal Medicine. doi:10.7326/0003-4819-137-5_Part_1-200209030-00005

Mittra, I., Mishra, G. A., Singh, S., Aranke, S., Notani, P., Badwe, R., . . . Shastri, S. S. (2010). A cluster randomized, controlled trial of breast and cervix cancer screening in mumbai, india: Methodology and interim results after three rounds of screening. Int J Cancer, 126(4), 976-984. doi:10.1002/ijc.24840

Morrell, S., Taylor, R., Roder, D., Robson, B., Gregory, M., & Craig, K. (2017). Mammography service screening and breast cancer mortality in New Zealand: A national cohort study 1999-2011. British Journal of Cancer, 116(6), 828-839. doi:10.1038/bjc.2017.6

Morrison, A.S., Brisson, J., Khalid, N. (1988). Breast cancer incidence and mortality in the breast cancer detection demonstration project. JNCI, 80(19), 1540-1547.

Moskowitz, C. S., Chou, J. F., Wolden, S. L., Bernstein, J. L., Malhotra, J., Friedman, D. N., . . . Oeffinger, K. C. (2014). Breast cancer after chest radiation therapy for childhood cancer. Journal of Clinical Oncology, 32(21), 2217-2223. doi:44601

Moss, S., Thomas, I., Evans, A., Thomas, B., Johns, L., & Trial Management Group. (2005). Randomised controlled trial of mammographic screening in women from age 40: Results of screening in the first 10 years. British Journal of Cancer, 92(5), 949-954. doi:6602396

Moss, S. (1999). 16-year mortality from breast cancer in the UK trial of early detection of breast cancer doi:https://doi.org/10.1016/S0140-6736(98)07412-1

Mushlin, A. I., Kouides, R. W., & Shapiro, D. E. (1998). Estimating the accuracy of screening mammography: A meta-analysis. American Journal of Preventive Medicine, 14(2), 143-153. doi:S0749379797000196

Myers, E. R., Moorman, P., Gierisch, J. M., Havrilesky, L. J., Grimm, L. J., Ghate, S., . . . Sanders, G. D. (2015). Benefits and harms of breast cancer screening: A systematic review. JAMA, 314(15), 1615-1634. doi:10.1001/jama.2015.13183

National Board of Health. (2005). National Cancer Plan II - Denmark 2005 - National Board of Health recommendations for improving cancer healthcare services. Retrieved from https://www.sst.dk/da/udgivelser/2005/national-cancer-plan-ii-denmark-2005#

National Cancer Institute. (2018). Alcohol and cancer risk fact sheet - national cancer institute. Retrieved from https://www.cancer.gov/about-cancer/causes-prevention/risk/alcohol/alcohol-fact-sheet

Nelson, H. D., Fu, R., Cantor, A., Pappas, M., Daeges, M., & Humphrey, L. (2016). Effectiveness of breast cancer screening: Systematic review and meta-analysis to update the 2009 U.S. preventive services task force recommendation. Annals of Internal Medicine, 164(4), 244-255. doi:10.7326/M15-0969

NHS Breast Screening Guidance. (2011). Breast screening: Quality assurance standards in radiology -. Retrieved from https://www.gov.uk/government/publications/breast-screening-quality-assurance-standards-in-radiology

Njor, S. H., Olsen, A. H., Bellstrøm, T., Dyreborg, U., Bak, M., Axelsson, C., . . . Lynge, E. (2003). Mammography screening in the county of Fyn. APMIS Supplementum, (110), 1-33.

Norman, S. A., Russell Localio, A., Weber, A. L., Coates, R. J., Zhou, L., Bernstein, L., . . . Nadel, M. R. (2007). Protection of mammography screening against death from breast cancer in women aged 40-64 years. Cancer Causes & Control : CCC, 18(9), 909-918. doi:10.1007/s10552-007-9006-8

Ohene-Yeboah, M., & Amaning, E. (2008). Spectrum of complaints presented at a specialist breast clinic in Kumasi, Ghana. Ghana Medical Journal, 42(3), 110-113.

Olsen, A. H., Agbaje, O. F., Myles, J. P., Lynge, E., & Duffy, S. W. (2006). Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. The Breast Journal, 12(4), 338-342. doi:TBJ272

Olsen, A. H., Lynge, E., Njor, S. H., Kumle, M., Waaseth, M., Braaten, T., & Lund, E. (2013). Breast cancer mortality in Norway after the introduction of mammography screening. International Journal of Cancer, 132(1), 208-214. doi:10.1002/ijc.27609

Olsen, A. H., Njor, S. H., Vejborg, I., Schwartz, W., Dalgaard, P., Jensen, M. B., . . . Lynge, E. (2005). Breast cancer mortality in Copenhagen after introduction of mammography screening: Cohort study. BMJ (Clinical Research Ed.), 330(7485), 220. doi:bmj.38313.639236.82

Otten, J. D.,M., Fracheboud,J., den Heeten, G.,J., Otto, S.,J., Holland,R., de Koning, H.,J., . . . Verbeek, A. L.,M. (2013). Likelihood of early detection of breast cancer in relation to false-positive risk in life-time mammographic screening: Population-based cohort study. Annals of Oncology, 24(10), 2501-2506. doi:10.1093/annonc/mdt227

Paci, E., & Duffy, S. (2005). Overdiagnosis and overtreatment of breast cancer: Overdiagnosis and overtreatment in service screening. Breast Cancer Research, 7(6), 266-270. doi:bcr1339

Paci, E., Miccinesi, G., Puliti, D., Baldazzi, P., De Lisi, V., Falcini, F., . . . Zorzi, M. (2006). Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening study in Italy. Breast Cancer Research : BCR, 8(6), R68. doi:bcr1625

Patel, S. (2018). Breast cancer: Lesser-known facets and hypotheses. doi:https://doi.org/10.1016/j.biopha.2017.12.087

Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., . . . Botstein, D. (2000). Molecular portraits of human breast tumours. Nature, 406(6797), 747-752. doi:10.1038/35021093

Perry N, Broeders M, de Wolf C, Tornberg S, Holland R, von Karsa L. (2006). European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis, European Commission.

Pettersson, A., Graff, R. E., Ursin, G., Santos Silva, I. D., McCormack, V., Baglietto, L., . . . Tamimi, R. M. (2014). Mammographic density phenotypes and risk of breast cancer: A meta-analysis. JNCI, 106(5), 10.1093/jnci/dju078. doi:10.1093/jnci/dju078

Pizot, C., Boniol, M., Mullie, P., Koechlin, A., Boniol, M., Boyle, P., & Autier, P. (2016). Physical activity, hormone replacement therapy and breast cancer risk: A meta-analysis of prospective studies. European Journal of Cancer, 52, 138-154. doi:10.1016/j.ejca.2015.10.063

Pollan, M., Ascunce, N., Ederra, M., Murillo, A., Erdozain, N., Ales-Martinez, J., & Pastor-Barriuso, R. (2013). Mammographic density and risk of breast cancer according to tumor characteristics and mode of detection: A Spanish population-based case-control study. Breast Cancer Research, 15(1), R9. doi:10.1186/bcr3380

Polyak, K. (2007). Breast cancer: Origins and Evolution. The Journal of Clinical Investigation, 117(11), 3155-3163. doi:10.1172/JCI33295

Polyak, K. (2011). Heterogeneity in breast cancer. The Journal of Clinical Investigation, 121(10), 3786-3788. doi:10.1172/JCI60534

Ponti, A., Anttila, A., Ronco, G., Senore, C., Basu, P., Segnan, N., . . . Vale D. (2017). Cancer screening in the European Union: Report on the implementation of the Council recommendation on cancer screening, European Commission.

Porter, P. (2008). "Westernizing" women's risks? breast cancer in lower-income countries. NEJM, 358(3), 213-216. doi:10.1056/NEJMp0708307

Porter, P. L., El-Bastawissi, A. Y., Mandelson, M. T., Lin, M. G., Khalid, N., Watney, E. A., . . . White, E. (1999). Breast tumor characteristics as predictors of mammographic

detection: Comparison of interval- and screen-detected cancers. JNCI, 91(23), 2020-2028.

Pradhan, M., & Dhakal, H. P. (2008). Study of breast lump of 2246 cases by fine needle aspiration. JNMA; Journal of the Nepal Medical Association, 47(172), 205-209.

Puliti, D., Duffy, S. W., Miccinesi, G., de Koning, H., Lynge, E., Zappa, M., . . . EUROSCREEN Working Group. (2012a). Overdiagnosis in mammographic screening for breast cancer in Europe: A literature review. Journal of Medical Screening, 19 Suppl 1, 42-56. doi:19/suppl_1/42

Puliti, D., Duffy, S. W., Miccinesi, G., de Koning, H., Lynge, E., Zappa, M., . . . EUROSCREEN Working Group. (2012b). Overdiagnosis in mammographic screening for breast cancer in Europe: A literature review. Journal of Medical Screening, 19 Suppl 1, 42-56. doi:19/suppl_1/42

Puliti, D., Zappa, M., Miccinesi, G., Falini, P., Crocetti, E., & Paci, E. (2009). An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. European Journal of Cancer, 45(18), 3166-3171. doi:10.1016/j.ejca.2009.06.014

Rajakariar, R., & Walker, R. A. (1995). Pathological and biological features of mammographically detected invasive breast carcinomas. British Journal of Cancer, 71(1), 150-154.

Rawshani, A., Svensson, A., Zethelius, B., Eliasson, B., Rosengren, A., & Gudbjörnsdottir, S. (2016). Association between socioeconomic status and mortality, cardiovascular disease, and cancer in patients with type 2 diabetes. JAMA Internal Medicine, 176(8), 1146-1154. doi:10.1001/jamainternmed.2016.2940

Renart-Vicens, G., Puig-Vives, M., Albanell, J., Castaer, F., Ferrer, J., Carreras, M., . . . Marcos-Gragera, R. (2014). Evaluation of the interval cancer rate and its determinants on the Girona health regional early breast cancer detection program. BMC Cancer, 14. doi:474210.1186/1471-2407-14-558

Ritte, R., Lukanova, A., Berrino, F., Dossus, L., Tjonneland, A., Olsen, A., . . . Kaaks, R. (2012). Adiposity, hormone replacement therapy use and breast cancer risk by age and hormone receptor status: A large prospective cohort study. Breast Cancer Research, 14(3), R76. doi:bcr3186

Roder, D., Houssami, N., Farshid, G., Gill, G., Luke, C., Downey, P., . . . Williamson, L. (2008). Population screening and intensity of screening are associated with reduced breast cancer mortality: Evidence of efficacy of mammography screening in Australia. Breast Cancer Research and Treatment, 108(3), 409-416. doi:10.1007/s10549-007-9609-5

Rue, H. (2015). Bayesian computing with INLA. the R-INLA project. Retrieved from http://www.r-inla.org/home

Ryerson, A. B., Miller, J., & Eheman, C. R. (2015). Reported breast symptoms in the national breast and cervical cancer early detection program. Cancer Causes & Control, 26(5), 733-740. doi:10.1007/s10552-015-0544-1

Sankaranarayanan, R., Ramadas, K., Thara, S., Muwonge, R., Prabhakar, J., Augustine, P., . . . Mathew, B. S. (2011). Clinical breast examination: Preliminary results from a cluster randomized controlled trial in India. JNCI, 103(19), 1476-1480. doi:10.1093/jnci/djr304

Sarkeala, T., Luostarinen, T., Dyba, T., & Anttila, A. (2014). Breast carcinoma detection modes and death in a female population in relation to population-based mammography screening. SpringerPlus, 3,. doi:147710.1186/2193-1801-3-348

Secretan, B., Straif, K., Baan, R., Grosse, Y., El Ghissassi, F., Bouvard, V., . . . WHO International Agency for Research on Cancer Monograph Working Group. (2009). A review of human carcinogens--part E: Tobacco, areca nut, alcohol, coal smoke, and salted fish. The Lancet.Oncology, 10(11), 1033-1034.

Seely, J. M., & Alhassan, T. (2018). Screening for breast cancer in 2018 "what should we be doing today? Current Oncology, 25(Suppl 1), S115-24. doi:10.3747/co.25.3770

Seitz, H. K., Pelucchi, C., Bagnardi, V., & La Vecchia, C. (2012). Epidemiology and pathophysiology of alcohol and breast cancer: Update 2012. Alcohol and Alcoholism, 47(3), 204-212. doi:10.1093/alcalc/ags011

Seltzer, M. (1992). The significance of breast complaints as correlated with age and breast cancer. The American Surgeon, 58(7), 413-417.

Shapiro, S. (1997). Periodic screening for breast cancer: The HIP randomized controlled trial. Health Insurance Plan. JNCI Monographs, (22)(22), 27-30.

Sinclair, N., Littenberg, B., Geller, B., & Muss, H. (2011). Accuracy of screening mammography in older women. American Journal of Roentgenology, 197(5), 1268-1273. doi:10.2214/AJR.10.5442

Skaane, P., Bandos, A.I., Niklason, L.T., Sebuødegård, S., Østerås, B.H., Gullien, R., Gur, D., Hofvind, S. (2019). Digital Mammography versus Digital Mammography Plus Tomosynthesis in Breast Cancer Screening: The Olso Tomosynthesis Screening Trial. Radiology. https://doi.org/10.1148/radiol.2019182394

Smith-Bindman, R., Ballard-Barbash, R., Miglioretti, D. L., Patnick, J., & Kerlikowske, K. (2005). Comparing the performance of mammography screening in the USA and the UK. Journal of Medical Screening, 12(1), 50-54. doi:10.1258/0969141053279130

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., . . . Borresen-Dale, A.L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA, 98(19), 10869-74.

Sobin LH, W. C. (2002). TNM classification of malignant tumours, 6th edition, UICC. Retrieved from https://www.uicc.org/tnm-classification-malignant-tumours-6th-edition

Sterns, E. E. (1992). Age-related breast diagnosis. Canadian Journal of Surgery. Journal Canadien De Chirurgie, 35(1), 41-45.

Surakasula, A., Nagarjunapu, G.C., Raghavaiah, K.V. (2014). A comparative study of pre- and post-menopausal breast cancer: Risk factors, presentation, characteristics and management. J Res Pharm Pract, 3(1), 12-18.

Tabár, L., Dean, P. B., & Pentek, Z. (1983). Galactography: The diagnostic procedure of choice for nipple discharge. Radiology, 149(1), 31-38. doi:10.1148/radiology.149.1.6611939

Tabár, L., Fagerberg, G., Duffy, S. W., Day, N. E., Gad, A., & Gröntoft, O. (1992). Update of the Swedish two-county program of mammographic screening for breast cancer. Radiologic Clinics of North America, 30(1), 187-210.

Tabár, L., Yen, A. M., Wu, W. Y., Chen, S. L., Chiu, S. Y., Fann, J. C., . . . Chen, T. H. (2015). Insights from the breast cancer screening trials: How screening affects the natural history of breast cancer and implications for evaluating service screening programs. The Breast Journal, 21(1), 13-20. doi:10.1111/tbj.12354

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

Taylor, R., Page, A., Bampton, D., Estoesta, J., & Rickard, M. (2004). Age-specific interval breast cancers in New South Wales and meta-analysis of studies of women aged 40-49 years. Journal of Medical Screening, 11(4), 199-206. doi:10.1258/0969141042467403

Taylor, R., Supramaniam, R., Rickard, M., Estoesta, J., & Moreira, C. (2002). Interval breast cancers in New South Wales, Australia, and comparisons with trials and other mammographic screening programmes. Journal of Medical Screening, 9(1), 20-25. doi:10.1136/jms.9.1.20

Tkaczuk, K. H. R., Kesmodel, S. B., & Feigenberg, S. J. (2017). Handbook of breast cancer and related breast disease. New York: Demos Medical Publishing.

Tornberg, S., Kemetli, L., Ascunce, N., Hofvind, S., Anttila, A., Seradour, B., . . . Ponti, A. (2010). A pooled analysis of interval cancer rates in six european countries. European Journal of Cancer Prevention, 19(2), 87-93. doi:10.1097/CEJ.0b013e32833548ed

van Dijck, J. A., Broeders, M. J., & Verbeek, A. L. (1997). Mammographic screening in older women. is it worthwhile? Drugs & Aging, 10(2), 69-79.

van Schoor, G., Moss, S. M., Otten, J. D. M., Donders, R., Paap, E., den Heeten, G. J., . . . Verbeek, A. L. M. (2010). Effective biennial mammographic screening in women aged 40–49, Eur J Cancer, 46(18), 3137-3140. doi:https://doi.org/10.1016/j.ejca.2010.09.041

Vinnicombe, S., Pinto Pereira, S. M., McCormack, V. A., Shiel, S., Perry, N., & Dos Santos Silva, I. M. (2009). Full-field digital versus screen-film mammography: Comparison within the UK breast screening program and systematic review of published data. Radiology, 251(2), 347-358. doi:10.1148/radiol.2512081235

Vitak, B., Stal, O., Manson, J. C., Thomas, B. A., Arnesson, L. G., Ekelund, L., . . . Bang, H. (1997). Interval cancers and cancers in non-attenders in the ostergotland mammographic screening programme. duration between screening and diagnosis, S-phase fraction and distant recurrence. Eur J Cancer, 33(9), 1453-1460. doi:S0959804997001858

Walker, L. G., Cordiner, C. M., Gilbert, F. J., Needham, G., Deans, H. E., Affleck, I. R., . . . Eremin, O. (1994). How distressing is attendance for routine breast screening? Psycho-Oncology, 3(4), 299-304. doi:10.1002/pon.2960030406

Walker, S., Hyde, C., & Hamilton, W. (2014). Risk of breast cancer in symptomatic women in primary care: A case-control study using electronic records. The British Journal of General Practice : The Journal of the Royal College of General Practitioners, 64(629), e788-93. doi:10.3399/bjgp14X682873

Waller, M., Moss, S., Watson, J., & Møller, H. (2007). The effect of mammographic screening and hormone replacement therapy use on breast cancer incidence in England and Wales. Cancer Epidemiology, Biomarkers & Prevention, 16(11), 2257-2261. doi:16/11/2257

Wang, H., Bjurstam, N., Bjørndal, H., Braaten, A., Eriksen, L., Skaane, P., . . . Thoresen, S. O. (2001). Interval cancers in the Norwegian breast cancer screening program: Frequency, characteristics and use of HRT. Int J Cancer, 94(4), 594-598. doi:10.1002/ijc.1511

WCRF. (2018). Diet, Nutrition, Physical activity and Cancer- a global perspective. Retrieved from https://www.wcrf.org/dietandcancer/about

Weber, R. J., van Bommel, R. M., Louwman, M.,W., Nederend, J., Voogd, A.C., Jansen, F.H., . . . Duijm, L. E. (2016). Characteristics and prognosis of interval cancers after

biennial screen-film or full-field digital screening mammography. Breast Cancer Res Treat, 158(3), 471-483.
Weedon-Fekjær, H., Romundstad, P. R., & Vatten, L. J. (2014). Modern mammography screening and breast cancer mortality: Population study. BMJ, 348, g3701. doi:10.1136/bmj.g3701
Welch, H. G., & Black, W. C. (2010). Overdiagnosis in Cancer. JNCI, 102(9), 605-613. doi:10.1093/jnci/djq099
Wilczek, B., Wilczek, H.E., Rasouliyan, L., Leifland, K. (2016). Adding 3D automated breast ultrasound to mammography screening in women with heterogeneously and extremely dense breasts: Report from a hospital-based, high-volume, single-center breast cancer screening program. Eur J Radiol., 85(9), 1554-1563. doi: 10.1016/j.ejrad.2016.06.004
Willett, A., Michell, M., Lee, M. (2010). Best practice diagnostic guidelines for patients presenting with breast symptoms, Mammography | Breast Cancer. Retrieved from https://www.scribd.com/document/140938243/Best-Practice-Diagnostic-Guidelines-for-Patients-Presenting-With-Breast-Symptoms
Williams, R. S., Brook, D., Monypenny, I. J., & Gower-Thomas, K. (2002). The relevance of reported symptoms in a breast screening programme doi:https://doi.org/10.1053/crad.2001.0910
World Health Organization. (2017). Guide to cancer early diagnosis. Geneva. Retrieved from http://www.who.int/cancer/publications/cancer_early_diagnosis/en/
Wu, Y., Zhang, D., & Kang, S. (2013). Physical activity and risk of breast cancer: A meta-analysis of prospective studies. Breast Cancer Research and Treatment, 137(3), 869-882. doi:10.1007/s10549-012-2396-7
Yu, X. Q. (2009). Socioeconomic disparities in breast cancer survival: Relation to stage at diagnosis, treatment and race. BMC Cancer, 9, 364-2407-9-364. doi:1471-2407-9-364
Zackrisson, S., Andersson, I., Janzon, L., Manjer, J., & Garne, J. P. (2006). Rate of over-diagnosis of breast cancer 15 years after end of Malmo mammographic screening trial: Follow-up study. BMJ, 332(7543), 689-692. doi:10.1136/bmj.38764.572569.7C
Zorzi, M., Guzzinati, S., Puliti, D., Paci, E., & IMPACT working group. (2010). A simple method to estimate the episode and programme sensitivity of breast cancer screening programmes. Journal of Medical Screening, 17(3), 132-138. doi:10.1258/jms.2010.009060

PUBLICATIONS

PUBLICATION I

Association of symptoms and breast cancer in population-based mammography screening in Finland

Singh, D., Malila, N., Pokhrel, A., Anttila, A.

Int J Cancer, 136(6), E630-E637
Doi: 10.1002/ijc.29170

IJC
International Journal of Cancer

Association of symptoms and breast cancer in population-based mammography screening in Finland

Deependra Singh[1], Nea Malila[2,3], Arun Pokhrel[2] and Ahti Anttila[2]

[1] School of Health Sciences, University of Tampere, Tampere, Finland
[2] Mass Screening Registry, Finnish Cancer Registry, Helsinki, Finland
[3] School of Health Sciences, University of Tampere, Tampere, Finland

The study purpose was to assess association of symptoms at screening visits with detection of breast cancer among women aged 50–69 years during the period 2006–2010. Altogether 1.2 million screening visits were made and symptoms (lump, retraction, secretion etc.) were reported either by women or radiographer. Breast cancer risk was calculated for each symptom separately using logistic regression [odds ratio (OR)] and 95% confidence intervals (CIs). Of the 1,198,410 screening visits symptoms were reported in 298,220 (25%) visits. Breast cancer detection rate for women with and without symptoms was 7.8 per 1,000 and 4.7 per 1,000 screening visits, respectively, whereas lump detected 32 cancers per 1,000 screens. Women with lump or retraction had an increased risk of breast cancer, OR = 6.47, 95% CI 5.89−7.09 and OR = 2.19, 95% CI 1.92–2.49, respectively. The sensitivity of symptoms in detecting breast carcinoma was 35.5% overall. Individual symptoms sensitivity and specificity ranged from, 0.66 to 14.8% and 87.4 to 99.7%, respectively. Of 5,541 invasive breast cancers, 1,993 (36%) reported symptoms at screen. Breast cancer risk among women with lump or retraction was higher in large size tumors (OR = 9.20, 95% CI 8.08–10.5) with poorly differentiated grades (OR = 5.91, 95% CI 5.03–6.94) and regional lymph nodes involvement (OR = 6.47, 95% CI 5.67–7.38). This study was done in a setting where breast tumors size is generally small, and symptoms sensitivity and specificity in diagnosing breast tumors were limited. Importance of breast cancer symptoms in the cancer prevention and control strategy needs to be evaluated also in other settings.

Early detection of breast cancer through organized screening in average risk women has reduced mortality from the disease.[1,2] In Finland, the national organized mammography screening program has been reported to reduce the incidence-based mortality from breast cancer by approximately 20–28% among those invited.[3] Many, even though not all, breast cancer screening programs include an examination of breasts done by the radiographer and/or reporting of symptoms by the woman at the screening visit.[1,4] Symptoms findings from such examination could convey to diagnostic work-up in the screening centers, as well as indicate a differential risk of breast cancer.

Over-diagnosis and unnecessary treatment of apparently healthy women in mammography screening raise the question about benefits versus harms of screening over clinical breast examination.[5] Many countries where mammography screening is not organized at population level but with the increasing awareness about breast cancer, patients may present with breast complaints.[6] Hence in such situation, detection of breast cancer cases mostly rely on breast complaints. Research on the possible symptoms can provide feedback for the clinicians and help in making decisions when reading screening films and in further investigations (recall or referral).[4] Few studies have highlighted the relevance of assessing symptoms at screening diagnostic mammography.[4,7–9]

So far, no studies till date have studied the association of symptom and breast cancer risk at population level. There is a possibility to learn about benefits of assessing symptoms during screening as well as to improve the procedures by reducing unnecessary diagnostics and false positive findings. Moreover, for developing countries, where high technology for detecting early cancer is not feasible, symptoms can be used as an indication for early diagnostics. Provided that adequate resources are available for confirmation and treatment, this could prevent late stage presentation of cancer.[7]

The aim of the study was to assess the association of symptoms with the occurrence of breast cancer and to

Key words: breast cancer symptoms, screening, clinical breast examination, lump, scar

Abbreviations: MSR: mass screening registry; OR: odds ratio; CI: confidence interval; CBE: clinical breast examination; BSW: Breast Test Wales; PPV: positive predictive value

DOI: 10.1002/ijc.29170
History: Received 4 Apr 2014; Accepted 19 Aug 2014; Online 27 Aug 2014

Correspondence to: Deependra Singh, School of Health Sciences, University of Tampere, Medisiinarinkatu 3, Tampere, Finland, Tel.: +358417060341, E-mail: deependra.singh@uta.fi or nep.singhd@gmail.com

What's new?

A key component of breast cancer screening programs is the collection of data on symptoms at the time of screening visit. In many cases, however, the data are not subsequently analyzed for relationships between symptoms and breast cancer diagnosis. Based on analysis of data from 1.2 million screening visits recorded in the Finnish Cancer Registry, the present report describes a significant association between breast cancer risk and symptoms either self-reported by patients or detected by radiographers. Risk was highest for breast lumps reported at screening. Importantly, the findings also highlight limitations regarding the clinical significance of symptoms.

analyse the cross-sectional clinical validity of symptoms among screened women under the organized breast cancer-screening program.

Materials and Methods

This study is based on breast cancer screening data provided by the Mass Screening Registry (MSR) of the Finnish Cancer Registry (http://www.cancer.fi/syoparekisteri/en/). The MSR receives information on the breast cancer screening program through the population files and the screening centers.[10] Registration is based on the law of personal data in the health-care and the Government Decree on Screenings, 1,339/2,011, and the respective recommendations published by the National Research and Development Centre for Welfare and Health.[11]

The Finnish breast cancer-screening program targets women aged 50–69 years every 2 years. A personal invitation letter is sent by mail with a prefixed time and place of screening. All women in the target age are invited with no exclusions. At the screening clinic woman may present with symptoms or no symptoms. Women are asked (or to fill in the form) whether they had any symptoms during the past 2 months. The nurse then examines the breast. Symptoms are recorded in the mammography screening form. After then, breasts are examined by mammography. After interpretation of the results those with mammography positives are recalled for further examination. Women who are mammography negative are sent home and invited after 2 years for the next biennial screening round. Physician examines the breast of the recalled women. Women may be healthy or referred for diagnostic workup at hospital. Those with cancers are followed up until death (mortality).[10]

For the current study, information on women aged 50–69 years who had breast cancer screening during the years 2006–2010 were retrieved. The first round of screening starts at the age of 50–51 years. The study is based on tabular information and originates from data recorded on the mammography screening form (http://www.cancer.fi/syoparekisteri/joukkotarkastusrekisteri/) for every woman who was screened during that period of time. Altogether 1,454,143 invitations were made during the period, of which 1,241,486 screening visits were made (attendance 85.4%). In all, 38,647 visits (3.11%) were excluded because of incomplete information on either the clinical examination or on self-reported symptoms. Furthermore, 4,429 (0.36%) visits were excluded because of not complying with the age range. The final data set contains 1,198,410 screening visits from all over Finland. Symptoms that were reported include lump, retraction, scar, secretion and mole. Outcome variables were histologically confirmed breast cancers (both invasive and in situ) and benign findings. Some tumor characteristics (tumors size and grade) were also available.

In the current analysis, women who had a given symptom at screen in either or both breasts were considered as symptomatic. Information on breast symptoms was dichotomized for any as well as for each individual symptom separately. The outcome was categorized as malignant (in-situ and invasive breast cancers) and benign finding (other histology). The age of screened women was categorized into four groups as "50–54," "55–59," "60–64" and "65–69." To do the homogeneity test of symptoms with age, age-groups were made as continuous variables where age-group 50–54 years indicate "0," 55–59 years indicate "1," 60–64 years indicate "2" and 65–69 years indicate "3." In classifying histologically confirmed tumors two categories of tumor size were made: "less than 20 mm" and "20–150 mm." Tumor grades were classified as "well-differentiated," "moderately differentiated" and "poorly differentiated." Tumor spreading was classified according to the TNM classification of tumors published by International Union Against Cancer (UICC) in 2002.[12]

Statistical analysis

Breast cancer detection rate (number of cancer cases detected divided by number of screening visits) was calculated for individual symptoms as well as for all possible pairwise combination of symptoms. Logistic regression model was used to calculate the crude and adjusted odds ratios (ORs) with 95% confidence interval (CI) using Wald statistics for individual terms. The univariate logistic regression model was used to estimate the age-adjusted association of symptom with the occurrence of breast cancer. For calculating the joint exposure effects and homogeneity analysis likelihood ratio statistics was used. Effects by individual and combined symptoms (self-reported and radiographer reported) as well as pairwise analysis of symptoms with all possible combinations were also estimated. All the statistical analyses were two-sided, and a p value ≤ 0.05 was considered to be statistically significant. The statistical analyses were carried out using STATA software release 11.0.

To analyze the cross-sectional validity of the symptoms, sensitivity and specificity were estimated. True positives are here those with a "positive" symptom and with breast cancer and vice-versa for true negatives. Breast cancer included invasive and in situ carcinoma of the breast, and analyses were done also separately for these two diagnosis categories when relevant. False positives are here those with a positive symptom but no breast cancer whereas false negatives are those with no symptom but had breast cancer. Sensitivity was here defined as the number of visits with screen-detected malignant cancers in those who had symptoms (true positives for symptoms) divided by the total number of visits with breast cancer. Specificity was the number of visits with no symptoms and no malignant finding (true negatives for symptoms), divided by the total number of visits with no malignant findings. CIs for sensitivity and specificity were produced with the Wilson score method.[13] The positive predictive value (PPV) is the likelihood of cancer detected among those who had symptoms. CI for PPV was calculated using the method described by Simel et al.[14] We considered lump or retraction as clinically relevant symptoms while reporting and analyzing symptoms information on histological confirmed tumors.

Results

A total of 1,198,410 screening visits were made in 2006–2010 and out of these, a histologically confirmed breast cancer (including in-situ cases) was diagnosed in 6,009 (0.5%) women at screen. In this period, the national decree of screening was given and women aged 60–69 years were also included into the target population if they were born in 1947 or later. Thus, the number of screened women increased year by year clearly between 2006 and 2010, *i.e.*, 192,892 and 264,678, respectively. Altogether 298,220 visits with at least one symptom out of 1,198,410 visits (24.9%) were reported in this period (Table 1). Lump was reported in 15,587 (1.30%) screening visits and retraction was reported in 20,880 (1.74%) visits. The percentage of women who reported symptoms (out of total screened) increased clearly by age of the women, 21.8% in age-group 50–54 years and 30% in age group 65–69 years, respectively. Screen positive women (who were recalled) were 30,392 (2.5%) out of which 9,659 (32%) reported any of the symptoms. The percentage of women out of total screening visits that were referred for further assessment was 0.75%.

Breast cancer detection rate of lump was 31.9 per 1,000 screening visits whereas detection rate of retraction and secretion was 11.6 per 1,000 and 10.8 per 1,000 screening visits, respectively. The age-adjusted risk of breast cancer in women who reported a lump was 6.61 (95% CI 6.03–7.26) times higher compared to those with no symptoms (Table 2). Similarly, the risk in women who reported retraction or secretion was more than twofold, OR = 2.11, 95% CI 1.86–2.41 and OR = 2.14, 95% CI 1.58–2.89, respectively, compared to women who reported no symptoms. Reporting a scar or mole indicated a small increase in the risk of breast cancer compared to those with no symptoms, *i.e.*, OR = 1.26, 95% CI 1.17–1.35 and OR = 1.16, 95% CI 1.09–1.25, respectively.

The risk of breast cancer in women who reported lump was higher in all age groups compared to women with other symptoms. Women who reported lump and/or retraction had a significant increase in breast cancer risk across age groups. Women who reported secretion had an increase in trend of breast cancer risk with age (Fig. 1). The joint effect of symptoms with two possible combinations was measured simultaneously. The cancer detection rate of lump and retraction combined was 102 per 1,000 screening visit whereas combined lump and secretion was 26 per 1,000 screening visits. Similarly, the combined cancer detection rate of retraction and scar was 12 per 1,000 screening visits. The combined effect of lump and retraction showed a 23-fold (OR = 22.6, 95% CI 16.5–30.8) increase in the risk of breast cancer compared to women with no lump or retraction. Similarly, the joint effect of lump and scar showed a sixfold (OR = 5.37, 95% CI 4.31–6.69) increase in the risk of breast cancer compared to women with no lump or scar (Table 3).

Overall, 2,314 women who had any of the symptoms were diagnosed with breast cancer at screen. The sensitivity to detect cancer for women with any of the symptoms was 35.5% (95% CI 34.3–36.6%) whereas specificity was 75.2% (95% CI 75.1–75.3%; Table 4). The sensitivity to detect cancer was 8% in women who had a lump whereas in case of retraction the sensitivity was 4%. However, the specificity was high for lump and retraction, 98.7 and 98.3%, respectively. Scar and mole both had a sensitivity of 15% each whereas specificity was low for these symptoms, 88.3 and 87.4%, respectively.

Altogether 5,541 invasive breast cancers were detected at screen out of which 1,993 (36%) were reported with symptoms at the time of screening and 652 (32.7%) reported lump or retraction only. In all, 70% of the invasive cancers were less than 20 mm in diameter. The presence of lump or retraction increased from 8% in tumors less than 20 mm of size to 22% in tumors of 20–150 mm in size (Table 5). The probability of having age-adjusted invasive breast cancer was significantly higher (OR = 4.31, 95% CI 3.96–4.69) in those who reported lump or retraction compared to those with no lump or retraction. Women with lump or retraction had a significantly higher age-adjusted risk for big tumors than nonsymptomatic women, OR = 2.84 (95% CI 2.53–3.19) in tumors less than 20 mm and OR = 9.20 (95% CI 8.08–10.5) in 20–150 mm size tumors. The probability of having poorly differentiated tumors was significantly higher (OR = 5.91, 95% CI 5.03–6.94) in women who reported symptoms than in those without symptoms. The probability of having tumors in regional lymph nodes was significantly greater in women with symptoms compared to those with no symptoms, OR = 6.47, 95% CI 5.67–7.38.

Discussion

The purpose of the study was to examine the association between symptoms at the screening visit and detection of breast cancer at screen. In addition, we described the size

Table 1. Symptoms[2] reported during the screening visits at different time-periods and by age, recall, or referral due to screening results

	Total screening visits	Lump (%)	Retraction (%)	Scar (%)	Mole (%)	Secretion (%)	Any of the symptoms (%)[1]
Year							
2006	192,892	2,570 (1.3)	3,425 (1.8)	22,239 (11.5)	27,411 (14.2)	726 (0.4)	50,402 (26.1)
2007	235,304	3,044 (1.3)	3,858 (1.6)	27,819 (11.8)	31,239 (13.3)	879 (0.4)	60,106 (25.5)
2008	237,389	3,011 (1.3)	4,297 (1.8)	27,821 (11.7)	30,937 (13.0)	797 (0.3)	60,117 (25.3)
2009	268,147	3,346 (1.2)	4,542 (1.7)	31,492 (11.7)	30,327 (11.3)	722 (0.3)	63,462 (23.7)
2010	264,678	3,616 (1.5)	4,758 (1.8)	30,766 (11.6)	31,115 (11.8)	841 (0.3)	64,133 (24.2)
Age							
50–54	469,594	6,932 (1.5)	6,794 (1.4)	44,587 (9.5)	52,734 (11.2)	2,081 (0.4)	102,538 (21.8)
55–59	339,635	4,095 (1.2)	6,368 (1.9)	41,029 (12.1)	42,885 (12.6)	943 (0.3)	85,715 (25.2)
60–64	306,227	3,622 (1.2)	6,012 (1.9)	42,552 (13.9)	42,343 (13.8)	721 (0.2)	85,177 (28.8)
65–69	82,954	938 (1.1)	1,706 (2.1)	11,969 (14.4)	13,067 (15.7)	220 (0.3)	24,790 (29.8)
Recall							
Yes	30,392	2,205 (7.2)	724 (2.4)	4,210 (13.8)	3,976 (13.1)	310 (1.0)	9,659 (31.7)
No	1,168,018	13,382 (1.1)	20,156 (1.7)	135,927 (11.6)	147,053 (12.6)	3,655 (0.3)	288,561 (24.7)
Referral							
Yes	8,093	613 (7.6)	278 (3.4)	1,248 (15.4)	1,169 (14.4)	87 (1.1)	2,876 (35.5)
No	1,073,462	13,979 (1.3)	19,429 (1.8)	129,050 (12.0)	141,470 (13.2)	3,559 (0.3)	276,507 (25.7)
Total	1,198,410	15,587 (1.3)	20,880 (1.7)	140,137 (11.7)	151,029 (12.6)	3,965 (0.3)	298,220 (24.9)

[1]Percentage (%) in the bracket means any of the symptoms out of total screening visits.
[2]Symptoms include women, radiographer reported or both.

Table 2. Age-adjusted odds ratios (OR) of breast cancer (including *in situ* and benign tumors) with 95% confidence intervals (CI) among women with symptoms[1] compared to women with no symptoms

	Cases (%)	Total (%)	Detection rate (per 1,000)	OR (95% CI)* adjusted with age
Lump				
Yes	497 (3.19)	15,587 (1.30)	31.9	6.61 (6.03–7.26)
No	6,027 (0.51)	1,189,601 (98.7)	5.07	Ref.
Retraction				
Yes	242 (1.16)	20,880 (1.74)	11.6	2.11 (1.86–2.41)
No	6,282 (0.53)	1,177,530 (98.3)	5.33	Ref.
Scar				
Yes	966 (0.69)	1,40,137 (11.7)	6.89	1.26 (1.17–1.35)
No	5,558 (0.53)	1,058,273 (88.3)	5.25	Ref.
Secretion				
Yes	43 (1.08)	3,965 (0.33)	10.8	2.14 (1.58–2.89)
No	6,481 (0.54)	1,194,364 (99.7)	5.43	Ref.
Mole				
Yes	963 (0.64)	1,51,029 (12.6)	6.38	1.16 (1.09–1.25)
No	5,561 (0.53)	1,047,299 (87.4)	5.31	Ref.
Total	6,524 (0.54)	1,198,410 (100.0)		

[1]Symptoms include women, radiographer reported or both.
Abbreviations: OR: odds ratio; CI: confidence interval; ref.: reference.

and grade of tumor in relation to symptoms at screen in women who attended screening and were diagnosed with a breast tumor. The large dataset of about 1.2 million screening visits allows studying breast cancer risk at the population level among women who reported symptoms at screening. The study found a significant association between all reported

symptoms and the occurrence of breast cancer. A breast lump at screen indicated the highest breast cancer risk.

In this study symptoms were either self-reported or radiographers reported and the rate of breast cancer associated with symptoms were calculated. Symptoms were reported in 25% of the screening exams. This is higher than previously reported in a study on postmenopausal women where the prevalence of symptoms was below 10%.[7] The explanation may be that in our study more symptoms were included and symptoms were considered valid whether reported by women or by the radiographer. In studies reporting symptoms at diagnostic mammography exams the prevalence has been more than 30%.[7,8,15] The reason for high prevalence of symptoms in diagnostic mammographic exams may be due to selection of women at increased risk of breast cancer[7,8,15] and premenopausal women in whom prevalence of symptom is higher.[6]

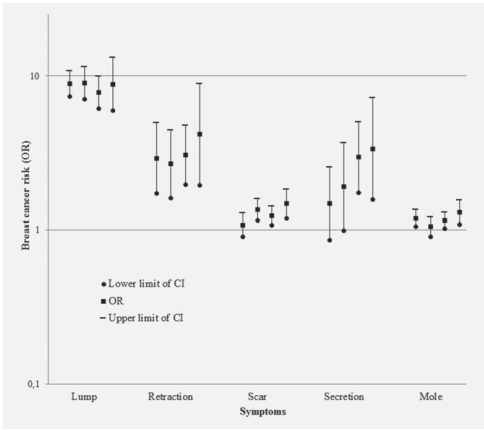

Figure 1. Breast cancer risk among women having symptoms reported by age groups.

The overall proportion of women with breast cancer among those reporting symptoms was 0.78% in our study. The recall rate (mammography positives) among women with symptoms was 3.24%, whereas only 1.73% of women with no symptoms were recalled. Similarly, the proportion of women who referred for further assessment was greater in those with symptoms compared to women with no symptoms, *i.e.*, 0.96 versus 0.44%, respectively. Aiello et al.[7] reported that 6.6% of women with symptoms at diagnostic examination and 1.3% of women at screening examination were diagnosed with breast cancer. Williams et al.[4] study on women who had mammography screening found the breast cancer rate of 0.5% in women with symptoms which is lower than in our study (0.8%). However, they evaluated only those women with "significant" breast symptoms as defined by Breast Test Wales (BSW) guidelines.[16] The Seltzer[9] study reported higher proportion (16%) of breast cancer diagnosed among women with symptoms or prior abnormal mammography that were referred for diagnostic examination. This study found the cancer rate of 7.6, 3.7 and 14.9% in patient with breast lump, retraction and scar, respectively, which is little lower than the study by Lumachi et al.[17] That study found a cancer rate of 3.2, 16.4 and 12.0%, respectively, in patient with breast pain, lump and nipple discharge.[17] In another study by Sterns,[18] breast cancer rate was 37, 11 and 3% in patients with breast mass, nipple discharge and lump, respectively. One reason for the differences may

Table 3. Odds ratios (ORs) of breast cancer with 95% confidence intervals (CI) for joint exposure to symptoms[1]

Symptoms	Screened women	Cancer cases (%)	Detection rate (per 1,000)	OR	95% CI
Lump#retraction	36,036	695 (1.93)	19.3		
0 0	1,162,374	5,829 (0.50)	5.01	Ref.	Ref.
0 1	20,449	198 (0.97)	9.68	1.94	1.68–2.24
1 0	15,156	453 (2.99)	28.9	6.15	5.55–6.74
1 1	431	44 (10.2)	102	22.6	16.5–30.8
Lump#scar	152,515	1,380 (0.90)	9.05		
0 0	1,045,895	5,144 (0.49)	4.92	Ref.	Ref.
0 1	136,928	883 (0.64)	6.45	1.31	1.22–1.41
1 0	12,378	414 (3.34)	33.4	7.0	6.32–7.75
1 1	3,209	83 (2.59)	25.9	5.37	4.31–6.69
Retraction#scar	157,121	1,162 (0.74)	7.4		
0 0	1,041,289	5,362 (0.51)	5.15	Ref.	Ref.
0 1	136,241	920 (0.68)	6.75	1.31	1.22–1.41
1 0	16,984	196 (1.15)	11.5	2.26	1.95–2.60
1 1	3,896	46 (1.18)	11.8	2.31	1.72–3.09

[1]Symptoms include women, radiographer reported or both.
Abbreviations: 0: absence of symptom; 1: presence of symptom; OR: odds ratio; CI: confidence interval; ref.: reference.

Table 4. Clinical validity of symptoms[1] in terms of sensitivity, specificity and positive predictive value (PPV)

Clinical validity	Lump	Retraction	Scar	Secretion	Mole	Any of the symptoms
True positives	497	242	966	43	963	2,314
True negatives	1,176,796	1,171,248	1,052,715	1,187,883	1,041,738	895,980
False positives	15,090	20,638	139,171	3,922	150,066	295,906
False negatives	6,027	6,282	5,558	6,481	5,561	4,210
Sensitivity %	7.62 (7.04–8.32)	3.71 (3.27–4.20)	14.8 (14.0–15.7)	0.66 (0.48–0.89)	14.8 (13.9–15.7)	35.5 (34.3–36.6)
Specificity %	98.7 (98.7–98.8)	98.3 (98.2–98.3)	88.3 (88.3–88.4)	99.7 (99.7–99.7)	87.4 (87.3–87.5)	75.2 (75.1–75.3)
Positive predictive value %	3.19 (2.92–3.48)	1.16 (1.02–1.32)	0.69 (0.65–0.73)	1.08 (0.79–1.47)	0.64 (0.60–0.68)	0.78 (0.74–0.81)

[1]Symptoms include women, radiographer reported or both.

be that our study was done among the general population and most women come for screening on a regular basis (once in every 2 years). Moreover, in our study women reported only symptoms that occurred in the past 2 months and radiographer reported those symptoms detected at the time of screening visit. The reason for higher rates in other studies[17,18] was that both studies were done among symptomatic women who had higher risk of developing cancer. However, due to variation in early detection program and collected symptoms information as well as varying age of the women at either screen or diagnostic examination than our study, results are not directly comparable with the current study.

In this study, the risk of breast cancer was found to be significantly associated with the occurrence of symptoms. The risk of developing breast cancer was sevenfold in women having a lump and the risk was almost similar across the age group. Aiello et al.[7] reported a risk of more than threefold in women who had a lump in the screening exam or diagnostic exam but no significant association between nipple discharge, breast pain and breast cancer risk. Moreover, our study showed a threefold increase in risk in women who had retraction in their breast and a small increase in risk in those who reported scar and mole. We are unaware of other epidemiological studies that would have examined the association between retraction, scar and the breast cancer risk. Two-way joint effects of symptoms showed a significant 23-fold breast cancer risk in women who reported lump and retraction and a 6-fold risk in women who reported lump and scar. The higher risk of breast cancer in our study may be due to the information about breast symptoms systematically collected. A study by Sarkeala et al.[3] in Finland found a 1.56 (95% CI = 1.25–1.91) times higher death rate in women who had no screening visits. The interval cancers, since screening visits are made once in every 2 years, can be more aggressive than screen-detected cancers. Hence, the risk might be even higher in women who had symptoms and are not screened.

In our study breast cancer rate among women with any of the given symptoms was 0.66% in age-group 50–59 years and 0.99% in age-group 60–69 years. Sterns'[18] study in symptomatic patients found the cancer rate to be significantly age-related, being 0.8% in women younger than 40 years and 5% in those between 41 and 55 years. Kerin et al.[19] evaluated the 585 symptomatic patients found breast cancer rate of 2.2% in patient aged 40–49 years, 4.5% in patient aged 50–59 years and 3.1% in patient aged more than 60 years of age. In our study women who reported a lump or retraction showed significantly higher risk of breast cancer in all age-groups compared to non-symptomatic women. Women with other symptoms had a nonsignificantly higher breast cancer risk across age groups. The p value test for homogeneity showed no age related breast cancer risk with an exception of secretion (p value <0.05).

The sensitivity of reporting any symptom in detection of invasive carcinoma was 35.5% in the present study, which is lower than that reported by others.[20–22] However, Harvey et al.[20] and Kerlikwoske et al.[21] measured sensitivity based on the mammography findings and Bobo et al.[22] based the sensitivity calculation on clinical breast examination. A community based study among asymptomatic women in United States reported lower sensitivity than found in our study, between 18.1 and 21.6% based on clinical breast examination.[23] Findings from a randomized controlled trial of breast cancer screening by clinical breast examination in India showed a moderate sensitivity and high specificity, 51.7 and 94.3%, respectively, but PPV was lower than found in our study.[24] In our study, the sensitivity of lump, retraction, scar and mole was 7.7, 3.7, 14.8 and 14.8%, respectively, while high specificity of 99% was reported by lump and retraction. We are not aware of any other studies that measured the clinical validity of symptoms at screen and hence our study findings are not directly comparable to other studies. The low sensitivity of any specific symptom in our study may be explained by the magnitude of diagnostic activities, several rounds of screening in the program, and access to mammography services outside the screening program. Thus, both the population and the tumors found by screening are different from those in the trial from India.[24]

Another purpose of our study was to assess tumor characteristics (size and grade) in relation with breast cancer symptoms. We found that close to 70% of invasive breast cancers detected by screening were less than 20 mm of size. Sankaranarayanan et al.[24] study reported a significantly lower percentage of tumors

Table 5. Characteristics of and probability (odds ratios, OR, with 95% confidence intervals, CI) of invasive breast cancer in women with symptoms[1] (lump or retraction only) compared to women with no symptoms at screen

Tumor characteristics	Total	Symptoms (lump or retraction)	OR (95% CI)* adjusted with age
Size in histology			
Less than 20 mm	3,841	311 (8.1 %)	2.84 (2.53–3.19)
20–150 mm	1,340	294 (21.9 %)	9.20 (8.08–10.5)
Missing[2]	360	47 (13.6 %)	
Total	5,541	652 (11.8%)	4.31 (3.96–4.69)
Grade			
Well differentiated	1,570	142 (9.05%)	3.18 (2.68–3.78)
Moderately differentiated	2,520	302 (12.0%)	4.39 (3.89–4.95)
Poorly differentiated	1,140	177 (15.5%)	5.91 (5.03–6.94)
Missing[2]	311	31 (10.0%)	
Total	5,541	652 (11.8%)	4.37 (4.02–4.76)
pN			
pN0	3,533	340 (9.6%)	3.42 (3.06–3.83)
pN1+	1,601	267 (16.7%)	6.47 (5.67–7.38)
pNX	40	4 (10.0%)	3.56 (1.27–10.0)
Missing[2]	367	41 (11.2%)	
Total	5,541	652 (11.8%)	4.34 (3.99–4.73)
pM			
pM0	4,213	484 (11.5%)	4.20 (3.81–4.62)
pM1	32	7 (21.9%)	9.04 (3.91–20.9)
pMX	772	109 (14.1%)	5.29 (4.32–6.48)
Missing[2]	5,24	52 (9.9%)	
Total	5,541	652 (11.8%)	4.41 (4.04–4.80)

[1]Symptoms include women, radiographer reported or both.
[2]Missing cases here are those referred for further assessment but with missing histological confirmation.
Abbreviations: pN: regional lymph nodes; pN0: no regional lymph node metastasis; pN1+: metastasis with ipsilateral lymph nodes; pNX: regional lymph nodes cannot be assessed; pM: distant metastases; pM0: no distant metastases; pM1: distant metastases; pMX: distant metastases cannot be assessed.

less than 20 mm in size compared to our study, 18.8% versus 69.3%, respectively. The high proportion of invasive cancers of small size highlights the importance of organized screening program where tumors can be detected at early stage of disease. Similarly, other studies have found quite significant difference in tumors characteristics between screen detected and clinical breast cancer cases.[25–29] A study by Miller et al.[5] among women with annual screening in age 40–59 found that 68% of the palpable cancers had a mean tumor size of 21 mm, which is significantly higher than in our study. The probability of detecting invasive tumors with poor differentiation (high grade) was significantly higher in those who reported symptom at screen compared to those with no symptom.

To the best of our knowledge, our study is the largest study done on breast cancer symptoms, either self-reported or radiographer reported, and breast cancer risk at screen. Our findings reinforce the importance of evaluating symptoms as a predictor of breast cancer and warrant extra consideration while evaluating mammograms of women with symptoms. Also, continual maintaining of the information about symptoms at screening visits is useful for the clinician as well as for epidemiological research.

This study was limited to those women who attended screening and the size of breast tumors was generally small. Thus, the sensitivity and specificity of diagnosing breast tumors based on symptoms were limited. Also, breast cancer cases detected outside screening were not included. It may be that women with symptoms also had other risk factors (like dense breasts or positive family history) which might confound the observed effect. A potential limitation of this cross-sectional study is the lack of descriptive information other than age so no adjustment for confounders such as breast density, family history of breast cancer or number of previous screens was possible in the multivariate analysis. The study was cross-sectional and no follow-up or subsequent round of screening was included. There was a possibility that knowing the symptom status may have already influenced the radiology result. Given the low sensitivity of symptoms in our study it is likely that a prevention program based on clinical examination would not provide sufficient benefit for breast cancer control in Finland and the mammography screening program is still justified. The study provides limited evidence that reporting symptoms at screen was associated with aggressive tumors, *i.e.*, tumors with poor prognosis. This study cannot say about the impact in low resource setting with currently no breast cancer screening services. However, considering the higher risk of breast cancer in women with symptoms, clinical breast examination together with the availability of diagnostic services could help in detecting large size tumors. Importance of breast cancer symptoms in the cancer prevention and control strategy needs to be evaluated also in other settings.

Acknowledgements

The original data comes from the Mass Screening Registry, Finnish Cancer Registry. The School of Health Sciences (HES), University of Tampere, has provided cooperation in preparation of the article.

References

1. International Agency for Research on Cancer (IARC). IARC handbook of cancer prevention, Vol. 7. Breast cancer screening. Oxford: Oxford University Press, 2002.
2. Njor S, Nyström L, Moss S, et al. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. *J Med Screen* 2012;19:33–41.
3. Sarkeala T, Heinävaara S, Anttila A. Organised mammography screening reduces breast cancer mortality: a cohort study from Finland. *Int. J. Cancer* 2008;122:614–9. Available at: http://www.

ncbi.nlm.nih.gov/pubmed/17847022 [Accessed May 7, 2013].
4. Williams RS, Brook D, Monypenny IJ, et al. The relevance of reported symptoms in a breast screening programme. *Clin Radiol* 2002;57:725–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12169283 [Accessed May 3, 2013].
5. Miller AB, Wall C, Baines CJ, et al. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ* 2014;348:1–10. Available at: http://www.bmj.com/content/348/bmj.g366.pdf%2Bhtml.
6. Morrow M. The Evaluation of Common Breast Problem. *Am Fam Physician* 2000;61:2371–78. Available at: http://www.aafp.org/afp/2000/0415/p2371 [Accessed May 3, 2013].
7. Aiello EJ, Buist DSM, White E, et al. Rate of breast cancer diagnoses among postmenopausal women with self-reported breast symptoms. *J Am Board Fam Med,* 2004;17:408–15. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15575032.
8. Barlow WE, Lehman CD, Zheng Y, et al. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. *J Natl Cancer Inst* 2002;94:1151–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12165640.
9. Seltzer MH. The significance of breast complaints as correlated with age and breast cancer. *Am Surg* 1992;58:413–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1616186 [Accessed May 3, 2013].
10. Finnish Cancer Registry. Breast Cancer Screening 2013. Available at: http://www.cancer.fi/syoparekisteri/en/mass-screening-registry/statistics/breast-cancer-screening/finland/.
11. Gissler M, Haukka J. Finnish health and social welfare registers in epidemiological research. *Norsk Epidemiologi* 2004;14:113–20.
12. Sobin LH, Wittekind C. International Union Against Cancer (UICC) TNM classification of malignant tumours. 6th ed. New York: Wiley, 2002.
13. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998;17:857–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9595616 [Accessed June 7, 2013].
14. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;44:763–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1941027 [Accessed June 7, 2013].
15. Geller BM, Barlow WE, Barbash RB, et al. Use of the American College of Radiology BI-RADS to report on the mammographic evaluation of women with signs and symptoms of breast disease. *Radiology* 2002;222:536–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11818625.
16. Breast Test Wales. Quality manual. Cardiff: Breast Test Wales, 1996.
17. Lumachi F, Ermani M, Brandes A, et al. Breast complaints and risk of breast cancer. Population-based study of 2,879 self-selected women and long-term follow-up. *Biomed Pharmacother* 2002;56:88–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12000140.
18. Sterns EE. Age-related breast diagnosis. *Can J Surg* 1992;35:41–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1739898 [Accessed May 3, 2013].
19. Kerin MJ, Hanlon DM, Khalid AA, et al. Mammographic assessment of symptomatic non suspicious breast. *Am J Surg* 1997;173:181–4.
20. Harvey SC, Geller B, Oppenheimer RG, et al. Increase in cancer detection and recall rates with independent double interpretation of screening mammography. *AJR* 2003;180:1461–7.
21. Kerlikwoske K, Grady D, Barclay J, et al. Effect of age, breast density, and family history on the sensitivity of first screening mammography. *JAMA* 1996;276:33–8. Available at: http://jama.jamanetwork.com/article.aspx?articleid=404660 [Accessed December 31, 2012].
22. Bobo JK, Lee NC, Thames SF. Findings From 752081 Clinical Breast Examinations Reported to National Screening Program From 1995 Through 1998. *J Natl Cancer Inst* 1999;92:971–6.
23. Fenton JJ, Rolnick SJ, Harris EL, et al. Screening clinical breast examination: how often does it miss lethal breast cancer? *J Natl Cancer Inst Monogr* 2005;35:67–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16287888 [Accessed December 28, 2012].
24. Sankaranarayanan R, Ramadas K, Thara S, et al. Clinical breast examination: preliminary results from a cluster randomized controlled trial in India. *J Natl Cancer Instit* 2011;103:1476–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21862730 [Accessed November 15, 2012].
25. Redondo M, Funez R. Cano FM, et al. Detection methods predict differences in biology and survival in breast cancer patients. *BMC Cancer* 2012;12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23244222 [Accessed December 26, 2012].
26. Bucchi L, Barchielli A, Ravaioli A, et al.; the SCREENREG Working Group. Screen-detected vs clinical breast cancer: the advantage in the relative risk of lymph node metastases decreases with increasing tumour size. *Br J Cancer* 2005;92:156–61. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2361732&tool=pmcentrez&rendertype=abstract [Accessed December 29, 2012].
27. Dawson SJ, Duffy SW, Blows FM, et al. Molecular characteristics of screen-detected vs symptomatic breast cancers and their impact on survival. *Br J Cancer* 2009;101:1338–44. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2768460&tool=pmcentrez&rendertype=abstract [Accessed December 28, 2012].
28. Kim J, Lee S, Bae S, et al. Comparison between screen-detected and symptomatic breast cancers according to molecular subtypes. *Breast Cancer Res Treat* 2012;131:527–40http://www.ncbi.nlm.nih.gov/pubmed/22042364 [Accessed December 28, 2012].
29. Ota J, Horino T, Taguchi T, et al. Mass screening for breast cancer: comparision of the clinical stages and prognosis of breast cancer detected by mass screening and in out-patient clinics. *Jpn J Cancer Res* 1989;80:1028–34.M.

PUBLICATION II

Cumulative risk of false positive test in relation to breast symptoms in mammography screening: a historical prospective cohort study

Singh, D., Pitkäniemi, J., Malila, N., Anttila, A.

Breast Cancer Res Treat, 159(2), 305-313
Doi: 10.1007/s10549-016-3931-8

Breast Cancer Res Treat (2016) 159:305–313
DOI 10.1007/s10549-016-3931-8

EPIDEMIOLOGY

Cumulative risk of false positive test in relation to breast symptoms in mammography screening: a historical prospective cohort study

Deependra Singh[1,2] · **Janne Pitkäniemi**[1] · **Nea Malila**[1,2] · **Ahti Anttila**[1]

Received: 27 July 2016 / Accepted: 27 July 2016 / Published online: 5 August 2016

Abstract Mammography has been found effective as the primary screening test for breast cancer. We estimated the cumulative probability of false positive screening test results with respect to symptom history reported at screen. A historical prospective cohort study was done using individual screening data from 413,611 women aged 50–69 years with 2,627,256 invitations for mammography screening between 1992 and 2012 in Finland. Symptoms (lump, retraction, and secretion) were reported at 56,805 visits, and 48,873 visits resulted in a false positive mammography result. Generalized linear models were used to estimate the probability of at least one false positive test and true positive at screening visits. The estimates were compared among women with and without symptoms history. The estimated cumulative probabilities were 18 and 6 % for false positive and true positive results, respectively. In women with a history of a lump, the cumulative probabilities of false positive test and true positive were 45 and 16 %, respectively, compared to 17 and 5 % with no reported lump. In women with a history of any given symptom, the cumulative probabilities of false positive test and true positive were 38 and 13 %, respectively. Likewise, women with a history of a 'lump and retraction' had the cumulative false positive probability of 56 %. The study showed higher cumulative risk of false positive tests and more cancers detected in women who reported symptoms compared to women who did not report symptoms at screen. The risk varies substantially, depending on symptom types and characteristics. Information on breast symptoms influences the balance of absolute benefits and harms of screening.

Keywords Breast cancer symptoms · False positive · True positive · Mammography · Screening · Lump

Abbreviations

IARC	International Agency for Research on Cancer
FCR	Finnish Cancer Registry
GLM	Generalized linear model
FP	False positive
CI	Confidence interval
UK	United Kingdom
USA	United States of America
BCSC	Breast Cancer Surveillance Consortium

Introduction

Organized screening programs for breast cancer have been estimated to reduce breast cancer mortality by about 23 % among those invited. On the other hand, however, it has also been shown to increase the risk of cumulative false positive results by about 20 % [1]. These estimates describe mainly screening programs that invite women aged 50–69 or 50–74 years. There is no clear evidence on effectiveness of systematic clinical breast examination without mammography or of breast self-examination [1, 2]. In addition to sole mammography as the screening test, some programs or trials have performed clinical or physical examination [3]. Clinical examination means systematic

Electronic supplementary material The online version of this article (doi:10.1007/s10549-016-3931-8) contains supplementary material, which is available to authorized users.

✉ Deependra Singh
deependra.singh@cancer.fi; nep.singhd@gmail.com

[1] Finnish Cancer Registry, Unioninkatu 22, 00130 Helsinki, Finland

[2] School of Health Sciences, University of Tampere, Arvo Building, Lääkärinkatu 1, 33014 Tampere, Finland

palpation by specifically trained staff [3]. However, the clinical examination in Finland is done by collecting self-reported information on symptoms during the screening examination as well as inspection of breasts by the radiographer.

Self-reported symptoms as well as radiographer reports on observations have been a part of the mammography screening program in Finland, since the program started in the late 1980s [3, 4]. Cross-sectional studies have indicated that symptoms have important consequences on the performance of screening [5–7]. There is a risk that harms of screening may increase, as information on symptoms launch further assessments not dealing with breast cancer. The findings of the physical examination may also relate to long-term patterns over several screening rounds.

The main purpose of this study was to estimate the cumulative probability of false positive mammography tests and true positives in women's visits with symptoms, compared with those visits with no reported symptoms at mammography screening in the Finnish programme in women aged 50–69 years. In addition, we estimated the risk of false positive test and true positive with accumulated same symptom or any symptoms in the screening history.

Materials and methods

Study design, setting, and data source

The current study is a register-based cohort study, which utilizes the screening visit history of women who attended the mammography screening program in Finland. The program invited women aged 50–69 years every second year for mammography screening in special organized clinics. Information on breast cancer screening has been registered at the Mass Screening Registry which is part of the Finnish Cancer Registry. The women were asked about breast symptoms at the visit. Any symptoms (lump, retraction, secretion, mole, and scar) women had during the past 2 months were recorded on the mammography form (http://www.cancer.fi/@Bin/44068785/Mammography+form_2006.pdf). The mammography screening examination was two-view for both breasts. The detailed mammography screening process has been described earlier [5]. The registration coverage increased with time, from 51.2 % in 1992 to 90 % in 1998 and virtually 100 % in 2005 and afterwards [4].

The current study population included 413,611 women who were invited for the first time at age 50–51 years in 1992–2004 and were followed up until 2012. Altogether, 2,627,256 invitations were identified during the period 1992–2012, out of which 2,283,706 (87 %) visits were made with an average of 5.5 visits per woman. Records with missing data on symptoms were excluded from the analysis (Table 1). The maximum number of visits per woman was 10, and visits exceeding 10 (145 visits) due to migration within the country were excluded from the current analysis.

Definition of variables

Test positives are those with primary mammography positive—they are recalled for further assessment (often more mammograms, ultrasound, and needle biopsy) at the screening clinic, if the mammogram indicated any abnormality. The assessment part is called an episode and those with a positive episode are referred to hospital for diagnostics/treatment. Test positives may be episode negative (no referral) or episode positives (referred) and those who are then diagnosed with cancer are true positives at all stages. False positive test are those with negative episode or with a positive episode but no cancer diagnosis at hospital. False positive mammography tests were further classified as at least one or first false positives depending on the screening history: 'at least one' if a woman was detected as false positive at any given screening visit irrespective of earlier visit findings and 'first' if a woman was detected as false positive at any given screening visit given that mammography in all previous visits was negative. False positive referrals are those with episode positive but no cancer diagnosis in hospital. The average number of visits per woman was defined as the total number of visits made at ages 50–69 years divided by the number of women screened during that period of age. Number of invitations per woman was counted as the number of subsequent invitations a woman received after the first invitation at age 50–51 years.

Women with symptoms reported either by the woman herself or by the radiographer were considered as symptomatic. Symptoms history variable for either lump or retraction or secretion, was created and defined as symptoms reported ever before or at the index visit. Here, index visit means the visit that resulted in a positive test result (either false positive test or true positive test). The possibility of reporting more than one symptom at a single screening visit was also considered. For that, combinations of two symptoms at a time were made as 'none,' 'either' and 'both.' Separate variables for each symptom reported once or more than once in the screening history were created and coded as '1 time' and 'more than 1 time.' A separate variable on the absolute number of visits (1–10) per woman was created to compare the probability of false positive test by screening visits, overall versus those with symptoms history.

Statistical analysis

Lump, retraction, and secretion, the most clinically relevant symptoms, were used for analysis. Let i be the index subjects $i = 1,\ldots, n$ and j be the index visits of ith subject $j = 1,\ldots, J_i$. We note by $P(Y_{ij} = 1; X_{ij})$ the probability of a false positive test for subject i at the jth screen given covariates X_{ij}. The cumulative risk of first outcome event after k rounds of screening is $q_k = 1 - \prod_{j=1}^{k}\{1 - P(Y_{ij} = 0; Y_{i(j-1)} = 0, \ldots, Y_1 = 0)\}$ [8]. Applying discrete-time hazard model with $logit\big(P(Y_{ij})\big) = X'_{ij}\beta$ an estimator for cumulative risk can be obtained. A standard logistic regression can be used to get an estimate of the logistic regression model parameters. Suppose that subject i had symptoms at the lth attended visit. For each subject i the visits can be divided into non-symptomatic $j = 1,\ldots, l-1$ visits and symptomatic visits $j = l,\ldots, J$ starting from the first symptomatic visit: $\big\{\big(y_{ij}, X_{ij} = 0\big); i = 1, \ldots, I; j = 1, \ldots, l-1\big\}$ and $\big\{\big(y_{ij}, X_{ij} = 1\big); i = 1, \ldots, I; j = l, \ldots, J\big\}$. Cumulative risk of false positive test and true positive (cancer diagnosis) was estimated as shown above. Generalized linear regression (GLM) model in R statistical software was used to estimate the effect of an individual symptom as well as combined symptoms on the false positive and true positive probabilities. Confidence intervals at 95 % were estimated using approximate Bayesian inference (INLA) [9].

Results

In 56,805 (2.5 %) visits at least one symptom was reported during the study period in 1992–2012 with a maximum follow-up of 21 years. A lump was reported in 26,145 (1.22 %) visits, retraction in 26,653 (1.59 %) visits, and secretion was reported in 5325 (0.24 %) visits (Fig. 1). There were combined symptoms, as well, with both lump and retraction at 557 visits, lump and secretion at 572 visits, and retraction and secretion at 207 visits. Overall, 48,873 visits (2.1 %) out of total visits had false positive tests. Of these, 44,541 false positive tests were confirmed one time and 4332 false positive test were confirmed more than one time in women screening history. The false positive test percentage at a given visit was 7.2 % (4063 visits) in women with symptoms compared to 2.0 % (44,810 visits) in women with no symptoms. Similarly, the true positive (breast carcinoma) percentage was 2.2 % (1230 visits) in women who reported symptoms compared to 0.4 % (9718 visits) in women with no symptoms (Fig. 1).

The percentage of women who reported a lump or secretion was higher in younger age groups compared to the older age groups (lump = 1.71 vs. 0.78 %; secretion = 0.32 vs. 0.04 % at 1st and 10th visit, respectively) (Table 1). The false positive proportion among women who reported any symptoms was significantly higher at every visit (order, 1–10) compared to those who did not report any symptoms, overall 7.2 vs. 1.5 %, respectively. False positive test probability based on the absolute number of woman's visits showed similar difference in women with symptom history compared to women with no history of symptoms (Fig. 2). However, false positive test probability was lower in women who had less (absolute) number of visits compared to those who had completed all possible (ten visits) screening visits. Similarly, the false positive referral and true positive proportions were higher among women who reported symptoms versus no reported symptoms, 2.8 vs. 0.6 % and 2.2 vs. 0.4 %, respectively.

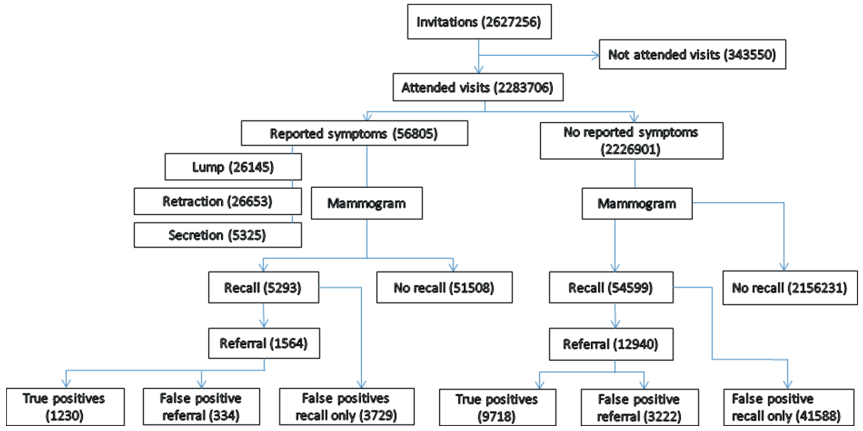

Fig. 1 Flow diagram of mammography screening program by symptom status

Table 1 Number and percentage of symptoms, false positive test and true positives (of symptoms) by number of visits

Symptoms	Number of visit	Number of visits with symptoms (%)	All visits	Missing visits	False positive test		False positive referral		True positives	
					Current symptoms (%)	No current symptoms (%)	Current symptoms (%)	No current symptoms (%)	Current symptoms (%)	No current symptoms (%)
Lump	1	6034 (1.71)	353,348	42,672	922 (15.9)	13,756 (3.98)	69 (1.14)	1196 (0.34)	238 (3.94)	1438 (0.41)
	2	4374 (1.26)	347,915	37,845	569 (13.5)	6890 (2.01)	32 (0.73)	444 (0.13)	146 (3.34)	1110 (0.32)
	3	3814 (1.09)	349,669	24,440	419 (11.4)	6057 (1.76)	33 (0.87)	400 (0.12)	127 (3.33)	1327 (0.38)
	4	3640 (1.08)	337,340	20,857	376 (10.7)	5244 (1.58)	17 (0.47)	317 (0.09)	132 (3.63)	1390 (0.42)
	5	3469 (1.10)	314,846	11,177	378 (11.3)	4564 (1.47)	27 (0.78)	301 (0.10)	118 (3.40)	1536 (0.49)
	6	2442 (1.11)	220,414	3844	270 (11.5)	3047 (1.41)	20 (0.82)	231 (0.11)	97 (3.97)	1233 (0.57)
	7	1522 (1.11)	136,540	2105	126 (8.61)	1868 (1.39)	11 (0.72)	138 (0.10)	59 (3.88)	825 (0.61)
	8	708 (1.07)	66,310	480	62 (9.13)	865 (1.33)	3 (0.42)	64 (0.10)	29 (4.10)	430 (0.66)
	9	122 (1.08)	11,317	36	14 (12.2)	181 (1.63)	0	17 (0.15)	7 (5.74)	79 (0.71)
	10	20 (0.78)	2549	0	4 (21.1)	55 (2.19)	0	6 (0.24)	1 (5.0)	22 (0.87)
	Overall	26,145 (1.22)	2,140,248	143,458	3140 (12.5)	42,527 (1.73)	212 (0.81)	3114 (0.15)	954 (3.65)	9390 (0.44)
Retraction	1	1446 (1.00)	145,113	250,907	81 (5.74)	5529 (3.87)	5 (0.35)	395 (0.27)	36 (2.49)	776 (0.54)
	2	2364 (1.13)	209,617	176,143	69 (2.95)	4271 (2.07)	7 (0.30)	247 (0.12)	27 (1.14)	790 (0.38)
	3	3519 (1.29)	273,512	100,599	109 (3.13)	4772 (1.78)	12 (0.34)	301 (0.11)	31 (0.88)	1169 (0.43)
	4	4597 (1.49)	307,512	50,685	110 (2.42)	4894 (1.62)	8 (0.17)	291 (0.10)	43 (0.94)	1339 (0.44)
	5	5383 (1.76)	306,059	19,964	109 (2.05)	4693 (1.57)	3 (0.06)	313 (0.10)	67 (1.24)	1537 (0.51)
	6	4542 (2.06)	220,381	3877	96 (2.14)	3220 (1.50)	5 (0.11)	246 (0.11)	54 (1.19)	1276 (0.59)
	7	3005 (2.20)	136,540	2105	56 (1.88)	1938 (1.46)	2 (0.07)	147 (0.11)	31 (1.03)	853 (0.64)
	8	1502 (2.27)	66,310	480	30 (2.01)	897 (1.39)	4 (0.27)	63 (0.10)	12 (0.80)	447 (0.69)
	9	244 (2.16)	11,317	36	6 (2.48)	189 (1.72)	0	17 (0.15)	2 (0.82)	84 (0.76)
	10	51 (2.0)	2549	0	2 (3.92)	57 (2.30)	0	6 (0.24)	0	23 (0.92)
	Overall	26,653 (1.59)	1,678,910	604,796	668 (2.54)	30,460 (1.85)	46 (0.17)	2026 (0.12)	303 (1.14)	8294 (0.50)
Secretion	1	1142 (0.32)	358,583	37,437	126 (11.1)	14,557 (4.09)	20 (1.75)	1235 (0.35)	6 (0.53)	1712 (0.48)
	2	1004 (0.28)	360,093	25,667	71 (7.13)	7557 (2.11)	13 (1.29)	483 (0.13)	8 (0.80)	1302 (0.36)
	3	909 (0.25)	360,910	13,201	68 (7.55)	6598 (1.84)	10 (1.10)	451 (0.13)	8 (0.88)	1506 (0.42)
	4	807 (0.23)	350,660	7537	63 (7.89)	5717 (1.64)	17 (2.11)	337 (0.10)	9 (1.12)	1588 (0.45)
	5	667 (0.21)	322,477	3546	54 (8.24)	4993 (1.56)	13 (1.95)	322 (0.10)	12 (1.80)	1688 (0.52)
	6	481 (0.22)	220,398	3860	25 (5.26)	3292 (1.51)	5 (1.04)	246 (0.11)	6 (1.25)	1324 (0.60)
	7	209 (0.15)	136,522	2123	9 (4.45)	1985 (1.47)	4 (1.91)	145 (0.11)	2 (0.96)	881 (0.65)
	8	92 (0.14)	66,297	493	5 (5.62)	922 (1.40)	1 (1.09)	66 (0.10)	3 (3.26)	456 (0.69)
	9	13 (0.11)	11,312	41	2 (15.4)	193 (1.72)	2 (15.4)	15 (0.13)	0	86 (0.76)
	10	1 (0.04)	2548	1	0	59 (2.34)	0	6 (0.24)	0	23 (0.90)
	Overall	5325 (0.24)	2,189,800	93,906	423 (8.03)	45,873 (2.11)	85 (1.60)	3306 (0.15)	54 (1.01)	10,566 (0.48)

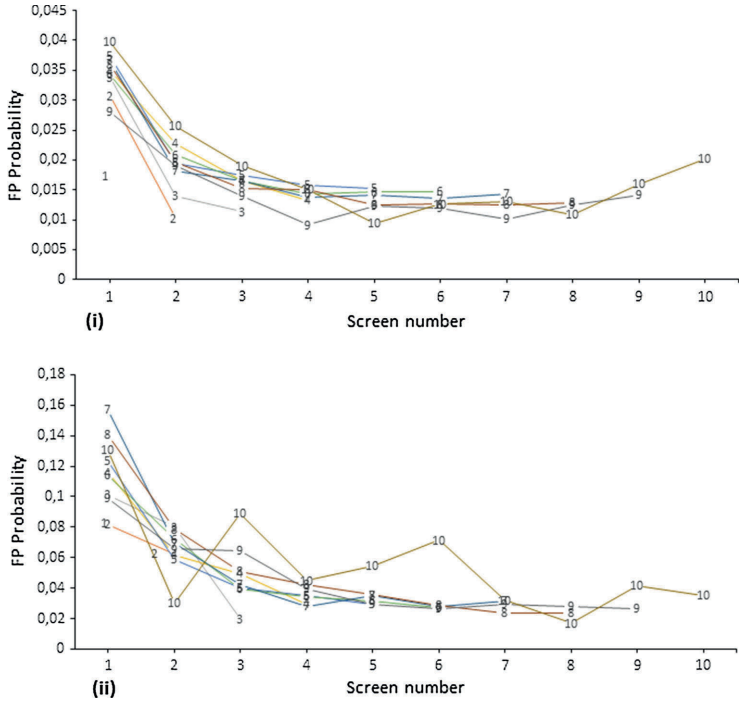

Fig. 2 False positive (FP) test probability; overall (**i**) and any symptoms (**ii**), by attended number of screening visits of women

Table 2 shows the at least one cumulative false positive test and true positive probability after 10 visits. The cumulative probabilities of at least one false positive test, false positive referral, and true positive were 18.2, 1.5, and 5.7 %, respectively, after 10 visits. The cumulative probability of first false positive test was 15.9 % (not shown in Table).

The cumulative probability of having at least one false positive test was significantly higher in those who had a history of lump compared to those with no history of lump, 45.2 vs. 17.2 % estimated for 10 visits. Cumulative probability of at least one false positive referral and true positive in women who reported any symptoms in screening history

Table 2 Cumulative probability of at least one false positive (FP) test, FP referral, and true positive after 10 screening visits

Screen number	FP test probability	Cumulative FP test probability	FP referral probability	Cumulative FP referral probability	True-positive probability	Cumulative probability of true-positive
1	0.0407	0.0407	0.0034	0.0034	0.0047	0.0047
2	0.0216	0.0614	0.0014	0.0048	0.0037	0.0084
3	0.0187	0.0790	0.0013	0.0061	0.0042	0.0125
4	0.0167	0.0944	0.0010	0.0071	0.0045	0.0170
5	0.0159	0.1089	0.0010	0.0081	0.0053	0.0222
6	0.0154	0.1226	0.0011	0.0092	0.0060	0.0281
7	0.0149	0.1357	0.0011	0.0103	0.0065	0.0344
8	0.0141	0.1479	0.0010	0.0113	0.0069	0.0411
9	0.0174	0.1627	0.0015	0.0128	0.0076	0.0483
10	0.0224	**0.1822**	0.0024	**0.0151**	0.0090	**0.0569**

Bold numbers indicate the final cumulative number

were 3.8 and 12.6 %, respectively, compared to 1.4 and 5.3 %, in women with no history of any symptom. (Table 3) There was some increase in the probability of false positive test before the visit with a lump compared to visits with no lump, though true positive probability did not differ (see supplementary table, S1). Women who reported lump or secretion more than one time had higher cumulative probability of at least one false positive test than women who reported lump or secretion once in screening history, 47.8 vs. 44.0 % for lump and 39.8 vs. 33.4 % for secretion, respectively. However, cumulative probability of true positive was lower in women who reported symptoms more than one time compared to one time in screening history.

The cumulative false positive probability in women who reported 'lump and retraction' was higher, 56.5 % (95 % CI 47.4–66.3) compared to those who did not report either symptom, 17.1 % (95 % CI 16.6–17.7) (Fig. 3). For Women who reported 'lump and secretion, the cumulative false positive test probability was 54.8 % (95 % CI 45.3–69.6).

Discussion

Our study found significantly higher cumulative false positive test and true positive probability among those who reported symptoms at screen compared to those who did not report any symptoms. The cumulative risk of false positive test (after 10 rounds) with any symptom was 38 % and that without was 17 %. Lump was associated with the highest cumulative false positive risk of 45 %, retraction 25 %, and secretion 35 %.

Table 3 Cumulative probability of at least one false positive (FP) test, FP referral, and true positive in women with a history of symptoms

Symptoms history	Cumulative probability of false positive test	Cumulative probability of false positive referral	Cumulative probability of true positive
Lump			
Yes	0.4516	0.0332	0.1630
1 time	0.4401	0.0357	0.2002
>1 time	0.4780	0.0268	0.0650
No	0.1721	0.0146	0.0531
Retraction			
Yes	0.2464	0.0261	0.0903
1 time	0.2662	0.0429	0.1868
>1 time	0.2342	0.0159	0.0368
No	0.1807	0.0204	0.0567
Secretion			
Yes	0.3477	0.0789	0.0638
1 time	0.3339	0.0677	0.0691
>1 time	0.3981	0.1239	0.0489
No	0.1811	0.0146	0.0569
Any symptom			
Yes	0.3843	0.0377	0.1262
1 time	0.3938	0.0422	0.1730
>1 time	0.3694	0.0315	0.0533
No	0.1699	0.0138	0.0530

The overall cumulative probability of at least one false positive test was 18 % after 10 screening visits at age 50–69 years and the false positive test probability was 3.6 % at the first visit at age 50–51 years. Our results are consistent or somewhat lower with that of previous studies from other European countries [10–17]. A study from Norway reported a higher cumulative false positive risk (23 %) than the current study [10]. Another study estimated a 21 % cumulative false positive probability projected after 10 screening visits, based on the results of three consecutive screening visits performed in four counties [11]. A retrospective cohort study from Spain projected the cumulative false positive risk to be 20.4 % after 10 screening visits [12]. Cumulative false positive probability from a randomized trial in the UK (2010) was 20.5 % over seven screening rounds [16]. A Danish study [14] made the prediction, based on 3–5 observed screening rounds, of cumulative false positive test probability slightly lower than that of our study. However, the false positive test probability at first screen was higher (5.7 %) in Copenhagen than that of the current study. In the Netherlands, Otten et al. (2013) found lower cumulative false positive risk after 13 consecutive screening examinations than that of our study, but they expected higher estimates after digital mammography was introduced in 2003 [18, 19]. Nonetheless, there were some variations between countries in the methodology and health service system, such as age at first invitation [10, 14, 18], projected estimates based on few observed rounds [11, 18], and lower recall proportion of <1 % at subsequent screens [18, 20] compared to 2.2 % in our study and <3 % in European guidelines [21], while estimating the false positive risk.

Studies conducted in the USA have reported much higher risk of cumulative false positive tests than that of the current study [8, 22–25]. In the US, Breast Cancer Surveillance Consortium (BCSC) data from all women ($n = 88{,}455$) first screened at age 50-69 years between 1996 and 2010 estimated the cumulative false positive risk to be 41.9 % after eight screens annually or biennially [24]. The reason for lower estimates in our study may be due to different program organizations in Finland than in USA as well as variation in age at first screening, definitions of recall, recording and coding of screening data, screening interval, etc. [26]. Also, the European quality standards [21] are adequately met by the Finnish screening program.

Together with the cumulative probability of 'at least one' false positive, this study also estimated the cumulative 'first' false positive test and true positive probability. The

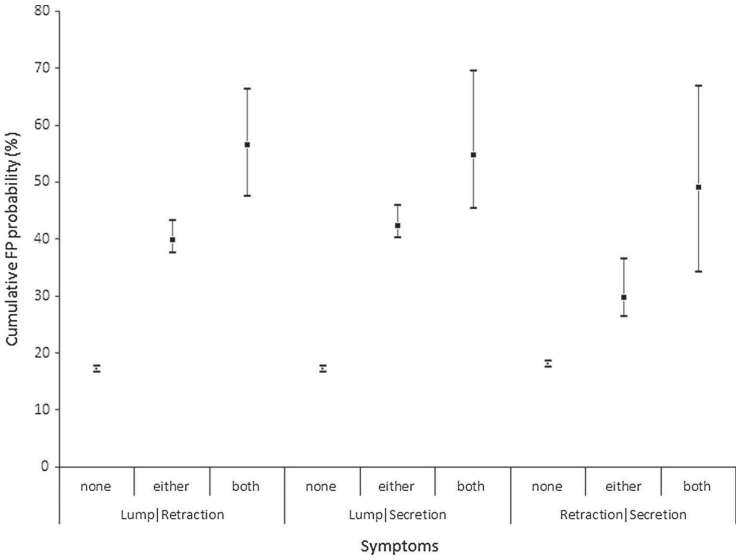

Fig. 3 Cumulative probability of at least one false positive (FP) test among women reported symptoms at screen

cumulative first false positive probability was 16 % as we considered only the first false positive mammography result, excluding later false positive findings of the same woman. Hence, the estimate is lower than the 'at least one' false positive estimate. Also, the lower probability of false positive in our study may be due to the exclusion of the first visits made at later age, hence removing contamination of newcomers at later visits with prevalent screens. Our study estimated the cumulative true positive probability to be 6 % after 10 screening visits. A study in the Netherlands estimated similar cumulative cancer detection risk after 13 consecutive screening examinations [18]. We are not aware of other studies on cumulative true positive estimates after 10 screening visits.

No prior studies have estimated the cumulative probability based on reported symptoms with a complete follow-up information. Women who reported having symptoms, especially lump and secretion at screening visit, current or at any previous visit, were significantly more likely to have a false positive test and true positive result than women with no symptoms reported. The cumulative false positive probability in women with a history of lump was 45 % compared to 17 % with no lump. When considering the full visit history of women with lump, before and after visit with lump, the higher probability of false positive test before the visit with a lump indicates that there was a possibility that some unspecific changes in the mammograms had been seen even several years before the visit when a lump was reported. On the other hand, after reporting the first symptom there was no increase in the probability of false positive test and true positive results in the later visits. This means that woman was treated and no cancer was detected in later visits. Women were more likely to be true positive if they reported symptoms at screen; cumulative true positive probability of 16 % was compared to 6.5 % with no reported lump. Similarly, women who reported both 'lump and retraction' in the same visit had cumulative false positive test probability of 56.5 % (95 % CI 47.4–66.3) compared to 17.1 % (95 % CI 16.6–18.3) without symptoms. Similar results were found in women with other possible combination of symptoms. Taking into account the information on breast symptoms, there is a concern for the radiologist whether or not to recall the symptomatic women. Also, variation in the false positive probability by symptom status, number of times symptom was reported, shows that not all symptoms are equally sensitive. At the same time, the findings also showed benefits of evaluating symptoms information on the performance (more cancers detected) of mammography screening program.

One of the limitations of this study is the missing information on some important risk factors such as hormone use, breast density, and family history of breast cancer, while estimating the cumulative false positive and

true positive probability in relation to symptoms. The missing information (1.2 % of total visits) on symptoms was due to incomplete reporting by some centers in the early years of the program. Women recalled but not referred to hospital and women referred but with no cancer in histological confirmation who may have had a cancer before the next screening visit (interval cancer), were not taken into account in this study. Other performance measures of screening program, including interval cancers and mortality as stated by Otten et al. [18] and Tornberg et al. [27], in relation to breast symptoms need to be evaluated thoroughly.

The current study is based on a large nationwide screening cohort with complete follow-up of the women up to maximum 10 visits (21 years). The high participation rate (>85 %) in the screening program and few opportunistic screening means false positive probability estimates over the 10 screens equals the lifetime risk of false positive test in Finland, which is similar to that reported by a Danish study [14]. The radiologists learning of the previous mammography results and the small difference between 'at least one' and 'first' cumulative probability estimates form the basis to conclude independence between false positive risks at subsequent screen.

In conclusion, the current study showed that information about breast symptoms, especially lump, cause harms in terms of extra false positive findings. The risk varies substantially, depending on symptom types and characteristics. At the same time, more cancers were detected in symptomatic women suggesting benefits of evaluating symptoms information in the program. Information on breast symptoms influences the balance of absolute benefits and harms of screening for the individual woman, and should be considered carefully in breast cancer screening programs.

Acknowledgments We thank Prof. Stephen Duffy, Wolfson Institute of Preventive Medicine, Queen Mary University of London and Dr. Rengaswamy Sankaranarayanan, Screening group, International Agency for Research on Cancer (IARC) for helpful comments.

Funding This work was supported by the Cancer Foundation Grant of Finland.

Compliance with ethical standrads

Conflict of interest The authors have no conflict of interest or disclosures.

References

1. Lauby-Secretan B, Scoccianti C, Loomis D et al (2015) Breast-cancer screening—viewpoint of the IARC working group. N Engl J Med 372(24):2353–2358
2. Kosters JP, Gotzsche PC (2003) Regular self-examination or clinical examination for early detection of breast cancer. Cochrane Database Syst Rev 2:CD003373
3. IARC (2002) Breast cancer screening—IARC handbook of cancer prevention. http://www.iarc.fr/en/publications/pdfs-online/prev/handbook7/index.php. Accessed 30 Sep 2015
4. Finnish Cancer Registry (2015) Breast cancer screening; mass screening programme—cancer society of Finland. http://www.cancer.fi/syoparekisteri/en/mass-screening-registry/breast_cancer_screening/screening_programme/. Accessed 30 Sep 2015
5. Singh D, Malila N, Pokhrel A, Anttila A (2015) Association of symptoms and breast cancer in population-based mammography screening in Finland. Int J Cancer 136(6):E630–E637
6. Brouckaert O, Schoneveld A, Truyers C et al (2013) Breast cancer phenotype, nodal status and palpability may be useful in the detection of overdiagnosed screening-detected breast cancers. Ann Oncol 24(7):1847–1852
7. Gill G, Luke C, Roder D (2006) Prognostic importance of palpability as a feature of screen-detected breast cancers. J Med Screen 13(2):98–101
8. Christiansen CL, Wang F, Barton MB et al (2000) Predicting the cumulative risk of false-positive mammograms. J Natl Cancer Inst 92(20):1657–1666
9. Rue H (2015) Bayesian computing with INLA. The R-INLA Project. http://www.r-inla.org/home. Accessed 20 Sep 2015
10. Roman M, Skaane P, Hofvind S (2014) The cumulative risk of false-positive screening results across screening centres in the Norwegian breast cancer screening program. Eur J Radiol 83(9):1639–1644
11. Hofvind S, Thoresen S, Tretli S (2004) The cumulative risk of a false-positive recall in the Norwegian breast cancer screening program. Cancer 101(7):1501–1507
12. Roman R, Sala M, Salas D et al (2012) Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. Ann Oncol 23(1):104–111
13. Roman M, Hubbard RA, Sebuodegard S et al (2013) The cumulative risk of false-positive results in the Norwegian breast cancer screening program: updated results. Cancer 119(22):3952–3958
14. Njor SH, Olsen AH, Schwartz W et al (2007) Predicting the risk of a false-positive test for women following a mammography screening programme. J Med Screen 14(2):94–97
15. Castells X, Molins E, Macia F (2006) Cumulative false positive recall rate and association with participant related factors in a population based breast cancer screening programme. J Epidemiol Community Health 60(4):316–321
16. Johns LE, Moss SM, Age Trial Management Group (2010) False-positive results in the randomized controlled trial of mammographic screening from age 40 ("age" trial). Cancer Epidemiol Biomark Prev 19(11):2758–2764
17. Hofvind S, Ponti A, Patnick J et al (2012) False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. J Med Screen 19(Suppl 1):57–66
18. Otten JD, Fracheboud J, den Heeten GJ et al (2013) Likelihood of early detection of breast cancer in relation to false-positive risk in life-time mammographic screening: population-based cohort study. Ann Oncol 24(10):2501–2506

19. Karssemeijer N, Bluekens AM, Beijerinck D et al (2009) Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. Radiology 253(2):353–358
20. Otten JD, Karssemeijer N, Hendriks JH et al (2005) Effect of recall rate on earlier screen detection of breast cancers based on the Dutch performance indicators. J Natl Cancer Inst 97(10): 748–754
21. Perry N, Broeders M, de Wolf C et al (2006) European guidelines for quality assurance in breast cancer screening and diagnosis. European Commission, Brussels
22. Braithwaite D, Zhu W, Hubbard RA et al (2013) 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 105(5):334–341
23. Hubbard RA, Kerlikowske K, Flowers CI et al (2011) Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Ann Intern Med 155(8):481–492
24. Jacobsen KK, Abraham L, Buist DS et al (2015) Comparison of cumulative false-positive risk of screening mammography in the United States and Denmark. Cancer Epidemiol 39(4):656–663
25. Elmore JG, Barton MB, Moceri VM et al (1998) Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med 338(16):1089–1096
26. Yankaskas BC, Klabunde CN, Ancelle-Park R et al (2004) International comparison of performance measures for screening mammography: can it be done? J Med Screen 11(4):187–193
27. Tornberg S, Kemetli L, Ascunce N et al (2010) A pooled analysis of interval cancer rates in six European countries. Eur J Cancer Prev 19(2):87–93

PUBLICATION III

Association of symptoms and interval breast cancers in the mammography-screening programme: population-based matched cohort study

Singh, D., Miettinen, J., Duffy, S., Malila, N., Pitkäniemi, J., Anttila, A.

Br J Cancer, 119(11), 1428-1435
Doi: 10.1038/s41416-018-0308-2

BJC British Journal of Cancer

www.nature.com/bjc

ARTICLE

Epidemiology

Association of symptoms and interval breast cancers in the mammography-screening programme: population-based matched cohort study

Deependra Singh[1,2], Joonas Miettinen[1], Stephen Duffy[3], Nea Malila[1,2], Janne Pitkäniemi[1] and Ahti Anttila[1]

BACKGROUND: We assessed the association between symptoms reported at breast cancer screening visits and interval cancers (ICs) in a prospective manner.

METHODS: This population-based matched cohort study uses data of the Finnish National Breast Cancer Screening Programme that invites women aged 50–69 years old during 1992–2012. Subjects who attended screening with symptoms were matched with asymptomatic reference cohorts based on age at screening visit, year of invitation, number of invited visits and municipality of invitation. The primary outcome was ICs.

RESULTS: Women with a lump had a threefold (hazard ratio 3.7, 95% confidence interval (CI) 3.0–4.6) risk of ICs and a higher risk (hazard ratio 1.7, 95% CI 1.4 to 2.0) at the subsequent visit compared with those without a lump. The fatal interval cancer risk increased by 0.39 per 1000 screens with a lump. The cumulative incidences of interval cancer increased within a month of a mammography-negative visit with a lump and after about 6 months of the visit with retraction or nipple discharge.

CONCLUSION: Women with breast symptoms have a clearly increased risk of interval breast cancer after the screening visit. Our findings indicate the need for different screening strategies in symptomatic women.

British Journal of Cancer (2018) 119:1428–1435; https://doi.org/10.1038/s41416-018-0308-2

BACKGROUND

Beyond the randomised trial environment, there is some uncertainty as to the underlying incidence of breast cancer and the rate of overdiagnosis in the screening population.[1,2] Along with screening performance,[3–5] rates and proportions of interval cancers (ICs) are important indicators for assessing the effectiveness and quality of screening.[2,6–8] The substantial proportion (about a third) of incident breast cancer diagnosed outside the mammography-screening programme[9,10] likely indicates that there is room for improvement in the detection capability of the mammography-screening programme. Earlier observational studies[7,11–15] have highlighted several reasons for the increased proportion of ICs in the screening programme. However, in terms of equity within the screening population, it is reasonable to aim for similar interval cancer rates or at least similar proportions of cancers arising as interval cases for the various heterogeneous groups participating in the screening. Furthermore, options to modify screening policies should be considered for high-risk groups. A shorter screening interval may be justified, for instance, if the interval cancer rate is significantly high.

Based on the European Union (EU) guidelines, screening is meant for unselected target population.[7] Earlier studies from Finland have indicated that a noticeable proportion (~ 2–3%) of women have clinically significant symptoms when they participate in breast cancer screening.[16,17] Most but not all symptomatic women will have further assessments with ultrasound, additional mammograms or other methods; if these return negative results, the women return to the normal, biennial screening interval. Studies on breast symptoms (such as a lump, retraction or nipple discharge) indicate an increased risk of breast cancer[16–19] at the cost of rise in false-positive findings. The relations of symptoms with interval breast cancers and screen-detected cancers (SDCs) at the subsequent visit have never been studied.

We investigated whether women reporting breast symptoms at screen are at a higher risk of developing subsequent breast cancers (ICs and cancers diagnosed at the next screen) than those without symptoms. To create foundations to modify the screening policies in high-risk groups, we estimated the cumulative incidence of ICs and fatal interval breast cancers, and compared the respective incidences in women with and without symptoms. The quality measures of screening mammography were compared between visits among subjects with and without symptoms to gather evidence for improving programme performance.

METHODS

Study design, data source and study population

Our matched cohort study design was based on the follow-up of the ongoing Finnish National Breast Cancer Screening Programme that began in 1987. Biennial screening visits made by women

[1]Mass Screening Registry, Finnish Cancer Registry, FI-00130 Helsinki, Finland; [2]Epidemiology group, Department of Health Sciences, University of Tampere, FI-33520 Tampere, Finland and [3]Wolfson Institute of Preventive Medicine, Queen Mary University of London, Charterhouse Square, London, England
Correspondence: Deependra Singh (deependra.singh@cancer.fi)

Received: 10 July 2018 Revised: 25 September 2018 Accepted: 26 September 2018
Published online: 7 November 2018

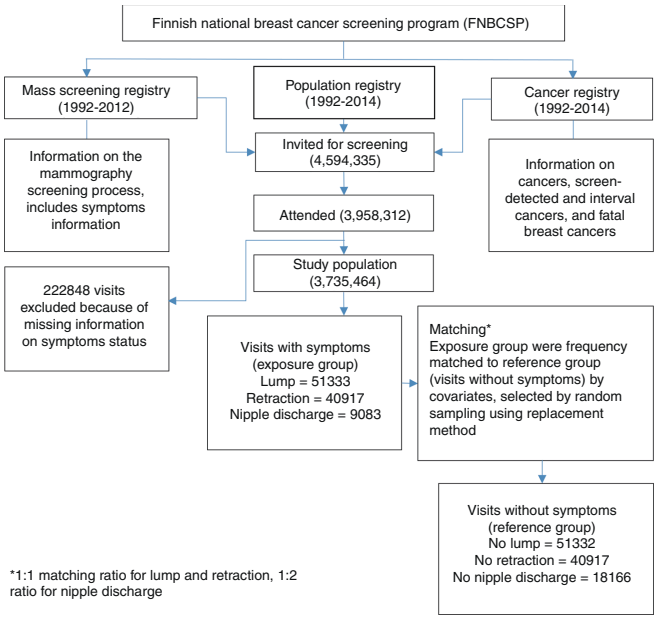

Fig. 1 Flow diagram of study settings

aged 50–69 years between 1991 and 2012 were selected. Three registries, the Finnish Cancer Registry (1953–2014), the Mass Screening Registry (1992–2012) and the Central Population Registry (1992–2014) were used to extract information on the study participants at the individual level. The Mass Screening Registry was used to extract information on demographic, symptomatic and screening procedure factors, including recalls and referral data that have been shown to be valid and of high quality. All individual visits were linked to the Finnish Cancer Registry database to retrieve information on breast cancers (screen-detected and ICs). This included histological findings and potential death from incident breast cancer. The Population Registry was used to identify possible dates of death or emigration, and where applicable, the cause of death was retrieved from Statistics Finland. Fig. 1 shows the flow diagram of the study design.

Exposure group (screening visits with symptoms)
The exposed group is defined as visits by women with breast symptoms (lump, retraction and nipple discharge) reported at a given screening round. This group contained all visits with at least one symptom reported. Different symptoms (lump, retraction, nipple discharge) were analysed separately, which in technical terms assumes that the first occurrence of any symptom was independent given the covariates. For example, if more than one symptom was reported at a single visit then each symptom was analysed separately. Here, the index visit meant any screening visit with any given symptom.

Reference group (screening visits without symptoms)
The reference group is defined as visits by women with no reported breast symptoms in the screening history before the index visit. The women from the reference group can later be the part of the exposed group if symptoms are reported at future screening visits. Thus, symptoms are time-dependent covariates. Individual sets of reference visits for each symptom—altogether three sets—were formed by matching.

Matching
The three exposed groups were frequency matched to the reference groups by age at the screening visit (within 2 years), year of invitation (2-year band), number of visits in the past and municipality of invitation. Visits with symptoms were then aggregated based on matching variables (and other covariates). Each symptom stratum was matched to the viable controls (reference visits) by random sampling. Random controls were selected, based on the matching variables above, as many times as the number of visits in each stratum of symptoms by the replacement sampling method. Hence, a single control had the possibility to be randomly selected more than once to the same stratum. Based on our assumption of an effect size and required power of 0.80, the exposed-to-reference-visits ratio was 1:1 for lump and retraction, and 1:2 for nipple discharge.

Outcome assessment
ICs were defined as breast cancers diagnosed in screened women before the next screening visit or within a period equal to a screening interval with (i) negative mammography at the index visit (i.e., test negative); (ii) positive mammography at the index visit, but negative further assessment (i.e., episode negative); and (iii) positive further assessment but a date of diagnosis > 6 months after mammography.[2] SDCs were defined as primary breast cancer diagnosed among the screening attendees within 6 months following an abnormal mammogram (test positive). The subsequent round SDCs were analysed following the index visits with or without symptoms if the women attended the subsequent round.

In addition, cancers were sub-grouped into in situ carcinomas and non-localised breast cancers. Fatal cancers were defined as those breast cancers that resulted in death during follow-up.

Follow-up

The follow-up time started from the index visit in 1 January 1992 to 31 December 2012 and ended at the date of emigration or death, upon diagnosis of interval cancer or at the end of the follow-up—i.e., 31 December 2014—whichever occurred first. Cancer cases diagnosed among those screened up to 31 December 2012 and followed up to 31 December 2014 (for those screened in 2011 and 2012) were divided into ICs and subsequent SDCs using Finnish Cancer Registry data. Considering possible delays in the diagnosis date after positive mammography findings, a screening episode of 6-month intervals was used in the definition of detection mode. Thus, the follow-up time for ICs started at 7 months for episode negative visits and at 1 month for test negative visits and ended at the date of the subsequent screening visit at 23 months.

Statistical analysis

We compared breast cancer risk and the risk of breast cancer death using Cox proportional hazard regression among women with and without reported symptoms at the index screen. Confidence intervals were computed exactly from parameter likelihoods. The analyses were adjusted for age at the screening visit. We calculated the incidence rate of interval cancer between the screens (from the index screen and the 6-month episode to the subsequent screening visit) separately for test negatives and episode negatives.

We also evaluated the programme characteristics using basic statistics. A test sensitivity was estimated as the number of visits with a positive mammography test and diagnosis of cancer at screen divided by the sum of SDCs plus ICs diagnosed after negative test results. Episode sensitivity was calculated as the number of visits with a diagnosis of cancer in a full diagnostic process in the screened population divided by all cancers detected in a screening round among attenders. Similarly, the positive predictive value (PPV) was assessed as the number of visits with a positive mammography test and diagnosis of cancer divided by the number of test positives. The negative predictive value (NPV) was estimated as the number of visits with a negative test result and no cancer diagnosed divided by the number of test negatives. All statistical analyses were performed using R-3.4.0.

RESULTS

Over the study period of 21 years, a lump was reported at 51,333 visits and retraction at 40,917 visits. These visits were matched to an equal number of asymptomatic visits. There were 9083 visits with nipple discharge, and they were matched with double the number of reference visits (i.e., 18,166 visits) without nipple discharge. Detailed numbers of the potentially eligible and the confirmed eligible population included in the study are shown in Fig. 1.

The mean age at a screening visit did not differ between visits with and without symptoms (mean age 55.7 vs 56.1 years for visits with and without a lump, respectively). Table 1 shows the characteristics of the final study cohort. About one in three women who reported a lump or nipple discharge and one in eight women who reported retraction were first-time attendees. The first and subsequent attendee's proportions were similar between visits with and without symptoms. More than 80% of ICs and the subsequent round's SDCs, irrespective of reported symptoms status, were not recalled for further assessment at the index visit.

Both the test and episode sensitivity of the mammography was higher for visits with a lump or retraction compared with those without such symptoms (82 and 75% vs 64 and 63% for a lump vs no lump; 77 and 76% vs 67 and 66% for retraction vs no retraction; Table 2). Likewise, the PPV of mammography was higher for retraction and a lump, compared with those without these symptoms. However, the specificity of mammography was clearly lower for visits with a lump (88%) than those without (98%). Some decrease in specificity was also seen for retraction and nipple discharge.

Incidence of screen-detected and ICs, subsequent SDCs and fatal ICs

In total, 1440 (2.8%) SDCs and 387 (0.7%) ICs (ICs) were diagnosed in those who reported a lump compared with 174 (0.3%) SDCs and 103 (0.2%) ICs in those without a lump, respectively (Table 3). The proportions of SDCs and ICs were higher also for retraction and nipple discharge compared with those without these symptoms. The age-adjusted risk of SDCs was significantly higher in those who reported a lump (adjusted hazard ratio 8.2, 95% CI 7.0–9.7), retraction (adjusted hazard ratio 2.3, 95% CI 2.0–2.8) or nipple discharge (adjusted hazard ratio 1.5, 95% CI 1.1–2.3) compared with those without symptoms. In addition, the age-adjusted risk of ICs was significantly higher for a lump (adjusted hazard ratio 3.7, 95% CI 3.0–4.6), retraction (adjusted hazard ratio 1.5, 95% CI 1.1–1.9), and nipple discharge (adjusted hazard ratio 2.4, 95% CI 1.6–3.7) compared with those without these symptoms. The risk of SDCs in the subsequent round was significantly higher only after visits with a lump compared with those without a lump (adjusted hazard ratio 1.6, 95% CI 1.3–2.0). The risk of in situ interval carcinomas or subsequent screen-detected carcinomas was also higher in visits with a lump and nipple discharge compared with a visit with no symptoms. The risk of non-localised interval breast cancer as well as SDCs in the subsequent round were also greater for all three symptoms in comparison with visits without symptoms.

The age-adjusted risk of dying from breast cancer was significantly higher in those who reported a lump and were diagnosed with invasive cancers (SDCs = adjusted hazard ratio 19, 95% CI 11–38; ICs = adjusted hazard ratio 2.0, 95% CI 1.1–3.4; subsequent round cancers = adjusted hazard ratio 2.7, 95% CI 1.6–4.1) compared with those without a lump (Table 3). In addition, the risk of dying was higher in those who reported retraction and were diagnosed with SDCs (adjusted hazard ratio 6.3, 95% CI 2.8–16) compared with those without retraction. Only a few deaths occurred during the follow-up in those who reported nipple discharge.

Cumulative incidence of breast cancers during the screening interval

The incidence of ICs after a visit with a lump or nipple discharge increased rather rapidly after the visit in all the studied progression types (Fig. 2 a–d). When using the cumulative incidence of the asymptomatic over the whole interval as a reference, the same level of ICs was reached in only about six months after visits with a lump and 12 months after visits with nipple discharge. Remarkably, in mammography-negative visits with a lump, the cumulative incidence curve detached from the no lump curve immediately after the first month of visit, whereas such a difference was not observed for the other symptom types.

DISCUSSION

In this population-based study with 21 years of screening (the follow-up is restricted to 23 months, although the study period was 21 years), we observed strong associations between symptoms and breast cancer risks. Women reporting a lump at a screening visit had a threefold risk of ICs compared with those with no symptoms, also including subsequent SDCs. The (cumulative) incidence of interval cancer was higher in those who reported a lump irrespective of the mammography findings

Table 1 Cohort characteristics

Characteristics		Lump ($n = 51,333$)		Retraction ($n = 40,917$)		Nipple discharge*	
		Yes (%)	No (%)	Yes (%)	No (%)	Yes (%) ($n = 9083$)	No (%) ($n = 18,166$)
Age at index visit (mean, SD)		55.7 (4.6)	56.1 (4.5)	57.7 (4.9)	56.6 (4.8)	54.9 (4.6)	56.1 (4.7)
Attendance	First attendance	18,305 (35)	17,954 (35)	5012 (12)	4959 (12)	2904 (32)	5643 (31)
	Subsequent attendance	33,028 (64)	33,379 (65)	35,905 (87)	35,958 (87)	6179 (68)	12,523 (68)
Previous round attendance	Yes	31,571 (61)	32,228 (62)	34,168 (83)	34,263 (83)	5842 (64)	12,069 (66)
	No	19,762 (38)	19,105 (37)	6749 (16)	6654 (16)	3241 (35)	6097 (33)
Period of visit	1992–1997	15,478 (30)	15,520 (30)	0	0	1134 (12)	2298 (12)
	1998–2002	7535 (14)	7493 (14)	3552 (8.7)	3552 (8.7)	1673 (18)	3316 (18)
	2003–2007	11,791 (23)	11,804 (23)	12,996 (31)	13,190 (32)	2872 (31)	5671 (31)
	2008–2012	16529 (32)	16,516 (32)	24,369 (59)	24,175 (59)	3404 (37)	6881 (37)
Recall (test positives)							
Invasive cancers	Screen-detected cancers	1440 (2.8)	174 (0.34)	461 (1.1)	192 (0.47)	66 (0.73)	83 (0.46)
	Interval cancers	70 (0.14)	7 (0.01)	16 (0.04)	5 (0.01)	12 (0.13)	6 (0.03)
	Subsequent screen at next round	34 (0.07)	6 (0.01)	11 (0.03)	7 (0.02)	4 (0.04)	3 (0.01)
Fatal breast cancers	Screen-detected cancers	215 (0.41)	11 (0.02)	38 (0.09)	6 (0.01)	4 (0.04)	2 (0.01)
	Interval cancers	8 (0.01)	2 (0.01)	2	2	0	0
	Subsequent screen at next round	1 (0.01)	0	1 (0.01)	0	0	0
No recall (test negatives)							
Invasive cancers	Screen-detected cancers	0	0	0	0	0	0
	Interval cancers	317 (0.62)	96 (0.19)	138 (0.34)	96 (0.23)	40 (0.44)	35 (0.19)
	Subsequent screen at next round	230 (0.45)	149 (0.29)	144 (0.35)	135 (0.33)	28 (0.31)	57 (0.31)
Fatal breast cancers	Screen-detected cancers	0	0	0	0	0	0
	Interval cancers	32 (0.06)	18 (0.03)	13 (0.03)	5 (0.01)	5 (0.05)	2 (0.01)
	Subsequent screen at next round	18 (0.03)	7 (0.01)	4 (0.01)	4 (0.01)	0	1 (0.01)

Values are numbers (percentage) unless stated otherwise
*For each visits with nipple discharge were matched with two visits without nipple discharge

Table 2 Performance quality measures of screening mammography in relation to symptoms status

Symptoms		Screening episode Sensitivity % (95% CI)	Mammography test Sensitivity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Specificity % (95%CI)
Lump	Yes	74 (71–78)	81 (80–83)	19 (19–20)	99 (99–99)	88 (88–88)
	No	62 (56–68)	64 (58–70)	12 (11–13)	99 (99–99)	97 (97–97)
Retraction	Yes	75 (72–79)	76 (73–80)	28 (26–29)	99 (99–99)	97 (96–97)
	No	65 (59–70)	66 (60–72)	17 (16–19)	99 (99–99)	97 (97–98)
Nipple discharge	Yes	55 (46–65)	62 (52–71)	8.1 (6.9–9.4)	99 (99–99)	91 (91–92)
	No	66 (57–75)	70 (61–78)	15 (13–17)	99 (99–99)	97 (97–97)

PPV positive predictive value, *NPV* negative predictive value, *CI* confidence interval

(test negatives or episode negatives) compared with those without symptoms at the index visit. Likewise, retraction and nipple discharge were significantly associated with increased risks of interval breast cancers. In absolute terms, per 1000 women who attended and reported a lump, seven women were diagnosed with invasive interval breast cancers, i.e., within 24 months compared with about two cancers diagnosed without symptoms. The fatal interval cancer risk increased by 0.39 per 1000 screens with a lump.

Women with symptoms had a clearly increased 'background' risk of breast cancer; the conventional screening performance measures—such as sensitivity, PPV and specificity—did not fully assess this aspect. The cumulative incidence patterns as well as the detection of cancers during the subsequent round provided direct evidence of the need for risk-adjusted screening and a better diagnostic work-up in the symptomatic women. The diagnostic work-up could include developing better reading and recall criteria in screening mammograms, a more detailed

Table 3 Frequency (per 1000 screening visits) and age-adjusted risk of cancer outcomes in those who reported symptoms compared with those without symptoms at screen

Outcomes	Lump ($n = 51,333$)			Retraction ($n = 40,917$)			Nipple discharge*		
	Yes (per 1000)	No (per 1000)	Age-adjusted hazard ratio (95% CI) with reference to no lump	Yes (per 1000)	No (per 1000)	Age-adjusted hazard ratio (95% CI) with reference to no retraction	Yes (per 1000) ($n = 9083$)	No (per 1000) ($n = 18,166$)	Age-adjusted hazard ratio (95% CI) with reference to no nipple discharge
Invasive cancers									
Screen-detected cancers	1440 (28)	174 (3.4)	8.3 (7.1–9.7)	461 (11)	192 (4.7)	2.4 (2–2.8)	66 (7.3)	83 (4.6)	1.7 (1.2–2.3)
Interval cancers	387 (7.5)	103 (2.0)	3.8 (3–4.7)	154 (3.8)	101 (2.5)	1.5 (1.2–1.9)	52 (5.7)	41 (2.3)	2.5 (1.7–3.8)
Subsequent screen at next round	264 (5.1)	157 (3.1)	1.7 (1.4–2.1)	156 (3.8)	158 (3.9)	1.1 (0.87–1.4)	32 (3.5)	72 (3.9)	1.0 (0.64–1.5)
In situ carcinomas									
Screen-detected cancers	61 (1.2)	38 (0.74)	1.6 (1.1–2.4)	35 (0.86)	31 (0.76)	1.1 (0.70–1.8)	15 (1.7)	16 (0.88)	1.9 (0.92–3.8)
Interval cancers	35 (0.68)	9 (0.18)	3.9 (1.9–8.6)	7 (0.17)	6 (0.15)	1.2 (0.39–3.6)	9 (0.99)	4 (0.22)	4.0 (1.3–14)
Subsequent screen at next round	159 (3.1)	102 (1.9)	1.5 (1.2–1.9)	76 (1.9)	66 (1.6)	1.2 (0.83–1.6)	32 (3.5)	32 (1.8)	2.0 (1.2–3.3)
Non-localised cancers									
Screen-detected cancers	693 (13)	57 (1.1)	12 (9.2–16.0)	230 (5.6)	61 (1.5)	3.5 (2.6–4.7)	25 (2.8)	27 (1.5)	1.9 (1.1–3.2)
Interval cancers	178 (3.5)	57 (1.1)	3.4 (2.5–4.6)	79 (1.9)	48 (1.2)	1.7 (1.2–2.4)	20 (2.2)	18 (0.99)	2.2 (1.2–4.2)
Subsequent screen at next round	649 (12)	438 (8.5)	1.5 (1.3–1.7)	244 (5.9)	182 (4.5)	1.4 (1.1–1.7)	110 (12)	181 (9.9)	1.2 (0.96–1.6)
Fatal cancers									
Screen-detected cancers	215 (4.2)	11 (0.21)	19 (11–38)	38 (0.93)	6 (0.15)	6.3 (2.9–16)	4 (0.44)	2 (0.11)	4.0 (0.78–28)
Interval cancers	40 (0.78)	20 (0.39)	2 (1.2–3.5)	15 (0.36)	7 (0.17)	2.1 (0.90–5.6)	5 (0.55)	2 (0.11)	5.0 (1.1–34)
Subsequent screen at next round	19 (0.37)	7 (0.14)	2.7 (1.7–4.1)	5 (0.12)	4 (0.07)	1.3 (0.45–2.7)	0	1 (0.05)	NA

*For each visits with nipple discharge were matched with two visits without nipple discharge; *CI* confidence interval, *NA* not available

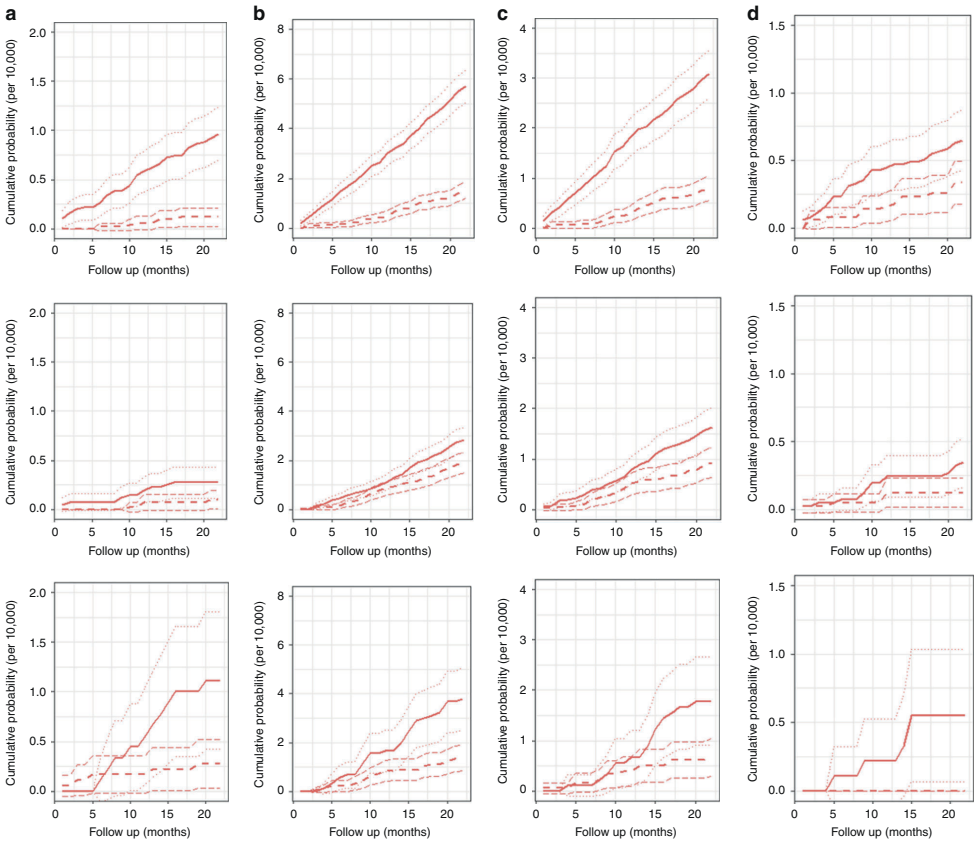

Fig. 2 **a–d** Cumulative incidence of invasive (per 1000): **a** recalled ICs; **b** not recalled ICs; **c** non-localised ICs; **d** fatal interval cancers. Note: the confidence intervals lines for cumulative incidence are indicated by light dotted lines in symptomatic and asymptomatic groups

indication for biopsies in the further assessment, and early recall for those with high-risk symptoms for whom the screening test or further assessment proves negative.

Strengths and limitations of the study

To our knowledge, this is the first population-based study to analyse the association between symptoms and interval and fatal breast cancers. Our study has several strengths. First, we included all screening visits with symptoms from the start of the mammography-screening programme (over 21 years) and compared them with visits without symptoms. Since this was a population-based service-screening programme, the selection bias was minimal. Second, we could use validated prospectively collected register data with no recall bias.[20] Validation of the cancer diagnosis (screen-detected and ICs) and death from cancer using national and covering data sources (Finnish Cancer Registry and Statistics Finland) and the use of unique personal identifier for individual-level linkage eliminates the possibility of selective misclassification. Third, we matched the symptomatic visits to asymptomatic visits by possible confounding baseline characteristics to minimise bias in the risk estimates. As a result, we found no significant difference in the background variables between visits with and without symptoms. Finally, the programme process and outcome indicator definitions used in this study adhere to those defined by WHO-IARC and mentioned in EU guidelines on breast cancer screening and diagnosis,[7,21] including the provision of relevant information describing the performance and also failures during the various steps of the screening process.

Our study also has potential limitations. The symptom information was based on the women's reporting in the past 2 or 6 months and a check by a radiographer at the visit. The radiographer's inspection is likely to be less comprehensive than a full clinical examination. The collection of symptom information is mainly done in order to support the interpretation of the mammograms. However, in almost every case—if not all—the radiographer or nurse examines the breast to confirm the presence of symptoms (mainly a lump and retraction) before the mammography is performed. Thus, there are reasons to consider the collected symptom information to be valid, albeit not perfect. A second potential limitation is that our estimates could have been confounded, because symptomatic women are more likely to attend than asymptomatic women. However, because of the high attendance rate in the Finnish mammography-screening programme (84% among the invited, the highest among any

existing mammography-screening programme), and as 97% of attendees are asymptomatic, the attendance bias caused by symptoms is likely to be small. Third, we did not use the background incidence of breast cancer to estimate the interval cancer rate in the absence of screening, but instead used a detection method[22] that takes into account cancers from the screening programme; thus, estimates are sensitive to over-diagnosis and lead time bias. Given the high coverage and attendance rate in the Finnish screening programme, it was not possible to find a comparable non-screened group to estimate the background incidence of breast cancer. In addition, there was no possibility to estimate background incidence of breast cancer in women with symptoms. We did not find any difference in the PPV of mammography, with and without symptoms, between the first and subsequent screening rounds (not shown in results); thus, the lead time bias because of prevalent screens is negligible. Furthermore, we used invasive breast cancers—and also advanced and fatal breast cancers, which would be less affected by overdiagnosis—to estimate the incidence rates and hazard ratios. Only 5% of all carcinomas in those with symptoms were in situ carcinomas.

Previous studies on the association between symptoms and the risk of breast cancer are limited because of factors such as study design, size, follow-up time and assessment of the possible outcome measures of breast cancer. Nonetheless, a few studies have found that the presence of a palpable lump is associated with a higher risk of SDCs.[18,19] We are not aware of any studies that have assessed the relationship of symptoms with other outcome measures.

Clinical and public health implications

Both the screening test and episode sensitivities tended to be higher in symptomatic women compared with the asymptomatic. Correspondingly, the screening specificity was lower in the symptomatic women. Of note, particularly in women with a lump, were the (episode) sensitivity losses in the further assessment during the index visits. Moreover, higher interval cancer incidence within six months after a negative mammography with a lump is a clear concern for the programme. This indicates that further assessment is needed more frequently (albeit with a potential loss of specificity), and there needs to be highly stringent diagnostic evidence for a decision not to carry out a full further assessment including a core biopsy in these cases. As most visits (~ 97%) were asymptomatic, the number of additional services would be rather small as well as the improvement of programme's overall performance and outcome. But still, high-quality and clinically appropriate services are important for women having symptoms at a screening visit. One option is to recall all women with symptoms even if the mammography result is negative, as practised in Norway (< 0.3% of all those screened were recalled with symptoms).[23] Doing this in Finnish programme would significantly lower the PPV of recall, as 2.5% of all screens had symptoms and only ~ 1 out of 10 symptomatic visits have been recalled.[17]

Taking into account the high incidence of advanced and fatal interval breast cancers in symptomatic women, it is likely that protection by biennial screening visit would clearly not be sufficient even after potential improvements in further assessments. We are not aware of recommendations for surveillance or follow-up of symptomatic women in the programme. Hence, we recommend a shorter screening interval for the symptomatic group so that the cumulative incidences of interval and fatal interval cancer would possibly become more equitable. For two of the studied symptoms (lump and nipple discharge), the interval cancer incidence increased so rapidly that the first follow-up visit could take place very shortly after the index visit. Finally, taking into account the probability of fatal screen-detected breast cancer is higher in women with a symptom already at the index visit, women need to be better informed about symptoms and made aware that if a symptom occurs, it is not a good idea to wait until the next invitation to the programme. Guidelines for further assessment in patients presenting with symptoms before having a scheduled invitation according to the programme—as developed by National Health Service in the UK[24]—could be useful also in other settings.

CONCLUDING REMARKS

Women with breast symptoms at visits within the population-based breast cancer screening programme have a clearly increased risk of breast cancer. The cumulative incidence of invasive, advanced and fatal breast cancers, as well as the detection of them at the next screen, provide direct evidence for the need for risk-adjusted screening in symptomatic women—e.g., tailoring the management procedures at index visits and shortening the screening intervals for these women. This study provides clear evidence to update and support the EU guidelines[7] recommendation that ensures sufficient attention being paid to symptomatic details provided by women. Our findings therefore have important implications for screening-aged women, radiologists, nurses and mammography-screening programmes overall.

ACKNOWLEDGEMENTS

This work was supported by the Cancer Foundation Grant of Finland. The funder had no role in the design, the conduct of the study and finalisation of the manuscript.

AUTHOR CONTRIBUTIONS

DS and AA designed the study. JM and DS did the statistical analysis and take responsibility of the accuracy of the data analysis. DS, AA, JM, SD, NM and JP interpreted the data and edited the manuscript.

ADDITIONAL INFORMATION

Competing interests: The authors have no conflict of interest or disclosures. All authors declare no financial or other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: This study comply with ethical standards.

REFERENCES

1. Taylor, R., Supramaniam, R., Rickard, M., Eostoesta, J. & Moreira, C. Interval breast cancers in New South Wales, Australia, and comparisons with trials and other mammographic screening programmes. *J. Med. Screen.* **9**, 20–25 (2002).
2. Houssami, N., Irwig, L. & Ciatto, S. Radiological surveillance of interval breast cancers in screening programmes. *Lancet Oncol.* **7**, 259–265 (2006).
3. Day, N. et al. Monitoring interval cancers in breast screening programmes: the East Anglian experience. *J. Med. Screen.* **2**, 180–185 (1995).
4. Day, N. E. The NHS breast cancer screening programme. *Br. J. Cancer* **89**, 1 (2003).
5. Warren, R. & Duffy, S. Interval cancers as an indicator of performance in breast screening. *Breast Cancer* **7**, 9–18 (2000).
6. Nystrom, L. et al. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* **359**, 909–919 (2002).
7. European Commission. European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis. 4th edn. In Perry N. et al. (eds). Office for Official Publications of the European Communities, Luxembourg, 2006.
8. Lauby-Secretan, B. et al. Breast-cancer screening - viewpoint of the IARC working group. *N. Engl. J. Med.* **372**, 2353–2358 (2015).
9. Sarkeala, T. et al. Episode sensitivity in association with process indicators in the Finnish breast cancer screening program. *Int. J. Cancer* **118**, 174–179 (2006).
10. Bennett, R. L., Sellars, S. J. & Moss, S. M. Interval cancers in the NHS breast cancer screening programme in England, Wales and Northern Ireland. *Br. J. Cancer* **104**, 571–577 (2011).
11. Giordano, L. et al. Mammographic screening programmes in Europe: organization, coverage and participation. *J. Med. Screen.* **19**, 72–82 (2012).
12. Finnish Cancer Registry. Breast cancer screening; Mass screening programme, 2018. Available at http://www.cancer.fi/syoparekisteri/en/mass-screening-registry/breast_cancer_screening/screening_programme/.

13. Waller, M., Moss, S., Watson, J. & Møller, H. The effect of mammographic screening and hormone replacement therapy use on breast cancer incidence in England and Wales. *Cancer Epidemiol. Biomark. Prev.* **16**, 2257–2261 (2007).
14. Ciatto, S., Visioli, C., Paci, E. & Zappa, M. Breast density as a determinant of interval cancer at mammographic screening. *Br. J. Cancer* **90**, 393–396 (2004).
15. Wang, H. et al. Interval cancers in the Norwegian breast cancer screening program: frequency, characteristics and use of HRT. *Int. J. Cancer* **94**, 594–598 (2001).
16. Singh, D., Malila, N., Pokhrel, A. & Anttila, A. Association of symptoms and breast cancer in population-based mammography screening in finland. *Int. J. Cancer* **136**, E630–E637 (2015).
17. Singh, D., Pitkaniemi, J., Malila, N. & Anttila, A. Cumulative risk of false positive test in relation to breast symptoms in mammography screening: a historical prospective cohort study. *Breast Cancer Res. Treat.* **159**, 305–313 (2016).
18. Brouckaert, O. et al. Breast cancer phenotype, nodal status and palpability may be useful in the detection of overdiagnosed screening-detected breast cancers. *Ann. Oncol.* **24**, 1847–1852 (2013).
19. Gill, G., Luke, C. & Roder, D. Prognostic importance of palpability as a feature of screen-detected breast cancers. *J. Med. Screen.* **13**, 98–101 (2006).
20. Leinonen, M. K., Miettinen, J., Heikkinen, S., Pitkaniemi, J. & Malila, N. Quality measures of the population-based Finnish cancer registry indicate sound data quality for solid malignant tumours. *Eur. J. Cancer* **77**, 31–39 (2017).
21. IARC Handbook of Cancer Prevention. Breast Cancer Screening. IARC, France, 2016. Available at http://publications.iarc.fr/Book-And-Report-Series/Iarc-Handbooks-Of-Cancer-Prevention/Breast-Cancer-Screening-2016.
22. Hakama, M., Pokhrel, A., Malila, N. & Hakulinen, T. Sensitivity, effect and overdiagnosis in screening for cancers with detectable pre-invasive phase. *Int. J. Cancer* **136**, 928–935 (2015).
23. Hofvind S. et al. The Norwegian Breast Cancer Screening Program, 1996–2016: celebrating 20 years of organised mammographic screening. In: Cancer in Norway 2016–Cancer incidence, mortality, survival and prevalence in Norway. Oslo, 2017.
24. Willett A., Michell M., Lee M. Best Practice Diagnostic Guidelines for patients presenting with Breast Symptoms. November 2010. Available at https://www.scribd.com/document/140938243/Best-Practice-Diagnostic-Guidelines-for-Patients-Presenting-With-Breast-Symptoms.

BJC British Journal of Cancer

www.nature.com/bjc

CORRECTION

Correction: Association of symptoms and interval breast cancers in the mammography-screening programme: population-based matched cohort study

Deependra Singh[1,2], Joonas Miettinen[1], Stephen Duffy[3], Nea Malila[1,2], Janne Pitkäniemi[1] and Ahti Anttila[1]

British Journal of Cancer (2019) 120:773–774; https://doi.org/10.1038/s41416-019-0417-6

Correction to: *British Journal of Cancer* (2018) **119**, 1428–1435; https://doi.org/10.1038/s41416-018-0308-2; www.bjcancer.com; published online 7 November 2018.

The authors report that the labels indicating the symptom types and no symptom lines in the original version of Fig. 2 were missing. The correct version of Fig. 2 with the labels included is provided below.

[1]Mass Screening Registry, Finnish Cancer Registry, FI-00130 Helsinki, Finland; [2]Epidemiology group, Department of Health Sciences, University of Tampere, FI-33520 Tampere, Finland and [3]Wolfson Institute of Preventive Medicine, Queen Mary University of London, Charterhouse Square, London, England
Correspondence: Deependra Singh (deependra.singh@cancer.fi)

Published online: 6 March 2019

Published by Springer Nature on behalf of Cancer Research UK

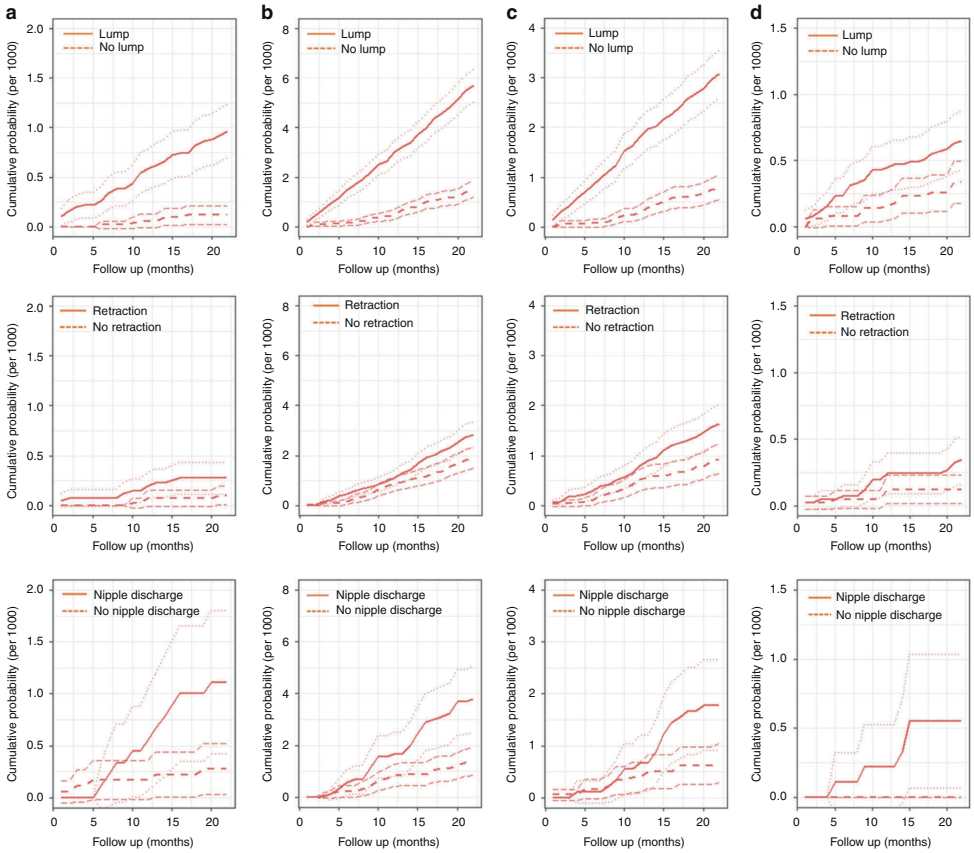

Fig. 2 **a**–**d** Cumulative incidence of invasive (per 1000): **a** recalled ICs; **b** not recalled ICs; **c** non-localised ICs; **d** fatal interval cancers. Note: the confidence intervals lines for cumulative incidence are indicated by light dotted lines in symptomatic and asymptomatic groups

PUBLICATION IV

Cancer incidence and mortality patterns in women with breast symptoms in the mammography screening programme: A matched cohort analysis

Singh, D., Malila, N., Pitkäniemi, J., Anttila, A.

Int J Cancer, 144(12), 2928-2935
Doi: 10.1002/ijc.32035

IJC
International Journal of Cancer

Cancer incidence and mortality patterns in women with breast symptoms in the mammography screening programme: A matched cohort analysis

Deependra Singh [1,2], Nea Malila[1,2], Janne Pitkäniemi[1] and Ahti Anttila[1]

[1]Mass Screening Registry, Finnish Cancer Registry, Helsinki, Finland
[2]Epidemiology Group, Department of Health Sciences, University of Tampere, Tampere, Finland

Cancer Epidemiology

Efforts to reduce mortality through early detection and diagnosis has intensified in the recent decade. An important risk factor, 'breast symptoms' reported by women during screening visit, remains overlooked. In this population based matched cohort study using Finnish National Breast Cancer Screening Program (FNBCSP), we assessed the association between breast symptoms reported at screening visit and the risk of cancer incidence and breast cancer mortality and all-cause mortality followed-up over a period of 24 years. For each visit with symptoms, non-symptomatic controls were matched (1:1 for lump and retraction; 1:2 for nipple discharge) based on age at screening visit, year of invitation, number of invited visits, and municipality of invitation. Women who reported lump or retraction had about two-fold risk of breast cancer incidence, three-fold risk of breast cancer mortality and all-cause mortality respectively as compared to women without respective symptoms (p-value<0.05). We found a substantial difference (p-value<0.05) in mortality rates throughout the follow-up period between symptomatic and asymptomatic group. In absolute terms, after the follow-up period for women who reported lump, 180 died from breast cancer as compared to 70 deaths in those without lump, per 10,000 person-years of follow-up, and 315 *versus* 160 all-cause deaths per 10,000 person-years in women with and without lump respectively. our study provides comprehensive evidence that women with breast symptoms remain in a higher risk of dying over a very long period. The findings indicate needs to develop improvements in the guidelines for screening and clinical services for women presenting with symptoms.

Introduction

Based on the evidence supported by randomized controlled trails and observational studies, World Health Organization, in 2016 concluded that women attending mammography screening have about 40% reduced risk of breast cancer mortality.[1] However, use of mammography screening remains debated.[2,3] Many studies have proposed personalized screening modalities based on important risk factors, such as family history of breast cancer,[4–6] breast density,[7–9] and gene mutation.[10] An important risk factor, 'breast symptoms' reported by clinical examination or by women during screening visit, remains overlooked. The single reason is that many population-based mammography-screening programmes have not collected or analyzed information on breast symptoms. In many programmes, on the other hand, most of the women presenting with symptoms such as a lump, retraction, or nipple discharge have further assessment, but are not systematically followed-up in the programme up to cancer diagnosis or death.[11,12] Also, the EU guidelines lack evidence on the risk of breast cancer or mortality in those women.[13]

Recent studies from Finland showed sufficient evidence that information on symptoms collected at women's mammography visit, increases the risk of advanced stage breast cancers.[14,15] Higher advanced cancer risk might be correlated with the higher risk of death from breast cancer. The relation of symptoms, as adjunct to screening mammography, with mortality have never been studied. Therefore, it is important to quantify *(or measure)* the discrepancy, if any exists, in mortality between women with and without symptoms.

Key words: breast cancer symptoms, cancer incidence, incidence-based mortality, lump, screening strategy

Abbreviations: EU: European Union; FCR: Finnish Cancer Registry; FNBCSP: Finnish National Breast Cancer Screening Program; HR: Hazard Ratio; IARC: International Agency for Research on Cancer; ICD: International Classification of Disease

The funder had no role in the design and the conduct of the study, and finalization of the study.

Conflict of interests: The authors have no conflict of interest or disclosures. All authors declare no financial or other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Our study comply with ethical standards.

Data sharing: No additional data available.

Grant sponsor: Cancer Foundation Grant of Finland

DOI: 10.1002/ijc.32035

History: Received 10 Sep 2018; Accepted 23 Nov 2018; Online 3 Dec 2018

Correspondence to: Deependra Singh, Mass Screening Registry, Finnish Cancer Registry, Unioninkatu 22, FI-00130, Helsinki, Finland, E-mail: deependra.singh@cancer.fi

What's new?

While breast cancer screening based on factors such as breast density and family history can significantly reduce breast cancer mortality, little is known about the importance of breast symptoms in screening programs. In this analysis of mammography screening data collected from 1992-2012 in Finland, women who reported breast symptoms at screening had significantly increased risks of breast cancer incidence and mortality. For women who reported a lump or retraction, breast cancer incidence was increased nearly two-fold and mortality three-fold. The findings suggest that improvements in screening and clinical services are needed in Finland to better serve women with breast symptoms.

To address these uncertainties, we conducted a prospective analysis of the association between breast symptoms reported at screening visit and the risk of cancer incidence and mortality followed-up over a period of 24 years.

Materials and Methods

Study population

We conducted a matched cohort study using prospective data of ongoing Finnish National Breast Cancer Screening Programme (FNBCSP). Since the beginning of the programme in 1987 every women in the target population aged 50–69 years are invited biennially for mammography screening.[16] Individual level information of the study participants is extracted using three registries, the Finnish Cancer Registry (1953–2015), the Mass Screening Registry (1992–2012), and the Central Population Registry (1992–2014). The study cohort includes 3,735,464 screening visits made by women during 1992–2012. Figure 1 shows the flow diagram of study settings and outcomes.

The exposure group (screening visits with symptoms) includes women with reported breast symptoms (lump, retraction, and nipple discharge) at a given screening visit. This group contained first symptomatic visit with at least one symptoms. Symptoms are analyzed separately meaning that first occurrence of any given symptom was independent of other symptom. Here, index visit means screening visit with any given symptom reported for the first time.

The reference group (screening visits without symptoms) means visits with no reported breast symptoms in the screening history before the index visit for any given symptom. The reference visit can later be the part of the exposure group if symptoms occurred at future screening visits. Thus, symptoms can be assumed of as a time-dependent covariate. Altogether three sets of reference visits for each three symptoms were formed by matching.

Matching

The frequency matching of the exposure groups to the reference groups were done by age at screening visit, year of invitation, number of previous screening visits, and municipality of invitation. Each visit with symptoms was matched to the viable controls by random sampling. Assuming the effect size and the power estimate of 0.80, the exposed-to-reference-visits ratio was 1:1 for lump and retraction, and 1:2 for nipple discharge.

Cause of death

Information on cause of death was derived from the Statistics Finland database that categorizes disease according to International Classification of Disease (ICD-7 until 1995 and ICD-10 from 1996) classification. The underline cause of death was categorized using topography codes in the analysis: death resulting from breast cancer (C50), and death resulting from other causes (all ICD codes except C50). Information on socio-economic status was only available for breast cancer cases and deaths.

Outcome assessment and follow-up

Incident breast cancer cases were defined as new breast cancers diagnosed in screened women after the first invitation by the programme. Breast cancer mortality means deaths due to breast cancer diagnosed during the screening age-period after the first invitation by the programme. Women having cancer diagnosed before the first invitation by the screening programme were excluded in the calculation of incidence (so-called incident-cancer) and mortality (so-called incidence-based mortality).[17,18] All-cause mortality means deaths due to any cause, including the breast cancer cases, whereas, cause-specific mortality (except breast cancer) refers to cases such as 'mortality from other cancer', and 'mortality from other cause'. Follow-up for incidence cases started from the index visit in 1 January 1992–31 December 2012 and ended at the date of cancer diagnosis or emigration, or at the end of the follow-up – i.e. 31 December 2012 – whichever occurred first. For interval cancer cases, follow-up time was extended up to 31 December 2014. Similarly, mortality follow-up started after the index visit in 1 January 1992–31 December 2012 and ended at the date of death (from breast cancer or death from other cause) or emigration, or at the end of follow-up – i.e. 31 December 2015. Deaths resulting from breast cancer diagnosed during 2013–2015 were excluded.

Statistical analysis

Age-adjusted Cox proportional hazard regression models were fitted to compare the risk of breast cancer incidence and mortality, separately for all-cause and cause-specific mortality, among women with and without reported symptoms at the index screen. The risks were estimated using hazard ratios (HRs) and 95% confidence intervals (CIs).

We also estimated the rates of breast cancer incidence, and all-cause and cause-specific mortality, and compared between the visits with and without symptoms. Incidence rates were calculated as number of breast cancer diagnosed divided by

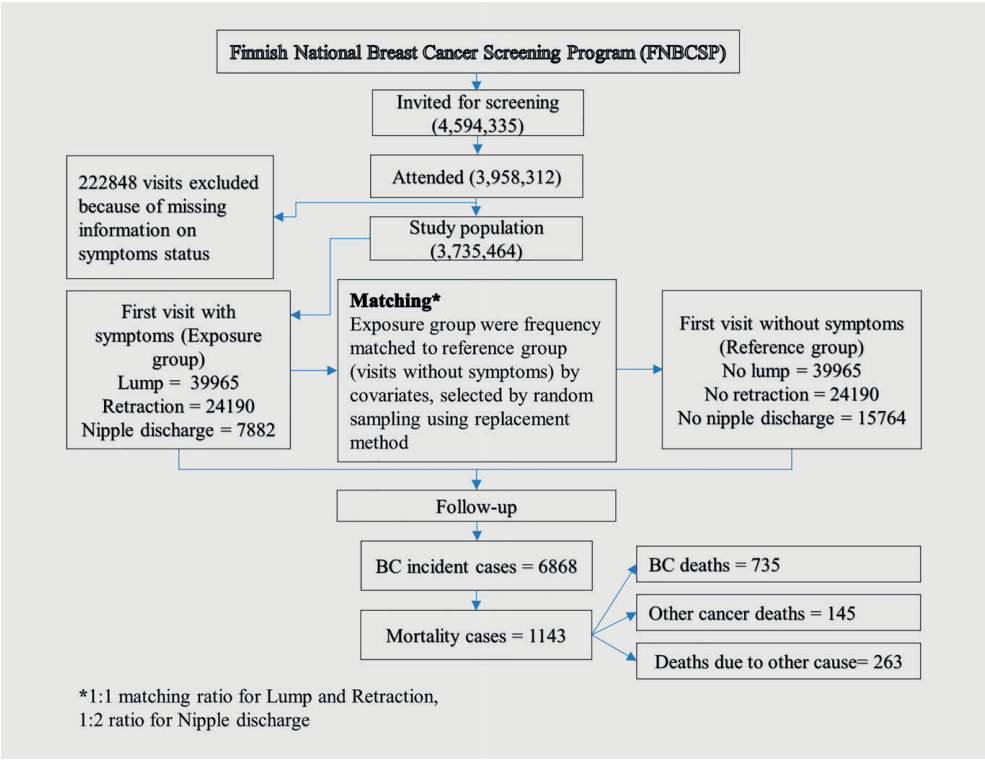

Figure 1. Flow diagram of study settings and outcomes. [Color figure can be viewed at wileyonlinelibrary.com]

the risk time, and compared between visits with and without symptoms. Likewise, mortality rates were calculated as number of deaths divided by the risk time, and compared between visits with and without symptoms. The rates and absolute difference in breast cancer incidence and mortality were reported per 10,000 person-years of follow-up at 95% confidence intervals. For this, we categorized the follow-up time into 5-years bands each. Instead, the cumulative hazards of breast cancer incidence and mortality were reported using years as underlying time units for risk time. The underlying time scale in all the analysis was time since the visits with or without breast symptom. The statistical difference in incidence and mortality rates were reported using two-sided *p*-value. All statistical analyses were conducted using Stata version 14.0 (STATA statistical software, release 14; Stata Corporation, TX).

Results

A total of 6868 (4.5%) visits, out of 151,956 visits, were diagnosed as incident breast cancers during 1992–2012. Of these, 735 women (10.7%) died (during 1992–2015) from breast cancer. The total follow-up time for incidence breast cancers was 1,303,484.2 person-years, with a median of 3.1 years and maximum of 20.7 years of follow-up, whereas, the breast cancer mortality follow-up time was 1,797,032.1 person-years, with a median of 8.8 years and maximum of 22.3 years of follow-up (Table 1). Table 1 shows the characteristics of the final study cohort.

The mean age at a screening visit, a cancer diagnosis, or a death did not differ between visits with and without symptoms. The follow-up time of breast cancer diagnosis or deaths resulting from breast cancer had wider variation between visits with and without symptoms (median follow-up time of 1.9 *versus* 6.4 years for diagnosis and 8.1 *versus* 11.4 years for deaths, for visit with and without lump). In women with a history of first cancer before the first invitation, there was no significant (p-value >0.05) difference between visits with and without symptoms. However, these women were later excluded to compute incidence or mortality rates and rate ratios.

Table 1. Cohort characteristics and outcomes

Characteristics			Lump		Retraction		Nipple discharge	
			Yes (%)	No (%)	Yes (%)	No (%)	Yes (%)	No (%)
i)	Attended previous round	Yes	20,735 (49.1)	21,536 (50.9)	17,983 (49.2)	18,571 (50.8)	4686 (32.3)	9798 (67.7)
		No	19,230 (51.1)	18,429 (48.9)	6207 (52.5)	5619 (47.5)	3196 (34.9)	5966 (65.1)
ii)	Age at visit	(mean, SD)	54.7 (4.3)	54.7 (4.3)	56.3 (5.2)	56.3 (5.2)	54.4 (4.5)	54.4 (4.5)
iii)	Time difference between index visit and breast cancer diagnosis, in years	Median	1.92	6.42	1.50	3.58	4.17	4.34
		Range	0.04 to 20.7	0.08 to 20.3	0.04 to 11.2	0.08 to 11.3	0.08 to 17.7	0.08 to 19.0
iv)	Time difference between index visit and breast cancer mortality, in years	Median	8.17	11.4	5.96	8.2	9.57	9.96
		Range	0.09 to 22.3	2.01 to 22.1	0.22 to 14.5	0.96 to 12.8	2.01 to 18.7	1.29 to 18.3
v)	Time difference between index visit and all-cause mortality, in years	Median	10.50	13.1	6.11	6.91	9.89	10.0
		Range	.09 to 23.7	1.93 to 23.7	0.22 to 14.5	0.74 to 12.8	2.01 to 22.1	1.29 to 22.6
vi)	History of breast cancer or other cancers before first invitation	Yes	18 (69.2)	8 (30.8)	5 (55.6)	4 (44.4)	5 (45.4)	6 (54.6)
		No	3185 (69.8)	1376 (30.2)	924 (63.6)	528 (36.4)	393 (45.9)	462 (54.1)
vii)	Stage at diagnosis	Localized	701 (82.6)	148 (17.4)	213 (71.9)	83 (28.1)	61 (55.9)	48 (44.1)
		Non-localized	723 (90.5)	76 (9.5)	234 (82.7)	49 (17.3)	39 (52.0)	36 (48.0)

SD, standard deviation.
Values are numbers (percentage) unless stated otherwise.

Breast cancer incidence and mortality rates in the longitudinal follow-up

The rates of incident breast cancer cases were significantly higher (p-value = 0.000) within 5-years after the screening visit (rate of 780 cases per 10,000 person-years *versus* 203 cases per 10,000 person-years, for visits with and without lump) (Table 2). Also, significant difference (p-value = 0.000) in overall rates (43.9 per 10,000 person-years, 95% CI 40.8 to 47.0; for visits with or without lump) were found between visits with and without symptoms during the follow-up time.

Table 2. BC incidence in women who reported symptoms compared to asymptomatic women at screen

Outcomes	Follow-up time, years	BC incidence						Incidence rate difference, per 10,000 pyrs. (95% CI)	p-value
		Yes			No				
		Pyrs.	n	rate[1]	Pyrs.	n	rate[1]		
Lump	0–5	26,811.3	2092	780.3	27,330.9	556	203.4	43.9 (40.8–47.0)	0.000
	More than 5–10	61,338.9	497	81.0	62,936.8	360	57.2		
	More than 10–15	69,490.4	379	54.5	72,622.2	291	40.1		
	More than 15–20	199,153.6	210	10.5	215,123.2	168	7.8		
	More than 20–25	64,665.7	7	1.1	56,388.9	1	0.2		
	Total	421,459.9	3185	75.6	434,402.0	1376	31.7		
Retraction[2]	0–5	28,635.9	745	260.2	28,876.0	354	122.6	31.0 (25.3–36.7)	0.000
	More than 5–10	64,727.6	165	25.5	66,430.5	157	23.6		
	More than 10–15	37,238.1	14	3.8	37,568.4	17	4.5		
	Total	130,601.6	924	70.7	132,874.9	528	39.7		
Nipple discharge	0–5	8048.6	222	275.8	16,123.6	251	155.7	27.1 (19.9–34.3)	0.000
	More than 5–10	17,745.2	103	58.0	35,476.7	132	37.2		
	More than 10–15	17,456.6	50	28.6	35,040.2	64	18.3		
	More than 15–20	15,732.9	18	11.4	33,915.2	15	4.4		
	More than 20–25	2080.0	0	0.0	3434.7	0	0.0		
	Total	61,063.3	393	64.4	123,990.4	462	37.3		

[1]Rate per 10,000 person years of follow-up.
[2]Information on retraction was started to collect during late 1990s.
Abbreviation: CI, confidence interval.

Similar differences (p-value = 0.000) in the incidence rates of cancer diagnosis or deaths were observed for retraction (31 per 10,000 person-years, 95% CI 25.3 to 36.7) and nipple discharge (27.1 per 10,000 person-years, 95% CI 19.9 to 34.3) as compared to respective asymptomatic visits.

The breast cancer mortality rates greatly differ (p-value = 0.000) within 5-years between visits with and without symptoms (77.7 deaths per 10,000 person-years *versus* 8.4 deaths per 10,000 person-years, for visits with and without lump) (Table 3). Likewise, the overall differences in mortality rates were significant (p-value = 0.000) for lump (5.3 per 10,000 person-years, 95% CI 4.4 to 6.1) and retraction (2.9 per 10,000 person-years, 95% CI 1.9 to 3.8) as compared to respective visits without symptom. But, no significant difference (p-value = 0.226) in mortality rate (0.75 per 10,000 person-years, 95% CI minus 0.51 to 2.0) was found for nipple discharge as compared to those without symptom. In addition, the all-cause mortality rate differ significantly (p-value<0.05) in all symptom types as compared to respective asymptomatic visits. Interestingly, we found more cases of other cause death than breast cancer, long (after 10 years of follow-up) after the visits with or without lump, which was quite opposite when the cause of death was a breast cancer.

Risk of breast cancer incidence and mortality

The risk of dying from breast cancer was significantly higher in those who reported lump (age-adjusted hazard ratio 3.1, 95% CI 2.5 to 3.7) and retraction (age-adjusted hazard ratio 3.8, 95% CI 2.4 to 6.2) compared to respective asymptomatic group (Table 4). For nipple discharge, the risk of breast cancer was not significant (age-adjusted hazard ratio 1.4, 95% CI 0.82 to 2.4), but the risk of all-cause mortality was significantly higher (age-adjusted hazard ratio 1.5, 95% CI 1.04 to 2.2) than those visits without nipple discharge.

Cumulative incidence and mortality rates per 10,000 person-years of follow-up

The incidence of breast cancer increased rather rapidly after the visit with a lump, whereas, the breast cancer mortality or all-cause mortality risks increased within 5-years after a visit with lump (Figs. 2*a*–2*c*). Similar patterns of increased (incidence and mortality) risks were found in visits with a retraction or nipple discharge (incidence only), as compared to visits without respective symptom. Considering the cumulative risk of breast cancer mortality in respective visits without lump and retraction as a reference, the same level of risk was reached within a half of the follow-up time in visits with lump and retraction, respectively. However, no such difference in mortality risk was evident for nipple discharge.

Discussion

Our study demonstrated that women who reported symptoms at screening visit had higher risk of breast cancer incidence and mortality. Women who reported lump or retraction had about two-fold risk of breast cancer incidence, three-fold risk of breast cancer mortality and all-cause mortality respectively as compared to women without respective symptoms. The cumulative mortality patterns, breast cancer mortality and all-cause mortality, remained higher in symptomatic women in the long follow-up time available in our study. These findings indicate an urgent need to develop improvements in the guidelines on screening and clinical services for women presenting with symptoms. To our knowledge, our study is the first to assess the association between symptoms and mortality risk.

Strength and limitations

Our study findings are based on valid collection of information with reliable linked data at national level. We used a prospective study design and included all-visits with symptoms and similar number of asymptomatic visits from the whole women population of the targeted age who attended for screening mammography. We followed women for 24 years and collected information on breast symptoms biennially, thus it was possible to accurately assess association of symptoms with the subsequent risk of breast cancer or death. Individual level data from several sources: screening reports, patient or hospital records, death reports, autopsy reports, etc. were linked using unique personal identifier. The exact date of reporting of symptoms, date of diagnosis or death or emigration were known for every subjects, thus, the calculated follow-up time intervals were precise. Matching symptomatic visits by background variables to asymptomatic visits have made the two groups comparable at any given period (maximum 10 visits) of screening visits. Because of the large cohort, the study power was sufficient to compute the difference in risk and rates for every individual symptoms with respective asymptomatic visits.

Our study has potential limitations. our study is based on the database research and information about important risk factors such as family history, breast density and socioeconomic status was not collected, and thus it was not possible to finely adjust for potential confounders. We do not know about delay in presentation of symptoms on whether women waited for her first (i.e. at age 50 years) or subsequent screening invitation. Our analysis of symptoms was based on women's reporting of symptom during the past two to six months and the examination by the radiographer or nurse at the screening visit. Thus, the collected symptom information is valid with negligible recall bias. The attendance bias because of symptomatic women were more likely to attend for screening than asymptomatic women cannot be ruled out. However, the high screening attendance proportion (84% among those invited) in the Finnish mammography screening programme and 99% of the attendees are asymptomatic, the attendance bias is likely to be negligible. In addition, we did not find any difference in the attendance proportion and risk of breast

Table 3. BC mortality and all-cause mortality rates in women who reported symptoms compared to asymptomatic women at screen

Outcomes	Follow-up time, years	Pyrs. of follow-up with symptom	Pyrs. of follow-up without symptom	BC mortality						All cause mortality					
				Yes		No		Mortality rate difference, per 10,000 pyrs. (95% CI)	p-value	Yes		No		Mortality rate difference, per 10,000 pyrs. (95% CI)	p-Value
				n	rate[1]	n	rate[1]			n	rate[1]	n	rate[1]		
Lump	0 to 5	17,382.4	16,702.1	135	77.7	14	8.4	5.31 (4.47–6.15)	0.000	162	93.2	24	14.4	7.27 (6.24–8.31)	0.000
	More than 5 to 10	78,878.3	76,721.6	116	14.7	46	6.0			153	19.4	58	7.6		
	More than 10 to 15	86,727.9	88,505.9	82	9.5	32	3.6			128	14.8	53	6.0		
	More than 15 to 20	94,410.8	96,695.7	82	8.7	34	3.5			145	15.4	66	6.8		
	More than 20 to 25	283,227.3	283,903.2	22	0.8	14	0.5			56	2.0	36	1.3		
	Total	560,626.7	562,528.5	437	7.8	140	2.5			644	11.5	237	4.2		
Retraction[2]	0 to 5	17,357.4	17,182.4	33	19.0	7	4.1	2.91 (1.95–3.87)	0.000	46	26.5	14	8.1	3.93 (2.76–5.10)	0.000
	More than 5 to 10	94,969.1	88,345.9	31	3.3	9	1.0			45	4.7	13	1.5		
	More than 10 to 15	94,182.2	101,711.1	17	1.8	5	0.5			26	2.8	9	0.9		
	Total	206,508.7	207,239.4	81	3.9	21	1.0			117	5.7	36	1.7		
Nipple discharge	0 to 5	3538.6	6767.7	4	11.3	5	7.4	0.75 (–0.51–2.01)	0.226	10	28.3	11	16.3	1.85 (–0.06–3.63)	0.034
	More than 5 to 10	25,919.9	48,724.5	9	3.5	13	2.7			14	5.4	20	4.1		
	More than 10 to 15	20,868.4	45,174.7	7	3.4	11	2.4			10	4.8	22	4.9		
	More than 15 to 20	22,970.6	45,524.2	3	1.3	4	0.9			12	5.2	8	1.8		
	More than 20 to 25	13,394.4	27,246.1	0	0.0	0	0.0			1	0.0	1	0.0		
	Total	86,691.9	173,437.2	23	2.7	33	1.9			47	5.4	62	3.6		

[1]Rate per 10,000 person-years of follow-up.
[2]Information on retraction was started to collect during late 1990s.
Abbreviations: Pyrs., person-years; BC, breast cancer; CI, confidence interval.

Table 4. Association between symptoms and mortality as compared to those without symptoms

	BC incidence	BC mortality	All-cause mortality
Symptom	Age-adjusted hazard ratio (95% CI)	Age-adjusted hazard ratio (95% CI)	Age-adjusted hazard ratio (95% CI)
Lump	2.37 (2.23–2.53)	3.14 (2.59–3.79)	2.72 (2.35–3.17)
Retraction	1.77 (1.59–1.97)	3.88 (2.40–6.27)	3.27 (2.25–4.75)
Nipple discharge	1.73 (1.51–1.97)	1.40 (0.82–2.39)	1.52 (1.04–2.22)

Abbreviations: BC, breast cancer; CI, confidence interval.

cancer between first and subsequent visit in women with or without symptoms.

The all-cause mortality rates were relatively higher than breast cancer mortality, and the substantial difference in rate was found between symptomatic and asymptomatic group throughout the follow-up period. In absolute terms, for lump, in every 10,000 person-years of follow up, 180 women died from breast cancer as compared to 70 women without lump, and 315 *versus* 160 all-cause deaths in women with and without lump respectively after 24 years of follow-up. We also found difference in the number of deaths in women who reported retraction or nipple discharge. The difference in all-cause mortality was not related clearly to other specific causes of death than breast cancer, but differences in deaths from other causes than breast cancer were related to several causes. It is likely that having a lower socio-economic status can partially affect the findings on symptoms, affecting therefore also to mortalities from breast cancer as well as of several other causes. We observed extra (all-cause) death cases in women who reported a lump in lower socio-economic class than higher class, and surplus number of deaths as compared to those without lump. In our study, the socio-economic status of the whole cohort is unknown, thus, the proportion of breast cancer deaths or all-cause deaths in those with or without symptoms might differ from our findings.

Clinical and public health implications

All three symptoms analyzed in our study provides new evidence on the risk of breast cancer incidence and mortality. Information on breast symptoms reported by the women should

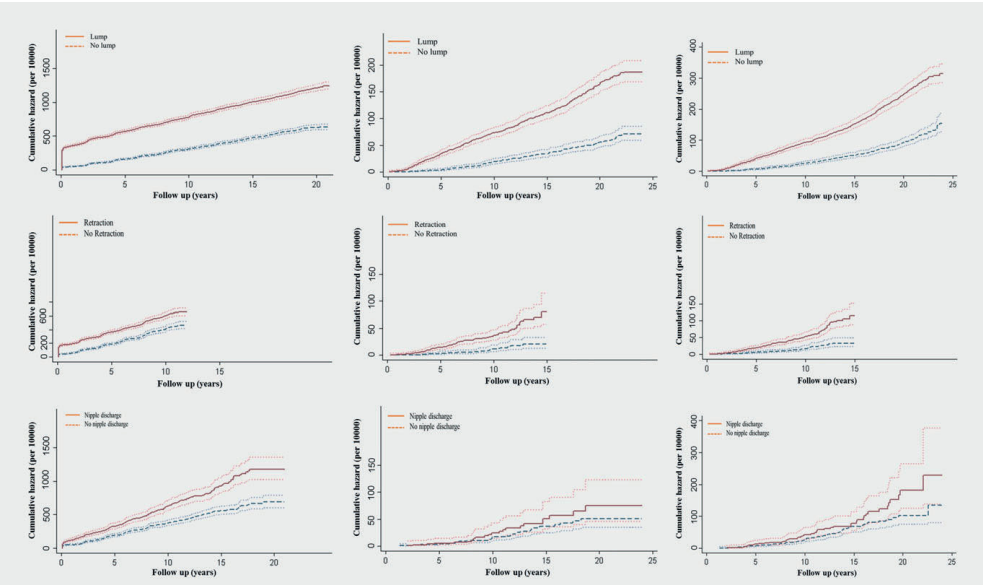

Figure 2. (*a-c*): Cumulative incidence of (*a*) breast cancer, (*b*) breast cancer mortality and (*c*) all-cause mortality per 10,000 person-years of follow up. Note: The confidence intervals lines for cumulative incidence are indicated by light dotted lines in symptomatic and asymptomatic groups. [Color figure can be viewed at wileyonlinelibrary.com]

not be ruled out or ignored. Neither the evidence should only be used to support the interpretation of the mammography findings as done until now. The substantial high mortality rates within 5-years after the visit with symptoms demands proper further assessment of screening visits with symptoms irrespective of the mammography findings. The radiographers or nurses should properly examine the women who present with breast symptoms.

The design of the Finnish programme to screen women with symptoms differs substantially from other programmes. The EU guidelines mention that screening invitation should be of unselected target population and sufficient attention is paid to symptomatic details provided by the women.[13] It would be very important to record findings of these women also to a separate report of the screening programme. Many screening programmes refer symptomatic women directly to hospital or special breast clinics, thus they do not consider findings for those referred women as a part of the screening programme.[11,12,19,20] However, clinical check-up still may miss cancers and may not have as good opportunities for systematical follow-up than the screening programme. One thing is clear that collection of symptoms information in the Finnish programme is completely based on the invitation to screening, thus, no diagnostic mammography results (based on clinical presentation of symptoms) are discussed here. Not considering these women as a part of the screening mammography by the other programmes might have important consequences, such as validity of programme evaluation since a significant proportion of women with symptoms might left out, and assessing those outside the programme may demand for additional resources.[21] A better option is to request for better recording and reporting of the symptoms information from the radiologists or nurses with additions of information (requesting women also) on duration and severity of pain[21] or other possible findings based on palpation. Also, awareness of breast symptoms and importance to seek diagnostic services already prior invitation to screening is important. On the other hand, referring all women with symptoms might cause additional false-positive cases, lower efficacy of the programme, and anxiety or other psychological distress to the women. The health economic assessment of providing such a comprehensive service to all screening visits with symptoms should be made, thus warrant attention for future research.

In conclusion, our study provided comprehensive evidence that women with breast symptoms such as lump, retraction remain in a higher risk of developing breast cancer during the rest of their lifetime. In addition, the risk of dying from breast cancer was in excess over a very long period; and, astonishingly also the overall mortality risk. The later finding might relate to low socio-economic status or other such barrier in the awareness and/or access to services. The collected symptoms information should extend to long-term evaluation also in other settings, and the development of guidelines to reduce these inequities.

References

1. Breast Cancer Screening. *IARC handbook of cancer prevention - breast cancer screening*, vol. 15. Lyon, France: IARC Press, 2016. http://publications.iarc.fr/Book-And-Report-Series/Iarc-Handbooks-Of-Cancer-Prevention/Breast-Cancer-Screening-2016.
2. Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2006;4:CD001877.
3. Kalager M, Zelen M, Langmark F, et al. Effect of screening mammography on breast-cancer mortality in Norway. *N Engl J Med* 2010;363:1203–10.
4. Ramsey SD, Yoon P, Moonesinghe R, et al. Population-based study of the prevalence of family history of cancer: implications for cancer screening and prevention. *Genet Med* 2006;8:571–5.
5. Pharoah PD, Day NE, Duffy S, et al. Family history and the risk of breast cancer: a systematic review and meta-analysis. *Int J Cancer* 1997;71:800–9.
6. Slattery ML, Kerber RA. A comprehensive evaluation of family history and breast cancer risk. The Utah population database. *JAMA* 1993;270:1563–8.
7. Ziv E, Shepherd J, Smith-Bindman R, et al. Mammographic breast density and family history of breast cancer. *J Natl Cancer Inst* 2003;95:556–8.
8. Boyd NF, Byng JW, Jong RA, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian national breast screening study. *J Natl Cancer Inst* 1995;87:670–5.
9. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;356:227–36.
10. Goodwin PJ, Phillips KA, West DW, et al. Breast cancer prognosis in BRCA1 and BRCA2 mutation carriers: an international prospective breast cancer family registry population-based cohort study. *J Clin Oncol* 2012;30:19–26.
11. Mansel R, Khonji N, Clarke D. Updating national guidelines for referral of breast problems. NHS Breast Screening Programme, 1999. https://www.guidelinesinpractice.co.uk/updating-national-guidelines-for-referral-of-breast-problems/300417.article.
12. Willett A, Michell M, Lee M. Best practice diagnostic guidelines for patients presenting with breast symptoms, 2010. https://www.scribd.com/document/140938243/Best-Practice-Diagnostic-Guidelines-for-Patients-Presenting-With-Breast-Symptoms
13. European Commission. European guidelines for quality assurance in breast cancer screening and diagnosis. In: Perry N, Broeders M, de Wolf C, et al., eds *Office for Official Publications of the European Communities, Luxembourg*, Belgium, 4th edn., 2006. http://ec.europa.eu/health/ph_projects/2002/cancer/fp_cancer_2002_ext_guid_01.pdf.
14. Singh D, Pitkaniemi J, Malila N, et al. Cumulative risk of false positive test in relation to breast symptoms in mammography screening: a historical prospective cohort study. *Breast Cancer Res Treat* 2016;159:305–13.
15. Singh D, Malila N, Pokhrel A, et al. Association of symptoms and breast cancer in population-based mammography screening in Finland. *Int J Cancer* 2015;136:E630–7.
16. Finnish Cancer Registry. Breast Cancer Screening; Mass Screening Programme. Cancer Society of Finland, 2018. http://www.cancer.fi/syoparekisteri/en/mass-screening-registry/breast_cancer_screening/screening_programme/
17. Hakama M, Pukkala E, Heikkila M, et al. Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *The BMJ* 1997;314:864–7.
18. Tabar L, Duffy SW, Yen MF, et al. All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point. *J Med Screen* 2002;9:159–62.
19. Barton MB, Elmore JG, Fletcher SW. Breast symptoms among women enrolled in a health maintenance organization: frequency, evaluation, and outcome. *Ann Intern Med* 1999;130:651–7.
20. Eberl MM, Phillips RL, Lamberts H, et al. Characterizing breast symptoms in family practice. *Ann Fam Med* 2008;6:528–33.
21. Walker S, Hyde C, Hamilton W. Risk of breast cancer in symptomatic women in primary care: a case-control study using electronic records. *Br J Gen Pract* 2014;64:e788–93.