UCSF
UC San Francisco Previously Published Works

Title

Comparing sensitivity and specificity of screening mammography in the United States and Denmark

Permalink

https://escholarship.org/uc/item/6wq2x4js

Journal

International Journal of Cancer, 137(9)

ISSN

0020-7136

Authors

Kemp Jacobsen, Katja
O'Meara, Ellen S
Key, Dustin
et al.

Publication Date

2015-11-01

DOI

10.1002/ijc.29593

Peer reviewed

Comparing sensitivity and specificity of screening mammography in the United States and Denmark

Katja Kemp Jacobsen[1], **Ellen S. O'Meara**[2], **Dustin Key**[2], **Diana SM Buist**[2], **Karla Kerlikowske**[3,4], **Ilse Vejborg**[5], **Brian L. Sprague**[6], **Elsebeth Lynge**[1], and **My von Euler-Chelpin**[1]

[1] Department of Public Health, University of Copenhagen, Copenhagen K, Denmark

[2] Group Health Research Institute, Seattle, WA 98101, United States

[3] Department of Epidemiology & Biostatistics, University of California, San Francisco, CA, United States

[4] General Internal Medicine Section, Department of Veterans Affairs, University of California, San Francisco, CA, United States

[5] Center of Diagnostic Imaging, Copenhagen University Hospital, Rigshospitalet, Denmark

[6] Department of Surgery, University of Vermont, Burlington, VT 05401

Abstract

Delivery of screening mammography differs substantially between the United States (US) and Denmark. We evaluate whether there are differences in screening sensitivity and specificity. We included screens from women screened at age 50-69 years during 1996-2008/2009 in the US Breast Cancer Surveillance Consortium (BCSC) (n=2,872,791), and from two population-based mammography screening programs in Denmark (Copenhagen, n=148,156 and Funen, n=275,553). Women were followed for one year. For initial screens, recall rate was significantly higher in BCSC (17.6%) than in Copenhagen (4.3%) and Funen (3.1%). Sensitivity was fairly similar in BCSC (91.8%) and Copenhagen (90.5%) and Funen (92.5%). At subsequent screens, recall rates were 8.8%, 1.8% and 1.4% in BCSC, Copenhagen and Funen, respectively. The BCSC sensitivity (82.3%) was lower compared to Copenhagen (88.9%) and Funen (86.9%), but when stratified by time since last screen, the sensitivity was similar. For both initial and subsequent screening, the specificity of screening in BCSC (83.2 and 91.6%) was significantly lower than in Copenhagen (96.6 and 98.8%) and Funen. (97.9 and 99.2%). Taking time since last screen into account, American and Danish women had the same probability of having their asymptomatic cancers detected at screening. However, the majority of women free of asymptomatic cancers experienced more harms in terms of false-positive findings in the US than in Denmark.

Correspondence to: Katja Kemp Jacobsen, Department of Public Health, University of Copenhagen, Øster Farimagsgade 5 opg. B, PO Box 2099, 1014 Copenhagen K, Denmark. Tel.:+ [45-35327753], kkja@sund.ku.dk).

Keywords

Mammographic performance; mass screening; sensitivity; specificity; Breast Cancer Surveillance Consortium

---

Introduction

International comparisons of performance between countries with different screening organization are important, notably to identify areas for improvement. Nevertheless, comparisons are often not straightforward.

The delivery of screening mammography differs substantially between the United States (US) and Denmark. Recommendations vary with respect to age of initiation and screening frequency, outreach and follow-up strategies, health insurance coverage and access and ad-hoc adoption vs. systematic program implementation. Although the US has national guidelines for starting and stopping ages and frequency, such as the US Preventive Services Task Force and American Cancer Society recommendations,[1,2] the actual use of screening is largely determined by the woman, medical practitioner and access to health care. Many women in the US are screened every one or two years from the age of 40 years onwards.[1-3] By contrast, in Denmark screening mammography takes place in organized programs, where all women aged 50-69 years are personally invited to screening every two years, with both screening and work-up free of charge.[4] However, previous comparative studies indicated that both recall rates and interval cancer rates were higher in the US than in Europe, [5-8] suggesting both lower sensitivity and lower specificity.

In the present study we compared performance between the US and Denmark using data from the Breast Cancer Surveillance Consortium (BCSC, http://breastscreening.cancer.gov/) in the US, and from two long-standing, organized, population-based screening programs in Copenhagen and Funen, Denmark.[4] We used standardized definitions and analytic methods to compare sensitivity and specificity between the two countries.

Materials and Methods

BCSC

The BCSC is a collaborative network of seven regional mammography registries covering a population with a composition comparable to that of all US women (http://breastscreening.cancer.gov/).[9] All mammography facilities in the BCSC are accredited by the US Food and Drug Administration, and they operate under the rules and regulations of the Mammography Quality Standards Act.[10] Furthermore, the BCSC data reflect the current practice of screening in the US, and contain data from counties that include 5% of the US population.[8,11]

Within the BCSC, screening is performed in a wide range of delivery systems, including traditional fee-for-service, solo and group radiology practices, managed care organizations, hospital-based radiology practices, free standing mammography centers and mobile van programs. The standard screening procedure is two-view mammography with a single

radiology reading. However, some facilities have double reading. The use of computer aided detection (CAD) increased in the US during the study period; in a Medicare population, CAD was used for 39% of screens in 2004 to 74% in 2008.[12] In the US, radiologists are required to read at least 960 mammograms every two years.[10] Where available, mammograms from previous rounds were used for comparison.

Copenhagen and Funen

The organized, population-based screening programs in Copenhagen and Funen started in 1991 and 1993, respectively. Women aged 50-69 are personally invited to biennial screening.[4] Both programs adhere to the European Guidelines for Quality Assurance in Mammographic Screening.[13] Women targeted by the two programs constituted around 20% of Danish women 50-69 years.

Screening in Copenhagen and Funen took place at two specialized clinics, supplemented by a mobile van in Funen. At first screen all women had two projections. At subsequent screens until 2004, women with fatty breast tissue had one projection, while women with mixed/dense breast tissue had two. From 2004 and onwards, all women had two projections. Independent double reading was performed. At least one of the readers was a senior radiologist, i.e., in accordance with the European Guidelines, a radiologist reading a minimum of 5000 mammograms a year. [13] Where available, mammograms from up to three previous rounds were used for comparison. CAD was not used. Disagreements were resolved by a senior radiologist.[14] We analyzed data separately for Copenhagen and Funen, as previous studies have found different performance of the programs, despite almost identical organization.[4,15] Work-up was undertaken at the coordinating radiology departments in Copenhagen and Funen, respectively.

Study population

In both countries, mammography register data included date of birth, date of screening, type of mammogram and screening result. We included screens undertaken in 1996-2009 in the BCSC (cancer follow-up until 2010) and in 1996-2008 in Copenhagen and Funen (cancer follow-up until 2009) in women aged 50-69 years at the time of screening.

In the BCSC, a mammogram was classified as a screening mammogram based on the indication reported by the radiologist.[9,16] In Denmark, all program mammograms were classified as a screening mammogram, excluding those undertaken for work-up of positive screens.

In both countries, we excluded women who at the time of the screening, a) had a history of invasive breast carcinoma or ductal carcinoma in situ (DCIS), b) had a unilateral mammogram, c) had a radiological exam within the previous 270 days, or d) had prior mastectomy. We also excluded BCSC women with breast implants, while in Denmark, the few women with breast implants were excluded only if the screening was not technically possible.

Definitions

We used standardized definitions and analytic methods to compare sensitivity and specificity between the two countries.

Based on the woman's screening history, screens were divided into initial screens, including only the first screen in her life, and subsequent screens, including all other screens, independently of whether the woman attended screening regularly or not. Subsequent screens were stratified by time since last screen into 9-17 months, 18-29 months, and 30+ months.

The BCSC radiologists used the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS).[17] A positive or negative screen referred to the result of the initial assessment which included screening views only. A screen was considered positive if the BI-RADS assessment was 0 (needs additional imaging evaluation), 4 (suspicious abnormality), 5 (highly suggestive of malignancy), or 3 (probably benign finding) *with* a recommendation for immediate work-up. In the present study, a screen was considered negative if the BI-RADS assessment was 1 (negative), 2 (benign finding), or 3 (probably benign finding) *without* a recommendation for immediate work-up. Screens with BI-RADS 3 and missing information on need for immediate work-up were included in the analysis as negative (n=1444). Denmark did not use BI-RADS; all screens that led to recall were defined as positive and recalled for work-up.

The term breast cancer was used to include both invasive breast cancer and ductal carcinoma in situ (DCIS).

Follow-up

For BCSC, incident breast cancers were identified from the regional Surveillance, Epidemiology, and End Results registries; state cancer registries; and pathology databases. Completeness of cancer ascertainment has been estimated to be >94.3%.[18] In Denmark, incident breast cancers were identified from the Danish Cancer Registry, the Danish Breast Cancer Cooperative Group, and the Danish Pathology Register. Reporting to the Danish Cancer Registry is mandatory by law in Denmark, and the registry is essentially complete for invasive breast cancers[19], and supplemented by the other registers above, for DCIS too.

Women were followed from the date of their initial screen, or their first subsequent screen, until their next screen, diagnosis of breast cancer, or for one year, whichever came first. A breast cancer was classified as screen-detected if it was diagnosed within one year of a positive screen (or before the next screen), and as interval cancer if it was diagnosed within one year of a negative screen (or before the next screen). If the same woman had both invasive breast cancer and DCIS, the earliest result was used if the two diagnoses were more than 60 days apart, otherwise invasive breast cancer combined with the earliest date was used.

Performance measures

We calculated the following performance measures: a) recall rate: number of screens resulting in recall for work-up as a proportion of all screens, b) screen-detection rate: number of screen-detected cases per 1000 screens, c) interval cancer rate: number of interval cancers per 1000 screens, d) sensitivity: proportion of screen-detected cases among all breast cancers (screen-detected/(screen-detected+interval cancer)),e) specificity: proportion of negative screens among all women free of breast cancer (true negative screens/(true negative screens+false-positive screens)), and also included f) proportion of invasive cancers   10 mm as proportion among all invasive breast cancers, including those with unknown size, and g) invasive proportion as proportion of invasive screen-detected cancers among all screen-detected cancers.

The results were stratified by time since last screen, and rates were age-standardized using the World Standard Population (WHO 2000-2025). For both crude and age-standardized rates we computed 95% exact confidence intervals (CI). The BCSC and Danish data were analyzed separately. Non-overlapping confidence intervals were interpreted as representing a statistically significant difference. Data were analyzed using SAS (version 9.3 and 9.4, © SAS Institute Inc.), R statistical software (version 3.0.3) and Stata (StataCorp. 2011. *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP).

Ethics

BCSC registries and the Statistical Coordinating Center received Institutional Review Board approval for active or passive consenting processes or a waiver of consent to enroll participants, link data, and perform analysis and a Federal Certificate of Confidentiality and other protections for the identities of women, physicians, and facilities. All procedures were Health Insurance Portability and Accountability Act compliant. For Danish registries, use of screening data and tumor-related information was approved by the Danish Data Inspection Agency (2008-41-2191).

Results

We included 2,872,791 screens from BCSC (1.8% were initial screens), 148,156 from Copenhagen (22.5% initial screens), and 275,553 from Funen (14.7% initial screens). More than 80% were film mammograms (Table 2).

Initial screens

For initial screens, the age-standardized recall rate was significantly higher in the BCSC (17.6%) than in Copenhagen (4.3%) and Funen (3.1%) (Table 3). The age-standardized screen-detection rates were fairly similar across the 3 sites; being 9.7 (95% CI 8.9-10.6); 9.1 (95% CI 7.0-11.3) and 10.2 (95% CI 7.5.-12.8) per 1000 screens. Interval cancer rates were similar, being 0.8 (95% 0.6-1.1); 1.0 (95% CI 0.3-1.6) and 0.7 (95% CI 0.1-1.2) per 1000 screens. Initial screening sensitivity was similar between areas at 91.8% in the BCSC, as compared to 90.5% in Copenhagen and 92.5% in Funen. The specificity of initial screens was significantly lower in BCSC (83.2%) than in Copenhagen (96.6%) and Funen (97.9%). At initial screens, the percentage of invasive cancers   10 mm was lower in BCSC (20.9%)

than in Copenhagen (28.5%) and Funen (21.6%). The percentage of invasive out of all screen-detected cancers was 81.4% in BCSC, which is similar to Copenhagen (86.5%) but significantly lower than in Funen (90.3%).

Subsequent screens

Recall rates decreased by half compared to initial screens and were 8.8%, 1.8% and 1.4% in the BCSC, Copenhagen and Funen, respectively (Table 4). Age-standardized screen-detection rate for subsequent screens was significantly lower in the BCSC, 4.3 (95% CI 4.3-4.4), than in Copenhagen, 5.8 (95% CI 5.4-6.3), and Funen, 5.5 (95% CI 5.2-5.8) per 1000 screens, and the interval cancer rates were fairly similar in the BCSC, 0.9 (95% CI 0.9-1.0), Copenhagen, 0.7 (95% CI 0.6-0.9), and Funen, 0.8 (95% CI 0.7-0.9) per 1000 screens. Sensitivity was significantly lower in the BCSC (82.3%) than in Copenhagen (88.0%) and Funen (86.9%), and also specificity was significantly lower in the BCSC (91.6%) than in Copenhagen (98.8%) and Funen (99.2%), not accounting for time between screens. The proportion of small invasive cancers was significantly lower in BCSC (27.3%) than in Copenhagen (39.9%) and Funen (34.5%), and invasive cancers constituted a significantly lower percentage of screen-detected cancers in BCSC (74.9%) than in Copenhagen (81.1%) and Funen (89.8%).

There were important differences in the time between screens. In BCSC, 66.5% of subsequent screens were undertaken 9-17 months after the previous screen; 22.2% 18-29 months after, and 11.3% 30+ months after. In the Danish programs very few women were rescreened within 9-17 months, and in Funen almost all women were rescreened within 18-29 months (Table 1). In BCSC, recall rates increased significantly by time since last screen and as did screen-detection rates, from 3.6 per 1000 screens for 9-17 months, to 4.7 for 18-29 months, and 6.8 for 30+ months, while interval cancer rates decreased slightly from 1.0 to 0.8 to 0.7 per 1000 screens (Table 5). Recall rates for subsequent screens in Copenhagen did not appear to differ between those screened 18-29 months or 30+ months previously, but the screen-detection rate increased from 5.6 to 7.2 per 1000 screens, while the interval cancer rate remained fairly stable. At 18-29 months after the last screen, the sensitivity was 85%; 87.5%; and 86.7%, respectively, in BCSC, Copenhagen and Funen, and with longer intervals it was 90.1% in BCSC, 90.2% in Copenhagen and 92.3% in Funen.

Sensitivity was significantly higher for initial screens as compared with subsequent screens in the BCSC; in contrast, there were only minor differences between initial and subsequent screens in Copenhagen and Funen (Figure 1).Specificity was significantly lower for initial than for subsequent screens, from 83.2% to 91.6% in BCSC, from 96.6% to 98.8% in Copenhagen and from 97.9% to 99.2% in Funen (Table 3).

Discussion

Main finding

Women screened in the US were at a four-six-fold increased risk of being recalled for work-up as compared with Danish women. At initial screens, the sensitivity was fairly similar in the US and Denmark and for subsequent screens when stratified by time since last screen.

The specificity of screening in the US was considerably lower than in Denmark at both initial and subsequent screens. The proportion of small invasive cancers was lower in the US than in Denmark. The percentage of DCIS was higher in the US than in Denmark.

Strengths and limitations

The strengths of this study included the use of comprehensive data from the BCSC, representing screening practice in the US in 1996-2009, and the two Danish programs covering all screens undertaken in 1996-2008, in women aged 50-69 years. To overcome the differences in procedures and terminology between the US and Denmark, we carefully applied inclusion and exclusion criteria and defined the screening outcomes in a comparable way. Furthermore, the cancer registries provided almost complete cancer data. Initial and subsequent screens were analyzed separately as the performance of screening tended to differ between the two.[6,11,20] This difference has been explained by a) the longer sojourn time of the tumors at initial screen and hence easier detection, and b) the larger proportion of young women receiving initial screens.[11]

Differences in the populations studied might remain, as we were not able to include potentially important variables such as postmenopausal hormone use, breast density and family history of breast cancer. Postmenopausal hormone use has been considerably higher in US than Danish women.[21] But hormone use is associated with increased breast density and thus decreased performance of mammography.[22,23] It should also be taken into account that screening and work-up was not free of charge in the US, and that many women were not covered by insurance. In the US, some women with a positive screen may therefore not have returned for work-up, while this was rare in Denmark. The US sensitivity might thus be underestimated.

International screening comparisons are important for assessing the effectiveness of different screening strategies, outreach, practice and intervals. In the US, 67% of screened women were rescreened within 9-17 months, while this was very rare in Denmark. As the follow-up for screen-detected cases and interval cancers stopped after the earlier of one year or the next screen, this means that the follow-up of the American women was on average shorter than the follow-up of Danish women. This difference might have increased the number of false-positives and decreased the number of false-negatives in the US data, which could lead to underestimation of specificity and overestimation of sensitivity in the US compared to Denmark. However, we expect only minor bias from this source because we excluded screens with a radiology exam within the previous 270 days and because only 1.8% of subsequent screens in the BCSC were performed within one year of the previous screen (data not shown).

The different screening strategies also explain why the outcome of 2-year follow-up, which might have been more reasonable for the Danish data, was not calculated. Almost the same number of interval cancers was registered in the second year as in the first year in the US, while in Denmark the number of interval cancers doubled from the first to the second year (data not shown). This could be due to the US data often being censored before two years, thus shortening the period where an interval cancer could occur. In the BCSC data, a screen-detected cancer could have been misclassified as an interval cancer if the next screening

examination occurred outside the BCSC. Hence, comparison of interval cancers beyond one year was not possible. Even a comparison based on rates per person-years at risk might be biased if those screened with longer intervals were not representative of all screened women in the US.

We did not stratify the analysis by mammography technology as the proportion of digital mammograms was low in Copenhagen. However, the distribution of mammograms by type was fairly similar in the BCSC and Funen at subsequent screens so we do not expect this to explain the differences in performance. Prior studies comparing digital with film-screen mammograms have produced inconsistent results on the association with performance.[24-27] One study based on a subset of our BCSC population found fairly similar performance of digital and film-screen mammograms among women aged 50-69 years.[24] Another study based on women aged 50-69 years enrolled in the Norwegian Breast Cancer Screening Program, which should be compatible to our Danish populations, found a lower recall rate among women screened with digital mammograms than among film-screened women.[27]

Comparison with other studies

Higher recall rates and somewhat lower screen-detection rates in subsequent screens in the US as compared to Europe were also observed in the large, classic study by Smith-Bindman et al.[28] covering the years 1996-1999, suggesting not much has changed since then in the US with respect to recall of screened women for work-up. This is probably due to a combination of less centralized screening in the US, less reading experience required of US radiologists, less use of double reading, different guidelines for acceptable levels of mammograms judged as abnormal (<5% in Europe) and (10% in US) and the litigious environment in the US.[13,17,29,30] Smith-Bindman et al.[28] included women aged 50+ years in the BCSC, the US National Breast and Cervical Cancer Early Detection Program, and in the UK National Health Service Breast Screening Program. Using age-standardization to the combined study population on subsequent screens, they found US recall rates of 8.0% and 6.8% and screen-detection rates of 3.6 and 3.4 per 1000. Our numbers for the US, age-standardized to the WHO population, were 8.8% and 4.3 per 1000, respectively. Danish recall rates of 1.8%/ 1.4% were only half of that found for the UK, 3.6%, while the Danish and UK screen-detection rates were similar at 5.8/5.5 and 5.4 per 1000. In another study that used data from 1997-2003 for women aged 50-69 years and age-standardization to the combined Vermont (US; one of the BCSC registries) and Norwegian population, Hofvind et al.[5] found the Vermont recall rate to be 9.8% and the screen-detection rate to be 4.01 per 1000, while in Norway the recall rate was 2.7% and the screen-detection rate 5.08 per 1000.

Higher screen-detection rates are expected in the European programs, with biennial or triennial screening, as compared to the predominantly annual screening in the US. This is due to an increasing pool of cancers for detection given longer time of development. In our study, the BCSC data on subsequent screens showed an increase in the screen-detection rate from 3.6 per 1000 for annual screens, 4.7 for biennial, and 6.8 for longer intervals. In the Vermont data reported by Hofvind et al.[5] these numbers were 3.53, 4.55, and 8.27, respectively. For Copenhagen, our data showed a screen-detection rate of 5.6 per 1000 for

biennial screening, and of 7.2 per 1000 for longer intervals. In the Norway data reported by Hofvind et al.[5] the rates were 4.96 and 9.13, respectively.

Both the recall rate and screen-detection rate reflect screening performance as well as the underlying risk of breast cancer. A purer picture of performance is therefore obtained by comparing screening sensitivity and specificity. Using a 1-year follow-up for the Vermont and Norway data on subsequent screens, Hofvind et al.[31] found a sensitivity of 83.8% and 91.0%, and a specificity of 90.6% and 97.8%, respectively. These data largely agree with ours, as we found the BCSC sensitivity to be 82.3% and the Danish sensitivity to be 88.0%/86.9%, and the specificities to be 91.6% and 98.8%/99.2%. For subsequent screens, the two studies thus showed a consistent pattern of 6-7% higher sensitivity (when not stratified for time since last screen) and 7% higher specificity in the organized Nordic programs as compared to the mixed setting of BCSC. However, a third study comparing interval cancers after subsequent screens with 1 year of follow-up in North Carolina (US; one of the BCSC registries) and Norway, Hofvind et al.[6] found rates of 1.29 and 0.54 per 1000 screens. In our data, per 1000 screens we get 0.9 for the BCSC and 0.7/0.8 for Denmark with overlapping confidence intervals.

We found a lower proportion of 10 mm invasive cancers in the BCSC than in the Danish programs. Assuming similar biology characteristics, this was surprising as tumor size is a time-dependent factor and higher tumor-size should be correlated with less frequent screening. However, using a cut-off of 15 mm, Hofvind et al.[5] found almost similar proportions of 62% and 67% for Vermont and Norway. The highest proportion of DCIS was detected in the BCSC, for subsequent screens 25% as compared to 19% in Copenhagen and 10% in Funen. Hofvind et al.[5] found 24% for Vermont and 18% for Norway. The higher rate of DCIS in the US could be due to the use of CAD which can increase the detection of DCIS.[32]

The difference in performance measures between Copenhagen and Funen might reflect a true performance difference between the programs. Nevertheless, it also reflects the well-known urban-rural gradient in screening participation[4], in combination with small differences between the programs that might influence on the screening population[15]. However, it suggests that comparisons of performance measures are not straightforward, even when comparing within a small country like Denmark.

Clinical implications

Cross-national performance comparisons are important because they allow us to identify differences in performance and relate these to specific differences in program organization, which might help to improve performance. International comparisons offer possibilities to study variations in performance beyond comparisons within nations. The most obvious difference in performance between the US and Denmark was the recall rates. Of 21 women recalled for work-up after a subsequent screen in the US, 1 was diagnosed with breast cancer within a year. In Denmark, these numbers were 1 out of 3. Recalling women who, on further work-up, are found not to have breast cancer induces increased costs and increased anxiety for the women involved.[33,34] These factors could affect women's general acceptance of screening mammography and hence their willingness to participate and have a large impact

on the quality and value of any screening program. It might though be difficult to map participation, especially in US settings where women can attend screening at different facilities and where there is no central database. Ultimately, high participation is necessary to ensure a favorable impact of screening on breast cancer mortality.

Conclusion

International comparisons of screening sensitivity and specificity are complex. The number of screen-detected cancers depends on time since the last screen, and interval cancers can be counted only until the next screen, factors that can differ by nation. By stratifying by time since the last screen and restricting follow-up to one year after the screen we sought to accommodate the different screening schemes in the US and Europe.

Our study showed that the American and Danish women taking time since last screen into account had the same probability of having their asymptomatic breast tumors detected at screening. However, the majority of women, who are tumor-free, pay a much higher price in terms of false-positive findings in the US than in Denmark.

This balance between US and European screening performance has persisted for more than ten years.

Acknowledgments

Abbreviations

BCSC	Breast Cancer Surveillance Consortium
US	United States
IBC	Invasive breast cancer
DCIS	Ductal Carcinoma in situ
CI	Confidence interval
ASR	World age-standardized (WHO 2000-2025) rate

References

1. Smith RA, Cokkinides V, Brawley OW. Cancer screening in the United States, 2012: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin. 2012

2. U.S Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009; 151:716–236. [PubMed: 19920272]

3. American Cancer Society. [January 30, 2015] Paying for breast cancer screening. 2014. Availible at http://www.cancer.org/cancer/breastcancer/moreinformation/breastcancerearlydetect ion/breast-cancer-early-detection-paying-for-br-ca-screening/

4. Jacobsen KK, Von Euler-Chelpin M. Performance indicators for participation in organized mammography screening. J Public Health (Oxf). 2012; 34:272–78. [PubMed: 22233570]

5. Hofvind S, Vacek PM, Skelly J, Weaver DL, Geller BM. Comparing screening mammography for early breast cancer detection in Vermont and Norway. J Natl Cancer Inst. 2008; 100:1082–91. [PubMed: 18664650]

6. Hofvind S, Yankaskas BC, Bulliard JL, Klabunde CN, Fracheboud J. Comparing interval breast cancer rates in Norway and North Carolina: results and challenges. J Med Screen. 2009; 16:131–9. [PubMed: 19805754]

7. Smith-Bindman R, Ballard-Barbash R, Miglioretti DL, Patnick J, Kerlikowske K. Comparing the performance of mammography screening in the USA and the UK. J Med Screen. 2005; 12:50–4. [PubMed: 15814020]

8. Yankaskas BC, Klabunde CN, Ancelle-Park R, Renner G, Wang H, Fracheboud J, Pou G, Bulliard JL. International comparison of performance measures for screening mammography: can it be done? J Med Screen. 2004; 11:187–93. [PubMed: 15624239]

9. Ballard-Barbash R, Taplin SH, Yankaskas BC, Ernster VL, Rosenberg RD, Carney PA, Barlow WE, Geller BM, Kerlikowske K, Edwards BK, Lynch CF, Urban N, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol. 1997; 169:1001–8. [PubMed: 9308451]

10. U.S.Food and Drug Administration. [January 30, 2015] Mammography Quality Standards Act and Program. 2015. Availible at http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/default.htm/

11. Yankaskas BC, Taplin SH, Ichikawa L, Geller BM, Rosenberg RD, Carney PA, Kerlikowske K, Ballard-Barbash R, Cutter GR, Barlow WE. Association between mammography timing and measures of screening performance in the United States. Radiology. 2005; 234:363–73. [PubMed: 15670994]

12. Rao VM, Levin DC, Parker L, Cavanaugh B, Frangos AJ, Sunshine JH. How widely is computer-aided detection used in screening and diagnostic mammography? J Am Coll Radiol. 2010; 7:802–5. [PubMed: 20889111]

13. Perry N, Broeders M, de WC, Tornberg S, Holland R, von KL. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition--summary document. Ann Oncol. 2008; 19:614–22. [PubMed: 18024988]

14. Utzon-Frank N, Vejborg I, Von Euler-Chelpin M, Lynge E. Balancing sensitivity and specificity: Sixteen year's of experience from the mammography screening programme in Copenhagen, Denmark. Cancer Epidemiol. 2011:393–8. [PubMed: 21239242]

15. Domingo L, Jacobsen KK, Von Euler-Chelpin M, Vejborg I, Schwartz W, Sala M, Lynge E. Seventeen-years overview of breast cancer inside and outside screening in Denmark. Acta Oncol. 2013; 52:48–56. [PubMed: 22943386]

16. Breast Cancer Surveillance Consortium. [January 30, 2015] BCSC Glossary of Terms. 2009. Availible at http://breastscreening.cancer.gov/data/bcsc_data_definitions.pdf/

17. Sickles, EA.; D'Orsi, CJ.; Bassett, LW., et al. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System.. American College of Radiology; Reston, VA: 2013. ACR BI-RADS® Mammography..

18. Ernster VL, Ballard-Barbash R, Barlow WE, Zheng Y, Weaver DL, Cutter G, Yankaskas BC, Rosenberg R, Carney PA, Kerlikowske K, Taplin SH, Urban N, et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. J Natl Cancer Inst. 2002; 94:1546–54. [PubMed: 12381707]

19. Sundhedsstyrrelsen. [January 30, 2015] Det moderniserede Cancerregister-metode og kvalitet. 2014. [cited 2015 Feb 4]. Availible at http://www.ssi.dk/Sundhedsdataogit/Registre/~/media/Indhold/DK%20-%20dansk/Sundhedsdata%20og%20it/NSF/Registre/Cancerregisteret/Det%20moderniserede%20Cancerregister%20%20metode%20og%20kvalitet.ashx.2014/

20. Tornberg S, Codd M, Rodrigues V, Segnan N, Ponti A. Ascertainment and evaluation of interval cancers in population-based mammography screening programmes: a collaborative study in four European centres. J Med Screen. 2005; 12:43–9. [PubMed: 15814019]

21. Von Euler-Chelpin M. Breast cancer incidence and use of hormone therapy in Denmark 1978-2007. Cancer Causes Control. 2011; 22:181–7. [PubMed: 21103920]

22. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, Geller BM, Abraham LA, Taplin SH, Dignan M, Cutter G, Ballard-Barbash R. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med. 2003; 138:168–75. [PubMed: 12558355]

23. Njor SH, Hallas J, Schwartz W, Lynge E, Pedersen AT. Type of hormone therapy and risk of misclassification at mammography screening. Menopause. 2011; 18:171–7. [PubMed: 20729766]

24. Kerlikowske K, Hubbard RA, Miglioretti DL, Geller BM, Yankaskas BC, Lehman CD, Taplin SH, Sickles EA. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. Ann Intern Med. 2011; 155:493–502. [PubMed: 22007043]

25. Karssemeijer N, Bluekens AM, Beijerinck D, Deurenberg JJ, Beekman M, Visser R, van ER, Bartels-Kortland A, Broeders MJ. Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. Radiology. 2009; 253:353–8. [PubMed: 19703851]

26. Skaane P, Hofvind S, Skjennald A. Randomized trial of screen-film versus full-field digital mammography with soft-copy reading in population-based screening program: follow-up and final results of Oslo II study. Radiology. 2007; 244:708–17. [PubMed: 17709826]

27. Skaane P. Studies comparing screen-film mammography and full-field digital mammography in breast cancer screening: updated review. Acta Radiol. 2009; 50:3–14. [PubMed: 19037825]

28. Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, Bobo JK, Lee NC, Wallis MG, Patnick J, Kerlikowske K. Comparison of screening mammography in the United States and the United kingdom. JAMA. 2003; 290:2129–37. [PubMed: 14570948]

29. Hofvind S, Geller B, Vacek PM, Thoresen S, Skaane P. Using the European guidelines to evaluate the Norwegian Breast Cancer Screening Program. Eur J Epidemiol. 2007; 22:447–55. [PubMed: 17594526]

30. Helvie M. Improving mammographic interpretation: double reading and computer-aided diagnosis. Radiol Clin North Am. 2007; 45:801–11. [PubMed: 17888770]

31. Hofvind S, Geller BM, Skelly J, Vacek PM. Sensitivity and specificity of mammographic screening as practised in Vermont and Norway. Br J Radiol. 2012; 85:e1226–e32. [PubMed: 22993383]

32. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. N Engl J Med. 2007; 356:1399–409. [PubMed: 17409321]

33. Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. Psychooncology. Nov. 2005; 14:917–38. [PubMed: 15786514]

34. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med. 2007; 146:502–10. [PubMed: 17404352]

What is new?

Women screened for breast cancer in the US were recalled for work-up four-six times more frequently than women screened in Denmark. American and Danish women had the same probability of having their asymptomatic cancers detected at screening. However, in the US only one out of 21 women recalled for work-up had breast cancer, while this was one out of three in Denmark.
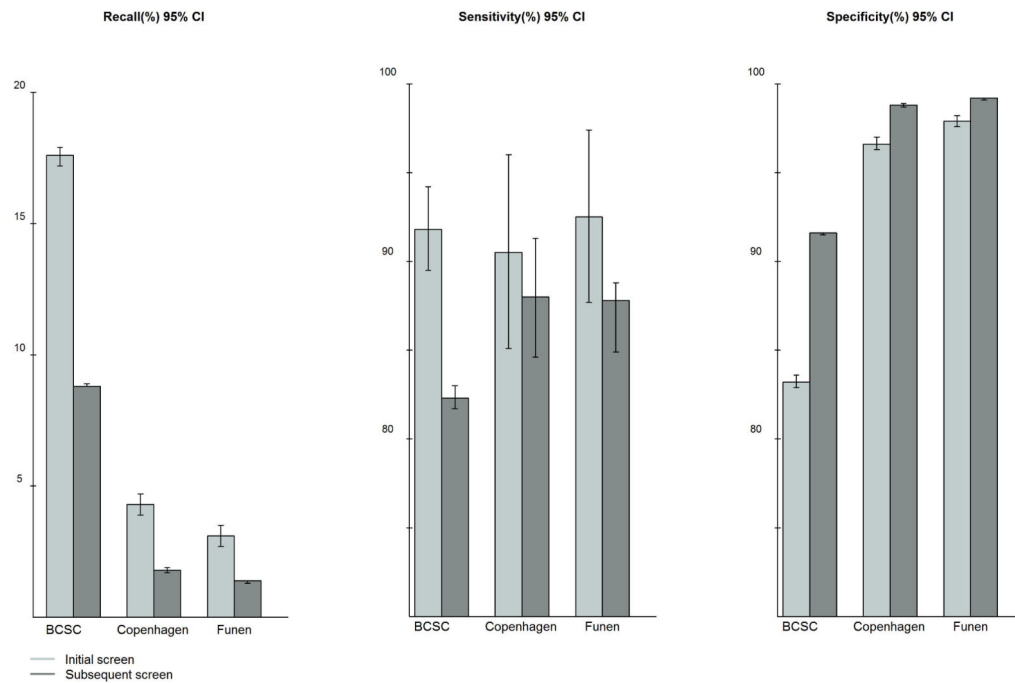
Figure 1.
Distribution of recall rates, sensitivity and specificity after initial and subsequent screens presented as age-standardized rates (World 2000-2025) for screening mammography in the Breast Cancer Surveillance Consortium (BCSC), United States, and in the organized programs in Copenhagen and Funen, Denmark
CI: Confidence interval.

Table 1

Screening mammography in the Breast Cancer Surveillance Consortium (BCSC) (1996-2009) and Denmark (1996-2008)

	BCSC, US	Copenhagen and Funen
Organization	Opportunistic screening (recommendation-based screening)	Organized, population-based screening programs
Invitation	Women self-refer or are referred by a health care provider	Women are personally invited
Target group	40 years (American Cancer Society)[1] 50-74 years (US Preventive Services Task Force)[2]	Women aged 50-69 years
Recommended screening interval	1-2 years	2 years
Facilities	Screening is performed within a wide range of delivery systems	Screening is performed at two centralized breast clinics, supplemented by a mobile van
Recommendation for minimum reading	960 mammograms per 24 months	5000 mammograms per 12 months
Reading	Primarily single reading	Independent double reading, with consensus
Use of Computer Aided detection (CAD)	Yes	No

[1] See reference 1 in the references.

[2] See reference 2 in the references.

Table 2

Characteristics of initial and subsequent screens for screening mammography in the Breast Cancer Surveillance Consortium (BCSC), United States (1996–2009) and in the organized programs in Copenhagen and Funen, Denmark (1996–2008). Women aged 50-69 years

	BCSC		Copenhagen		Funen	
	Number	Column %	Number	Column %	Number	Column %
Initial screens, total	51,445		33,325		40,406	
Mammogram type						
Film	45,534	88.5	32,006	96.0	35,402	87.6
Digital	4,017	7.8	1,319	4.0	5,004	12.4
Film or digital (NOS)	1,894	3.7	0	0.0	0	0.0
Age at the screen						
50-54 years	21,455	41.7	27,757	83.3	37,246	92.2
55-59 years	11,764	22.9	3,312	9.9	1,305	3.2
60-64 years	9,097	17.7	1,544	4.6	1,150	2.9
65-69 years	9,129	17.7	712	2.1	705	1.7
Subsequent screens, total	2,821,346		114,831		235,147	
Mammogram type						
Film	2,253,587	79.9	107,242	93.4	206,488	87.8
Digital	412,642	14.6	7,589	6.6	28,659	12.2
Film or digital (NOS)	155,117	5.5	0	0.0	0	0.0
Age at the screen						
50-54 years	916,265	32.5	15,096	13.2	48,289	20.5
55-59 years	780,460	27.7	40,800	35.5	75,451	32.1
60-64 years	612,438	21.7	33,009	28.8	62,418	26.5
65-69 years	512,183	18.1	25,926	22.6	48,989	20.8
Time since last screen[1]						
9-17 months	1,753,548	66.5	884	0.8	1,936	0.8

	BCSC		Copenhagen		Funen	
	Number	Column %	Number	Column %	Number	Column %
18-29 months	585,545	22.2	91,254	79.5	225,374	95.9
30+ months	296,583	11.3	22,693	19.8	7,837	3.3

NOS: Not specified in data.

[1] In BCSC 185,670 (6.6% of 2,821,346) screens have missing data on time since the previous screen.

Table 3

Distribution of performance measures after initial screens presented as crude and age-standardized rates (ASR) for screening mammography in the Breast Cancer Surveillance Consortium (BCSC), United States, and in organized programs in Copenhagen and Funen, Denmark

	BCSC			Copenhagen			Funen		
	Numerator (denominator)	Crude rate per 1000 screens or percentage (95%CI)	ASR (world)[1] per 1000 screens or percentage (95%CI)	Numerator (denominator)	Crude rate per 1000 screens or percentage (95%CI)	ASR (world)[1] per 1000 screens or percentage (95%CI)	Numerator (denominator)	Crude rate per 1000 screenings or percentage (95%CI)	ASR (world)[1] per 1000 screenings or percentage (95%CI)
Recall rate (%)	9,160(51,445)	17.8(17.5-18.1)	17.6(17.2-17.9)	1,472(33,325)	4.4(4.2-4.6)	4.3(3.9-4.7)	1,139(40,406)	2.8(2.7-3.0)	3.1(2.7-3.5)
Screen-detection rate (per 1000 screens)	481(51,445)	9.3(8.5-10.2)	9.7(8.9-10.6)	244(33,325)	7.3(6.4-8.3)	9.1(7.0-11.3)	240(40,406)	5.9(5.2-6.7)	10.2(7.5-12.8)
Interval cancer rate (per 1000 screens)	43(51,445)	0.8(0.6-1.1)	0.8(0.6-1.1)	26(33,325)	0.8(0.5-1.1)	1.0(0.3-1.6)	31(40,406)	0.8(0.5-1.1)	0.7(0.1-1.2)
Sensitivity (%)	481(524)	91.8(89.1-94.0)	91.8(89.5-94.2)	244(270)	90.4(86.2-93.6)	90.5(85.1-96.0)	240(271)	88.6(84.2-92.3)	92.5(87.7-97.4)
Specificity (%)	42,242(50,921)	83.0(82.6-83.3)	83.2(82.9-83.6)	31,827(33,055)	96.3(96.1-96.5)	96.6(96.3-97.0)	39,236(40,135)	97.8(97.6-97.9)	97.9(97.6-98.2)
Invasive cancers ≤10 mm (%)	85(393)	21.6(17.7-26.0)	20.9(16.9-25.0)	66(200)	33.0(26.5-40.0)	28.5(19.4-37.6)	51(204)	25.0(19.2-31.5)	21.6(12.6-30.6)
Invasive out of all screen-detected cancers (%)	393(481)	81.7(78.0-85.1)	81.4(77.9-84.9)	200(244)	82.0(76.5-86.6)	86.5(80.4-92.7)	204(240)	85.0(80.9-89.8)	90.3(85.2-95.3)

CI: Confidence interval.

[1] World age-standardized (WHO 2000-2025) rate.

Table 4

Distribution of performance measures after subsequent screens presented as crude and age-standardized (ASW) for screening mammography in the Breast Cancer Surveillance Consortium (BCSC), United States, and in organized programs in Copenhagen and Funen, Denmark

	BCSC			Copenhagen			Funen		
	Numerator (denominator)	Crude rate per 1000 screens or percentage (95%CI)	ASR (world)[1] per 1000 screens or percentage (95%CI)	Numerator (denominator)	Crude rate per 1000 screens or percentage (95%CI)	ASR (world)[1] per 1000 screens or percentage (95%CI)	Numerator (denominator)	Crude rate per 1000 screenings or percentage (95%CI)	ASR (world)[1] per 1000 screenings or percentage (95%CI)
Recall rate (%)	249,167(2,821,346)	8.8(8.8-8.9)	8.8(8.8-8.9)	2,095(114,831)	1.8(1.8-1.9)	1.8(1.7-1.9)	3,304(235,147)	1.4(1.4-1.5)	1.4(1.3-1.4)
Screen-detection rate (per 1000 screens)	12,226(2,821,346)	4.3(4.3-4.4)	4.3(4.3-4.4)	757(114,831)	6.6(6.1-7.1)	5.8(5.4-6.3)	1367(235,147)	5.8(5.5-6.1)	5.5(5.2-5.8)
Interval cancer rate (per 1000 screens)	2,586(2,821,346)	0.9(0.9-1.0)	0.9(0.9-1.0)	95(114,831)	0.8(0.7-1.0)	0.7(0.6-0.9)	187(235,147)	0.8(0.7-0.9)	0.8(0.7-0.9)
Sensitivity (%)	12,226(14,812)	82.5(81.9-83.1)	82.3(81.7-83.0)	757(852)	88.9(86.5-90.9)	88.0(84.6-91.3)	1,367(1,554)	88.0(86.2-89.5)	86.9(84.9-88.8)
Specificity (%)	2,569,593(2,806,534)	91.6(91.5-91.6)	91.6(91.5-91.6)	112,641(113,979)	98.8(98.8-98.9)	98.8(98.7-98.9)	231,656(233,593)	99.2(99.1-99.2)	99.2(99.1-99.2)
Invasive cancers 10 mm (%)	2,540(9,217)	27.6(26.6,28.5)	27.3(26.4-28.2)	269(632)	42.6(38.5-46.4)	39.9(34.1-45.7)	434(1,225)	35.4(32.8-38.2)	34.5(31.4-37.5)
Invasive out of all screen-detected cancers (%)	9,217(12,226)	75.4(74.6-76.2)	74.9(74.2-75.7)	632(757)	83.5(80.7-86.1)	81.1(76.6-85.6)	1,225(1,367)	89.6(87.9-91.2)	89.8(88.0-91.6)

CI: Confidence interval.

[1] World age-standardized (WHO 2000-2025) rate.

Table 5

Distribution of performance measures after subsequent screens presented as crude and age-standardized rates (World 2000-2025), stratified by time since last screen for screening mammography in the Breast Cancer Surveillance Consortium (BCSC), United States, and in the organized programs in Copenhagen and Funen, Denmark

	BCSC			Copenhagen			Funen		
	Numerator (denominator)	Crude rate per 1000 screens or percentage (95%CI)	ASR (world)[1] per 1000 screens or percentage (95%CI)	Numerator (denominator)	Crude rate per 1000 screens or percentage (95%CI)	ASR (world)[1] per 1000 screens or percentage (95%CI)	Numerator (denominator)	Crude rate per 1000 screenings or percentage (95%CI)	ASR (world)[1] per 1000 screenings or percentage (95%CI)
Recall rate (%)									
9-17 months	141,744(1753,548)	8.1(8.0-8.1)	8.1(8.1-8.1)	25(884)	2.8(1.8-4.2)	3.0(1.8-4.2)	28(1,936)	1.4(1.0-2.1)	1.4(0.9-2.0)
18-29 months	52,648 (585,545)	9.0(8.9-9.1)	9.0(8.9-9.0)	1,647(91,254)	1.8(1.7-1.9)	1.8(1.7-1.9)	3,117(225,374)	1.4(1.3-1.4)	1.4(1.3-1.4)
30+ months	32,694(296,583)	11.0(10.9-11.1)	10.9(10.8-11.1)	423(22,693)	1.9(1.7-2.1)	1.6(1.4-1.8)	159(7,837)	2.0(1.7-2.4)	2.0(1.7-2.4)
Screen-detection rate (per 1000 screens)									
9-17 months	6,381(1753,548)	3.6(3.6-3.7)	3.6(3.5-3.7)	6(884)	6.8(2.5-14.7)	7.3(1.4-13.2)	11(1,936)	5.7(2.8-10.1)	5.8(2.3-9.3)
18-29 months	2,706(585,545)	4.6(4.4-4.8)	4.7(4.5-4.8)	569(91,254)	6.2(5.7-6.7)	5.6(5.1-6.0)	1,293(225,374)	5.7(5.4-6.1))	5.5(5.2-5.8)
30+ months	1,965(296,583)	6.6(6.3-6.9)	6.8(6.5-7.1)	182(22,693)	8.0(6.9-9.3)	7.2(5.7-8.9)	63(7,837)	8.0(6.2-10.3)	8.1(5.8-10.4)
Interval cancer rate (per 1000 screens)									
9-17 months	1,743(1753,548)	1.0(0.9-1.0)	1.0(0.9-1.0)	2(884)	2.3(0.3-8.1)	1.7(−0.6-4.0)	3(1,936)	1.5(0.3-4.5)	1.4(−0.2-3.0)
18-29 months	471(585,545)	0.8(0.7-0.9)	0.8(0.7-0.9)	77(91,254)	0.8(0.7-1.1)	0.8(0.6-0.9)	180(225,374)	0.8(0.7-0.9)	0.8(0.7-0.9)
30+ months	214(296,583)	0.7(0.6-0.8)	0.7(0.6-0.8)	16(22,693)	0.7(0.4-1.2)	0.7(0.2-1.3)	4(7,837)	0.5(0.1-1.3)	0.6(−0.1-1.3)
Sensitivity (%)									
9-17 months	6,381(8,124)	78.5(77.6-79.4)	78.2(77.3-79.1)	6(8)	75.0(35.0-96.8)	NA	11(14)	78.6(49.2-95.3)	84.8(70.3-99.3)
18-29 months	2,706(3,177)	85.2(83.9-86.4)	85.0(83.7-86.2)	569(646)	88.1(85.3-90.5)	87.5(83.8-91.2)	1,293(1,473)	87.8(86.0-89.4)	86.7(84.7-88.8)
30+ months	1,965(2,179)	90.2(88.9-91.4)	90.1(88.8-91.3)	182(198)	91.9(87.2-95.3)	90.2(82.2-98.1)	63(67)	94.0(85.4-98.4)	92.3(84.0-100.3)
Specificity (%)									
9-17 months	1,610,061(1,745,424)	92.2(92.2-92.3)	92.2(92.2-92.3)	857(876)	97.8(96.7-98.7)	97.7(96.6-98.8)	1,905(1,922)	99.1(98.6-99.5)	99.2(98.8-99.6)
18-29 months	532,426(582,368)	91.4(91.4-91.5)	91.5(91.4-91.5)	89,530(90,608)	98.8(98.7-98.9)	98.8(98.7-98.9)	222,077(223,901)	99.2(99.2-99.2)	99.2(99.1-99.2)
30+ months	263,675(294,404)	89.6(89.5-89.7)	89.7(89.6-89.8)	22,254(22,595)	98.9(98.8-99.1)	99.1(98.9-99.2)	7,674(7,770)	98.8(98.5-99.0)	98.8(98.5-99.1)

	BCSC			Copenhagen			Funen		
	Numerator (denominator)	Crude rate per 1000 screens or percentage (95%CI)	ASR (world)[1] per 1000 screens or percentage (95%CI)	Numerator (denominator)	Crude rate per 1000 screens or percentage (95%CI)	ASR (world)[1] per 1000 screens or percentage (95%CI)	Numerator (denominator)	Crude rate per 1000 screenings or percentage (95%CI)	ASR (world)[1] per 1000 screenings or percentage (95%CI)
Invasive cancers ≤10 mm (%)									
9-17 months	1,419(4,725)	30.0(28.7,31.4)	29.7(28.3-31.0)	2(5)	40.0(5.3-85.3)	NA	4(11)	36.4(10.9-69.2)	29.8(13.2-46.4)
18-29 months	566(2,053)	27.6(25.6,29.6)	27.3(25.3-29.2)	201(477)	42.1(37.7-46.7)	37.2(31.4-43.0)	405(1,152)	35.2(32.4-38.0)	34.1(31.0-37.2)
30+ months	342(1,552)	22.0(20.0,24.2)	21.8(19.8-23.9)	65(150)	43.3(35.2-51.7)	52.7(37.9-67.5)	25(62)	40.3(28.1-53.6)	37.6(23.6-51.7)
Invasive out of all screen-detected cancers (%)									
9-17 months	4,725 (6,381)	74.0(73.0-75.1)	73.4(72.3-74.5)	5(6)	83.3(35.9-99.6)	NA	11(11)	100(71.5-100)	NA
18-29 months	2,053 (2,706)	75.9(74.2-77.5)	75.5(73.8-77.1)	477(569)	83.8(80.5-86.8)	82.3(77.6-87.0)	1,152(1,293)	89.1(87.3-90.7)	89.3(87.5-91.2)
30+ months	1,552 (1,965)	79.0(77.1-80.8)	78.9(77.0-80.7)	150(182)	82.4(76.1-87.7)	75.1(62.6-87.5)	62(63)	98.4(91.5-100)	96.0(88.6-103.4)

CI: Confidence interval.

NA: Not available due to the low number of screens in this screening interval.

[1] World age-standardized (WHO 2000-2025) rate.