Title

Comparison of cumulative false-positive risk of screening mammography in the United States and Denmark

Journal

Cancer Epidemiology, 39(4)

ISSN

1877-7821

Authors

Jacobsen, Katja Kemp
Abraham, Linn
Buist, Diana SM
et al.

Publication Date

2015-08-01

DOI

10.1016/j.canep.2015.05.004

Peer reviewed

Published in final edited form as:

Cancer Epidemiol. 2015 August ; 39(4): 656–663. doi:10.1016/j.canep.2015.05.004.

Comparison of Cumulative False-Positive Risk of Screening Mammography in the United States and Denmark

K.K. Jacobsen[1,*], **L. Abraham**[2], **D.S.M. Buist**[2], **R. A. Hubbard**[3], **E.S. O'Meara**[2], **B.L. Sprague**[4], **K. Kerlikowske**[5,6], **I. Vejborg**[7], **M. von Euler-Chelpin**[1], and **S. Njor**[1]

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

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

[3]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, United States

[4]Department of Surgery, University of Vermont, Burlington, VT, United States

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

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

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

Abstract

Background—Studies have shown that in the United States (US) about one-half of women screened with annual mammography have 1 false-positive test after ten screens. The estimate for European women screened ten times biennially is much lower. However, these estimates were found in different organizations and used different statistical methods. This study evaluates to what extent screening interval, mammogram type and statistical methods can explain the reported discrepancies between United States and Europe.

Patients and methods—We used data from US Breast Cancer Surveillance Consortium (BCSC), and from two population-based mammography screening programs in Denmark. We included all screens from women first screened at age 50–69 in 1996–2010 in BCSC (1–13 screens/ woman), in 1991–2012 in Copenhagen (1–8 screens/ woman), and in 1993–2013 in Funen (1–10 screens/ woman). Empirical cumulative risks were stratified by screening interval and mammogram type. Model-based cumulative risks were computed for the entire sample using two statistical methods (Hubbard, Njor) previously used to estimate false-positive risks in US and Europe, respectively.

Results—We included 99,455 screens from BCSC, 230,452 from Copenhagen and 400,204 from Funen. Empirical cumulative risk of 1 false-positive test after eight (annual or biennial) screens

[*]Correspondence to: Mrs Katja Kemp Jacobsen, University of Copenhagen, Øster Farimagsgade 5 opg. B, PO Box 2099, Denmark. Tel [+45-35327735]; kkja@sund.ku.dk.

disclosure

The authors have declared no conflicts of interest.

was 41.9% in BCSC, 16.1% in Copenhagen and 7.4% in Funen. Variation in screening interval and mammogram type did not explain the differences in cumulative false-positive risk by country. We only found small differences between model-based and empirical cumulative false-positive risks and between estimates using the two different statistical methods. Using the Hubbard method, model-based cumulative risks after eight screens was 45.1% in BCSC, 9.6% in Copenhagen, and 8.8% in Funen. Using the Njor method these risks were estimated to be 43.6%, 11.2%, 8.0%.

Conclusion—Choice of statistical method, screening interval and mammogram type does not explain the substantial differences in cumulative false-positive risk between US and Europe.

Keywords

breast cancer; false-positive; cumulative risk; statistical methods; screen

introduction

False-positive tests are an unavoidable consequence of mammography screening.

Information on the burden of false-positive tests expected from screening is needed for women in order to make informed decisions about screening participation. From the woman's perspective, it is not only the risk of a false-positive test after attending one screen that is important, but her expected risk of a false-positive test after participating in the multiple rounds of screens called for by a screening program.

Studies from the United States (US), following women with ten years of annual mammography screening, have reported cumulative false-positive risks ranging from 43% to 63% [1–4]. Studies from European mammography screening programs report considerably lower risks, ranging from 8% to 21% after ten biennial screens [5–8]. When comparing estimates of false-positive tests differences in screening organization and choice of statistical methods should be taken into account since these can affect the estimates. Organization of mammography screening differs considerably between the US and Europe. In the US, there are conflicting guidelines [9,10] for screening, so that age at first screen, screening interval and number of screens in a woman's lifetime vary significantly. European screening programs typically offer biennial screening, but also vary in age range, organization and overall program performance [11].

To our knowledge, this is the first study to compare cumulative false-positive risk of mammography screen between the US and Europe using standardized definitions and statistical methods and long-term follow-up. This study had two objectives: to compare empirical cumulative false-positive risk in different settings and to evaluate whether choice of statistical model results in differences in model-based cumulative false-positive risk. To do this, we applied standard definitions and analysis methods to data from the National Cancer Institute-funded Breast Cancer Surveillance Consortium (BCSC) in the US and from the two long-standing, organized, population-based mammography screen programs in Denmark.

materials and methods

The National Cancer Institute-funded Breast Cancer Surveillance Consortium (BCSC, http://breastscreening.cancer.gov/) [12] is a collaborative network of seven regional mammography registries, with catchment representative of the US female population of mammography screening age. The BCSC reflects screening practice in the US and contains data from slightly more than 5% of the female population of screening age [13].

The organized, population-based screening programs in Copenhagen [14] and Funen [14] started in 1991 and 1993, respectively, inviting women aged 50–69 years to biennial screening. Women covered by the two screening programs constitute 20% of Danish women 50–69 years. Unlike in the US, all service, including assessment and treatment for Danish women, is free of charge.

Data from both countries were collected at facilities at the time of screening. In BCSC, breast cancers were obtained by linking mammography data to one or more of three sources: regional Surveillance, Epidemiology, and End Results (SEER) registries, state cancer registries, and pathology databases. Completeness of cancer ascertainment is estimated to be >94.3% [15]. In Copenhagen and Funen, breast cancers were obtained by linking mammography data to the Danish Cancer Registry, the Danish Breast Cancer Cooperative Group, and the Danish Pathology Register. Reporting cancer diagnoses to the Danish Cancer Registry is mandatory by law in Denmark and the registry is almost 100% complete [16].

study population

We included women who had their first screen at age 50–69 years during 1996–2010 in BCSC, 1991–2012 in Copenhagen and 1993–2013 in Funen. This covered 1–13 screens in BCSC, 1–10 screens in Copenhagen and 1–10 screens in Funen. We excluded screens from women with breast implants, a previous mastectomy or diagnosis of invasive breast carcinoma or ductal carcinoma in situ (BCSC: n=31,111 women, Copenhagen: n=3,511 women, Funen: n=3,025). In the Danish data, women with breast implants were only excluded if screening was not technically possible, as data on breast implants were not available.

definitions

According to the BCSC's standard definition, a mammograms was classified as a screening mammogram based on the indication reported by the radiologist [17]. To avoid misclassifying diagnostic mammograms as screening mammograms, we excluded mammograms that were unilateral or obtained within 270 days after a radiological examination. In Denmark, all program mammograms were classified as screening mammogram. Based on the women's screening history, screens were divided into first screens including only the first screen for a given woman, and subsequent screens including all other screens.

To classify a screen as positive or negative, BCSC radiologists use the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) [18]. Screens are coded as BI-RADS 0–5, indicating the level of suspicion of malignancy, and were considered

positive if the initial BI-RADS assessment was 0 (needs additional imaging evaluation), 4 (suspicious abnormality), 5 (highly suggestive of malignancy), or 3 (probably benign finding) when accompanied by a recommendation for immediate evaluation [18], and negative if the initial BI-RADS assessment was 1 (negative), 2 (benign finding), or 3 (probably benign finding) without a recommendation for immediate evaluation.[18] Denmark does not use BI-RADS, therefore all screens that lead to recall for further assessment were referred to as positive screens without further specification.

A false-positive test was defined as a positive screen where no invasive breast carcinoma or ductal carcinoma in situ was diagnosed within one year or prior to the next screen (if this took place before one year).

Subsequent screens were stratified by time since last screen into screening intervals 9–17 months (annual), 18–30 months (biennial) and >30 months (triennial).

statistical analysis

We computed empirical false-positive risks as the proportion of false-positive tests for each number of completed screens. Empirical cumulative risk of 1 false-positive were computed as the proportion of 1 false-positive among women who had completed 1–10 screens in BCSC and Funen, and 1–8 screens in Copenhagen. Proportions were stratified by screening interval (annual or biennial) and mammogram type (film or digital). Data were censored for four reasons: 1) information about time since last screen differed from self-reported information by 6 months (to censor screens in women who were screened outside BCSC); 2) time since last screen was >36 months; 3) BI-RADS assessment or result were missing; 4) when stratifying by screening interval we censored screens where the screening interval differed from previous screen intervals. Similarly, when stratifying data by mammogram type, we censored screens with a different mammogram type compared to previous screens.

Model-based cumulative false-positive risks were estimated using two methods, one developed by Hubbard et al. [4], allowing for variation in false-positive risk among women choosing to attend versus not attend mammography screen, and another method developed by Njor et al. [6] not allowing for this variation. In contrast to the Hubbard method [4], the Njor method [6] assumes independence between screens; meaning, a woman with a false-positive test has the same false-positive risk at the next screen as women without a false-positive test. This assumption only makes sense if radiologists learn from previous mammograms and therefore do not evaluate a new mammogram as suspicious based on findings that have previously caused a false-positive test. The Hubbard method censors women after the first false-positive test. Although the Njor method does not require censoring at first false-positive test, we obtained estimates both with and without censoring at first false-positive test in order to ensure that this did not bias our estimates.

For empirical false-positive risk estimates, 95% confidence intervals (CI) were computed using a normal approximation to the binomial distribution. 95% CI for cumulative risks from the Hubbard and Njor methods are based on 2.5^{th} and 97.5^{th} percentiles of the cumulative risk computed for 1,000 bootstrapped samples. Data were analyzed using SAS (version 9.3 and 9.4) © SAS Institute Inc. and R statistical software (version 3.0.3).

results

Our study included 730,111 screens. Of those, 221,737 were first screens (30.4%). In all three areas, the majority of mammograms were film, but Funen had a smaller rate of film mammograms than BCSC and Copenhagen. Age distribution did not differ much between the three areas. Among first screens at ages 50–69, women in the Danish areas were younger than women from BCSC. The distribution of number of screens varied considerably between the populations, as 78.6% of women in BCSC only had one or two screens compared to 48.5% in Copenhagen and 33.4% in Funen. (Table 1a).

The most frequent screening interval was annual (9–17 months) in BCSC (64.6% of subsequent screens) and biennial (18–30 months) in Danish settings (88.8% of subsequent screens in Copenhagen and 96.4% of subsequent screens in Funen. (Table 1a).

Biennial screens in BCSC had a slightly larger proportion of film mammograms, a younger population, and a larger proportion of women with only one screen compared to biennial screens in the Danish areas. Annual screens in BCSC had a larger proportion of film mammograms, an older population, and a smaller proportion of women with only one screen compared to annual screens in the Danish areas (Table 1b, supplementary appendix).

Of 99,455 screens in BCSC 11,967 were false-positives (12.0%). In Copenhagen 5,347 of 230,452 screens were false-positives (2.3%) and 4,752 of 400,204 screens in Funen were false-positives (1.2%). In all areas false-positive risk at subsequent screens were significantly lower compared to first screens. False positive risk decreased slightly after the second screen In BCSC and Copenhagen; this was not seen in Funen. False-positive risks were significantly higher at all screen rounds in BCSC compared to Copenhagen and Funen (Table 2).

Empirical cumulative false-positive risk after 8 completed screens, the highest number available for comparison, was 41.9% in BCSC, more than twice the estimate for Copenhagen (16.1%) and more than five times that for Funen (7.4%) (Table 2). Stratifying by screening interval only slightly changed cumulative false-positive risks. However we were only able to estimate risks for up to six screens for biennial screens in BCSC and three screens for annual screens in Copenhagen and Funen. Film mammograms had a slightly lower cumulative false-positive risk in BCSC and Funen and a higher cumulative false-positive risk in Copenhagen, compared to digital mammograms (Table 3a–3b).

Model-based versus empirical cumulative false-positive risks differed most for Copenhagen. This is due to a considerably higher proportion of false-positives in the first invitation rounds of the Copenhagen screen program [6]. These keep affecting the empirical cumulative estimates especially for higher number of completed screens where women who participated in the first invitation rounds constitute a high proportion. When excluding first invitation round (1991–93) empirical cumulative false-positive risk for eight screens decreased from 16.1% (95% CI:14.6%–17.6%) to 10.0 % (95%CI: 2.4–17.6%), much closer to the model-based estimate. In BCSC and Funen the opposite trend was seen, probably reflecting a slightly lower proportion of false-positives in early invitation rounds.

When estimating model-based cumulative false-positive risk we only found minor differences between the Hubbard [4] and the Njor methods [6] (Figure 1). Computing Njor [6] estimates without censoring for first false-positive resulted in non-significantly increased estimates for BCSC, while this did not affect estimates from Danish areas (data not shown).

Discussion

Differences in screening interval and mammogram type cannot explain the substantial differences in cumulative false-positive risk between BCSC and Copenhagen or Funen. Empirical cumulative false-positive risk, for women entering screening at age 50–69 years and participating in 8 screens, was considerably higher for BCSC compared to Copenhagen and Funen. Women aged 50–69 years when entering BCSC screening, had a higher proportion of annual screens and film mammograms than women of similar age entering screen in Copenhagen or Funen. Two statistical methods previously used for estimating cumulative false-positive risks in the US and Europe provided fairly similar estimates.

Previous studies have estimated cumulative false-positive risks from the US [1–4,19] and Europe [5–8,20–22]. Elmore et al. [3] studied US women aged 40–69 years at entry and estimated a cumulative false-positive risk of 49.1% after ten annual screens [3]. Hubbard et al. [1] reported an estimate of 61.3% after 10 years of annual screen for women age 50 years at first examination and an estimate of 41.6% for women of the same age screened biennially for ten years. Another US study [19] based on 359 women participating in ten screens reported a substantially lower risk of 29.2% [19]. Hubbard et al. [1] note that their higher estimates may be due to underestimation of false-positive risks in prior studies, which used data only from the sub-group of women who completed all ten screens. Our study and the Hubbard et al. study are based on an overlap of women from the same seven BCSC registries. However, this study includes a larger cohort of women due to the addition of more years of screen. The lower estimate of false-positive risk in our study could be due to a higher specificity in BCSC in recent years, e.g. due to a higher proportion of comparison mammograms available. Furthermore, our estimate is based on both annual and biennial screens. Studies from Europe that include women from screen organizations similar to Copenhagen and Funen reported cumulative false-positive risks in the range 7.3%–20.4%.

Comparison of false-positives between studies is challenging due to a wide-variety of factors that differ between countries, e.g. underlying risk, technology, program organization, screening interval and ages. Prior studies have found that risk is highest in younger women [1], biennial screen tends to lower cumulative false-positive risks after 10 years [1,23], and the absence of previous mammograms tends to increase the risk [1,24,25]. Moreover, studies have demonstrated considerable variation in false-positive risks depending on individual risk factors [2,26].

Our comparisons were based on populations of similar age with no history of breast cancer, similar definitions and similar methods for estimation of false-positive risks. Another strength is the large study population, with longitudinal follow-up. This allowed us to obtain empirical false-positive estimates and compare those with model-based estimates over a

period of screens similar to that called for in some screening recommendations and used by most European mammography screening programs.

However, organization and delivery of screening in the US and Denmark differ considerably. Stratifying by annual versus biennial screening interval and digital versus film mammography did not explain the differences in cumulative risks between BCSC and Denmark. Importantly, these comparisons are based on relatively small sample sizes, especially in the Danish areas. Comparison of women who choose different screening patterns might be confounded, as e.g. women who have a family history of breast cancer and women who have heterogeneously dense breasts have a higher false positive risk [1], and might choose annual screening more often than other women.

Differences in populations being studied might remain, as we were not able to adjust for hormone use, breast density, family history of breast cancer and race, which might influence false-positive risks [27–31]. We have no reason to believe that breast density and family history of breast cancer differ between the US and Denmark. However, during the study period hormone therapy use was considerably higher among American compared to Danish women [32]. False-positive risk in BCSC could be affected by the fact that mammography screen and follow-up is not free in the US, and many women are not covered by insurance. This might lower the number of women with a positive screen who return for the recommended diagnostic evaluation in BCSC, leading to misclassification of true-positive as false-positive tests. However, given the relative frequency of cancer diagnoses compared to false-positive tests, this probably only had a minor influence on our results. In the Danish data, the number of women not receiving recommended follow-up is very small.

The majority of US participants start screen before age 50 years, while we only included women who started screen after age 50 in this study. These groups of women might differ according to false-positive risks. However, Hubbard et al. [4] found that , in the US women, starting screen at 50 years was only associated with a higher false-positive risk at first screen compared to starting at age 40 years. Overall estimates for starting at age 40 or 50 years did not differ. Another factor that might influence estimates is that only a minor subset of the population contributed to the higher number of completed screens, especially in the BCSC population. Estimates of cumulative false-positive risk were fairly similar using the Hubbard method and the Njor method. The Hubbard method allows for an association between false positive risk and censoring time due to external factors (e.g. family history of breast cancer). In a study from Virginia, US, Wilson et al. [33] found that daughters of mothers with breast cancer were significantly more likely to have had mammography screen compared to women from a more general population. This indicates that there is an association between false positive risk and censoring time due to family history of breast cancer in US. Whether there is a similar association in Denmark is unknown. The Hubbard method [4] assumes conditional independence between screens, while the Njor method [6] assumes unconditional independence between screens. It has previously been shown by Njor et al. [6], that it is reasonable to assume independence between subsequent screens in the Danish programs. It might be that independence between screens is a more acceptable assumption in Europe, where the use of previous mammograms for comparison is more frequent compared

to the US. Nevertheless, in this study we found little variation in risk estimates regardless of the method used.

In conclusion, we found that the considerable difference in false-positive risk between the US and Europe reported in prior studies is not due to choice of statistical method, difference in screening interval or mammogram type. Furthermore, we demonstrated that both statistical models appear appropriate for these cohorts, as model-based and empirical estimates were highly similar.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Cancer Institute-funded grants (P01CA154292, R03CA182986, U54CA163303) and the Breast Cancer Surveillance Consortium (HHSN261201100031C). The collection of cancer data was supported in part by several US cancer registries (http://breastscreen.cancer.gov/work/acknowledgement.html). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or National Institutes of Health. We thank the participating women, mammography facilities, and radiologists for the data they have provided. A list of BCSC investigators and procedures for requesting BCSC data for research purposes are provided at: http://breastscreening.cancer.gov/.

References

1. Hubbard RA, Kerlikowske K, Flowers CI, et al. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Ann Intern Med. 2011; 155:481–92. [PubMed: 22007042]
2. Christiansen CL, Wang F, Barton MB, et al. Predicting the cumulative risk of false-positive mammograms. J Natl Cancer Inst. 2000; 92:1657–66. [PubMed: 11036111]
3. Elmore JG, Barton MB, Moceri VM, et al. Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med. 1998; 338:1089–96. [PubMed: 9545356]
4. Hubbard RA, Miglioretti DL, Smith RA. Modelling the cumulative risk of a false-positive screening test. Stat Methods Med Res. 2010; 19:429–49. [PubMed: 20356857]
5. Hofvind S, Thoresen S, Tretli S. The cumulative risk of a false-positive recall in the Norwegian Breast Cancer Screening Program. Cancer. 2004; 101:1501–7. [PubMed: 15378474]
6. Njor SH, Olsen AH, Schwartz W, et al. Predicting the risk of a false-positive test for women following a mammography screening programme. J Med Screen. 2007; 14:94–7. [PubMed: 17626709]
7. Puliti D, Miccinesi G, Zappa M. More on screening mammography. N Engl J Med. 2011; 364:284–5. [PubMed: 21247319]
8. Salas D, Ibanez J, Roman R, et al. Effect of start age of breast cancer screening mammography on the risk of false-positive results. Prev Med. 2011; 53:76–81. [PubMed: 21575653]
9. 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]
10. Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. CA Cancer J Clin. 2007; 57:75–89. [PubMed: 17392385]
11. Giordano L, von KL, Tomatis M, et al. Mammographic screening programmes in Europe: organization, coverage and participation. J Med Screen. 2012; 19(Suppl 1):72–82. [PubMed: 22972813]
12. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol. 1997; 169:1001–8. [PubMed: 9308451]

13. Yankaskas BC, Klabunde CN, Ancelle-Park R, et al. International comparison of performance measures for screening mammography: can it be done? J Med Screen. 2004; 11:187–93. [PubMed: 15624239]
14. Jacobsen KK, Euler-Chelpin MV. Performance indicators for participation in organized mammography screening. J Public Health (Oxf). 2012
15. Ernster VL, Ballard-Barbash R, Barlow WE, et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. J Natl Cancer Inst. 2002; 94:1546–54. [PubMed: 12381707]
16. Sundhedsstyrrelsen. Det moderniserede Cancerregister-metode og kvalitet. 2014.
17. Breast Cancer surveillance Consortium. BCSC Glossery of Terms 2009. Jan 5.2012
18. American college of Radiology. Breast Imaging and Data System (BIRADS). 2003.
19. Blanchard K, Colbert JA, Kopans DB, et al. Long-term risk of false-positive screening results and subsequent biopsy as a function of mammography use. Radiology. 2006; 240:335–42. [PubMed: 16864665]
20. Hofvind S, Ponti A, Patnick J, et al. False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. J Med Screen. 2012; 19(Suppl 1):57–66. [PubMed: 22972811]
21. Otten JD, Fracheboud J, den Heeten GJ, et al. Likelihood of early detection of breast cancer in relation to false-positive risk in life-time mammographic screening: population-based cohort study. Ann Oncol. 2013; 24:2501–6. [PubMed: 23788759]
22. Castells X, Molins E, Macia F. Cumulative false positive recall rate and association with participant related factors in a population based breast cancer screening programme. J Epidemiol Community Health. 2006; 60:316–21. [PubMed: 16537348]
23. Braithwaite D, Zhu W, Hubbard RA, et al. Screening outcomes in older US women undergoing multiple mammograms in community practice: does interval, age, or comorbidity score affect tumor characteristics or false positive rates? J Natl Cancer Inst. 2013; 105:334–41. [PubMed: 23385442]
24. Elmore JG, Miglioretti DL, Reisch LM, et al. Screening mammograms by community radiologists: variability in false-positive rates. J Natl Cancer Inst. 2002; 94:1373–80. [PubMed: 12237283]
25. Esserman L, Cowley H, Eberle C, et al. Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst. 2002; 94:369–75. [PubMed: 11880475]
26. Roman R, Sala M, Salas D, et al. Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. Ann Oncol. 2012; 23:104–11. [PubMed: 21430183]
27. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology. 1992; 184:613–7. [PubMed: 1509041]
28. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med. 2003; 138:168–75. [PubMed: 12558355]
29. Holland R, Mravunac M, Hendriks JH, et al. So-called interval cancers of the breast. Pathologic and radiologic analysis of sixty-four cases. Cancer. 1982; 49:2527–33. [PubMed: 7074569]
30. Njor SH, Hallas J, Schwartz W, et al. Type of hormone therapy and risk of misclassification at mammography screening. Menopause. 2011; 18:171–7. [PubMed: 20729766]
31. O'Meara ES, Zhu W, Hubbard RA, et al. Mammographic screening interval in relation to tumor characteristics and false-positive risk by race/ethnicity and age. Cancer. 2013; 119:3959–67. [PubMed: 24037812]
32. 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]
33. Wilson DB, Quillin J, Bodurtha JN, et al. Comparing screening and preventive health behaviors in two study populations: daughters of mothers with breast cancer and women responding to the behavioral risk factor surveillance system survey. J Womens Health (Larchmt). 2011; 20:1201–6. [PubMed: 21671767]

Key message

Cumulative false-positive risk during eight (annual or biennial) mammography screens is more than three times higher in the United States compared to Denmark. Choice of statistical method, different screening intervals and mammogram type does not explain the substantial difference in cumulative false-positive risk between the United States and Denmark.

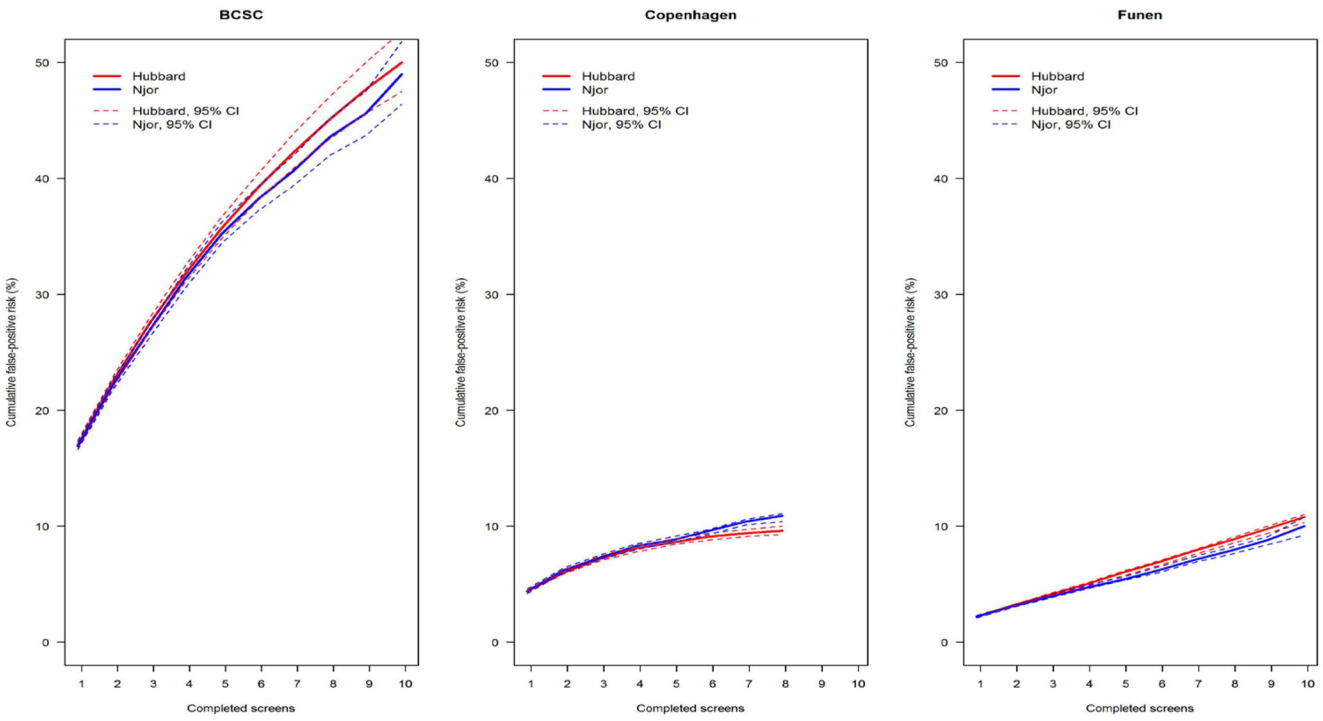

Figure 1.

Model-based estimates for cumulative false-positive risks (%) by screen number namong women age 50–69 years at their first screen in the Breast Cancer Surveillance Consortium (BCSC), United States (US), Copenhagen and Funen, Denmark.

The following censoring mechanisms are used: 1) if information about time since last mammogram from database differs from self-report information; 2) if time since last mammogram >36 months; 3) if the mammogram had a missing BI-RADS assessment or result; 4) at the first false-positive test.

The Hubbard method is adjusted for mammography registry.

95% CI: Confidence interval are based on the 2.5th and 97.5th percentiles of cumulative risk computed from1,000 bootstrapped samples.

Table 1a

Characteristics of all screens among women aged 50–69 years in the Breast Cancer Surveillance Consortium (BCSC) 1996–2010, United States (US) and Copenhagen 1991–2012 and Funen 1993–2013, Denmark, supplementary Table 1b, available online

	BCSC, US		Copenhagen, Denmark		Funen, Denmark	
	All screens (% of stratum)	First screens (% of stratum)	All screens (% of stratum)	First screens (% of stratum)	All screens (% of stratum)	First screens (% of stratum)
Total number of women	51,477	51,477	75,455	75,455	94,805	94,805
Total number of screens	99,455	51,477	230,452	75,455	400,204	94,805
Film mammograms	85,031 (85.5)	45,499 (88.4)	188,612 (81.8)	65,967 (87.4)	273,635 (68.4)	75,067 (79.2)
Digital mammograms	11,090 (11.2)	4,040 (7.8)	41,840 (18.2)	9,488 (12.6)	126,569 (31.6)	19.738 (20.8)
Missing type	3,334 (3.4)	1,938 (3.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Age at the screen						
50–54 years	30,819 (31.0)	21,495 (41.8)	71,228 (30.9)	44,449 (58.9)	126,140 (31.5)	63,758 (67.3)
55–59 years	27,123 (27.3)	11,766 (22.9)	67,749(29.4)	11,805 (15.7)	108,447 (27.1)	12,442 (13.1)
60–64 years	21,318 (21.4)	9,087 (17.7)	51,522 (22.4)	9,784 (13.0)	92,354 (23.1)	10,840 (11.4)
65–69 years	20,195 (20.3)	9,129 (17.7)	39,953(17.3)	9,417 (12.5)	73,263 (18.3)	7,765 (8.2)
Number of screens per women						
1	31,697 (61.6)		23,296 (30.9)		17,182 (18.2)	
2	8,774 (17.0)		13,270 (17.6)		14,363 (15.2)	
3	4,281 (8.3)		12,636 (16.8)		12,519 (13.2)	
4	2,481 (4.8)		7,274 (9.6)		10,363 (10.9)	
5	1,606 (3.1)		7,653(10.1)		9,298 (9.8)	
6	982 (1.9)		6,152 (8.2)		9,072 (9.6)	
7	706 (1.4)		2,957 (3.9)		8,630 (9.1)	
8	415 (0.8)		2,217 (2.9)		7,321 (7.7)	
9	273 (0.5)				5,183 (5.5)	
10	138 (0.3)				874 (0.9)	
>10	124 (0.2)					

	BCSC, US		Copenhagen, Denmark		Funen, Denmark	
	All screens (% of stratum)	First screens (% of stratum)	All screens (% of stratum)	First screens (% of stratum)	All screens (% of stratum)	First screens (% of stratum)
Screen interval[a]						
9–17 months	30,989 (64.6))		3,537 (2.3)		2,912 (1.0)	
18–30 months	14,456 (30.1)		137,709 (88.8)		294,268 (96.4))	
30+ months	2,533 (5.3)		13,751 (8.9)		8,219 (2.7)	

The following censoring mechanisms are used: 1) if information about time since last mammogram from database differs from self-report information; 2) if time since last mammogram >36 months; 3) if the mammogram had a missing BI-RADS assessment or result.

[a]First screens are excluded.

Table 2a

Proportion of false-positive tests and cumulative false-positive risks by number of completed screens, for women aged 50–69 years in the Breast Cancer Surveillance Consortium (BCSC), United States (US) and Copenhagen and Funen, Denmark

Completed screens	False-positive tests		First false-positive tests		Cumulative false-positive tests	
	Numerator (denominator)	Proportion of false-positive tests (95%CI)	Numerator (denominator)	Proportion of first false-positive tests (95%CI)	Numerator (denominator)	Cumulative false-positive risk (95%CI)
BCSC, US, 1996–2010						
1	8,714(51,477)	16.9(16.6–17.3)	8,714(51,477)	16.9(16.6–17.3)	8,714(51,477)	16.9(16.6–17.3)
2	1,424(19,780)	7.2(6.8–7.6)	1,068(16,368)	6.5(6.2–6.9)	4,480(19,780)	22.6(22.1–23.2)
3	736(11,006)	6.7(6.2–7.2)	493(8,501)	5.8(5.3–6.3)	2,998(11,006)	27.2(26.4–28.1)
4	452(6,725)	6.7(6.1–7.3)	303(4,912)	6.2(5.5–6.9)	2,116(6,725)	31.5(30.4–32.6)
5	283(4,244)	6.7(5.9–7.4)	164(2,906)	5.6(4.8–6.5)	1,502(4,244)	35.4(34.0–36.8)
6	158(2,638)	6.0(5.1–6.9)	78(1,730)	4.5(3.6–5.6)	986(2,638)	37.4(35.5–39.2)
7	87(1,656)	5.3(4.2–6.3)	43(1062)	4.0(2.9–5.4)	637(1,656)	38.5(36.1–40.8)
8	56(950)	5.9(4.4–7.4)	29(581)	5.0(3.4–7.1)	398(950)	41.9(38.8–45.0)
9	29(535)	5.4(3.5–7.3)	11(320)	3.4(1.7–6.1)	226(535)	42.2(38.1–46.4)
10	19(262)	7.3(4.4–10.4)	10(160)	6.3(3.0–11.2)	112(262)	42.7(36.8–48.7)
>10	9(182)	4.9(1.8–8.1)	5(113)	4.4(1.5–10.0)	74(182)	40.7(33.5–47.8)
Total	11,967(99,455)	12.0(11.8–12.2)	10,918(88,130)	12.4(12.2–12.6)		
Copenhagen, Denmark 1991–2012						
1	3,319(75,455)	4.4(4.3–4.6)	3,319(75,455)	4.4(4.3–4.6)	3,319(75,455)	4.4(4.3–4.6)
2	1,001(52,159)	1.9(1.8–2.0)	951(49,784)	1.9(1.1–1.3)	3,375(52,159)	6.5(6.3–6.7)
3	461(38,889)	1.2(1.1–1.3)	425(36,301)	1.2(1.1–1.3)	3,084(38,889)	7.9(7.7–8.2)
4	289(26,253)	1.1(1.0–1.2)	253(24,036)	1.1(0.9–1.2)	2,555(26,253)	9.7(9.4–10.1)
5	121(18,979)	0.6(0.5–0.8)	102(17,108)	0.6(0.5–0.7)	2,060(18,979)	10.9(10.4–11.3)
6	91(11,326)	0.8(0.6–1.0)	79(10,073)	0.8(0.6–1.0)	1,407(11,326)	12.4(11.8–13.0)
7	49(5,174)	0.9(0.7–1.2)	40(4,510)	0.9(0.6–1.2)	755(5,174)	14.6(13.6–15.6)
8	16(2,217)	0.7(0.4–1.1)	11(1,899)	0.6(0.2–0.9)	357(2,217)	16.1(14.6–17.6)
Total	5,347(230,452)	2.3(2.3–2.4)	5,180(219,166)	2.4 (2.3–2.4)		
Funen, Denmark 1993–2013						

Completed screens	False-positive tests		First false-positive tests		Cumulative false-positive tests	
	Numerator (denominator)	Proportion of false-positive tests (95%CI)	Numerator (denominator)	Proportion of first false-positive tests (95%CI)	Numerator (denominator)	Cumulative false-positive risk (95%CI)
1	2,122(94,805)	2.2(2.1–2.3)	2,122(94,805)	2.2(2.1–2.3)	2,122(94,805)	2.2(2.1–2.3)
2	709(77,623)	0.9(0.8–1.0)	684(76,045)	0.9(0.8–1.0)	2,287(77,623)	2.9(2.8–3.1)
3	548(63,260)	0.9(0.8–0.9)	528(61,469)	0.9(0.8–0.9)	2,358(63,260)	3.7(3.6–3.9)
4	401(50,741)	0.8(0.7–0.9)	393(48,935)	0.8(0.7–0.9)	2,234(50,741)	4.5(4.3–4.6)
5	309(40,378)	0.8(0.7–0.9)	291(38,635)	0.8(0.7–0.8)	2,074(40,378)	5.1(4.9–5.4)
6	259(31,080)	0.8(0.7–0.9)	239(29,539)	0.8(0.7–0.9)	1,824(31,080)	5.9(5.6–6.1)
7	222(22,008)	1.0(0.7–1.2)	204(20,780)	1.0(0.8–1.1)	1,471(22,008)	6.7(6.4–7.0)
8	112(13,378)	0.8(0.7–1.0)	101(12,517)	0.8(0.7–1.0)	996(13,378)	7.4(7.0–7.9)
9	59(6,057)	1.0(0.7–1.2)	54(5,659)	1.0(0.7–1.2)	471(6,057)	7.8(7.1–8.5)
10	11(874)	1.3(0.5–2.0)	11(822)	1.3(0.6–2.1)	64(874)	7.3(5.6–9.0)
Total	4,752(400,204)	1.2(1.2–1.2)	4,627(389,206)	1.2(1.2–1.2)		

The following censoring mechanisms are used: 1) if information about time since last mammogram from database differs from self-report information; 2) if time since last mammogram >36 months; 3) if the mammogram had a missing BI-RADS assessment or result.

CI: confidence interval.

Table 3a

Cumulative false-positive tests by number of completed screes, stratified by annual (9–17 months) and biennial (18–30 months) screen interval in the Breast Cancer Surveillance Consortium (BCSC), United States (US), Copenhagen and Funen, Denmark

	BCSC, US Cumulative false-positive risk (95%CI)		Copenhagen, Denmark Cumulative false-positive risk (95%CI)		Funen, Denmark Cumulative false-positive risk (95%CI)	
Completed screens	**Annual**	**Biennial**	**Annual**	**Biennial**	**Annual**	**Biennial**
1[a]	16.9 (16.6–17.3)	16.9 (16.6–17.3)	4.4 (4.3–4.6)	4.4 (4.3–4.6)	2.2 (2.1–2.3)	2.2 (2.1–2.3)
2	24.2 (23.4–25.0)	20.7 (19.7–21.6)	3.6 (2.8–4.5)	6.7 (6.5–7.0)	3.1 (2.1–4.1)	2.9 (2.8–3.1)
3	27.8 (26.5–29.0)	25.1 (23.0–27.2)	7.4 (2.1–12.8)	8.3 (8.0–8.6)	14.3 (–11.6–40.2)	3.7 (3.5–3.8)
4	31.9 (30.2–33.7)	30.7 (26.8–34.6)	16.7 (–13.2–46.2)	10.5 (10.1–11.0)		4.3 (4.2–4.5)
5	36.4 (34.1–38.8)	31.7 (25.9–37.5)		12.3 (11.7–12.9)		5.1 (4.8–5.3)
6	38.7 (35.8–41.6)	33.3 (24.7–42.0)		14.2 (13.4–15.2)		5.8 (5.5–6.1)
7	41.2 (37.5–44.9)	30.0 (15.8–44.2)		14.4 (12.7–16.2)		6.7 (6.3–7.0)
8	43.3 (38.6–48.1)	28.6 (0.0–62.0)		16.5 (13.2–19.9)		7.4 (7.0–8.5)
9	41.2 (35.1–47.4)					7.8 (7.0–8.5)
10	43.6 (35.4–51.8)					6.7 (4.6–8.7)
>10	41.4 (32.9–49.9)					

The following censoring mechanisms are used: 1) if information about time since last mammogram from database differs from self-report information; 2) if time since last mammogram >36 months; 3) if the mammogram had a missing BI-RADS assessment or result; 4) if the screen has a different screen interval compared to interval reported at second screen.

CI: Confidence interval.

[a]To compute the false-positive risk for the first screen, all first screens were included.

Table 3b

Cumulative false-positive tests by number of completed screens among women aged 50–69 years at the first screen, stratified by digital and film mammograms in the Breast Cancer Surveillance Consortium (BCSC), United States (US) and Copenhagen and Funen, Denmark

	BCSC, US Cumulative false-positive risk (95%CI)		Copenhagen, Denmark Cumulative false-positive risk (95%CI)		Funen, Denmark Cumulative false-positive risk (95%CI)	
Completed screens	**Film mammograms**	**Digital mammograms**	**Film mammograms**	**Digital mammograms**	**Film mammograms**	**Digital mammograms**
1	16.3 (16.0, 16.7)	21.6 (20.4, 22.9)	4.8 (4.6–5.0)	1.6 (1.4–1.9)	2.0 (1.9–2.1)	3.3(3.0–3.5)
2	22.0 (21.4, 22.7)	26.2 (23.3, 29.1)	7.4 (7.1–7.6)	2.8 (2.3–3.3)	2.8 (2.7–2.9)	4.0(4.2–5.3)
3	26.7 (25.8, 27.6)	30.2 (24.8, 35.5)	9.2 (8.9–9.5)	6.5 (3.3–9.7)	3.6 (3.4–3.8)	4.8(4.2–5.3)
4	30.9 (29.6, 32.1)	31.7 (22.8, 40.7)	10.8 (10.4–11.2)		4.3 (4.1–4.5)	5.1(1.7–8.6)
5	34.5 (32.9, 36.2)	40.0 (24.8, 55.2)	12.1 (11.6–12.6)		5.1 (4.8–5.3)	
6	36.2 (34.0, 38.4)	57.1 (31.2–83.1)	13.4 (12.7–14.1)		5.5 (5.2–5.8)	
7	37.5 (34.5, 40.4)	50.0 (1.0–99.0)	14.6 (13.6–15.5)		6.1 (5.4–6.9)	
8	38.1 (34.0, 42.2)		16.2 (14.7–17.8)			
9	38.9 (33.1, 44.7)					
10	39.6 (30.3, 48.9)					
>10	34.5 (22.3–46.7)					

The following censoring mechanisms are used: 1) if information about time since last mammogram from database differs from self-report information; 2) if time since last mammogram >36 months; 3) if the mammogram had a missing BI-RADS assessment or result; 4) if the screen has another mammogram type compared to the mammogram type reported at the previous screen.

95% CI: 95% Confidence interval.