

Predicting the risk of a false-positive test for women following a mammography screening programme

Sisse Helle Njor, Anne Helene Olsen, Walter Schwartz, Ilse Vejborg and Elsebeth Lynge

J Med Screen 2007;14:94-97

See end of article for authors' affiliations

Correspondence to:
Sisse H Njor, Institute of Public Health, University of Copenhagen, Øster Farimagsgade 5, opg. B, DK-1014 Copenhagen K, Denmark;
S.Njor@pubhealth.ku.dk

Accepted for publication
26 February 2007

Objectives The objectives of this study was to provide a simple estimate of the cumulative risk of a false-positive test for women participating in mammography screening. To test the method, we used data from two well-established, organized mammography screening programmes offering biennial screening to women aged 50-69 years in Copenhagen and Fyn, Denmark.

Methods We defined the outcome from a screen as being either a false-positive test or not a false-positive test. We then tested whether the outcomes from subsequent screens were independent, and afterwards estimated the risk over 10 screens of a false-positive test, i.e. the risk of getting at least one false-positive test for a woman participating in all 10 screens typically offered in Europe.

Results The outcomes of subsequent screens were found to be independent. After completion of screening rounds 3-5, the risk of a false-positive test over 10 screens was predicted to be 15.8-21.5% for a woman participating in the programme in Copenhagen, and 8.1-9.6% for a woman participating in the programme in Fyn.

Conclusions Our study showed that a relatively robust prediction of the risk of a false-positive test over 10 screens can be calculated in a simple way relatively early after the start of a mammography screening programme.

INTRODUCTION

Screening is testing of healthy people for a potential unrecognized disease, and the screening test necessarily has to be quick and cheap. A screening programme will therefore always imply that some participants test positive, although they turn out to be disease free after thorough assessment. These tests are called false-positive tests, and they constitute one of the inevitable negative side-effects of screening.

In mammography service screening programmes, the proportion of false-positive tests at first screen has varied from 0.7% in the Netherlands to 17.5% in Navarra, Spain,^{1,2} reflecting differences in screening policies, where the use of extra resources for assessment and the risk of false-positive tests are weighted against the risk of leaving undetected, false-negative cases in the screened population.

However, the false-positive tests have to be also seen from the perspective of the individual woman. Her perspective is not that of the single invitation round, but one of participation in the entire programme. In Europe, this typically means being screened every second year from the age of 50 years to the age of 70 years, in total 10 times.

Although the burden of false-positive tests expected from participating in a screening programme forms part of the background information needed for women to decide on participation, only a few studies have addressed this problem.^{3,4,5,6} In studies from health centres in Boston, USA, the cumulative risk of a false-positive result of a breast cancer screening test was estimated based on conditional probabilities.^{3,4,5} A study from the organized screening programme in Norway calculated the cumulative risk by assuming independence and limiting the number of possible outcomes considered.⁶ It would in fact be much easier to calculate the cumulative risk of a false-positive test if a

woman's risk of getting a false-positive test at a given screen did not depend on the outcome of her previous screens. We therefore tested the independence hypothesis based on data from the first five rounds of two organized mammography screening programmes in Denmark. Furthermore, we considered whether the independence hypothesis seems reasonable in view of the radiologists' practice of comparing new and old mammograms. Finally, we tested the robustness of the cumulative risk estimated after completion of an increasing number of screens, as it is desirable early on in a screening programme to be able to provide targeted women with an estimate of the risk of experiencing a false-positive test.

MATERIAL AND METHODS

There are two well-established, organized mammography screening programmes in Denmark. The first programme started in the municipality of Copenhagen in April 1991, and the small municipality of Frederiksberg was added in June 1994. The second programme started in the county of Fyn in November 1993. Both programmes are population based and offer biennial screening to women aged 50-69 years. Two-view mammography is standard for the prevalence screen.^{7,8} Women with dense breast tissue, about 60% of all participants, have two-view mammography at subsequent screens, while the rest have one-view mammography.⁷ Independent double reading is used for all screening mammograms, and new mammograms at subsequent screens are compared with those taken previously. During the first invitation round of the Copenhagen programme, 6.8% of the screened women were recalled for assessment. When the Fyn programme started about two-and-a-half years later, it was decided to aim for a lower recall rate,

which ended up being 2.7% during the first invitation round. The false-positive rates have consequently varied between the two programmes, being 5.6% during the first invitation round in Copenhagen and 1.8% during the first invitation round in Fyn. Both programmes operate with fixed invitation rounds of two years during which all women in the target age groups are invited.

We defined the outcome of a screen as being either a false-positive test or not a false-positive test. A screening test was defined as false-positive if a woman was referred to assessment without being diagnosed with either invasive breast cancer or ductal carcinoma *in situ*. We then examined whether there was independence between the outcomes of subsequent screens. Namely, was the probability of getting a false-positive test influenced by the fact that a woman did have or did not have a false-positive test at the previous screen?

The number of women with 10 screens and at least one false-positive test follows a binomial distribution with parameter p . If the outcomes of subsequent screens are independent, then this parameter can be estimated by $p_{10}^* = 1 - [(1-p_1)(1-p_2)...(1-p_{10})]$, where p_1 to p_{10} are the number of false-positive tests divided by the number of participants at screen numbers one to 10. To test our hypothesis of independence, we compared the observed proportion of women with at least one false-positive test in the first five screens with $p_5^* = 1 - [(1-p_1)(1-p_2)(1-p_3)(1-p_4)(1-p_5)]$. To test the robustness of our estimate, we calculated cumulative risks for at least one false-positive test based on the outcome of an increasing number of completed screens, assuming p_i for future screens to be equal to that of the last completed screen.

However, not all screening programmes will have sufficiently detailed data to calculate the p_i 's by screen number. Normally, the false-positive rate will be known only for all women participating in a given invitation round. Therefore, we also calculated cumulative risks for at least one false-positive test based on the outcome of an increasing number of completed invitation rounds, where the p_i 's were calculated based on all participating women aged 50–69 years in each invitation round, and assuming p_i for future invitation rounds to be equal to that of the last completed invitation round.

RESULTS

The proportion of women with a false-positive mammogram decreased over screen number. For Copenhagen, the proportion decreased regularly from 5.7% at first screen to 0.9% at fifth screen. For Fyn, the proportion of women with a false-positive test was only 1.8% at first screen, leaving less room for a later decline. This may also explain why the decline was less regular in Fyn, where the proportions were 0.8% at second, 0.9% at third, 0.8% at fourth and 0.7% at fifth screen (Table 1).

At least one false-positive screen was experienced by 1041 of the 21,261 women participating in all five invitation rounds in Fyn, equal to 4.9% (95% confidence interval 4.6–5.2%). Assuming independence between the outcomes of subsequent screens, the probability of getting at least one false-positive test for a woman participating in all five invitation rounds in Fyn was 4.9%; actual was 4.87% and thus different from the observed 4.90%. As 4.9% is within the 95% confidence interval of the observed 4.9%, the hypothesis of independence between subsequent screens

was accepted ($P=0.84$). Similarly, for Copenhagen, at least one false-positive screen was experienced by 1151 of the 9039 women participating in all five invitation rounds, equal to 12.7% (95% confidence interval 12.1–13.4%). Assuming independence between the outcomes of subsequent screens, the probability of getting at least one false-positive test for a woman participating in all five invitation rounds was 12.9%, and the hypothesis of independence was consequently accepted also for Copenhagen ($P=0.71$). In Fyn, the estimated proportion of women with at least one false-positive mammogram was below the observed proportion, whereas the opposite was true for Copenhagen, thus strengthening our hypothesis on independence.

Based on the false-positive risk observed during the first screen in Fyn, we predicted the risk of a false-positive test over 10 screens for women participating in the mammography screening programme in Fyn to be 16.6% (Table 2). By also including the observed false-positive risk experienced during the second screen, the estimated risk over 10 screens was 8.5%; including the third screen gave 9.6%; including the fourth screen gave 8.6% and including the fifth screen gave 8.1%. Based on the false-positive risk observed during the first screen in Copenhagen, the predicted risk of a false-positive test over 10 screens for women participating in the mammography screening programme in Copenhagen was 43.8%. When the observed risk during the second screen was also included, the predicted risk over 10 screens was 30.2%; including the

Table 1 Proportion of participants with false-positive test at the first five screens in two organized mammography screening programmes in Denmark

Invitation round, screen number*	Fyn 1993–2003		Copenhagen 1991–2001	
	(%)	95% CI	(%)	95% CI
1,1	1.84	(1.68–2.00)	5.66	(5.32–6.00)
2,2	0.79	(0.68–0.90)	3.26	(2.86–3.44)
3,3	0.93	(0.81–1.05)	1.88	(1.61–2.09)
4,4	0.75	(0.64–0.86)	1.78	(1.53–2.17)
5,5	0.65	(0.54–0.76)	0.93	(0.66–1.40)

*Women screened for the first time in the first invitation round; screened for the second time in the second invitation round; screened for the third time in the third invitation round; screened for the fourth time in the fourth invitation round and screened for the fifth time in the fifth invitation round. The calculation therefore includes only women aged 50–61 years at first invitation; 52–63 years at second invitation; 54–65 years at third invitation; 56–67 years at fourth invitation and 58–69 years at fifth invitation

Table 2 Cumulative risk over 10 screens of a false-positive result estimated after completion of an increasing number of screens* in two organized mammography screening programmes in Denmark

Number of completed screens	Fyn 1993–2003 (%)	Copenhagen 1991–2001 (%)
1	16.6	43.8
2	8.5	30.2
3	9.6	21.5
4	8.6	20.7
5	8.1	15.8

*With one completed screen, the expected value at screen numbers 2–10 are set to the observed value at first screen. With two completed screens, the expected value at screen numbers 3–10 are set to the observed value at second screen. With three completed screens, the expected value at screen numbers 4–10 are set to the observed value at third screen. With four completed screens, the expected value at screen numbers 5–10 are set to the observed value at fourth screen. With five completed screens, the expected value at screen numbers 6–10 are set to the observed value at fifth screen

third screen gave 21.5%; including the fourth screen gave 20.7% and including the fifth screen gave 15.8%.

Both screen number and invitation round are expected to influence the risk of a false-positive mammogram. The false-positive risk is expected to be particularly high at the first screen, where the radiologist has no previous mammograms to check with. The false-positive risk could also be particularly high during the first invitation round, where all personal groups are inexperienced. In both Copenhagen and Fyn, the screen number clearly influenced the risk of a false-positive mammogram, but the number of the invitation round had a limited impact on the false-positive risk in Copenhagen and no systematic impact on the false-positive risk in Fyn (Table 3). Therefore, the estimated cumulative risk of a false-positive mammogram is expected to be almost the same regardless of whether the data came from the first years after the start of a screening programme, or from a later point in time.

Using the crude proportions of participants with a false-positive test from each of the first five invitation rounds, the predicted risk of at least one false-positive test over 10 invitation rounds for women participating in a mammography screening programme in Fyn became 9.9% and 22.6% for women participating in the mammography screening programme in Copenhagen.

DISCUSSION

Our results showed that the false-positive rate depended strongly on the screen number, being considerably higher for the first and perhaps for the second screen than for subsequent screens. This means that the cumulative risk over 10 screens can only be predicted after completion of the third screening round, unless data already collected in other screening programmes can be used. Our study showed that we could assume independence between the outcomes of subsequent screens. This was found for two screening programmes with very different percentages of false-positive mammograms, illustrating the robustness of the finding. It is therefore possible to predict the risk of a false-positive test over 10 screens as a simple function of the outcome of each screening round, and a relatively robust estimate can be obtained after the first three rounds. After the fifth screen, the predicted risk of a false-positive test during participation in the entire programme was 8% for Fyn and 16% for

Copenhagen. As opportunistic screening is rare in Denmark, the risk of a false-positive test over the 10 screens offered in the organized screening programme equals the lifetime risk of a false-positive test.

Using the crude proportions of all 50–69-year-old women with a false-positive test in each invitation round as the basis for calculating the cumulative risk led to an overestimation of the risk. The estimate for Fyn changed from 8% to 10%, and the estimate for Copenhagen changed from 16% to 23%. This overestimation is due to the contamination in later invitation rounds from newcomers with prevalent screens. Wherever possible, it is therefore important to base estimates on data specified by screen number.

Törnberg *et al.*⁹ found that only 53% of women invited to five consecutive screens in the Stockholm screening programme accepted all five. Women accepting all invitations could therefore constitute a selected group. In this case, the cumulative risk over 10 rounds of a false-positive mammogram calculated for participating women might not be a valid estimate of the risk for women considering participation. However, our results showed that the cumulative risk could be estimated from the independent outcomes of the individual screens, and women accepting all invitations were therefore not a selected group concerning the false-positive risk. This is also illustrated by the fact that our estimation methods worked both for Fyn, where 76% of targeted women were faithful users, and in Copenhagen, where only 53% of targeted women were faithful users.¹⁰

Various estimates of the cumulative risk of getting at least one false-positive result after 10 mammograms have been reported based on data from 2400 women screened in a health plan in Boston, USA.^{3,4,5} Elmore *et al.*³ estimated the cumulative risk of getting at least one false-positive result after 10 mammograms to be 49%. This estimate was based on the assumption that the conditional probabilities of having a false-positive result at the j th screening, given no false-positive result for the first $j-1$ screenings, followed a beta distribution. Gelfand and Wang⁴ assumed a slightly different distribution for these conditional probabilities and estimated the cumulative risk of getting at least one false-positive result after 10 mammograms to be 48%. Christiansen *et al.*⁵ using another model based on the same conditional probabilities, estimated the cumulative risk of getting at least one

Table 3 Proportion of participants with false-positive test by invitation round and screen number in two organized mammography screening programmes in Denmark

Invitation round	Screen number					Total (%)
	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	
<i>Copenhagen</i>						
1	5.6	—	—	—	—	5.6
2	6.9	3.3	—	—	—	3.9
3	5.1	2.0	1.9	—	—	2.5
4	5.0	1.8	1.5	1.8	—	2.4
5	4.5	1.6	1.3	1.0	0.9	1.8
Total	5.5	2.7	1.8	1.6	0.9	3.4
<i>Fyn</i>						
1	1.8	—	—	—	—	1.8
2	2.6	0.8	—	—	—	1.1
3	2.5	0.9	0.9	—	—	1.1
4	2.3	0.7	1.0	0.8	—	1.0
5	1.9	0.8	0.6	0.8	0.7	0.9
Total	2.0	0.8	0.9	0.8	0.7	1.2

false-positive result after nine mammograms to be 43%. Based on data from 83,416 women participating in the first three rounds of an organized mammography screening programme in Norway, Hofvind *et al.*⁶ predicted the cumulative risk of at least one false-positive test over 10 screens to be 21%. They assumed independence between some outcomes of subsequent screens and assumed that the risk of having a false-positive test at screen numbers four to 10 equalled the risk of having a false-positive test at screen number three.

Independence between outcomes of subsequent screens means that a woman with a false-positive test has the same risk of getting a false-positive test in the next screening round as women without a false-positive test. This can only be so if the radiologists learn from the previous mammograms. A mammogram is evaluated to be positive based on some suspicious finding. If the radiologists did not take previous mammograms into consideration when evaluating new mammograms, a woman with a false-positive screen in a previous round would have a higher risk than other women of getting a false-positive screen also in the new round. If, for instance, the first false-positive result was due to a misinterpreted scar from a previously removed benign lesion, then this misinterpretation would be repeated at the next screen if the previous mammogram was not considered. In the two Danish mammography screening programmes, new mammograms are always compared with those taken previously. In this way, radiologists avoid referring women for further assessment based on suspicious findings already assessed as non-malignant after previous screens. The learning from previous mammograms thus forms the basis for the observed independence between false-positive risks at subsequent screens.

As noted previously by Fletcher and Elmore,¹¹ there is a striking difference between the proportion of participants expected to experience at least one false-positive test in Europe and in the USA. The estimated proportions of participants expected to experience at least one false-positive test was 9% for Fyn, 16% for Copenhagen, 21% for Norway and 49% for Boston, USA. The definition of a false-positive test in the Boston study differed slightly from the one used in our study. In the Boston study, a woman had a false-positive test if her mammogram was interpreted as indeterminate, aroused a suspicion of cancer or prompted recommendation for additional workup, and she was not diagnosed with breast cancer within a year. However, if the same definition had been followed for the mammography screening programme in the county of Fyn, Denmark, 1726 false-positive tests would have occurred during the first three screening rounds, where we found 1724 false-positive tests using our definition. So differences in definitions cannot explain the very large span in cumulative risk after 10 screens in Fyn and Boston. It should be noted, however, that when assessing the overall quality of screening programmes, the false-positive rates form only one element, and the detection rate, tumour size, interval cancers, etc. should also be taken into account. For instance, the difference in false-positive rates between the two Danish programmes is reflected in a higher interval cancer rate in Fyn than in Copenhagen.⁷

CONCLUSIONS

Compared with the experiences from Boston, the risk of having a false-positive test was low in the Nordic service screening programmes. Anyhow, 9–16% of Danish women participating in the offered screening programmes will experience a false-positive test, and it is extremely important to inform future participants about this risk. Our study showed that a relatively robust estimate of this risk of getting a false-positive test over 10 screens can be predicted upon completion of the first three screening rounds, and it can probably be predicted even earlier by using data already collected in other screening programmes. It is therefore possible early on in a screening programme to include risk of a false-positive test during participation in the programme in the background information provided to targeted women.

ACKNOWLEDGEMENTS

This study was financially supported by the EU Commission, DG Sancho.

Authors' affiliations

Sisse Helle Njor, *Statistician*, Institute of Public Health, University of Copenhagen, Øster Farimagsgade 5, opg. B, DK-1014 Copenhagen K, Denmark

Anne Helene Olsen, *Statistician*, Institute of Public Health, University of Copenhagen, Øster Farimagsgade 5, opg. B, DK-1014 Copenhagen K, Denmark

Walter Schwartz, *Chief physician*, Mammography Screening Clinic, University Hospital Odense, Klørvænget 10, DK-5000 Odense, Denmark

Ilse Vejborg, *Chief physician*, Department of Diagnostic Imaging, University Hospital Copenhagen, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark

Elsebeth Lynge, *Professor*, Institute of Public Health, University of Copenhagen, Øster Farimagsgade 5, opg. B, DK-1014 Copenhagen K, Denmark

REFERENCES

- Fracheboud J, de Koning HJ, Boer R, *et al.* Nationwide breast cancer screening programme fully implemented in the Netherlands. *Breast* 2001;**10**:6–10
- Ascunce N, del Moral A, Murillo A, *et al.* Early detection programme for breast cancer in Navarra, Spain. *Eur J Cancer Prev* 1994; (Suppl 1):41–8
- Elmore JG, Barton MB, Mocerri VM, *et al.* Ten year risk of false positive screening mammograms and clinical breast examination. *N Eng J Med* 1998;**338**:1089–96
- Gelfand AE, Wang F. Modelling the cumulative risk for a false positive under repeated screening events. *Stat Med* 2000;**19**:1865–79
- Christiansen CL, Wang F, Barton MB, *et al.* Predicting the cumulative risk of false positive mammograms. *J Natl Cancer Inst* 2000;**92**:1657–66
- 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
- Njor SH, Olsen AH, Bellstrøm T, *et al.* Mammography screening in the county of Fyn November 1993–December 1999. *APMIS* 2003; (Suppl 110):1–33
- Vejborg I, Olsen AH, Jensen MB, *et al.* Mammography screening in Copenhagen 1991–99. *J Med Screen* 2002;**9**:115–19
- Törnberg S, Kemetli L, Svane G, *et al.* Pattern of participation in a cohort aged 50–60 years at first invitation to the service-screening programme with mammography in Stockholm county, Sweden. *Prevent Med* 2005;**41**:728–33
- von Euler-Chelpin M, Olsen AH, Njor S, *et al.* Women's patterns of participation in mammography screening in Denmark. *Eur J Epidemiol* 2006;**21**:203–9
- Fletcher SW, Elmore JG. False-positive mammograms – can the USA learn from Europe? *Lancet* 2005;**365**:7–8