

Cumulative false positive recall rate and association with participant related factors in a population based breast cancer screening programme

Xavier Castells, Eduard Molins, Francesc Macià

J Epidemiol Community Health 2006;**60**:316–321. doi: 10.1136/jech.2005.042119

See end of article for authors' affiliations

Correspondence to:
Dr X Castells, Health
Services Research Unit,
Institut Municipal
d'Investigació Mèdica
(IMIM-IMAS), Passeig
Marítim 25–29, 08003,
Barcelona, Spain;
xcastells@imas.imim.es

Accepted for publication
28 November 2005

Study objective: To investigate the cumulative false positive recall rate throughout the period of participation in a population based breast cancer screening programme and to examine its association with women related factors.

Design: Analysis of a database to estimate the cumulative false positive recall rate after 10 biennial mammograms in a cohort of women. Cumulative risk after 10 rounds was calculated by projecting forward the information available on the four rounds. Logistic regression was used to evaluate the association between the cumulative risk of false positive recall and women related factors.

Setting: Population based breast cancer screening programme in Barcelona City (Spain).

Participants: 8502 women aged 50–69 years who participated in four consecutive screening rounds. Eligible women had received a mammogram in the first screening round between 1 December 1995 and 31 December 1996.

Main results: The false positive recall rate in the first screening for women who entered the screening programme at the age of 50–51 years was assessed at 10.6% (95% CI 8.9, 12.3). In the second screening this risk decreased to 3.8% (95% CI 2.7, 4.9) and remained almost constant in subsequent rounds. After 10 mammograms, the cumulative false positive recall rate was estimated at 32.4% (95% CI 29.7, 35.1). The factors associated with a higher cumulative risk of false positive recall were: previous benign breast disease (OR = 8.48; CI 7.39, 9.73), perimenopausal status (OR = 1.62; CI 1.12, 2.34), body mass index above 27.3 (OR = 1.17; CI 1.02, 1.34), and age 50–54 years (OR = 1.15; CI 1.00, 1.31).

Conclusions: One third of women could have at least one false positive recall over 10 biennial screens. Women participating in screening programmes should be informed about this risk, especially those with associated factors.

Extensive use of screening mammograms increases the risk of participants' experiencing at least one adverse effect. One of the most important disadvantages, or adverse effects, concerns the risk of false positive recall—that is, the recommendation for further assessments because of an abnormal screening mammogram without a diagnosis of breast cancer. A false positive recall leads to additional tests (some invasive), increases costs and, in particular, provokes anxiety^{1–2} in women before malignancy is ruled out and may affect subsequent screening attendance.^{3–4}

As a result of the debate about the effectiveness of breast cancer screening in the past few years, several agencies have recommended that the associated risks be estimated to provide women with more reliable information (International Agency for Research on Cancer, Danish Council of Ethics), especially on the psychological and social effects of screening. Moreover, the benefits and risks associated with breast cancer screening programmes should be weighed up, not only when comparing each round but also those accumulated during a woman's life span, that is, after several screening rounds.^{5–7}

Few studies have assessed the cumulative risk of false positive recalls and their results show substantial differences. These discrepancies can be explained mainly by the differences between the health systems in the USA and Europe, as well as by protocol related factors and radiologists' experience, which hamper comparisons.⁸ Our programme is based on European guidelines and complies with quality standards. However, in Spain, the cumulative risk of false positive recalls has not previously been estimated. The aim of this study was

to estimate the cumulative risk of at least one false positive recall in women throughout the period of participation in a population based breast cancer screening programme and to assess the association between this risk and women's characteristics.

METHODS

Setting

This study was carried in a cohort of women participating in a population based breast cancer screening programme in Barcelona City (Spain), which began in 1995 and had completed four screening rounds. The programme was based on the European guidelines for quality assurance in mammographic screening^{9–10} and its results met the Europe Against Cancer standards.

The programme invited women aged 50–69 years to undergo a mammogram. Thus, it allowed women who began the breast cancer screening programme at the age of 50–51 years to have up to 10 mammograms over two decades. All women in the target population received information on the programme (especially on the benefits of early breast cancer detection) and were contacted by surface mail.¹¹ Women not attending the screening were reminded by surface mail and finally by telephone. All mammograms were located at the same radiology unit and readings were performed by the

Abbreviations: FNAC, fine needle aspiration cytology; CNB, core needle biopsy; OB, open biopsy; HRT, hormone replacement therapy; BMI, body mass index

same team of radiologists. The same technical mammography equipment was used in all screening rounds. Mediolateral oblique and craniocaudal views were available for each breast. All mammograms were read by two radiologists and, when double readings led to different assessments, a third radiologist served as a tie breaker. In each round, the radiologists had information about each woman at their disposal, obtained from a clinical and epidemiological survey carried out through face to face interview with the same technicians (with specific training in interviews) who then performed the mammograms; the information was then entered into a computer. Previous mammograms were available during the reading except in the first screening. Further assessments took place at the same radiology department and a definitive diagnosis of breast cancer was always histopathologically confirmed (invasive carcinomas and ductal carcinoma in situ).

Study population

Of the 19 458 potentially eligible women who participated in the first screening round between 1 December 1995 and 31 December 1996, 3193 were excluded from the analysis because they could not complete all four rounds: 251 because they were diagnosed with breast cancer, 1499 because they moved city or died, and 1443 because they were 64 years old at the baseline round and would be over 69 years of age after four rounds.

Of the remaining 16 265 women, 8502 (52.27%) participated in all four rounds.

Definition of false positive recall

Two possible mammogram results were included in our programme: a negative result (follow up at two years is recommended) or a positive result requiring a recall for further assessment to rule out malignancy. As proposed by European guidelines, the programme did not include the possibility of an early recall figuring as a result of mammography (that is, women requiring another screening mammogram—for example, at 6 or 12 months—before the interval corresponding to the normal sequence). Further assessments could include both non-invasive (additional mammogram, ultrasound) and invasive procedures (fine needle aspiration cytology (FNAC), core needle biopsy (CNB), and open biopsy (OB)). Recall because of insufficient technical quality of the mammogram was not included as a positive result. A positive result was a true positive if, after further assessments, breast cancer was found (in situ or invasive). Otherwise, the result was considered to be a false positive recall.

Cumulative risk of at least one false positive recall

The aim of this study was to assess the cumulative false positive recall risk for women who entered the screening programme aged 50–51 years and who participated in 10 consecutive routine follow up rounds until the age of 68–69 years. The probability of undergoing at least one false positive recall was estimated through the method described in detail by Hofvind *et al.*⁶ We used the false positive recall rate of each of four rounds to project those risks to 10 rounds in women who began screening in their early 50s. The probability was assessed by adding the false positive recall rate of each round, subtracting the false positive recall intersections of two rounds and adding the intersection of three rounds. Because data from only four rounds were available, we followed up the cohort until the age of 56–57 years and calculated the false positive recall rates until the age of 69 years using the remaining age pairs in the fourth screening round (numbers shown in bold face in table 1). The same methodology was used to calculate the cumulative risk of additional invasive procedures (FNAC, CNB, OB) after 10 rounds.

Analogously, we estimated the cumulative risk for women in the cohort who began screening aged 52–53, 54–55, 56–57, 58–59, 60–61, or 62–63 years and who participated in nine, eight, seven, six, five, and four screening rounds respectively.

Association between women’s characteristics and false positive recall

The participant related variables included in the analysis that might be associated with false positive recalls were: age, use of hormone replacement therapy (HRT), menopausal status, body mass index (BMI), symptomatology in the previous year (lumps, pain, skin changes, nipple retraction, nipple secretion, and ulceration), previous benign breast disease (including benign biopsies), and a familial history of breast cancer (mother, sister, daughter plus grandmother or aunt).

As most variables could change in women between two rounds and many combinations of variables were possible, the variables were held constant throughout the cohort (that is, as cross sectional data). Regarding age and BMI, we included values from the baseline screening round. Menopausal status, HRT, symptomatology, previous benign breast disease, and a familial history of breast cancer were re-coded into a value of 1 if women responded affirmatively in any of the four rounds, or into 0 if women responded negatively in all four rounds. In the case of HRT, in the first round, women were asked if they had taken HRT for more than two years at any time in their lives (women who responded affirmatively were coded as positive).

A logistic regression model was fitted using the GENMOD procedure in SAS 8.0. Validation of the models was based on

Table 1 Proportions of false positive recalls by age and screening round

Age (y)	First screening round		Second screening round		Third screening round		Fourth screening round	
	Number screened	FP recalls (%) (95% CI)	Number screened	FP recalls (%) (95% CI)	Number screened	FP recalls (%) (95% CI)	Number screened	FP recalls (%) (95% CI)
50–51	1193	10.6 (8.9, 12.3)						
52–53	1090	7.4 (5.8, 9.0)	1193	3.8 (2.7, 4.9)				
54–55	1159	7.8 (6.3, 9.3)	1090	4.0 (2.8, 5.2)	1193	3.5 (2.5, 4.5)		
56–57	1084	8.6 (6.9, 10.3)	1159	3.2 (2.2, 4.2)	1090	3.2 (2.2, 4.2)	1193	3.3 (2.3, 4.3)
58–59	1274	7.1 (5.7, 8.5)	1084	3.0 (2.0, 4.0)	1159	2.8 (1.9, 3.7)	1090	2.7 (1.7, 3.7)
60–61	1490	7.4 (6.1, 8.7)	1274	1.8 (1.1, 2.5)	1084	3.3 (2.2, 4.4)	1159	2.2 (1.4, 3.0)
62–63	1212	7.4 (5.9, 8.9)	1490	2.9 (2.0, 3.8)	1274	2.6 (1.7, 3.5)	1084	3.0 (2.0, 4.0)
64–65			1212	3.0 (2.0, 4.0)	1490	3.5 (2.6, 4.4)	1274	3.1 (2.1, 4.1)
66–67					1212	3.3 (2.3, 4.3)	1490	2.6 (1.8, 3.4)
68–69							1212	3.4 (2.4, 4.4)
Total	8502	8.0 (7.4, 8.6)	8502	3.1 (2.7, 3.5)	8502	3.2 (2.8, 3.6)	8502	2.9 (2.5, 3.3)

FP recalls, false positive recalls; 95% CI confidence interval.

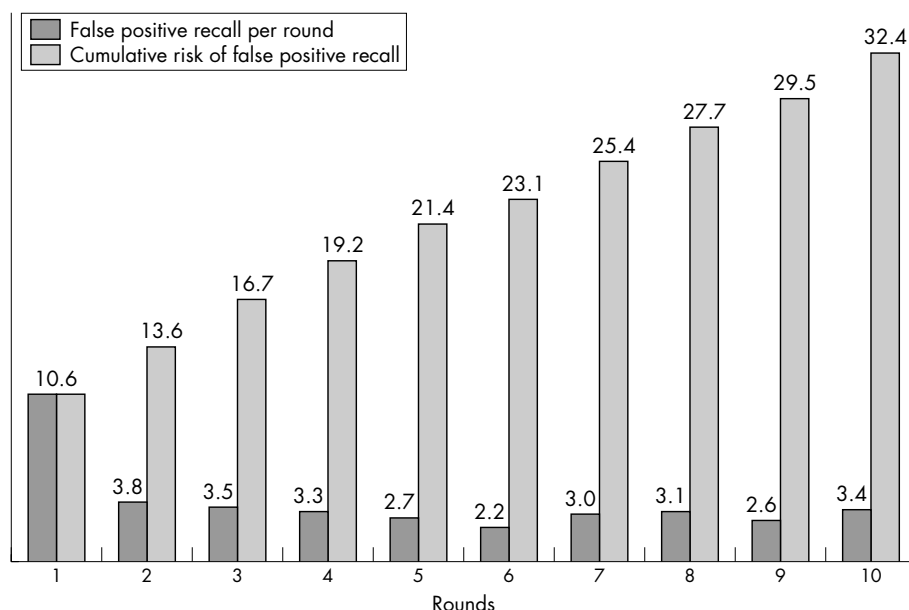


Figure 1 False positive recall in women aged 50–51 years in the first round who participated in 10 rounds. Risk per round and cumulative risk (%).

deviance and over-dispersion. Based on the results of the logistic regression analysis, the highest risk profiles of false positive recalls were used to estimate the cumulative risk at the first, fourth, and tenth screening rounds.

Ethical issues

This study was performed in accordance with the national and international guidelines stated at the Declaration of Helsinki and complies with the legal requirements regarding confidentiality (Law 15/1999 of 13 December concerning Personal Data Protection).

RESULTS

Of the 8502 women included in the study who participated in four rounds, 2860 were aged 50–54 years, 2940 were aged 55–59 years, and 2702 were aged 60–63 years. Overall, 7196 (84.6%) had no false positive recalls during the four rounds, 1171 (13.8%) had one false positive recall, 119 (1.4%) had

two false positive recalls, 16 (0.2%) had three false positive recalls, and none had four false positive recalls. As a whole, 1306 women (15.4%) had at least one false positive recall. Women who had already had a false positive recall were at higher risk for a second false positive recall than those who had never had a false positive recall (RR = 1.35 (95%CI 1.14, 1.59)) (data not shown).

Table 1 shows the false positive recall rates by screening rounds. False positive recalls were distributed as follows: 8.0% (95% CI 7.4, 8.6%), 3.1% (95% CI 2.7%, 3.5%), 3.2% (95% CI 2.8, 3.6%), and 2.9% (95% CI (2.5, 3.3%)) in the first, second, third, and fourth screening round respectively. Figure 1 shows the per-round and cumulative false positive recall risk for women aged 50–51 years in the first screening round. Based on a false positive rate of 10.6% in the first round, the cumulative risk increased by 2.4%, on average, in each round. The cumulative risk after 10 mammograms was estimated at 32.4% (95% CI 29.7, 35.1).

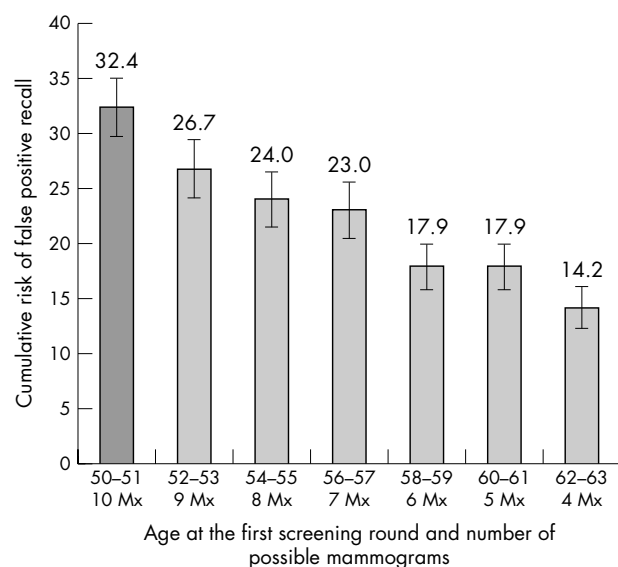


Figure 2 Cumulative risk of false positive recall by age at the first screening round (%). Mx, mammogram.

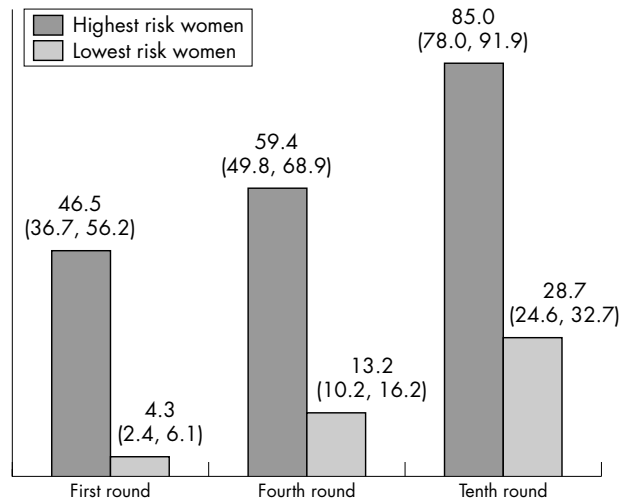


Figure 3 Cumulative risk of at least one false positive recall by patient risk in women aged 50–51 years old in the first screening round (%). Highest risk variables (n = 101): previous benign breast disease and BMI higher than 27.3. Lowest risk variables (n = 486): no previous benign breast disease and BMI up to 27.3.

Table 2

Cumulative risk of false positive recall by participant related factors

	Screening women	% Cumulative false positive	$p>\chi^2$	OR	95% CI		OR Adjusted	95% CI	
Age at the first screening round									
50–54	2860	16.40	0.0648	1.12	0.99	1.27	1.15	1.00	1.31
>54	5642	14.80	–	1.00	–	–	1.00	–	–
HRT									
Yes	672	15.00	0.804	0.97	0.78	1.21	0.88	0.69	1.12
No	7830	15.40	–	1.00	–	–	1.00	–	–
Menopause									
Perimenopausal	200	22.50	0.005	1.62	1.16	2.27	1.62	1.12	2.34
Menopausal	8302	15.20	–	1.00	–	–	1.00	–	–
BMI									
Overweight or obesity	5367	15.70	0.2471	1.08	0.95	1.22	1.17	1.02	1.34
Normal weight	3135	14.80	–	1.00	–	–	1.00	–	–
Symptomatology									
Yes	908	15.70	0.0576	1.19	0.99	1.43	0.91	0.74	1.11
No	7594	15.10	–	1.00	–	–	1.00	–	–
Personal previous benign breast disease									
Yes	1219	47.80	<0.0001	8.32	7.26	9.53	8.48	7.39	9.73
No	7283	9.90	–	1.00	–	–	1.00	–	–
Previous familial breast cancer									
Yes	978	15.20	0.9155	0.99	0.82	1.19	0.92	0.75	1.12
No	7504	15.40	–	1.00	–	–	1.00	–	–

HRT, hormone replacement therapy; BMI, body mass index (overweight or obesity, BMI>27.3).

Because women of distinct ages participated in the first screening round (range 50–63 years) when the programme started, we were interested in determining the cumulative false-positive recall risk for each age group. Figure 2 shows the cumulative risk until the age of 68–69 years in women who entered the cohort at ages 52–53, 54–55, 56–57, 58–59, 60–61, and 62–63 years.

Further assessments were conducted using invasive procedures in 349 of 681 women with a positive result (51%) in the first screening round. This figure decreased to 101 of 261 women (39%) in the second screening round, 89 of 270 women (33%) in the third screening round, and 95 of 245 women (39%) in the fourth screening round. Overall, 54 women had more than one additional invasive test. Estimates of 10 rounds for the percentage of invasive procedures in women aged 50–51 years in the first screening round showed that 11.7% of women would undergo at least one FNAC, 4.5% at least one CNB, and 0.9% at least one OB (data not shown).

To analyse the association between women’s characteristics and the cumulative risk of false positive recall, a logistic regression model was used, based on data from the first four rounds (table 2). Women with previous benign breast disease (OR = 8.48; CI 7.39, 9.73), perimenopausal status (OR = 1.62; CI 1.12, 2.34), BMI over 27.3 (OR = 1.17; CI 1.02, 1.34), and age 50–54 years old (OR = 1.15 CI 1.00–1.31) had a higher risk of false positive recall. The remaining variables were not statistically significant. Based on these results, the cumulative risk of false positive recall was projected to 10 rounds bearing in mind the highest risk profiles. In the group of women aged 50–51 years at the first screening round, the cumulative risk after 10 rounds in women with both previous benign breast disease and a BMI higher than 27.3 was estimated at 85.0% while women with opposite categories in those variables had an estimated risk of 28.7% (fig 3).

DISCUSSION

Of the women participating in the population based breast cancer screening programme, 15.4% had at least one false positive recall in the four rounds analysed. Using these data, we estimated that roughly one of every three women (32.4%) who started the screening programme at the age of 50 years and who participated in 10 consecutive screening rounds

What this paper adds

- There is wide variability in the results of the few studies that have been published on the cumulative risk of false positive recall throughout the entire period of participation in a breast cancer screening programme.
- Like other authors, we found a high cumulative false positive recall rate. The results of this study suggest that because of their organisational characteristics, programmes carried out within national health systems have a lower false positive recall rate.

(until the age of 69 years) would have at least one false positive recall. Otherwise, in the cohort analysed, women aged 50–51 years had the highest rate of false positive recalls in the first screening. This rate substantially decreased between the first and second rounds and subsequently continued to decrease slightly with age. Factors that may increase the risk of false positive recall are overweight, perimenopause, and especially previous benign breast disease. However, HRT and a familial history of breast cancer did not modify the risk of false positive recall.

Policy implications

- Because one of the most important harms related to false positive results is the anxiety provoked in women before malignancy is ruled out, women participating in a breast cancer screening programme should be informed of this risk, especially those presenting related factors.
- Inevitably, screening programmes will differ in their organisational models and protocol characteristics. However, it would be advisable to promote those that guarantee the best results.

Cumulative risk of false positive recall

In a study published by Elmore *et al.*,⁵ based on a health maintenance organisation in the USA, the cumulative risk of a false positive recall over 10 mammograms was estimated at 47.3% in women aged 50–79 years. Another study of a Norwegian programme by Hofvind *et al.*⁶ estimated a risk of 20.8% in women aged 50–69 years. The differences seen can generally be attributed both to the organisational models of the screening programmes and to the characteristics of the protocols.¹² In Europe, screening programmes are population, publicly funded, and adhere to guidelines guaranteeing quality (European guidelines,⁹ International Agency for Research on Cancer¹⁰), while in the USA the organisational models are decentralised, funded by private companies, and lack the coordination that would evaluate the process and its results. Concerning the characteristics of the protocol, in addition to mammogram quality, the experience of the radiologists as a factor in ensuring an appropriate level of accuracy,^{13–15} and the adaptation of the BI-RADS scale by each programme, other radiological criteria also influence the recall rate and consequently the risk of false positives. Thus, for example, if the mammogram is considered to be positive when the size of a nodule is greater than 1 cm instead of 0.8 cm, the differences in the recall rate can vary substantially. Protocols can also differ in the system of double reading and the tie breaker method, in the number of mammographic views, and in the percentage of screening clinical breast examinations.

Another protocol related feature that can substantially affect estimation of the false positive rate concerns the distinct ways of managing early recall mammograms or intermediate mammograms. Although the European guidelines recommend a value of the maximum standard for early recalls of less than 1% (desirable 0%), some programmes include this possibility, even as a direct result of mammography. In this case, they are not usually counted as a positive result and probably underestimate false positive recalls. Because this percentage is sometimes substantial, the false positive rates can show wide variability among programmes, greatly hampering comparisons among them. Therefore, in any study, the specifications of the protocol concerning early recalls and their role in the definition of a false positive result should be made clear, which is not always the case.

The probability of undergoing a false positive recall in our study was higher in the first screening than in subsequent screenings, in which the risk was similar. Thus, the risk of a false positive result in the second or subsequent rounds was about half that of such a result in the initial round, probably because of the risk inherent to the prevalent screening round and to the lack of a previous mammogram in the first round when radiologists interpreted the result. A noticeable difference between the first and subsequent rounds is a common finding in all programmes, although the magnitude of this difference may vary because of the distinct protocols used.^{16, 17}

In this study, no clear association was found between age and the risk of a false positive result, a finding that is in agreement with those of other studies that also analysed cohorts of women undergoing successive screening rounds.⁶ Although table 2 shows a positive odds ratio on the threshold of statistical significance for women aged 50–54 years old and the risk of having at least one false positive result, the tendency concerning the probability of a false positive result in table 1 shows a highly similar profile in each age subcohort at the first screening. Thus, the risk of a false positive result is much higher in the first (or prevalent) screening than in subsequent screenings but substantial reductions from the second round onwards were not seen, despite women's increasing age. Therefore, the association between age and

false positive recall rate seen in several cross sectional studies could be overestimated^{18, 19} because of the proportion of young women in the first screening round (that is, when radiologists evaluate prevalent cases without a previous mammogram) whereas, when a cohort is analysed, the results suggest that age is not as relevant as the first screening round itself.

Women related factors associated with the cumulative risk of false positive recall

Another important issue is the association between the cumulative risk of false positive recall and women's characteristics. The characteristics analysed are of interest to radiologists because they can influence their interpretation when reading screening mammograms and many are related to breast tissue. We found that previous benign breast disease (for example, biopsies) presented a close association, equal to that found by other authors,²⁰ as, to a lesser extent, did perimenopausal status and higher BMI. As previously described, all these variables can modify the radiological image, thus increasing variability when reading the mammogram. Unlike the findings of other studies, the lack of association with a familial history of breast cancer in our study is a reasonable result, as the availability of this information may influence the radiologist but does not modify the radiological image or reduce the accuracy of reading. When the combinations of profiles of women with a higher risk were compared, we found that women with both benign breast disease and a BMI higher than 27.3 had a cumulative false positive risk after 10 rounds of 85%, while this figure was 29% in low risk women. These results suggest that women's characteristics have a pronounced impact on the rate of false positives in high risk women and an intrinsic impact in low risk women. However, interpretation of these results is difficult because the association between a false positive result and women's characteristics are in turn related to the density of breast tissue. This possible confounder could hamper interpretation of the effects of these variables.

Methodological issues and limitations

In addition to aspects associated with screening programmes and women's characteristics, published studies show methodological differences in the selection of the cohort of women analysed. American authors have approached these differences by taking into account that the time interval between two mammograms could vary among women and in individual women. They propose an analysis based on the assumption that drop outs do not depend on previous false positive,^{5, 20, 21} whereas other authors take this association into account.^{22, 23} Like other authors, we included only women participating in all rounds^{6, 7} because we believe that the consistency of our results is attributable to the large sample size (8502 women and 34 008 mammograms without drop outs). Excluding women not participating in all rounds probably produced a bias but it is difficult to know whether we are overestimating or underestimating the cumulative risk.

A possible limitation of this study is that some participant related factors, such as genetic characteristics, were not included in the analysis and could have influenced the positive predictive value. Further studies are required to elucidate this issue.

Conclusions

One of the most important adverse effects of a false positive result is the anxiety provoked in women before malignancy is ruled out. As a not inconsiderable percentage of women will receive more than one result of this type, further studies are required to evaluate the emotional impact. This subject is sufficiently important to warrant analysis of possible inter-

ventions to reduce the negative impact of the cumulative false positive recall risk. Women invited to participate in a breast cancer screening programme should be able to decide whether they wish to attend or not after being fully informed not only of the obvious benefits of screening but also of the possible adverse effects, some of which are unrelated to breast malignancies. Potential participants should be informed of the frequency of possible recalls and consequent additional assessments, some of which are invasive.

Authors' affiliations

X Castells, E Molins, Evaluation and Clinical Epidemiology Department, Institut Municipal d'Assistència Sanitària (IMAS); Health Services Research Unit, Institut Municipal d'Investigació Mèdica (IMIM-IMAS) Barcelona, Spain

F Macià, Evaluation and Clinical Epidemiology Department, Institut Municipal d'Assistència Sanitària (IMAS); Research Unit on Clinical and Molecular Epidemiology of Cancer (IMIM-IMAS), Barcelona, Spain

Funding: this work was supported by grants from the Catalan Health System (CATSALUT ST(391/03)) and the Catalan Agency for Health Technology Assessment and Research (AATRM-GENDECAT 087/18/2004). Eduard Molins received partial funding from the ISCIII (Red de Centros RCESP C03/09).

Conflicts of interest: none.

REFERENCES

- 1 Bull AR, Campbell MJ. Assessment of the psychological impact of a breast screening programme. *Br J Radiol* 2001;**64**:510–15.
- 2 Lerman C, Trock B, Rimer BK, et al. Psychological and behavioral implications of abnormal mammograms. *Ann Intern Med* 1991;**114**:657–61.
- 3 McCann J, Stockton D, Godward S. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer Res* 2002;**4**:R11.
- 4 Lampic C, Thurfjell E, Sjoden PO. The influence of a false-positive mammogram on a woman's subsequent behaviour for detecting breast cancer. *Eur J Cancer* 2003;**39**:1730–7.
- 5 Elmore JG, Barton MB, Moceris VM, et al. Ten-year risk of false-positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998;**338**:1089–96.
- 6 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.
- 7 Barratt A, Howard K, Irwing L, et al. Model of outcomes of screening mammography: information to support informed choices. *BMJ* 2005;**330**:936.
- 8 Fletcher SW, Elmore JG. False positive mammograms-can the USA learn from Europe? *Lancet* 2005;**365**:7–8.
- 9 Perry N, Broeders M, de Wolf C, et al. *European guidelines for quality assurance in mammography screening*. Luxembourg: Office for Official Publications of the European Communities, 2001.
- 10 Vainio H, Bianchini F, eds. *IARC handbooks of cancer prevention*. Vol 7. *Breast cancer screening*. Lyon: IARC Press, 2002:119–34.
- 11 Thornton H, Edwards A, Baum M. Women need better information about routine mammography. *BMJ* 2003;**327**:101–3.
- 12 Smith-Bindman R, Ballard-Barbash R, Miglioretti DL, et al. Comparing the performance of mammography screening in the USA and the UK. *J Med Screen* 2005;**12**:50–4.
- 13 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.
- 14 Kan L, Olivetto IA, Warren Burhenne LJ, et al. Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. *Radiology* 2000;**215**:563–7.
- 15 Beam CA, Conant EF, Sickles EA. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. *J Natl Cancer Inst* 2003;**95**:282–90.
- 16 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.
- 17 Lynge E, Olsen AH, Fracheboud J, et al. Reporting of performance indicators of mammography screening in Europe. *Eur J Cancer Prev* 2003;**12**:213–22.
- 18 Kerlikowske K, Grady D, Barclay J, et al. Positive predictive value of screening mammography by age and family history of breast cancer. *JAMA* 1993;**270**:2444–50.
- 19 Lidbrink E, Elfving J, Frisell J, et al. Neglected aspects of false positive findings of mammography in breast cancer screening: analysis of false positive cases from the Stockholm trial. *BMJ* 1996;**312**:273–6.
- 20 Christiansen CL, Wang F, Barton MB, et al. Predicting the cumulative risk of false-positive mammograms. *J Natl Cancer Inst* 2000;**92**:1657–66.
- 21 Gelfand AE, Wang F. Modelling the cumulative risk for a false-positive under repeated screening events. *Stat Med* 2000;**19**:1865–79.
- 22 Baker SG, Erwin D, Kramer BS. Estimating the cumulative risk of false positive cancer screenings. *BMC Med Res Methodol* 2003;**3**:1–6.
- 23 Xu J-L, Fagerstrom RM, et al. Estimating the cumulative risk of false-positive test in a repeated screening program. *Biometrics* 2004;**60**:651–60.