Eur Radiol (2011) 21:2083–2090
DOI 10.1007/s00330-011-2160-0

BREAST

Effect of radiologist experience on the risk of false-positive results in breast cancer screening programs

Raquel Zubizarreta Alberdi · Ana B. Fernández Llanes · Raquel Almazán Ortega · Rubén Roman Expósito · Jose M. Velarde Collado · Teresa Queiro Verdes · Carmen Natal Ramos · María Ederra Sanz · Dolores Salas Trejo · Xavier Castells Oliveres · and the CFPR (Cumulative False Positive Risk) group(1)

Received: 15 December 2010 / Revised: 14 March 2011 / Accepted: 11 April 2011 / Published online: 4 June 2011

Abstract
Objectives To evaluate the effect of radiologist experience on the risk of false-positive results in population-based breast cancer screening programmes.
Methods We evaluated 1,440,384 single-read screening mammograms, corresponding to 471,112 women aged 45–69 years participating in four Spanish programmes between 1990 and 2006. The mammograms were interpreted by 72 radiologists.
Results The overall percentage of false-positive results was 5.85% and that for false-positives resulting in an invasive procedure was 0.38%. Both the risk of false-positives overall and of false-positives leading to an invasive procedure significantly decreased ($p<0.001$) with greater reading volume in the previous year: OR 0.77 and OR 0.78, respectively, for a reading volume 500–1,999 mammograms and OR 0.59 and OR 0.60 for a reading volume of >14,999 mammograms with respect to the reference category (<500). The risk of both categories of false-positives was also significantly reduced ($p<0.001$) as radiologists' years of experience increased: OR 0.96 and OR 0.84, respectively, for 1 year's experience and OR 0.72 and OR 0.73, respectively, for more than 4 years' experience with regard to the category of <1 year's experience.
Conclusion Radiologist experience is a determining factor in the risk of a false-positive result in breast cancer screening.

Keywords Breast neoplasm · Mass screening · Mammography · False-positive reactions · Observer variation

Introduction

Biannual mammographic screening in women aged 50–69 years is widely recognised to reduce breast cancer mortality by an estimated 24–29% [1–3]. Evidence of the effectiveness of this preventive technique is that most European countries have breast cancer screening programmes, although organ-

R. Z. Alberdi · A. B. F. Llanes · R. A. Ortega
Galician Breast Cancer Screening Programme,
Public health & Planning Directorate, Health Office,
Galicia, Spain

R. R. Expósito · J. M. V. Collado · X. C. Oliveres
Institut Municipal d'Investigació Mèdica-Parc de Salut Mar.
CIBERESP, Department of Epidemiology and Evaluation,
Barcelona, Spain

T. Q. Verdes
Galician Agency for Health Technology Assessment,
Public health & Planning Directorate, Health Office,
Galicia, Spain

C. N. Ramos
Principality of Asturias Breast Cancer Screening Programme,
Principality of Asturias, Spain

M. E. Sanz
Navarra Breast Cancer Screening Programme,
Public Health Institute,
Pamplona, Spain

D. S. Trejo
Valencia Breast Cancer Screening Programme, General Directorate Public Health & Centre for Public Health Research (CSISP),
Valencia, Spain

R. Z. Alberdi (✉)
Servicio de Programas Poboacionais de Cribado,
Dirección Xeral de Saude Pública e Planificación,
Edificio Administrativo da Consellería de Sanidade,
San Lázaro s/n,
15703 Santiago de Compostela, Galicia, Spain
e-mail: raquel.zubizarreta.alberdi@sergas.es

isational models differ [4]. However, screening is not free from adverse effects, one of the most frequent and important being the risk of a false-positive result. The specificity of mammography is limited and further examinations, sometimes invasive, that do not reveal malignancy are not infrequent. The risk of a false-positive result has been estimated to vary within a range between 20% and 50% for women participating in ten screening rounds, i.e. women participating in biannual screens between the ages of 50 and 69 years, as recommended by the European Union Recommendations [5–7]. False-positive results are influenced by multiple factors. Some are associated with women's characteristics such as age, breast density, a history of breast disease, etc. and cannot be modified [8–10]. However, others, including the number of views, type of reading and radiologist experience, can potentially be modified and should be taken into account when establishing organisational models in screening programmes [11].

The effect of radiologist experience on the false-positive rate has been little studied and the results obtained to date are contradictory. Some studies suggest that higher reading volume reduces this rate and provides greater diagnostic accuracy, improving the sensitivity and specificity of mammography [12–14]. However, this association has not been found by other authors, who have concluded that the radiologist effect reflects a complex, multifactorial process [15–17]. Importantly, many of these studies were not performed in the context of population-based programmes but under experimental conditions and sometimes included screening and diagnostic mammograms randomly selected by several radiologists and with a proportion of tumours in the sample that sometimes reached 43% [14, 16, 17], thus failing to reflect normal reading conditions in screening programmes.

In Spain and most other European countries, breast cancer screening programmes are population-based and follow the European Guidelines For Quality Assurance in Mammographic Screening, which stipulate that radiologists must interpret a minimum of 5,000 screening mammograms yearly to ensure quality [4, 18, 19].

The aim of this study was to evaluate the effect of radiologist experience on the risk of a false-positive result in population-based breast cancer screening programmes. Other factors that could also influence this risk, such as women's characteristics and protocol-related factors, were also taken into account.

Materials and methods

Design

Information was retrospectively available from a cohort of women participating in four Spanish population-based breast cancer screening programmes. This information corresponded to the subset of screening mammograms interpreted by a single reading, extracted from the database of the Cumulative False-Positive Risk (CFPR) Project. This project is a publicly-funded research study performed by the Carlos III Health Institute and includes information on ten Spanish screening programmes. The organisational characteristics of these programmes differ somewhat, such as the type of reading (single or double) and the age range of the target population (from 45 years in some programmes and from 50 years in others), but all the programmes have features in common such as biannual screening, all evaluate their results following the indicators and standards proposed by the European Guidelines. Of the ten regional programmes participating in the project, we selected those that systematically identified the reading radiologist and used single reading as one of its reading methods. Four of these programmes met this requirement, although all four used both single and double reading. Only examinations with a single reading were selected since, for the present study, false-positive results could not be attributed to a specific radiologist when double reading was used.

The study period was from March 1990 to December 2006 and the study population consisted of women aged 45 to 69 years participating in at least one screening round. The radiologist did not know that this study was ongoing provided that the information was obtained retrospectively.

At each screening, the following information on the mammography examination was available: type of reading and its result, number of views (one or two), mammogram type (analog or digital, the latter being considered only if performed and read in a digital format), the participant's age, screen type (first or successive screen) and the radiology unit, i.e. where the mammogram was performed; the code identifying the radiologist(s) interpreting the mammogram; its result; information on additional procedures performed (invasive or non-invasive) and the final result of the screen, i.e. the absence or presence of histopathologically confirmed cancer. In addition, to take the period effect into account, the date when the mammogram was performed was used, grouped by two-yearly periods. In the successive screen the previous studies were available.

Study population

The initial study population consisted of 1,657,288 screening mammograms corresponding to 497,597 women. The number of excluded women and mammograms and the reasons for exclusion are shown in Fig. 1. The final sample consisted of 1,440,384 single-read mammograms corresponding to 471,112 women who participated in at least one round of the four selected programmes.

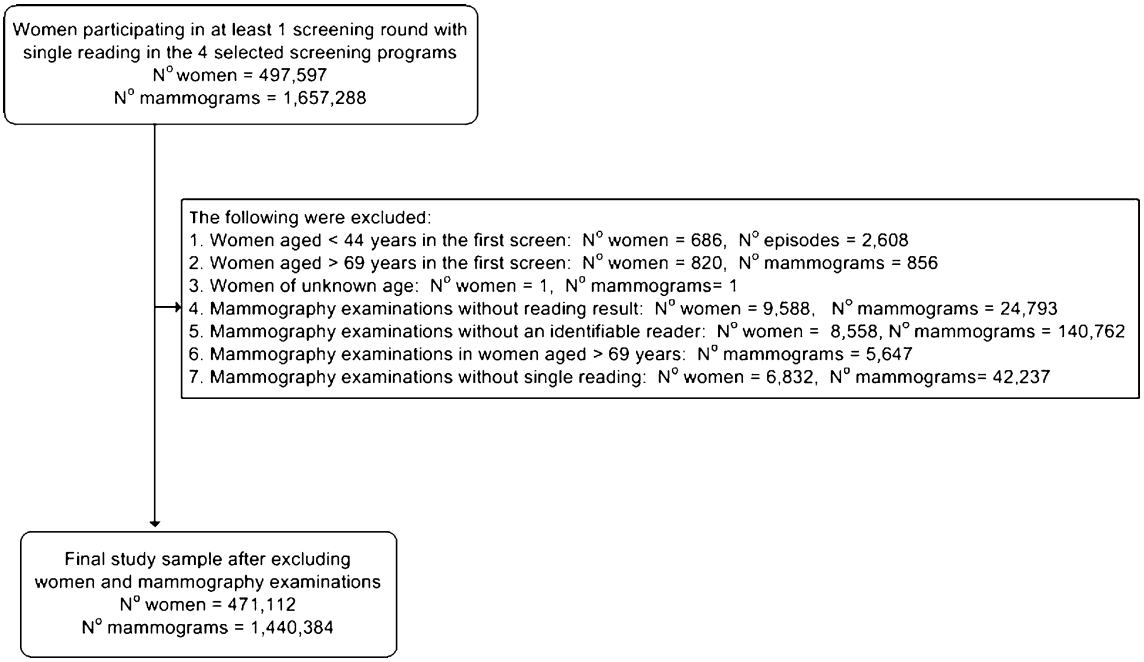

Fig. 1 Initial study population, final sample, number of excluded women and mammograms with the reasons for exclusion

Definition of a false-positive result

Three possible results were considered: negative (recommendation of routine screening examination at 2 years), positive (additional procedures were required to exclude malignancy) and short interval follow-ups (women requiring an intermediate mammogram at 6 or 12 months before the routine screening interval). Short interval follow-ups were not considered as false-positive results, unless they led to further assessment with a negative result. Indication of a repeat mammogram due to inadequate technical quality was not considered a positive result.

A false-positive result was defined as the absence of histologically confirmed breast cancer after a positive screening mammogram and additional procedures. False-positive results were divided into two types: (i) overall false-positive results, consisting of mammograms requiring additional assessment of any type (invasive and/or non-invasive) to exclude malignancy and (ii) false-positive results leading to at least one invasive procedure.

Measurement of radiologist experience

Two variables were used to determine radiologist experience at each reading: (i) reading volume in the previous year as a measure of recent experience and (ii) length of service in the breast cancer screening programme as a measure of cumulative experience. Because experience is constantly modified over time, each radiologist's experience was recalculated for each mammogram interpreted. The reading volume attributed to each radiologist refers to the mammograms read by the same radiologist within the screening program in the 365 days before the examination under consideration, independently of whether the examination was single- or double-read. Double-read mammograms were not included in the sample but were included as radiologist experience. Annual reading volume was divided into the following six categories: less than 500, 500–1,999, 2,000–4,999, 5,000–9,999, 10,000–14,999 and more than 14,999. The cut-off points selected in our study to evaluate reading volume in the previous year ranged from less than 500 to more than 15,000. No other cut-off points were found in the literature that could be used as a reference and the choice of these limits allowed us to evaluate the effect of radiologist experience on false-positive results, both for very small and for large reading volumes. These limits also allowed us to determine whether the greater the radiologist's experience of reading mammograms, the greater the reduction in the risk of false-positive results. Length of service in the screening programme was divided into less than 1 year, 1, 2, 3 or 4 years and more than 4 years. Only years in which radiologists interpreted at least 500 mammograms were included.

Statistical analysis

Two multilevel logistic regression models were performed, one with each of the measures of radiologist experience as the main explanatory variable: reading volume in the previous 365 days and the number of years' experience as a reading radiologist. The dependent variable was the false-positive result. The adjustment variables were those that

could influence false-positive results: the number of views (one or two), mammogram type (analogue or digital), screen type (first or successive), period when the screening mammogram was performed, and the participant's age. As a random effect, the radiology unit where the mammogram was carried out was included to control for the correlation among mammograms from units with distinct characteristics. The multilevel model was specifically used to take this random effect into account.

For the statistical analysis, the GLIMMIX module of version 9.1 of the SAS statistical package was employed. Validation of the model was based on deviance and on Akaike's information criterion.

Results

Of the total number of mammograms analysed, 385,436 were first screens and 1,054,948 were successive screens. Mammograms corresponding to these screens were read by 72 radiologists from 19 radiological units in the four screening programmes selected for the study.

During the study period, there were 84,320 screening mammograms with a false-positive result (total false-positive results 5.85%); of these, 5,435 led to an invasive assessment (false-positive results leading to an invasive procedure 0.38%).

Table 1 shows the percentages of overall false-positive results, false-positive results leading to an invasive assessment and the number of examinations for each of the categories of the following variables: reading volume in the previous year, the radiologist's years of experience in the programme, number of views, mammogram type, screen type and participant's age. All the variables studied in the univariate analysis statistically significantly ($p<0.001$) influenced the percentage of overall false-positive results and of false-positive results leading to an invasive assessment. The percentage of mammography examinations with false-positive results leading to some type of further assessment varied from 7.16% to 4.94% according to the radiologist's experience measured by reading volume in the previous year; the highest percentage of false-positive results was found for mammograms read by radiologists with the least experience. For the distinct categories of length of radiologist experience in the screening programme, the range of mammograms with a false-positive result leading to some type of further assessment varied from 7.58% for radiologists with less than 1 year's experience to 4.92% for those with more than 4 years' experience. The same tendency was observed for false-positive results leading to an invasive procedure in both measures of radiologist experience.

The results of the univariate and multivariate analyses to evaluate the risk of false-positive results associated with reading volume in the previous year and the radiologist's length of service in the programme are shown in Tables 2 and 3. The results were analysed separately for false-positive results leading to further assessment of any type and false-positive results leading to an invasive procedure and were expressed as crude (OR) risks and risks adjusted by variables that might have influenced false-positive results.

The risk of a false-positive result significantly decreased with increasing reading volume in the previous year when adjusted by the participant's age, screen type, number of views, mammogram type, radiology unit and the period effect. This tendency decreased from the reference category (less than 500 readings in the previous year) to the highest category (more than 14,999) and affected both overall false-positive results and false-positives leading to an invasive procedure. The reduction in the risk of overall false-positive results was more evident after the category of 10,000–14,999 mammograms in the previous year, with no overlap with the confidence intervals of the OR with the previous category (Table 2). The risk of a false-positive result leading to an invasive assessment was also reduced by a higher reading volume in the previous year and with a magnitude similar to that found for overall false-positive results: OR 0.60 (95% CI 0.51 to 0.70) for the category of more than 14,999 readings in the previous year compared with the category of less than 500 readings per year.

Longer experience in the screening programme also statistically significantly reduced the frequency of overall false-positive results and false-positives resulting in an invasive test, although to a lesser extent. A decreasing tendency was observed for overall false-positive results after the first year of experience but the greatest risk reduction was found in radiologists with more than 4 years' experience: OR 0.72 (95% CI 0.70 to 0.74). For false-positives leading to an invasive procedure, no clear tendency was observed among the distinct categories of years of experience, and the 95% CI of the OR overlapped; however, overall, radiologists' years of experience also reduced the risk of a false-positive result compared with mammograms interpreted by radiologists with less than 1 year's experience in the programme (Table 3).

Discussion

The results obtained in this study show how radiologist experience reduces the risk of false-positive results, both overall and those leading to an invasive procedure. Once the two measures of radiologist experience were adjusted by other variables, that which most reduced the frequency of a false-positive result was reading volume in the previous year.

Table 1 Percentage of overall false-positive (FP) results, according to variables influencing these results

	Number of examinations (%)	% overall FP (95% CI[a]) p value	% FP resulting in an invasive procedure (95% CI[a])
Mammograms read in the previous 365 days		$p<0.001$	$p<0.001$
0–499	31,527 (2.2)	7.16 (6.87, 7.44)	0.62 (0.53, 0.71)
500–1,999	101,981 (7.1)	5.32 (5.18, 5.45)	0.46 (0.41, 0.50)
2,000–4,999	196,420 (13.6)	4.83 (4.73, 4.92)	0.42 (0.39, 0.45)
5,000–9,999	337,032 (23.4)	6.92 (6.83, 7.00)	0.49 (0.47, 0.51)
10,000–14,999	363,087 (25.2)	6.50 (6.42, 6.58)	0.33 (0.31, 0.34)
>15,000	410,337 (28.5)	4.94 (4.87, 5.00)	0.27 (0.26, 0.29)
Length of radiologist experience in the programme		$p<0.001$	$p<0.001$
<1 year	138,737 (9.6)	7.58 (7.44, 7.72)	0.73 (0.69, 0.78)
1 year	200,368 (13.9)	7.78 (7.66, 7.89)	0.54 (0.51, 0.57)
2 years	208,723 (14.5)	5.70 (5.60, 5.80)	0.31 (0.29, 0.34)
3 years	157,782 (11)	5.97 (5.85, 6.08)	0.38 (0.35, 0.41)
4 years	141,932 (9.9)	5.44 (5.32, 5.56)	0.35 (0.31, 0.38)
>4 years	592,824 (41.2)	4.92 (4.87, 4.98)	0.27 (0.26, 0.28)
Number of views		$p<0.001$	$p<0.001$
One	463,759 (32.2)	8.36 (8.28, 8.44)	0.24 (0.23, 0.26)
Two	976,625 (67.8)	4.66 (4.62, 4.71)	0.44 (0.43, 0.45)
Mammogram type		$p<0.001$	$p<0.001$
Analogue	1,404,446 (97.5)	5.93 (5.89, 5.97)	0.38 (0.37, 0.39)
Digital	35,938 (2.5)	3.01 (2.83, 3.18)	0.21 (0.17, 0.26)
Women's age		$p<0.001$	$p<0.001$
44–49 years	290,413 (20.2)	9.41 (9.30, 9.51)	0.69 (0.66, 0.72)
50–54 years	348,750 (24.2)	6.33 (6.25, 6.41)	0.41 (0.39, 0.43)
55–59 years	345,162 (24)	4.68 (4.61, 4.75)	0.28 (0.26, 0.30)
60–64 years	311,433 (21.6)	4.26 (4.19, 4.33)	0.25 (0.23, 0.26)
65–69 years	144,626 (10)	3.81 (3.71, 3.91)	0.19 (0.16, 0.21)
Screen type		$p<0.001$	$p<0.001$
First	385,436 (26.8)	9.37 (9.27, 9.46)	0.76 (0.73, 0.79)
Successive	1,054,948 (73.2)	4.57 (4.53, 4.61)	0.24 (0.23, 0.25)

[a] *CI* Confidence Interval

A decreasing tendency in the risk of overall false-positive results was found as the reading volume in the previous year increased. Specific estimations of risk revealed a cut-off point above 10,000 readings in the previous year, with a lower limit of the confidence interval that did not overlap with any of the categories of less than 10,000 readings per year. The reduced risk of a false-positive result with greater reading volume was also observed with a similar magnitude for false-positives resulting in an invasive procedure but without a clearly differentiated cut-off point, as the confidence intervals of the OR overlapped between categories.

To a lesser extent than reading volume, radiologists' length of service in the screening programme also reduced the risk of a false-positive result. As with reading volume in the previous year, this reduction was of a similar magnitude for overall false-positive results and for false-positives leading to an invasive procedure. However, with overall false-positive results, the risk tended to decrease as the radiologist's length of service in the programme increased, which was reflected in smaller confidence intervals for the OR with little overlap between categories.

These results agree with those of other authors who found a reduced false-positive rate with greater radiologist experience measured as reading volume in the previous year [12, 13]. Smith-Bindman found wide variability in false-positive results among radiologists, with those reading between 2,500 and 4,000 mammograms per year, having approximately 50% fewer false-positive results than those interpreting between 481 and 750 mammograms yearly

Table 2 Effect of reading volume in the previous 365 days on the risk of overall false-positive (FP) results and FP results leading to an invasive procedure

Reading volume in the previous year	Overall FP				FP resulting in an invasive procedure			
	Univariate analysis (OR, 95% CI[a])	*p*	Multivariate[b] analysis (OR, 95% CI[a])	*p*	Univariate analysis (OR, 95% CI[a])	*p*	Multivariate[b] analysis (OR, 95% CI[a])	*p*
0–499	Ref		Ref		Ref		Ref	
500–1,999	0.77 (0.73, 0.81)	<0.001	0.77 (0.73, 0.81)	<0.001	0.79 (0.67, 0.93)	0.006	0.78 (0.66, 0.92)	0.004
2,000–4,999	0.70 (0.66, 0.73)	<0.001	0.71 (0.68, 0.75)	<0.001	0.76 (0.65, 0.90)	0.001	0.78 (0.66, 0.92)	0.003
5,000–9,999	0.74 (0.71, 0.79)	<0.001	0.76 (0.72, 0.80)	<0.001	0.71 (0.61, 0.83)	<0.001	0.75 (0.64, 0.87)	<0.001
10,000–14,999	0.61 (0.58, 0.64)	<0.001	0.62 (0.59, 0.65)	<0.001	0.50 (0.42, 0.58)	<0.001	0.56 (0.47, 0.65)	<0.001
>15,000	0.54 (0.52, 0.57)	<0.001	0.59 (0.57, 0.62)	<0.001	0.53 (0.46, 0.63)	<0.001	0.60 (0.51, 0.70)	<0.001

[a] *CI* Confidence Interval

[b] Adjusted by women's age, screen type (first or successive), number of views (1 or 2), mammogram type (analogue or digital) period effect and radiology unit (as a random effect)

[12]. Many radiologists working in screening programmes in Spain spend a large part of their working day interpreting screening mammograms, which in our study was reflected by the finding that a high percentage of screening mammograms (53.7%) were interpreted by a reading radiologist who had evaluated more than 10,000 mammography examinations in the previous year and 41.2% were evaluated by radiologists with more than 4 years' experience in the programme.

Other studies, although with aims distinct from our own, such as evaluation of the validity of mammography as a screening method, also found that mammographic accuracy increased with greater radiologist experience, measured as reading volume in the previous year [14, 20]. Other authors also found that radiologists' years of experience influenced diagnostic accuracy, measured as sensitivity and specificity and that greater experience lowered the false-positive rate but did not find the same association with reading volume in the previous year [21–23].

On the contrary, our results, although they also show a direct relationship between length of experience in the programme and a reduction in false-positive risk, establish that this effect is of greater magnitude for reading volume in the previous year, recent experience thus having a greater effect than cumulative experience. However, some studies have failed to find an association between mammographic accuracy and reading volume or years' experience and conclude that increasing volume requirements or experience of interpreting mammograms is unlikely to improve overall mammography performance [15–17].

The methodology used in other studies evaluating the effect of radiologist experience differs from that used in the present study. Some of these publications employed a sample of mammograms read by two radiologists with

Table 3 Effect of radiologists' length of experience in the screening programme on risk of overall false-positive (FP) results and FP results leading to an invasive procedure

Years' experience in the programme	Overall FP				FP leading to an invasive procedure			
	Univariate analysis (OR, 95% CI[a])	*p*	Multivariate[b] analysis (OR, 95% CI[a])	*p*	Univariate analysis (OR, 95% CI[a])	*p*	Multivariate[b] analysis (OR, 95% CI[a])	*p*
<1	Ref		Ref		Ref		Ref	
1	0.96 (0.94, 0.99)	0.006	0.96 (0.93, 0.99)	0.002	0.83 (0.76, 0.91)	<0.001	0.84 (0.77, 0.91)	<0.001
2	0.78 (0.76, 0.80)	<0.001	0.86 (0.84, 0.89)	<0.001	0.54 (0.48, 0.59)	<0.001	0.62 (0.56, 0.69)	<0.001
3	0.73 (0.71, 0.75)	<0.001	0.86 (0.83, 0.89)	<0.001	0.61 (0.55, 0.67)	<0.001	0.77 (0.69, 0.85)	<0.001
4	0.68 (0.66, 0.70)	<0.001	0.79 (0.77, 0.82)	<0.001	0.62 (0.55, 0.69)	<0.001	0.75 (0.67, 0.84)	<0.001
>4	0.64 (0.62, 0.65)	<0.001	0.72 (0.70, 0.74)	<0.001	0.61 (0.55, 0.67)	<0.001	0.73 (0.66, 0.80)	<0.001

[a] *CI* Confidence Interval

[b] Adjusted by women's age, screen type (first, successive), number of views (1 or 2) mammogram type (analogue or digital), period effect and radiology unit (as a random effect)

distinct experience to compare variability in the false-positive rate and other measures of diagnostic accuracy. This experimental context is far removed from routine practice in which the proportion of cases is much higher than that found in screening programmes [14, 16, 24]. The high prevalence of cases in the sample does not affect sensitivity or specificity but does reduce the false-positive rate to the extent that it affects the positive predictive value of mammography.

Our results, in contrast, are drawn from a retrospective cohort study analysing mammography examinations performed in routine practice in screening programmes, all of which were population-based and had common quality criteria defined in the European Guidelines for Quality Assurance, thus more closely reflecting the real risk of a false-positive result [19].

Another novel feature of this study is the distinction made between false-positive results and false-positives leading to an invasive procedure, given that invasive procedures cause greater psychological distress, a higher risk of complications due to the procedure and greater resource utilisation than non-invasive procedures.

As expected, we found that the percentage of false-positive results (both overall and those leading to invasive procedures) decreased as women's age increased. This percentage also decreased in successive rounds compared with the first round (Table 1). We also found that the percentage of false-positives was greater with screen-film than with digital mammography. Given the recent introduction of digital mammography in screening programmes, the tendency observed in this study should be investigated in future studies. The percentage of false-positive results overall decreased when two views were used rather than a single view. However, the opposite effect was found for false-positive results leading to invasive procedures, a finding for which we have no explanation.

Additionally, the effect of the experience-related variables studied in the univariate analysis was similar to that observed in the multivariate analysis. This finding indicates that radiologist experience has a clear effect on false-positive results and is independent of the adjustment variables (women's age, screening type, number of views, mammography type, radiology unit and period effect), allowing us to state the radiologist experience per se affects the risk of a false-positive result.

Another advantage of our study is that the measure of radiologist experience employed was based purely on objective data collected from screening programme databases, while most of the literature reviewed used other, less suitable methods such as questionnaires completed by the radiologists themselves, years of experience defined as the number of years since the radiologist became medically qualified or approximations according to radiologists' age [21, 23]. Moreover, the length of the period evaluated, 1990–2006, allowed us to identify the evolution of radiologists' experience, which was recalculated for each mammography examination studied, and to observe how this evolution reduced the risk of a false-positive result.

While the objectivity with which experience was measured is a strong point of our study, a weak point is that radiologist experience outside the screening programme was not taken into account. Thus radiologists' overall experience in mammogram interpretation may sometimes have been underestimated. The main aim of breast cancer screening programmes is to reduce the morbidity and mortality associated with this disease by minimising adverse effects so that the risk-benefit ratio is as favourable as possible. One of the most frequent and important adverse effects of screening programmes are false-positive results, causing women without cancer to undergo a series of additional tests, some of which can be invasive, to exclude a diagnosis of malignancy. This process has physical and psychological repercussions that would have been avoided if these women had not participated in the programme. Minimising false-positive results should be one of the main aims of screening programmes and consequently study of their determining factors is essential. Identification of these factors has clear implications for the organisational models of these screening programmes, which should guarantee their quality.

In conclusion, the results of this study highlight the importance of breast imaging specialists and of their length of service in screening programmes and establish indicators for the minimum reading volume per year. Compliance with these indicators would help to reduce false-positive results, thus favouring the risk-benefit ratio of these programmes.

Acknowledgements We would like to thank the regional screening-programme's staff who have provided data for the preparation of this study.

(1) CFPR (Cumulative False Positive Risk) group (alphabetical order): Department of Epidemiology and Evaluation, Institut Municipal d'Investigació Mèdica-Parc de Salut Mar, Barcelona:Francesc Belvis, Xavier Castells, Francesc Macià, Rubén Román, María Sala, José María Velarde Galician breast cancer screening programme. Public health & Planning Directorate. Health Office, Galicia: Raquel Almazán, Ana Belén Fernández, María Teresa Queiro, Raquel Zubizarreta. Navarra Breast Cancer Screening Programme Public Health Institute, Pamplona: Nieves Ascunce, María Ederra, Marcela Guevara, Juana Vidan. General Directorate Public Health & Centre for Public Health Research (CSISP), Valencia: Josefa Ibáñez, Dolores Salas. Valencian Health Agency & Centre for Public Health Research (CSISP), Valencia: Dolores Cuevas. Servicio Canario de la Salud, Canary Islands: María Obdulia De la Vega, Isabel Díez de la Lastra. Health Research Institute, Reus, Tarragona: Jaume Galceran. Program & Analysis Unit. Health Office, Asturias: Carmen Natal. La Rioja Breast Cancer Screening Programme. Fundacion Rioja Salud, Logroño: Araceli Baroja. Cancer Screening and Epidemiology Department, UDIAT-CD. Corporació Parc Taulí-Institut Universitari Parc Taul* (UAB),Sabadell: Marisa Baré. Castilla-Leon Breast Cancer Screening Programme. Dirección General de Salud Pública ID e I. SACYL. Castilla y León:Isabel González.

References

1. Deck W, Kakuma R (2006) Screening mammography: a reassessment. Montréal (Canada): Agence d'Évaluation des Technologies et des Modes d'Intervention en Santé (AETMIS). Available via http://collections.banq.qc.ca/ark:/52327/bs57227. Accessed 14 May 2011
2. Gotzsche PC, Nielsen M (2009) Screening for breast cancer with mammography. Cochrane Database Syst Rev 1:CD001877. doi:10.1002/14651858.CD001877.pub4
3. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE (2002) Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet 359:909–919
4. von Karsa L, Anttila A, Ronco G, et al (2008) Cancer screening in the European Union. Report on the Implementation of the Council Recommendation on cancer screening—First Report. European Commission (2008). Available via http://ec.europa.eu/health/ph_determinants/genetics/documents/cancer_screening.pdf. Accessed 25 Nov 2009
5. Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW (1998) Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med 338:1089–1096
6. Hofvind S, Thoresen S, Tretli S (2004) The cumulative risk of a false-positive recall in the Norwegian Breast Cancer Screening Program. Cancer 101:1501–1507
7. Council of the European Union (2003) Council Recommendation of 2 December 2003 on Cancer Screening (2003/878/EC) Off J Eur Union. 16 Dic 327:34–38
8. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V (1996) Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA 276:33–38
9. Carney PA, Miglioretti DL, Yankaskas BC et al (2003) Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 138:168–175
10. Lehman CD, White E, Peacock S, Drucker MJ, Urban N (1999) Effect of age and breast density on screening mammograms with false-positive findings. AJR Am J Roentgenol 173:1651–1655
11. Taplin S, Abraham L, Barlow WE et al (2008) Mammography facility characteristics associated with interpretive accuracy of screening mammography. J Natl Cancer Inst 100:876–887
12. Smith-Bindman R, Chu P, Miglioretti DL et al (2005) Physician predictors of mammographic accuracy. J Natl Cancer Inst 97:358–367
13. Kan L, Olivotto IA, Warren Burhenne LJ, Sickles EA, Coldman AJ (2000) Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. Radiology 215:563–567
14. Esserman L, Cowley H, Eberle C et al (2002) Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst 94:369–375
15. Barlow WE, Chi C, Carney PA et al (2004) Accuracy of screening mammography interpretation by characteristics of radiologists. J Natl Cancer Inst 96:1840–1850
16. Beam CA, Conant EF, Sickles EA (2003) Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. J Natl Cancer Inst 95:282–290
17. Molins E, Macia F, Ferrer F, Maristany MT, Castells X (2008) Association between radiologists' experience and accuracy in interpreting screening mammograms. BMC Health Serv Res 8:91
18. Castells X, Sala M, Ascunce N, Salas D, Zubizarreta RMC (2007) Descripción del cribado del cáncer en España. Proyecto DESCRIC. In: Plan de Calidad para el Sistema Nacional de Salud, ed.: Ministerio de Sanidad y Consumo. Agència d'Avaluació de Tecnologia i Recerca Mèdiques de Cataluña. Available via http://www.gencat.cat/salut/depsan/units/aatrm/pdf/in0601es.pdf. Accessed 14 May 2011
19. European Commission (2006) European guidelines for quality assurance in breast cancer screening and diagnosis, 4th edn. In: Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L, et al. (eds) Office for Official Publications of the European Communities, Luxembourg
20. Woodard DB, Gelfand AE, Barlow WE, Elmore JG (2007) Performance assessment for radiologists interpreting screening mammography. Stat Med 26:1532–1551
21. Elmore JG, Miglioretti DL, Reisch LM et al (2002) Screening mammograms by community radiologists: variability in false-positive rates. J Natl Cancer Inst 94:1373–1380
22. Miglioretti DL, Smith-Bindman R, Abraham L et al (2007) Radiologist characteristics associated with interpretive performance of diagnostic mammography. J Natl Cancer Inst 19 (99):1854–1863
23. Tan A, Freeman DH Jr, Goodwin JS, Freeman JL (2006) Variation in false-positive rates of mammography reading among 1067 radiologists: a population-based assessment. Breast Cancer Res Treat 100:309–318
24. Castells X, Molins E, Macia F (2006) Cumulative false positive recall rate and association with participant related factors in a population based breast cancer screening programme. J Epidemiol Community Health 60:316–321