Annals of Oncology

Annals of Oncology 23: 104–111, 2012
doi:10.1093/annonc/mdr032
Published online 23 March 2011

Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening

R. Román[1,2], M. Sala[1,2], D. Salas[3], N. Ascunce[2,4], R. Zubizarreta[5] & X. Castells[1,2,6]* Cumulative False Positive Risk Group

[1]Department of Epidemiology and Evaluation, Institut Municipal d'Investigació Mèdica-Parc de Salut Mar Barcelona; [2]CIBER de Epidemiología y Salud Pública (CIBERESP), Barcelona; [3]General Directorate of Public Health and Centre for Public Health Research, Valencia; [4]Navarra Breast Cancer Screening Programme, Public Health Institute, CIBERESP, Pamplona; [5]Galician Breast Cancer Screening Programme, Public Health and Planning Directorate, Health Office, Santiago de Compostela; [6]Department of Pediatrics, Obstetrics and Gynecology, Preventive Medicine and Public Health, Universitat Autònoma de Barcelona (UAB), Bellaterra (Barcelona), Spain

Received 23 November 2010; revised 17 January 2011; accepted 19 January 2011

Background: Reducing the false-positive risk in breast cancer screening is important. We examined how the screening-protocol and women's characteristics affect the cumulative false-positive risk.

Methods: This is a retrospective cohort study of 1 565 364 women aged 45–69 years who underwent 4 739 498 screening mammograms from 1990 to 2006. Multilevel discrete hazard models were used to estimate the cumulative false-positive risk over 10 sequential mammograms under different risk scenarios.

Results: The factors affecting the false-positive risk for any procedure and for invasive procedures were double mammogram reading [odds ratio (OR) = 2.06 and 4.44, respectively], two mammographic views (OR = 0.77 and 1.56,

*Correspondence to: Dr X. Castells, Department of Epidemiology and Evaluation, Mar Teaching Hospital, 25-29 Passeig Marítim, 08003 Barcelona, Spain. Tel: +34-93-248-32-88; Fax: +34-93-248-32-54; E-mail: xcastells@parcdesalutmar.cat

respectively), digital mammography (OR = 0.83 for invasive procedures), premenopausal status (OR = 1.31 and 1.22, respectively), use of hormone replacement therapy (OR = 1.03 and 0.84, respectively), previous invasive procedures (OR = 1.52 and 2.00, respectively), and a familial history of breast cancer (OR = 1.18 and 1.21, respectively). The cumulative false-positive risk for women who started screening at age 50–51 was 20.39% [95% confidence interval (CI) 20.02–20.76], ranging from 51.43% to 7.47% in the highest and lowest risk profiles, respectively. The cumulative risk for invasive procedures was 1.76% (95% CI 1.66–1.87), ranging from 12.02% to 1.58%.

Conclusions: The cumulative false-positive risk varied widely depending on the factors studied. These findings are relevant to provide women with accurate information and to improve the effectiveness of screening programs.

Key words: breast cancer, false positive, invasive procedures, risk factors, screening, variability

introduction

Reducing the false-positive risk, and therefore its associated factors, is a major goal of breast cancer screening as it would improve the balance of benefits and harms of this preventive modality [1]. The negative effects of false-positive results have been widely described and include anxiety, additional physician visits and diagnostic tests, and excision biopsies [2, 3] and may also affect adherence to subsequent mammographic screening [4].

The benefit of screening is usually measured as mortality reduction after participation in several screening rounds, while the false-positive risk is usually assessed for each round, thus underestimating the cumulative negative effect of participation in several rounds. Some studies have estimated the cumulative risk of a false-positive result during a woman's life span ranging from 20% to 50% after 10 screening rounds [5–10]. These estimates were based on different methodologies but the wide variation observed could also be explained by differences in the screening setting (opportunistic or population based with quality standards) and in the cohort of women analyzed.

False-positive recall rates may be affected by screening-protocol characteristics that are potentially modifiable, such as double or single mammogram reading [11, 12], the type of mammography (digital or film-screen) [13] and the number of images taken [14]. Other factors affecting these rates are women's personal characteristics, such as age, use of hormone replacement therapy (HRT), and a familial history of breast cancer.

A false-positive result leading to an invasive procedure (fine-needle aspiration, core biopsy, and open biopsy) produces greater anxiety in women and a higher cost to the health system than additional imaging tests. The association between false-positive determinants and whether invasive or noninvasive procedures are carried out has not been sufficiently evaluated. This evaluation would provide greater knowledge of breast cancer screening and its distinguishing features.

The aim of this study was to estimate the cumulative false-positive risk for all procedures and for invasive procedures throughout the period of participation in a population-based breast cancer screening program and to determine the effect of women's personal variables and screening-protocol characteristics on this risk.

methods

setting

All women resident in Spain aged 50–69 are actively invited to participate in the population-based screening program by written letter every 2 years. A screening mammogram is offered, allowing women who begin screening at 50–51 years up to a maximum of 10 screening mammograms. Breast cancer screening in Spain adheres to the European Guidelines for Quality Assurance in Mammographic Screening [15] and its results meet the required standards [16, 17]. Each of the 17 administrative regions in Spain is responsible for the local application of the screening program in its area. Population-based breast cancer screening in Spain started in 1990 in one region and became nationwide in 2006. Data from eight regions, representing 44% of the Spanish target population in 2005, were collected. The selection criterion for including the regions in the study was completion of at least three screening rounds by December 2006. Each region has one or several radiology units that carry out screening. Local application of the screening program can vary in the target population and in the mammographic screening protocol used [17].

This study included variables related to the mammographic screening-protocol and women's personal characteristics. All information was collected from each participant at each attendance. The variables related to the screening protocol included the number of views [one (craniocaudal) or two (mediolateral oblique and craniocaudal) images were taken for each breast], reading method (single reading by one radiologist or double reading by two radiologists, with or without consensus or arbitration), and mammography type (film-screen or digital). The variables related to women's personal characteristics were age, use of HRT at screening or in the previous 6 months, menopausal status (pre- or postmenopausal), previous invasive procedures with a benign result, and the presence or absence of a first-degree familial history of breast cancer. In some regions, however, data on women's personal variables were either not routinely gathered or data collection did not meet the protocol's requirements before the specified date.

study population

A total of 1 586 762 eligible women participated in at least one screening round in any of the eight regions from March 1990 to December 2006 (Table 1). These women underwent a total of 4 797 609 screening mammograms. However, 19 055 women were excluded because their mammographic screening result was unknown, 2246 because their age at first screening was not in the 44- to 69-year interval, and 97 because their age was unknown. The total number of screened women analyzed was 1 565 364, with 4 739 498 mammographic screening tests carried out in 74 distinct radiology units.

definition of a false-positive result

Women with a positive mammographic reading were recalled for further assessments. A positive mammogram reading was considered a false-positive result if, after further assessments, breast cancer was not diagnosed. Additional evaluation to rule out malignancy included both noninvasive (additional mammography, magnetic resonance imaging, ultrasonography, etc.) and invasive procedures (fine-needle aspiration cytology, core-needle biopsy and open surgical biopsy). The diagnostic work-up for further assessments took place within a maximum of 2 months after screening.

Women with a negative result (at mammographic reading or after further assessments) were recalled for a new screening mammography 24 months after the previous screen. A definitive diagnosis of breast cancer was always histopathologically confirmed (invasive ductal carcinoma or carcinoma *in situ*).

Two definitions of false-positive results were used: false-positive results leading to any procedure (noninvasive and/or invasive further assessments) and false-positive results leading to invasive procedures (at least one invasive further assessment was carried out). Screening mammograms repeated due to insufficient technical quality (<0.2%) were not included as a positive result.

statistical analysis

To calculate the risk of a false-positive result and of cancer detection, discrete-time hazard models were fitted, as described in detail by Singer and Willett [18]. This methodology uses a logistic regression approach to compute these particular survival models with discrete time intervals. Two sorts of predictors were introduced in the model: 'time indicators', given by the women's screening round (acting as multiple intercepts), and 'substantive predictors' for the effect of covariates on the model. The event of interest was defined as the occurrence of a first false-positive result. Subsequent observations were censored in the statistical models to avoid correlation among repeated participations. As data were collected at each attendance, time-changing variables could be included in the models. The models were adjusted by a time period effect (calendar years) as the start date of the radiology units differed. To improve interpretation of the results of the regression models in terms of the risks and benefits of screening, the breast cancer detection model was also included in the tables.

The radiology unit was introduced as a random effect in the models because of the correlation structure among observations in the same radiology unit. The GLIMMIX procedure in SAS 9.1 (SAS Institute, Cary, NC) was used. The models had a multilevel structure component in which mammographic screenings (level 1) were nested within radiology units (level 2, random effect). Residual pseudo-likelihood estimation was used in all the models. Two models were computed to ascertain the effect of substantive predictors. A full database model with the screening-protocol variables was computed as this information was always available. The model was then extended by adding women's personal variables with the subset of screening mammograms for which this information was complete. This subset accounted for 2 777 429 (58.6%) screening mammograms from 45 radiology units. To evaluate possible differences between the initial study population and the subset with complete information, we compared the overall false-positive rate and the age distribution among missing and non-missing data for each personal variable (see supplemental Appendix 1, available at *Annals of Oncology* online). Univariate analysis carried out to evaluate the collinearity of women's personal variables showed a stable association of these factors with the false-positive risk.

cumulative risk of a false-positive result

The false-positive risk was projected forward to 10 screening mammograms for women aged 50–51 years at their first screening round. This 10-screening projection allowed us to ascertain the risk of a false-positive result for the entire period women are invited to participate in screening programs. Projections were carried out assuming that the hazard of the 7th to 10th mammograms was similar to that of the 6th mammogram. Mammograms from the 7th to 10th screening were not used for projection because they represented only 2% of overall screening mammograms and this information was only available in 12 of the 74 participating radiology units. From the estimated risk at each screening mammogram obtained from the regression models, cumulative risk was calculated as the risk for each screening mammogram multiplied by the proportion of women without a false-positive result up to that screening; the cumulative risk up to the previous screening mammogram was then added. Confidence intervals (CIs) for the cumulative risk of a false positive were calculated using Greenwood's approximation [19].

Two extreme risk profiles were defined for projection based on the results of multivariate analysis. The highest risk profile was defined as a woman with all the factors associated with an increased false-positive risk. The lowest risk profile was defined as a woman without any of the factors corresponding to increased risk.

results

A total of 4 739 498 screening mammograms carried out in 1 565 364 women were analyzed (see Table 1). Of these participating women, 1 205 943 (77.04%) had a second

Table 1. Screening information description by screening period

	1990–1992, *n* (%)	1993–1994, *n* (%)	1995–1996, *n* (%)	1997–1998, *n* (%)	1999–2000, *n* (%)	2001–2002, *n* (%)	2003–2004, *n* (%)	2005–2006, *n* (%)	Total, *n*
Screening tests	67 806 (1.4)	233 407 (4.9)	371 033 (7.8)	485 800 (10.3)	714 981 (15.1)	849 415 (17.9)	932 861 (19.7)	1 084 195 (22.9)	4 739 498
Women screened (first screening)	61 746 (3.9)	178 245 (11.4)	198 190 (12.7)	198 721 (12.7)	298 747 (19.1)	240 817 (15.4)	196 470 (12.6)	192 428 (12.3)	1 565 364
Screening test (subsequent screening)	6060 (0.2)	55 162 (1.7)	172 843 (5.4)	287 079 (9.0)	416 234 (13.1)	608 598 (19.2)	736 391 (23.2)	891 767 (28.1)	3 174 134
Further assessments	13 037 (3.4)	24 013 (6.4)	35 070 (9.3)	41 886 (11.1)	68 603 (18.1)	64 991 (17.2)	63 945 (16.9)	66 515 (17.6)	378 060
Women with a FP[a]	10 175 (3.9)	18 992 (7.2)	27 727 (10.5)	30 077 (11.4)	46 024 (17.5)	43 707 (16.6)	42 278 (16.0)	44 627 (16.9)	263 607
Women with a FP (invasive)[b]	566 (2.3)	2532 (10.4)	2471 (10.1)	3075 (12.6)	4511 (18.5)	4259 (17.4)	3687 (15.1)	3306 (13.5)	24 407
Radiology units[c]	9	21	33	41	63	68	71	74	74

[a]An FP result for any procedure (invasive or noninvasive).
[b]An FP result for an invasive procedure.
[c]Expressed as number of radiology units running in that screening period.
FP, false positive.

screening mammogram, 867 160 (55.40%) had a third and 156 414 (9.99%) a sixth. Mammographic screenings were carried out by 74 distinct radiology units, with an average of 64 047 screening tests (10th to 90th percentile: 9159–117 988) and 21 154 women screened per radiology unit (10th to 90th percentile: 3424–38 268).

Of the 1 565 364 women who participated in at least one screening round, 467 910 were first screened at 44–49 years, 477 177 at 50–54 years, 300 901 at 55–59 years, 260 223 at 60–64 years, and 59 153 at 65–69 years. Table 2 shows the false-positive rate for all procedures and for invasive tests and the cancer detection rate for first and subsequent screening mammograms.

Adjusted odds ratios (ORs) for the false-positive risk for all procedures, false-positive risk for invasive procedures and the cancer detection rate related to the screening-protocol variables are shown in Table 3. Double reading mammograms conferred a higher risk (OR = 2.06; 95% CI 2.00–2.13) than single reading. This risk was higher for invasive procedures (OR = 4.44; 95% CI 4.08–4.84). Two mammographic views had a protective effect for the false-positive risk for all procedures (OR = 0.77; 95% CI 0.76–0.79) but was a risk factor for the false-positive risk for invasive procedures (OR = 1.56; 95% CI 1.48–1.64). Digital mammography had a protective effect on the false-positive risk for invasive procedures (OR = 0.83; 95% CI 0.72–0.96), but this effect was not statistically significant for the false-positive risk for all procedures.

The model including the women's personal variables is shown in Table 4. A higher risk for the false-positive risk for all procedures and false-positive risk for invasive procedures was observed in the youngest women (OR = 1.50; 95% CI 1.46–1.54 and OR = 1.44; 95% CI 1.30–1.58), women with previous invasive procedures (OR = 1.52; 95% CI 1.49–1.56 and OR = 2.00; 95% CI 1.89–2.12), a familial history of breast cancer (OR = 1.18; 95% CI 1.15–1.20 and OR = 1.21; 95% CI 1.13–1.30) and premenopausal women (OR = 1.31; 95% CI 1.29–1.33 and OR = 1.22; 95% CI 1.16–1.29). HRT conferred a lower false-positive risk for invasive procedures (OR = 0.84; 95% CI 0.78–0.90).

The overall cumulative risk of a false-positive result for all procedures and for invasive procedures in women aged 50–51 years at the first screening when projected forward to the 10th screening was 20.39% (95% CI 20.02–20.76) and 1.76% (95% CI 1.66–1.87), respectively. Figures 1 and 2 show the estimated cumulative risk for women aged 50–51 years, with the highest and lowest risk profiles. The cumulative risk after 10 consecutive rounds in high-risk women was estimated at 51.43% (95% CI 51.02–51.84), while women without these risk factors had an estimated risk of 7.47% (95% CI 7.23–7.72) (Figure 1). The differential risk between the highest and the lowest risk profiles was 43.96%. Protocol characteristics explained 54.2% of this differential risk, while women's personal characteristics explained the remaining 45.8%. The cumulative risk of a false-positive result for invasive procedures in high-risk women was 12.02% (95% CI 11.75–12.30) while that in the lowest risk group was 1.58% (95% CI 1.48–1.69) (Figure 2). The differential risk between the highest and the lowest risk profiles was 10.44%. Women's personal characteristics explained 73.3% of this differential risk.

discussion

Estimation of the cumulative risk of a false positive aims to provide the maximum available information to women invited to participate in breast cancer screening. Nowadays, false-positive results are a noteworthy adverse effect of screening. If mortality reduction as a benefit of screening is analyzed in terms of a sequence of multiple screening participations, adverse effects should be studied in a similar way.

We estimated that one in every five women who participated in 10 screening rounds had a false-positive result. These results are consistent with findings in Norway [7] and the UK [20], where screening programs' organization is similar, but are much lower than the 49.1% observed in the United States [6, 10]. These differences were also observed in a comparison between the United States and the UK [20]. An explanation for these findings could be that breast cancer screening in the United States is not government sponsored and organized, whereas in Europe programs must meet quality standards involving lower false-positive rates [14,20–22].

Importantly, the cumulative risk of a false-positive result involving a biopsy or other invasive procedures was 10-fold or less lower than for any procedure. Despite its lower risk, the adverse effect of a false-positive result leading to an invasive procedure is higher in terms of the physical impact to women and involves a higher cost than imaging procedures and a delay in informing women of the results.

Previous studies have found a higher cumulative risk of a false-positive result leading to invasive procedures [7, 20] in the European context and an even higher risk in the United States [6, 20]. However, further studies are required to analyze the variability found in the estimated cumulative risk within the European context.

Several factors have previously been described as influencing the false-positive recall rate, including the reading method, the number of mammographic views, mammogram quality and the radiologist experience [23–25]. In line with the results of several previous studies [21, 26, 27] we found that double reading was associated with a higher recall rate (OR = 2.06) and a higher

Table 2. False positives and cancer detection outcomes (by screening mammogram)

Outcome	First screening		Subsequent screening		Overall	
	n	Percentage (95% CI)	*n*	Percentage (95% CI)	*n*	Percentage (95% CI)
False positive	134 757	8.6 (8.56–8.65)	130 044	4.10 (4.08–4.12)	264 801	5.59 (5.57–5.61)
False positive (invasive)	15 894	1.02 (1.00–1.03)	8542	0.27 (0.26–0.28)	24 436	0.52 (0.51–0.52)
Cancer detection	7065	0.45 (0.44–0.46)	9464	0.30 (0.29–0.30)	16 529	0.35 (0.34–0.35)

CI, confidence interval.

Table 3. False-positive risk and cancer detection by screening-protocol characteristics (N = 4 739 498)

	Screening mammograms	Multivariate analysis (OR, 95% CI)[a]		
		False-positive risk (all procedures)	False-positive risk (invasive procedures)	Cancer detection
Reading method				
Single reading	1 734 930	Ref.	Ref.	Ref.
Double reading	3 004 568	2.06 (2.00–2.13)[b]	4.44 (4.08–4.84)[b]	1.08 (1.04–1.12)[b]
Number of views				
One	1 482 503	Ref.	Ref.	Ref.
Two	3 256 995	0.77 (0.76–0.79)[b]	1.56 (1.48–1.64)[b]	1.02 (0.97–1.06)
Mammography type				
Film-screen	4 676 138	Ref.	Ref.	Ref.
Digital	63 360	0.96 (0.92–1.01)	0.83 (0.72–0.96)[b]	1.26 (1.10–1.45)[b]

[a]Multivariate analysis adjusted by women's screening number, radiology unit (random effect), screening period and age.
[b]Significant at the 95% CI.
OR, odds ratio; CI, confidence interval.

Table 4. False-positive risk and cancer detection by women's characteristics (adjusted by screening-protocol characteristics) (N = 2 777 429)

	Screening mammograms	Multivariate analysis (OR, 95% CI)[a]		
		False-positive risk (all procedures)	False-positive risk (invasive procedures)	Cancer detection
Age at screening (years)				
44–49	469 047	1.50 (1.46–1.54)[b]	1.44 (1.30–1.58)[b]	0.39 (0.35–0.43)[b]
50–54	699 256	1.26 (1.23–1.29)[b]	1.26 (1.15–1.37)[b]	0.48 (0.44–0.52)[b]
55–59	695 921	1.13 (1.10–1.16)	1.06 (0.97–1.16)	0.67 (0.62–0.73)[b]
60–64	633 845	1.06 (1.03–1.09)	0.96 (0.88–1.06)	0.84 (0.77–0.90)[b]
65–69	279 360	Ref.	Ref.	Ref.
HRT				
No	2 485 550	Ref.	Ref.	Ref.
Yes	291 879	1.03 (1.01–1.05)[b]	0.84 (0.78–0.90)[b]	0.86 (0.80–0.94)[b]
Menopause				
Menopausal	2 157 627	Ref.	Ref.	Ref.
Premenopausal	619 802	1.31 (1.29–1.33)[b]	1.22 (1.16–1.29)[b]	1.16 (1.07–1.25)[b]
Previous invasive procedure				
No	2 585 871	Ref.	Ref.	Ref.
Yes	191 558	1.52 (1.49–1.56)[b]	2.00 (1.89–2.12)[b]	1.31 (1.20–1.42)[b]
Familial breast cancer				
No	2 581 981	Ref.	Ref.	Ref.
Yes	195 448	1.18 (1.15–1.20)[b]	1.21 (1.13–1.30)[b]	1.66 (1.55–1.79)[b]

Menopause: pre-/perimenopausal or menopausal status; previous invasive procedure: personal previous invasive procedure; familial breast cancer: first-degree familial history of breast cancer previously described.
[a]Multivariate analysis adjusted by women's screening number, screening period, radiology unit (random effect) and reading-protocol variables (reading method, number of views, mammography type).
[b]Significant at the 95% CI.
OR, odds ratio; CI, confidence interval; HRT, hormone replacement therapy.

cancer detection rate (OR = 1.08) than single reading. However, there is a wide variability in the balance found in previous studies between the risk and the benefits of double reading over single reading [11, 14, 21, 22].

Some studies have reported that the increase in recall rate associated with double reading was reduced when consensus or arbitration was used over non-consensus double reading [11, 12, 28]. In our study, although the use of consensus and arbitration did not constitute study variables, 84.8% of double readings involved consensus or arbitration, while only 15.2% were double readings without consensus.

Although the European guidelines recommend two views, in our study some radiology units carried out one view, mainly for first screening. Our results are in agreement with those of previous studies that the use of two views reduces the false-positive risk for all procedures [14], but we also found that the use of two views increased the false-positive risk for invasive procedures. We observed a higher detection rate and a lower risk of false-positive results with digital mammography. A higher detection rate in younger women has been previously described [29, 30], while a reduction in overall false-positive rates has been found in some studies [13, 31] but not in others [32].

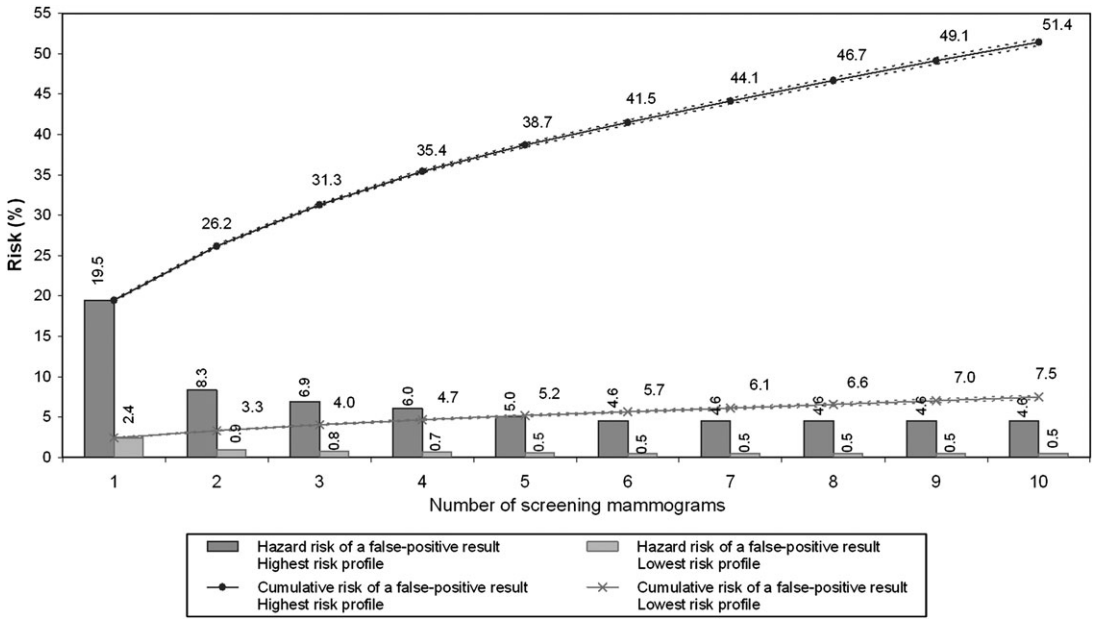

Figure 1. Cumulative risk and hazard risk of a false-positive result for any procedure for women starting screening at age 50–51 years. Highest risk (double reading, one view, film-screen mammography, premenopausal status, previous invasive procedures, and familial breast cancer) versus lowest risk profiles (opposite categories).

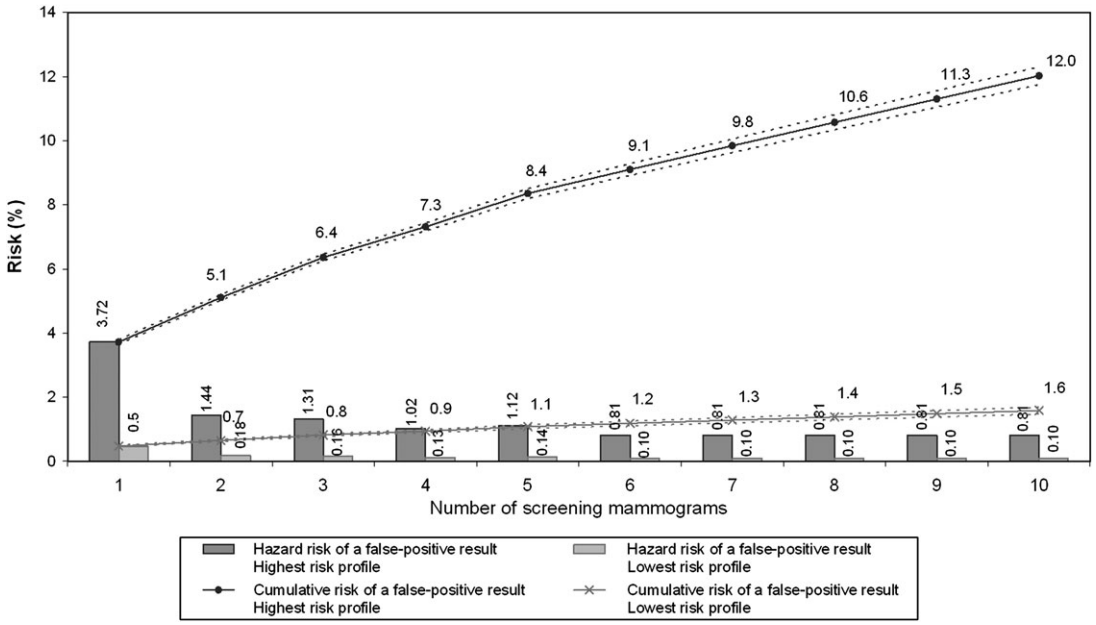

Figure 2. Cumulative risk and hazard risk of a false-positive result for invasive procedures for women starting screening at age 50–51 years. Highest risk (double reading, two views, not using HRT, premenopausal status, previous invasive procedures, and familial breast cancer) versus lowest risk profiles (opposite categories). HRT, hormone replacement therapy.

Our results on the influence of women's characteristics are in agreement with those of previous studies. The risk of a false-positive result is higher in younger women, adjusted by screening round, which probably reflects certain age-related features such as breast density, which we could not study because information on this factor is not routinely collected. HRT use was not associated with a higher false-positive risk, which seems contradictory given the relationship of this treatment with breast density and breast cancer. However, this finding might be explained by the lower use of the combination of estrogens plus progestin, which is associated with breast density [33, 34], in Spain compared with current recommendations in other European countries [35]. As expected, previous invasive procedures and familial breast cancer were also risk factors both for false-positive risk for all procedures and false-positive risk for invasive procedures.

A wide range was observed in the estimated cumulative risk of a false-positive result among the different risk profiles

defined, based on women's personal and protocol-related characteristics. The false-positive risk over 10 screening rounds for the highest and the lowest risk profiles ranged from 51.4% to 7.5% (maximum–minimum ratio: 6.8). The reading-protocol variables were responsible for over half of the risk range between the highest and the lowest risk profiles. A similar proportion in the range (1.58% to 12.0%) was observed for invasive procedures (ratio: 7.6). The lowest risk value obtained (1.58%) was close to the estimated baseline risk (1.76%) due to the small impact of the protective factors obtained from the regression models. Women's characteristics played a major role and explained 73.3% of this variability. Obviously, women's personal factors, except HRT use, are unmodifiable, but evaluating its impact provides essential information about the risk–benefit balance of breast cancer screening.

This study has some limitations. The information on women's personal variables was not always available or complete in all the radiology units. Although the age distribution between missing and non-missing data related to women's variables was similar, we found a moderately lower false-positive risk for all procedures and a moderately higher false-positive risk for invasive procedures in missing data. We analyzed a subsample with the maximum available information, which allowed us to control for reading-protocol and women's characteristics together. Information on radiologist experience inside and outside the program could not be obtained. The European guidelines recommend that radiologists read at least 5000 mammograms/year and most of the radiologists reading within the screening program achieved this volume.

In conclusion, our study uses information from a screening program with distinct screening protocols and at different stages of development and experience, this being one of the largest cohorts of screened women ever analyzed. We found that the screening-protocol and women's characteristics strongly affected the cumulative risk of a false positive for all procedures and for invasive procedures after 10 screening mammograms. Understanding the sources of variability may lead to more effective screening programs. The adverse effects of cancer screening could be reduced by taking modifiable variables into account when the risks and benefits of screening are analyzed and more accurate information could be provided to participating women.

acknowledgements

The authors would like to thank the entire 'Cumulative False Positive Risk Group' for their support and dedication to this study. Cumulative False Positive Risk Group (alphabetical order): Almazán R; Ascunce N; Barcos A; Baré M; Baroja A; Belvis F; Castells X; Cuevas D; De la Vega M; Díez de la Lastra I; Ederra M; Erdozain N; Fernández AB; Galceran J; González-Román I; Ibáñez J; Macià F; Natal C; Queiro MT; Román R; Sala M; Salas D; Velarde JM; Zubizarreta R. The funding sources had no role in the performance of the study or in the preparation of the manuscript.

funding

Instituto de Salud Carlos III-FEDER (PI061230).

disclosure

The authors declare no conflict of interest.

references

1. Kalager M, Zelen M, Langmark F, Adami HO. Effect of screening mammography on breast-cancer mortality in Norway. N Engl J Med 2010; 363: 1203–1210.
2. Lerman C, Trock B, Rimer BK et al. Psychological and behavioral implications of abnormal mammograms. Ann Intern Med 1991; 114: 657–661.
3. Bull AR, Campbell MJ. Assessment of the psychological impact of a breast screening programme. Br J Radiol 1991; 64: 510–515.
4. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med 2007; 146(7): 502–510.
5. Barratt A, Howard K, Irwig L et al. Model of outcomes of screening mammography: information to support informed choices. BMJ 2005; 330: 936.
6. Elmore JG, Barton MB, Moceri VM et al. Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med 1998; 338: 1089–1096.
7. Hofvind S, Thoresen S, Tretli S. The cumulative risk of a false-positive recall in the Norwegian Breast Cancer Screening Program. Cancer 2004; 101: 1501–1507.
8. Castells X, Molins E, Macia F. Cumulative false positive recall rate and association with participant related factors in a population based breast cancer screening programme. J Epidemiol Community Health 2006; 60: 316–321.
9. Njor SH, Olsen AH, Schwartz W et al. Predicting the risk of a false-positive test for women following a mammography screening programme. J Med Screen 2007; 14: 94–97.
10. Christiansen CL, Wang F, Barton MB et al. Predicting the cumulative risk of false-positive mammograms. J Natl Cancer Inst 2000; 92: 1657–1666.
11. Ciatto S, Ambrogetti D, Risso G et al. The role of arbitration of discordant reports at double reading of screening mammograms. J Med Screen 2005; 12: 125–127.
12. Duijm LE, Groenewoud JH, Fracheboud J et al. Introduction of additional double reading of mammograms by radiographers: effects on a biennial screening programme outcome. Eur J Cancer 2008; 44: 1223–1228.
13. Sala M, Comas M, Macia F et al. Implementation of digital mammography in a population-based breast cancer screening program: effect of screening round on recall rate and cancer detection. Radiology 2009; 252: 31–39.
14. Elmore JG, Nakano CY, Koepsell TD et al. International variation in screening mammography interpretations in community-based programs. J Natl Cancer Inst 2003; 95: 1384–1393.
15. Perry N, Broeders M, de Wolf C et al. European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis. Luxembourg: Office for Official Publications of the European Communities 2006.
16. Von Karsa L, Anttila A, Ronco G et al. Cancer screening in the European Union. Report on the implementation of the Council Recommendation on cancer screening – First Report. Luxembourg: Services of the European Commission 2008.
17. Ascunce N, Salas D, Zubizarreta R et al. Cancer screening in Spain. Ann Oncol 2010; 21 (Suppl 3): iii43–iii51.
18. Singer J, Willett JB. Fitting basic discrete-time hazard models. In Applied Longitudinal Data Analysis: Modelling Change and Event Occurrence. New York: Oxford University Press 2003; 357–406.
19. Singer J, Willett JB. Describing discrete-time event occurrence data. In Applied Longitudinal Data Analysis: Modelling Change and Event Occurrence. New York: Oxford University Press 2003; 325–356.
20. 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: 50–54.
21. Hofvind S, Vacek PM, Skelly J et al. Comparing screening mammography for early breast cancer detection in Vermont and Norway. J Natl Cancer Inst 2008; 100: 1082–1091.
22. 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: 187–193.
23. Esserman L, Cowley H, Eberle C et al. Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst 2002; 94: 369–375.

24. Kan L, Olivotto 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: 563–567.
25. 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: 282–290.
26. Harvey SC, Geller B, Oppenheimer RG et al. Increase in cancer detection and recall rates with independent double interpretation of screening mammography. Am J Roentgenol 2003; 180: 1461–1467.
27. Kopans DB. Double reading. Radiol Clin North Am 2000; 38: 719–724.
28. Shaw CM, Flanagan FL, Fenlon HM, McNicholas MM. Consensus review of discordant findings maximizes cancer detection rate in double-reader screening mammography: Irish National Breast Screening Program Experience. Radiology 2009; 250: 354–362.
29. Pisano ED, Gatsonis C, Hendrick E et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005; 353: 1773–1783.
30. Pisano ED, Hendrick RE, Yaffe MJ et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology 2008; 246: 376–383.
31. Del Turco MR, Mantellini P, Ciatto S et al. Full-field digital versus screen-film mammography: comparative accuracy in concurrent screening cohorts. Am J Roentgenol 2007; 189: 860–866.
32. Vigeland E, Klaasen H, Klingen TA et al. Full-field digital mammography compared to screen film mammography in the prevalent round of a population-based screening programme: the Vestfold County Study. Eur Radiol 2008; 18: 183–191.
33. Topal NB, Ayhan S, Topal U, Bilgin T. Effects of hormone replacement therapy regimens on mammographic breast density: the role of progestins. J Obstet Gynaecol Res 2006; 32: 305–308.
34. Noh JJ, Maskarinec G, Pagano I et al. Mammographic densities and circulating hormones: a cross-sectional study in premenopausal women. Breast 2006; 5: 20–28.
35. Benet M, Carvajal A, Garcia Del Pozo J et al. [Hormonal replacement therapy in Spain] Tratamiento Hormonal Sustitutivo En Espana. Un Estudio De Utilizacion. Med Clin (Barc) 2002; 119: 4–8.