TITLE:

Risk of breast cancer two years after a benign biopsy depends on the mammographic feature prompting recall

CONTRIBUTING AUTHORS: Maria Vernet-Tomás, MD, PhD[a], Javier Louro[a,b], MSc, Marta Román, PhD[a,b], Francina Saladié, MD[c], Margarita Posso, PhD[a,b], Miguel Prieto, MD[d], Ivonne Vázquez, MD[a], Marisa Baré, MD[e], Lupe Peñalva, MD[f], Carmen Vidal, MD[g], Xavier Bargalló, MD, PhD[h], Mar Sánchez, MD[i], Joana Ferrer, MD[j], Josep A Espinàs, MD, PhD[k], M Jesús Quintana, MD, PhD[l], Ana Rodríguez-Arana, MD[a], Xavier Castells, MD, PhD[a,b], on behalf of the BELE study group.

a. Institut Hospital del Mar d'Investigacions Mèdiques (IMIM). Dr. Aiguader 88, 08003 Barcelona, Spain
b. Research Network on Health Services in Chronic Diseases (REDISSEC), Barcelona, Spain
c. Fundació Lliga per a La Investigació i Prevenció del Càncer (FUNCA). Avinguda Josep Laporte, 2,43204 Reus, Spain
d. Consejería de Sanidad, Gobierno de Asturias. Calle Ciriaco Miguel Vigil, 9, 33005 Oviedo, Spain
e. Consorci Corporació Sanitaria Parc Taulí. Parc Taulí, 1, 08208 Sabadell, Spain
f. Hospital General de Granollers. Av. Francesc Ribas, s/n, 08402 Granollers, Spain
g. Programa de Prevenció i Control del Càncer de l'Institut Català d'Oncologia. Gran Via de l'Hospitalet, 199-203, 08908 L'Hospitalet de Llobregat, Spain
h. Centro de Diagnóstico por la Imagen Clínic (CDIC) del Hospital Clínic de Barcelona. Calle Villarroel 170, 08036 Barcelona, Spain
i. Dirección General de Salud Pública del Gobierno de Cantabria. C/ Federico Vial 13, 39009 Santander, Spain
j. Hospital de Santa Caterina, Carrer del Dr. Castany, s/n, 17190 Salt, Girona, Spain
k. Pla Director d'Oncologia del Departament de Salut de la Generalitat de Catalunya, Travessera de les Corts, 131-159, 08028 Barcelona, Spain
l. Departament d'Epidemiologia de l'Hospital de la Santa Creu i Sant Pau, c/ San Antoni M. Claret 167 08025 Barcelona, Spain

CORRESPONDING AUTHOR:

Maria Vernet-Tomás, MD, PhD, on behalf of the research team

Breast Functional Unit, Hospital del Mar, Barcelona (Spain)

Passeig Maritim 25-29, 08003 Barcelona, Spain

Tel: 00 34 932483129

e-mail: mvernet@hospitaldelmar.cat

Conflict of interest: The authors declare no potential conflicts of interest.

ABSTRACT:

Objective: We aimed to explore whether the type of mammographic feature prompting a false-positive recall (FPR) during mammography screening influences the risk and timing of breast cancer diagnosis, particularly if assessed with invasive procedures.

Study design: We included information on women screened and recalled for further assessment in Spain between 1994 and 2015, with follow-up until 2017, categorizing FPRs by the assessment (noninvasive or invasive) and mammographic feature prompting the recall.

Main outcome measures: Breast cancer rates in the first two years after FPR (first period) and after two years (second period).

Results: The study included 99,825 women with FPRs. In both periods, the breast cancer rate was higher in the invasive assessment group than in the noninvasive group (first period 12 ‰ vs 1.9 ‰, p< 0.001; second period 4.4‰ vs 3.1‰, p<0.001). During the first period, the invasive assessment group showed diverse breast cancer rates for each type of mammographic feature, with a higher rate for asymmetric density (31.9‰). When the second period was compared with the first, the breast cancer rate decreased in the invasive assessment group (from 12‰ to 4.4‰, p<0.001) and increased in the noninvasive assessment group (from 1.9‰ to 3.1‰, p<0.001).

Conclusion: In the context of mammography screening, the risk of breast cancer diagnosis during the first two years after FPR was particularly high for women undergoing invasive assessment; importantly, the risk was modified by type of mammographic feature prompting the recall. This information could help to individualize follow-up after exclusion of malignancy.

KEYWORDS

Breast cancer screening; false-positive recall; mammographic feature; breast cancer risk; invasive assessment.

1. Introduction

There is ongoing research on how to reduce the undesired effects of population-based mammography screening, such as overdiagnosis and overtreatment [1]. Researchers seek to identify groups of women who are at low or high risk of developing breast cancer and who could benefit from tailored screening recommendations. Women with a false-positive recall (FPR) at screening (a recall caused by an anomaly found on screening mammography that finally excludes malignancy) are at increased risk of breast cancer diagnosis [2–5]. These women could be considered candidates for a specific intervention. The risk may be even higher for women requiring invasive assessment during the FPR [2,6].

Based on these and other data, some clinical guidelines recommend an intensive two-year follow-up period after a benign breast biopsy prompted by a mammographic finding [7,8]. No such recommendation exists in the European Guidelines for population-based mammography screening [9], which recommend that women return to regular biennial screening after an assessment ruling out malignancy.

In the present study, we aimed to investigate whether the type of mammographic feature assessed in an FPR influences the risk and timing of a subsequent breast cancer diagnosis, particularly in women undergoing invasive assessment. This information would enable the selection of women for intensive follow-up after FPRs in the context of population-based mammography screening.

2. Methods

Our study was based on data from the Spanish population of women undergoing breast cancer screening. Breast cancer screening in Spain follows the European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis [10,11] and has been described in detail elsewhere [12]. In brief, women aged 50-69 years are invited to undergo a two-view mammogram, every two years. Trained radiologists read and classify mammograms according to the Breast Reporting and Data System (BI-RADS scale) [13,14]. Women with a BI-RADS score of 0, 3, 4 or 5 are recalled for further assessment within two months after mammography. During the assessment, additional imaging methods with or without invasive procedures are used. The additional imaging methods comprise further mammographic projections, ultrasound, and magnetic resonance imaging (MRI) if necessary. Invasive procedures include fine-needle aspiration cytology (FNAC), core-needle biopsy, vacuum-assisted biopsy, percutaneous excisional biopsy, and surgical biopsy. The samples obtained are analyzed in local pathology laboratories. As part of a standardized protocol, the screening offices in each region actively search for interval cancers. Information is systematically collected in each region and includes the mammographic features prompting the recall, additional imaging and invasive procedures performed during the assessment, pathology results, and interval cancer diagnosis.

2.1. Study population

We conducted a retrospective study using data from 10 different centers involved in the Spanish breast cancer screening program. The cohort included women aged 50-69 years who were screened between 1994 and 2015, followed up until 2017, and recalled for further assessment. If a woman had more than one recall during the study period, only the most recent was included in the analyses. The study was approved by the institutional review boards of all participating institutions. Informed consent from the patients was not necessary because the study was retrospective, and the data were anonymized.

2.2 Additional assessment and mammographic features

We classified the cohort in two groups according to the assessments performed during the recall: *i)* noninvasive assessment group, which underwent only additional imaging procedures, and *ii)* invasive assessment group, which underwent at least one invasive procedure.

For the noninvasive assessment group, if more than one imaging method was performed at recall, we included only the most complex imaging modality, considering MRI as the most complex, followed by ultrasound, further mammographic projections, and unreported imaging. For the invasive assessment group, if more than one invasive procedure was performed at recall, we included the method that we considered the most informative (i.e., a larger tissue sample was processed by the pathology exam), with surgical biopsy being the most informative, followed by imaging-guided biopsy (including core-needle biopsy, vacuum-assisted biopsy and percutaneous excisional biopsy), FNAC, and unreported invasive procedures.

We categorized the mammographic features prompting recalls as follows: mass, architectural distortion, calcification, asymmetric density, and multiple features [15]. For women included in the study with more than one mammographic feature assessed at different screens, we included only the information on the most recent screen.

Pathology results from invasive procedures were grouped as follows: malignancy, negative for malignant cells on cytology, non-proliferative benign breast disease, and proliferative benign breast disease with or without atypia.

2.3. Time to diagnosis and laterality of breast cancer after an FPRl

For FPRs, we assessed the number of breast cancer cases diagnosed during the follow-up period, in subsequent screens or as interval cancers, for each assessment group (invasive and noninvasive) and each category of mammographic features within each group.

We divided the follow-up period into two groups: within two years after the assessment date of the FPR (first period) and beyond two years after the assessment date of the FPR (second period). The first period included interval cancers diagnosed before the subsequent screen and cancers diagnosed in the subsequent screen due to the two-year periodicity of the Spanish population-based screening.

Because rates are affected by the length of time a woman has been at risk for breast cancer, rates were calculated based on person-years at risk in both groups. Person-years were defined from the date of mammographic reading until a diagnosis of breast cancer or the end of the follow-up period, whichever came first.

We gathered information on the laterality of the tumors diagnosed during the follow-up period and its relationship with the mammographic features assessed at FPRs.

2.4. Statistical analysis

We analyzed the number of women recalled and their distribution in the noninvasive and invasive assessment groups by the type of mammographic feature and by the pathology result in the invasive assessment group. To check screening quality, we calculated the sensitivity of the screening program (overall and for each mammographic feature) as standardized in the European Guidelines[9–11]; namely, the number of screen-detected cancers was divided by the number of screen-detected cancers plus interval cancers.

For FPRs, we computed the rates of breast cancer diagnosed during the follow-up period in the first and second periods for each mammographic feature within each group and overall. We compared breast cancer rates between the noninvasive and invasive assessment groups (overall) and between both periods (overall and for each mammographic feature within each group).

We compared both periods regarding the proportion of interval cancers and the proportion of cancers ipsilateral to the mammographic feature prompting the FPR overall and for each mammographic feature within each group.

As the study period was very long and both technological and protocol changes have undoubtfully occurred meanwhile, we performed a sensitivity analysis to assess the differences in the main outcomes (breast cancer rates and interval cnacers in the noninvasive and invasive assessment groups in both periods defined, overall) for three different time terms: 1994 to 2001, 2002 to 2008 and 2009 to 2015.

We analyzed differences between groups and categories using a chi-square test performed with IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp and R version 3.3.2 (www.r-project.org).

3. Results

3.1. Cohort description and recall performance at screening

During the study period, 784 145 women were screened and 107 417 women were recalled: 83,887 (78.1%) underwent a noninvasive assessment, and 23,694 (21.9%) underwent invasive assessment. Figure 1 shows the flowchart of the study population and the types of imaging and invasive procedures used.

Table 1 shows the number and distribution of mammographic features across the noninvasive and invasive assessment groups together with the distribution of the pathology results in the invasive assessment group: 99,825 had a benign pathology result, and were consequently considered FPRs. Calcification was the most frequently biopsied mammographic feature and most frequently showed malignancy and benign proliferative disease.

The number of interval cancers diagnosed before subsequent rounds was 668, and the overall sensitivity of screening mammography was 93.4%. The screening sensitivity for each mammographic feature is detailed in Figure 2.

3.2 Breast cancer diagnosis after FPRs

The overall number of breast cancer cases diagnosed during the follow-up period was 1904, including 756 during the first period and 1148 during the second period. The person-year figures, number of breast cancers and breast cancer rates after FPRs for each period of follow-up, and each mammographic feature within each group are detailed in Table 2. In both periods, the crude breast cancer rates were higher in the invasive assessment group than in the noninvasive assessment group, but the difference was greater during the first period (12 ‰ vs 1.9 ‰, $p< 0.001$ and 4.4‰ vs 3.1‰, $p<0.001$, respectively).

All categories of mammographic features in the invasive assessment group showed higher breast cancer rates during the first period than during the second period, and the differences were significant overall and for all mammographic features except for architectural distortion. In contrast, all categories of mammographic features in the noninvasive assessment group showed lower rates during the first period than during the second period; the differences were significant overall and for all mammographic features except for architectural distortions and asymmetric densities. The highest risk of a breast cancer diagnosis corresponded to asymmetric densities in the invasive assessment group during the first period (31‰). These differences are shown in Figure 2.

Table 3 summarizes the data on the breast cancer detection method during the follow-up period after an FPR. During the first follow-up period, most breast cancers in the noninvasive assessment group were detected by screening, while most breast cancers in the invasive assessment group were interval cancers; the differences were significant overall and for each mammographic feature. During the second follow-up period, most breast cancers were detected by screening in both groups, and the differences between both groups overall and for each mammographic feature were not significant.

Information on the laterality of breast cancers diagnosed during the follow-up period was available for 476 of 756 cancers diagnosed during the first period and for 744 of 1147 cancers diagnosed during the second period. For both assessment groups, cancers ipsilateral to the mammographic feature causing the FPR were more frequent during the first period than during the second period except for architectural distortions in the invasive assessment group, which showed a slightly higher number of ipsilateral breast cancers during the second period than during the first. In the invasive assessment group, the differences between the two periods regarding laterality were significant overall and for each mammographic feature except for architectural distortions and asymmetric density. In the noninvasive assessment group, differences in laterality were significant only overall and for mass. Data on laterality are summarized in Table 4.

The sensitivity analyses to assess differences attributable to the long study period showed that the results for each term period were the same that the results for the whole 21 years period (see Supplemental Table 1 and Supplemental Table 2).

4. Discussion

In the present study, we observed that in the context of population-based mammography screening, women with FPRs who underwent an invasive procedure had a high risk of a breast cancer diagnosis within the first two years after the recall; importantly, this risk was modulated by the type of mammographic feature prompting the recall, and breast cancer was most frequently diagnosed as an interval cancer. During the first period, the highest breast cancer risk was conferred by asymmetric densities in the invasive assessment group. To the best of our knowledge, this study is the first to identify this risk determinant in the context of mammography screening.

Importantly, after an FPR, all mammographic features evaluated with an invasive assessment showed a higher risk of breast cancer diagnosis during the follow-up period than those evaluated with noninvasive assessments, particularly during the following two years. A history of a benign breast biopsy is a well-known risk factor for developing breast cancer later in life [16], and is an item included in different models for breast cancer risk calculation, such as the Breast Cancer Risk Assessment Tool, the IBIS model [17] and others [18,19]. In a previous study, our group confirmed that a previous benign breast biopsy was a risk factor in the context of population-based mammography screening [2], and those findings are supported by the present results.

The National Comprehensive Cancer Network (NCCN) guidelines recommend that mammographic features with a benign result on biopsy should be followed up with mammography every 6-12 months for one or two years [20]. However other institutions consider that such a policy is not cost-effective [21]. Our results strongly support the principles recommended by the NCCN guidelines for screening programs, particularly when the mammographic feature assessed with an invasive procedure is asymmetric density.

The increased breast cancer risk of women with FPRs could be explained by several factors. Some studies provide evidence that, for some recalled women, cancer may be incorrectly excluded during assessment; this situation is termed “misclassification” by some authors [5,22,23]. The tumor may eventually be diagnosed because of related symptoms or through

subsequent screening. According to Goosens et al, the percentage of misclassified tumors could be as high as 41% of all tumors diagnosed during follow-up [5]. Part of the higher breast cancer rate found for mammographic features evaluated with an invasive assessment in our study could be related to misclassification, as the rate was higher during the first two-year period than subsequently, and cancers were more frequently diagnosed as interval cancers ipsilateral to the mammographic feature causing the FPR. Screening programs should prioritize strategies to minimize this situation, to avoid not only delays in diagnosis but also the negative psychological impact of an interval breast cancer revealed shortly after a benign breast biopsy. Promising tools to minimize misclassification are new imaging or biopsy methods such as contrast-enhanced mammography, MRI, and vacuum-assisted biopsy [2, 29, 30]

Notably, misclassification may only partially explain the higher risk of breast cancer found in women with FPRs. Most frequently, additional assessments at recall correctly rule out malignancy but identify benign breast disease with a high risk of subsequent breast cancer [6,23,24]. The risk is higher for women with a familial history of breast cancer [6,25]. The mammographic feature that most accurately identifies a high-risk mammary gland may be calcification. In a previous study by our group, calcifications showed the highest long-term risk of breast cancer [15]. In the present study, calcifications were the mammographic feature most frequently showing proliferative benign breast disease, a condition that has already been associated with high breast cancer risk [25]. New trials should evaluate whether women whose screening mammograms show calcifications might benefit by a particular preventive approach.

Despite the higher risk of breast cancer within the first two years after an FPR requiring biopsy in our study, women in this situation should be reassured that they are unlikely to be diagnosed with breast cancer during that period, as breast cancer screening programs take measures to obtain optimal results. The highest incidence was seen for asymmetric densities. Even so, in this study, the screening sensitivity for this mammographic feature was as high as 96.06% (data not shown).

The results presented here are also reassuring regarding recall performance at screening, as the percentage of recalled women and screening sensitivity were within the European quality standards [12,26]; therefore, the high breast cancer rates found after FPRs evaluated with invasive assessments during the first follow-up period should not be attributed to deficient screening sensitivity. The two most frequently biopsied mammographic features (calcifications and multiple features) were those most often showing malignancy, thus confirming the high accuracy of radiologists in handling mammography findings.

Our study has some strengths and limitations. The first limitation is that all women included in population-based mammography screening in Spain are between 50 and 69 years old and mammography is performed every two years; therefore, the results presented here may be of value only for the population in this age range participating in biennial mammography screening, as is the case of most European Countries [11].

Secondly, the decision to recall women could be somewhat subjective, depending on radiologists' personal experience and, consequently, the recall rate may vary among the screening offices; however, screening radiologists in Spain are trained to overcome subjectivity as far as possible and recall rates show no significant differences for most regions in the country [27].

Notably, we included 10 centers representative of the Spanish breast cancer screening program, and we analyzed a large number of screened women with more than 100,000 recalls for further assessment with an extended follow-up of up to 20 years. Although an extended follow-up is of value and allows analysis of a large sample, it also has some flaws, the most important being that, during the 20-year period, mammographic technique and interpretation underwent enormous changes; however, our sensitivity analysis showed the same results across three different terms of the study period, supporting the consistency of our findings.

Of note, although the number of women with FPRs included in the study was high (99,825), the number of subsequent breast cancers diagnosed among them was insufficient for some subgroup analyses. Another limitation is that the study was retrospective, even though the databases were prospectively entered. In addition, we were unable to retrieve information on laterality for all cases of cancer due to the heterogeneity of the initial databases; however, we assumed that the losses were random and did not influence the overall results.

Finally, we could not include information on breast density or body weight, both of which affect the accuracy of radiologic assessment [28,29]. Density in particular is a well-established factor affecting the performance of screening [30–32]: Moshina et al reported that recall rates for women with non-dense versus dense breasts were 2.7% and 3.6% [33]. We were not able to include information on density because for most of the 20-year study period, it was not systematically included in the databases of the participating centers. We opted to perform the present analysis by including a large cohort of patients who were followed up for a long time because we considered that the information obtained might be highly valuable. We are now conducting further studies in this research area that include information on breast density and weight.

In conclusion, our study shows that women assessed with invasive procedures at an FPR had a high risk of breast cancer diagnosis in the first two years after the assessment; the risk varied depending on the mammographic feature prompting the recall. This information could enable selection of women for intensive follow-up after FPRs in the context of population-based mammography screening.

Contribution of each co-author:

Maria Vernet-Tomás: conceptualization, methodology, validation, investigation, writing- original draft, visualization and funding acquisition. **Javier Louro**: methodology, software, validation, formal analysis, and writing- original draft. **Marta Román**: methodology, software, validation, formal analysis, investigation and writing- original draft. **Francina Saladié**: validation, resources and writing- review&editing, **Margarita Posso**: validation, resources and writing- review&editing. **Miguel Prieto**: validation, resources and writing- review&editing. **Ivonne Vázquez**: validation, resources and writing- review&editing. **Marisa Baré**: validation, resources and writing- review&editing. **Lupe Peñalba**: validation, resources and writing- review&editing. **Carmen Vidal**: validation, resources and writing- review&editing. **Xavier Bargalló**: validation, resources and writing- review&editing. **Mar Sánchez**: validation, resources and writing- review&editing. **Joana Ferrer**: validation, resources and writing- review&editing. **Josep A Espinàs**: validation, resources and writing- review&editing. **MJesús Quintana**: validation, resources and writing- review&editing. **Ana Rodríguez-Arana**: validation, resources and writing- review&editing. **Xavier Castells**: conceptualization, methodology, validation, investigation, resources, writing- review&editing, project administration and funding acquisition.

Conflict of interest statement

All the authors declare no conflicts of interest.

Funding

This work was supported by grants from the *Instituto de Salud Carlos III* FEDER, (PI11/01296 and PI15/00098) and from the Network for Research in Healthcare in Chronic Diseases of *Instituto de Salud Carlos III*, REDISSEC (RD16/0001/0016).

Ethical approval

The present study did not include experimentation with human beings. The study was approved by the institutional review boards of all participating institutions. Informed consent from patients was not necessary because the study was retrospective, and the data anonymized. All data were retrieved and handled in accordance with the General Data Protection Regulation.

Statement on the welfare of animals**:** This article does not contain any studies involving animals performed by any of the authors.

Research data (data sharing and collaboration)

There are no linked research data sets for this paper. The authors will be pleased to share data on request.

Acknowledgments

The authors acknowledge the rest of the Benign Lesion (BELE) Study Group, listed here in alphabetical order and grouped by institution: (a) IMIM (Hospital Del Mar Medical Research Institute), Barcelona, Spain: Andrea Burón, Laia Domingo, Maria Sala, Sònia Servitja; (b) Parc Taulí University Hospital, Sabadell, Spain: Xavier Andreu; (c) Catalan Institute of Oncology, Cancer Prevention and Monitoring Program, Barcelona, Spain: Llucia Benito; (d) Hospital de la Santa Creu i Sant Pau, Epidemiology Department, Barcelona, Spain: Judit Solà-Roca; (g) Fundació Lliga per a La Investigació i Prevenció Del Càncer, Universitat

Rovira i Virgili, Tarragona, Spain: Jaume Galceran; and (k) Hospital Clínic, Preventive Medicine and Epidemiology Department, Barcelona, Spain: Isabel Torá-Rocamora.

References

[1] D.L. Monticciolo, M.A. Helvie, R.E. Hendrick, Current Issues in the Overdiagnosis and Overtreatment of Breast Cancer, Am. J. Roentgenol. 210 (2018) 285–291. https://doi.org/10.2214/AJR.17.18629.

[2] X. Castells, M. Román, A. Romero, J. Blanch, R. Zubizarreta, N. Ascunce, D. Salas, A. Burón, M. Sala, Cumulative False Positive Risk Group, Breast cancer detection risk in screening mammography after a false-positive result, Cancer Epidemiol. 37 (2013) 85–90. https://doi.org/10.1016/j.canep.2012.10.004.

[3] L.M. Henderson, R.A. Hubbard, B.L. Sprague, W. Zhu, K. Kerlikowske, Increased Risk of Developing Breast Cancer after a False-Positive Screening Mammogram, Cancer Epidemiol. Biomarkers Prev. 24 (2015) 1882–1889. https://doi.org/10.1158/1055-9965.EPI-15-0623.

[4] M. Román, X. Castells, S. Hofvind, M. von Euler-Chelpin, Risk of breast cancer after false-positive results in mammographic screening, Cancer Med. 5 (2016) 1298–1306. https://doi.org/10.1002/cam4.646.

[5] M.C. Goossens, I. De Brabander, J. De Greve, E. Vaes, C. Van Ongeval, K. Van Herck, E. Kellen, Breast cancer risk is increased in the years following false-positive breast cancer screening, Eur. J. Cancer Prev. 26 (2017) 396–403. https://doi.org/10.1097/CEJ.0000000000000313.

[6] L.C. Hartmann, T.A. Sellers, M.H. Frost, W.L. Lingle, A.C. Degnim, K. Ghosh, R.A. Vierkant, S.D. Maloney, V.S. Pankratz, D.W. Hillman, V.J. Suman, J. Johnson, C. Blake, T. Tlsty, C.M. Vachon, L.J. Melton, D.W. Visscher, Benign Breast Disease and the Risk of Breast Cancer, N. Engl. J. Med. 353 (2005) 229–237. https://doi.org/10.1056/NEJMoa044383.

[7] E. Bonaccio, K.E. Calhoun, M. Camp, M.B. Daly, C. Dobbins Lehman, W.B. Farrar, J.E. Garber, R. Gray, R.E. Harris, T. Helsten, L. Hodgkiss, ϕ St Jude, T.L. Hoyt, J.G. Huff, M.S. Jochelson, B.L. Niell, C.C. Parker, M. Pearlman, L. Philpotts, D. Plecha, J.K. Plichta, R.M. Strigel, M. ϕ, L. Tumyan, C. Williams, ϕ Fred, P.S. Buffett Cancer Center Nicole Winkler, C. Young, N. Mary Anne Bergman Rashmi Kumar, Continue NCCN Guidelines Panel Disclosures NCCN Guidelines Version 1.2019 Breast Cancer Screening and Diagnosis, 2019. https://www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf (accessed May 8, 2020).

[8] D.L. Monticciolo, M.S. Newell, R.E. Hendrick, M.A. Helvie, L. Moy, B. Monsees, D.B. Kopans, P.R. Eby, E.A. Sickles, Breast Cancer Screening for Average-Risk Women: Recommendations From the ACR Commission on Breast Imaging, J. Am. Coll. Radiol. 14 (2017) 1137–1143. https://doi.org/10.1016/j.jacr.2017.06.001.

[9] N. Perry, M. Broeders, C. de Wolf, S. Törnberg, R. Holland, L. von Karsa, European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition--summary document., Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 19 (2008) 614–22. https://doi.org/10.1093/annonc/mdm481.

[10] N. Perry, M. Broeders, C. de Wolf, S. Törnberg, R. Holland, L. von Karsa, eds., European guidelines for quality assurance in breast cancer screening and diagnosis, 4th ed., Office for Official Publications of the European Communities, Brussels, 2006.

[11] N. Perry, M. Broeders, C. de Wolf, S. Törnberg, R. Holland, L. von Karsa, eds., European guidelines for quality assurance in breast cancer screening and diagnosis, 4th edition supplements, Office for Official Publications of the European Union, Brussels, 2013.

[12] N. Ascunce, D. Salas, R. Zubizarreta, R. Almazan, J. Ibanez, M. Ederra, Network of Spanish Cancer Screening Programmes (Red de Programas Espanoles de Cribado de Cancer), Cancer screening in Spain, Ann. Oncol. 21 (2010) iii43–iii51. https://doi.org/10.1093/annonc/mdq085.

[13] M.B. Mainiero, A. Lourenco, M.C. Mahoney, M.S. Newell, L. Bailey, L.D. Barke, C. D'Orsi, J.A. Harvey, M.K. Hayes, P.T. Huynh, P.M. Jokich, S.-J. Lee, C.D. Lehman, D.A. Mankoff, J.A. Nepute, S.B. Patel, H.E. Reynolds, M.L. Sutherland, B.G. Haffty, ACR Appropriateness Criteria Breast Cancer Screening, J. Am. Coll. Radiol. 13 (2016) R45–R49. https://doi.org/10.1016/j.jacr.2016.09.021.

[14] C.J. D'Orsi, M.S. Newell, BI-RADS Decoded: Detailed Guidance on Potentially Confusing Issues, Radiol. Clin. North Am. 45 (2007) 751–763. https://doi.org/10.1016/j.rcl.2007.06.003.

[15] X. Castells, I. Torá-Rocamora, M. Posso, M. Román, M. Vernet-Tomas, A. Rodríguez-Arana, L. Domingo, C. Vidal, M. Baré, J. Ferrer, M.J. Quintana, M. Sánchez, C. Natal, J.A. Espinàs, F. Saladié, M. Sala, BELE Study Group, Risk of Breast Cancer in Women with False-Positive Results according to Mammographic Features, Radiology. 280 (2016) 379–386. https://doi.org/10.1148/radiol.2016151174.

[16] M.H. Gail, J. Benichou, Validation studies on a model for breast cancer risk., J. Natl. Cancer Inst. 86 (1994) 573–5. http://www.ncbi.nlm.nih.gov/pubmed/8179704 (accessed April 2, 2018).

[17] J.C. Boughey, L.C. Hartmann, S.S. Anderson, A.C. Degnim, R.A. Vierkant, C.A. Reynolds, M.H. Frost, V.S. Pankratz, Evaluation of the Tyrer-Cuzick (International Breast Cancer Intervention Study) Model for Breast Cancer Risk Prediction in Women With Atypical Hyperplasia, J. Clin. Oncol. 28 (2010) 3591–3596. https://doi.org/10.1200/JCO.2010.28.0784.

[18] V.S. Pankratz, A.C. Degnim, R.D. Frank, M.H. Frost, D.W. Visscher, R.A. Vierkant, T.J. Hieken, K. Ghosh, Y. Tarabishy, C.M. Vachon, D.C. Radisky, L.C. Hartmann, Model for Individualized Prediction of Breast Cancer Risk After a Benign Breast Biopsy, J. Clin. Oncol. 33 (2015) 923–929. https://doi.org/10.1200/JCO.2014.55.4865.

[19] R.D. Frank, S.J. Winham, R.A. Vierkant, M.H. Frost, D.C. Radisky, K. Ghosh, K.R. Brandt, M.E. Sherman, D.W. Visscher, L.C. Hartmann, A.C. Degnim, C.M. Vachon, Evaluation of 2 breast cancer risk models in a benign breast disease cohort., Cancer. (2018). https://doi.org/10.1002/cncr.31528.

[20] NCCN guidelines, (n.d.).

[21] J. Daroles, I. Borget, V. Suciu, C. Mazouni, S. Delaloge, C. Balleyguier, Value of a short-term imaging follow-up after a benign result in a one-stop breast unit: Is it still useful?, Eur. J. Cancer. 85 (2017) 23–30. https://doi.org/10.1016/j.ejca.2017.08.002.

[22] S. Ciatto, N. Houssami, D. Ambrogetti, S. Bianchi, R. Bonardi, B. Brancato, S. Catarzi, G.G. Risso, Accuracy and Underestimation of Malignancy of Breast Core Needle Biopsy: the Florence Experience of Over 4000 Consecutive Biopsies, Breast Cancer Res. Treat. 101 (2007) 291–297. https://doi.org/10.1007/s10549-006-9289-6.

[23] L.E.M. Duijm, J.H. Groenewoud, H.J. de Koning, J.W. Coebergh, M. van Beek, M.J.H.H. Hooijen, L. V. van de Poll-Franse, Delayed diagnosis of breast cancer in women recalled for suspicious screening mammography, Eur. J. Cancer. 45 (2009) 774–781. https://doi.org/10.1016/j.ejca.2008.10.020.

[24] M. von Euler-Chelpin, M. Kuchiki, I. Vejborg, Increased risk of breast cancer in women with false-positive test: The role of misclassification, Cancer Epidemiol. 38 (2014) 619–622. https://doi.org/10.1016/j.canep.2014.06.006.

[25] X. Castells, L. Domingo, J.M. Corominas, I. Torá-Rocamora, M.J. Quintana, M. Baré, C. Vidal, C. Natal, M. Sánchez, F. Saladié, J. Ferrer, M. Vernet, S. Servitja, A. Rodríguez-Arana, M. Roman, J.A. Espinàs, M. Sala, Breast cancer risk after diagnosis by screening mammography of nonproliferative or proliferative benign breast disease: a study from a population-based screening program, Breast Cancer Res. Treat. 149 (2015) 237–244. https://doi.org/10.1007/s10549-014-3208-z.

[26] S.H. Njor, A.H. Olsen, W. Schwartz, I. Vejborg, E. Lynge, Predicting the risk of a false-positive test for women following a mammography screening programme, J. Med. Screen. 14 (2007) 94–97. https://doi.org/10.1258/096914107781261891.

[27] Programas de cribado de cáncer de mama 1, n.d. www.cribadocancer.es (accessed May 8, 2020).

[28] S.H. Njor, M. von Euler-Chelpin, A. Tjønneland, I. Vejborg, E. Lynge, Body weight and sensitivity of screening mammography, Eur. J. Cancer. 60 (2016) 93–100. https://doi.org/10.1016/j.ejca.2016.02.028.

[29] A.S. Tagliafico, M. Calabrese, G. Mariscotti, M. Durando, S. Tosto, F. Monetti, S. Airaldi, B. Bignotti, J. Nori, A. Bagni, A. Signori, M.P. Sormani, N. Houssami, Adjunct Screening With Tomosynthesis or Ultrasound in Women With Mammography-Negative Dense Breasts: Interim Report of a Prospective Comparative Trial, J. Clin. Oncol. 34 (2016) 1882–1888. https://doi.org/10.1200/JCO.2015.63.4147.

[30] M. KE, S. J, Role of Breast Density Measurement in Screening for Breast Cancer, Climacteric. 21 (2018). https://doi.org/10.1080/13697137.2018.1424816.

[31] N. SS, M. P, An Overview of Mammographic Density and Its Association With Breast Cancer, Breast Cancer. 25 (2018). https://doi.org/10.1007/S12282-018-0857-5.

[32] L. J, L. K, A Review of Breast Density Implications and Breast Cancer Screening, Clin. Breast Cancer. (2020). https://doi.org/10.1016/J.CLBC.2020.03.004.

[33] M. N, S. S, L. CI, A. LA, T. KM, E. JG, H. S, Automated Volumetric Analysis of Mammographic Density in a Screening Setting: Worse Outcomes for Women With Dense Breasts, Radiology. 288 (2018). https://doi.org/10.1148/RADIOL.2018172972.

FIGURE CAPTIONS

Figure 1

Flowchart of screened women recalled for further assessment included in the study.

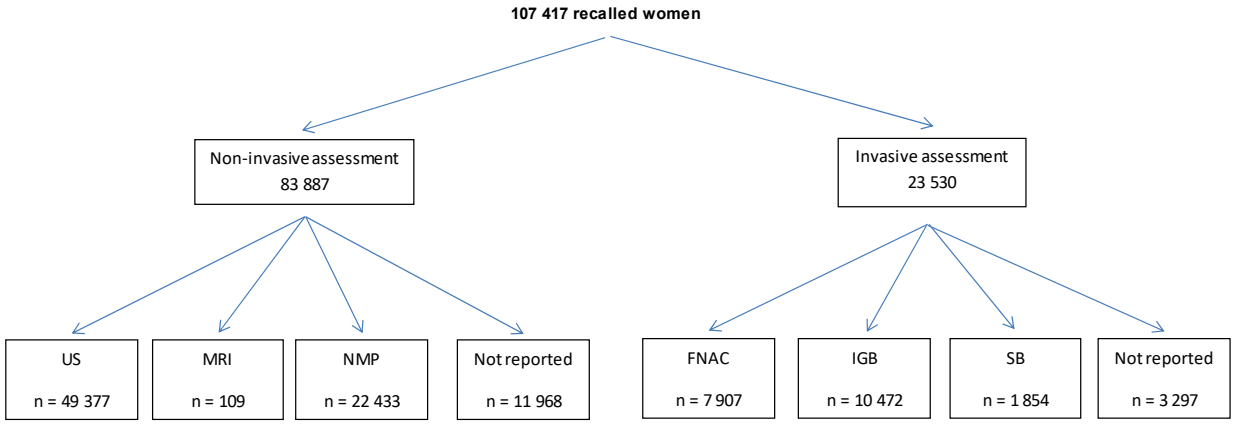
Figure 2

Sensitivity of the screening program.

Figure 3

Longitudinal analysis of breast cancer rates among women with false-positive recalls, for each mammographic feature assessed during the recall and each assessment type (invasive or noninvasive).

Figure 1.



US: Ultrasound, MRI: Magnetic resonance imaging, NMP: New mammographic projection, FNAC: Fine needle aspiration cytology, IGB: Imaging guided biopsy, SB: Surgical Biopsy

Figure 2:

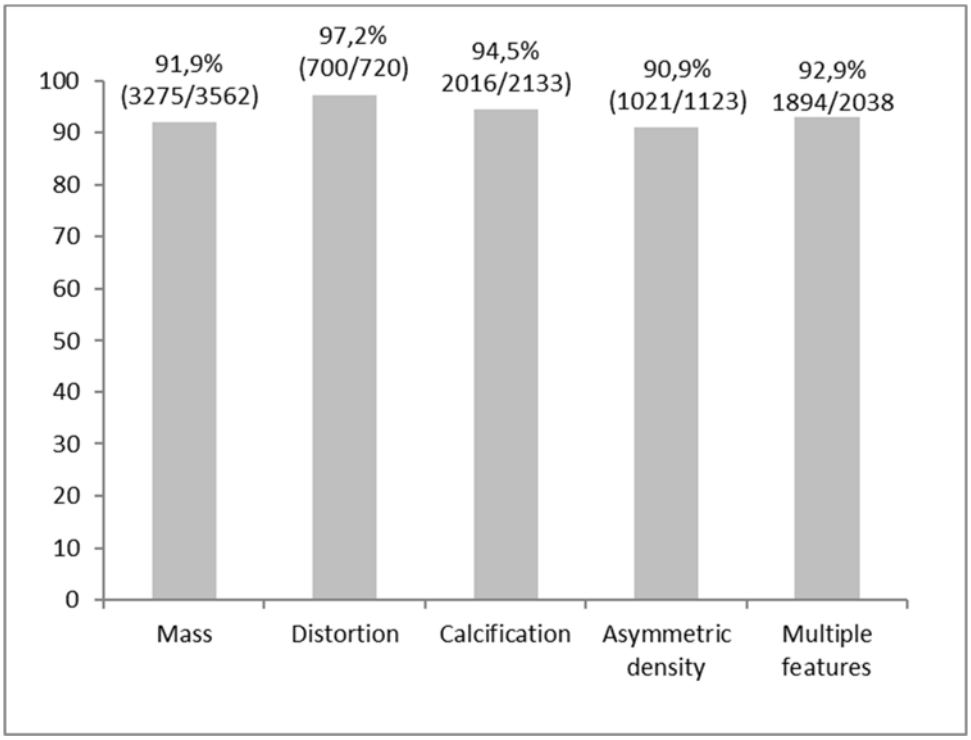

91,9%
(3275/3562)
97,2%
(700/720)
94,5%
2016/2133)
90,9%
(1021/1123)
92,9%
1894/2038
100
90
80
70
60
50
40
30
20
10
0
Mass
Distortion
Calcification
Asymmetric density
Multiple features

Figure 3.

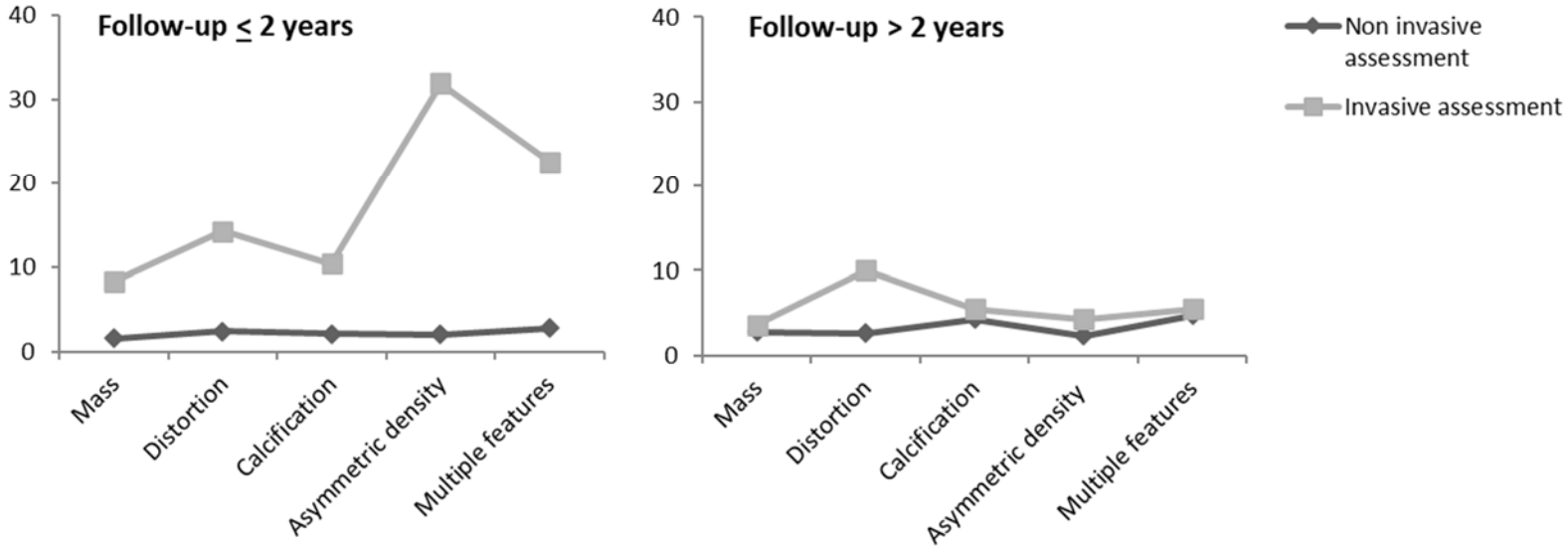

Follow-up ≤ 2 years
40
30
20
10
0
Mass
Distortion
Calcification
Asymmetric density
Multiple features
Follow-up > 2 years
40
30
20
10
0
Mass
Distortion
Calcification
Asymmetric density
Multiple features
Non invasive assessment
Invasive assessment

Table 1

	Non-invasive assessment		**Invasive assessment**								**Total**
			Negative for malignant cells		Non-proliferative BBD		Proliferative BBD		Tumour		
	N	%	N	%	N	%	N	%	N	%	N
Mass	41 407	77.8%	3 305	6.2%	5 020	9.4%	745	1.4%	2 719	5.1%	53 196
Distortion	4 382	81.4%	96	1.8%	160	3.0%	107	2.0%	636	11.8%	5 381
Calcification	7 555	58.6%	584	4.5%	2 089	16.2%	863	6.7%	1 796	13.9%	12 887
Asymmetric density	21 781	92.9%	398	1.7%	354	1.5%	122	0.5%	794	3.4%	23 449
Multiple features	8 762	70.1%	440	3.5%	1 258	10.1%	397	3.2%	1 647	13.2%	12 504
Total	83 887	78.1%	4 823	4.5%	8 881	8.3%	2 234	2.1%	7 592	7.1%	107 417

Type of assessment and pathology results according to mammographic features. BBD: benign breast disease

Table 2

Mammographic feature	Assessment group	Follow-up ≤ 2 years: Person year	Breast cancers	Rate[a]	Follow-up ≥ 2 years: Person year	Breast cancers	Rate[a]	p-value[b]
Mass	Non-invasive	89 502	140	1.6‰	139 844	393	2.8‰	<0.001
	Invasive	19 526	163	8.3‰	38 427	137	3.6‰	<0.001
Distortion	Non-invasive	9 492	23	2.4‰	14 165	37	2.6‰	0.78
	Invasive	773	11	14.2‰	1 297	13	10‰	0.39
Calcification	Non-invasive	16 473	34	2.1‰	34 994	151	4.3‰	<0.001
	Invasive	7 579	79	10.4‰	13 271	73	5.5‰	<0.001
Asymmetric density	Non-invasive	47 220	96	2.0‰	70 025	164	2.3‰	0,27
	Invasive	1 821	58	31.9‰	2 553	11	4.3‰	<0.001
Multiple features	Non-invasive	18 909	52	2.8‰	27 646	130	4.7‰	<0.001
	Invasive	4.429	100	22.6‰	7 086	39	5.5‰	<0.001
Overall	Non-invasive	181 595	345	1.9‰	286 674	875	3.1‰	<0.001
	Invasive	34 127	411	12‰	62 635	273	4.4‰	<0.001

Breast cancers diagnosed during the follow-up among women with false-positive recalls, categorized by the assessment group (invasive and non-invasive) and by the mammographic feature that caused the recall.

[a]Crude rates were calculated dividing the number of breast cancers by the number of person-year

[b]p-values obtained with the two-sided Chi-square test when comparing the first with the second period.

Table 3

Mammographic feature	Assessment group	Follow-up ≤ 2 years: Breast cancers n	Interval cancers n(%)	Screen-detected cancers n(%)	p-value[a]	Follow-up > 2 years: Breast cancers n	Interval cancers n(%)	Screen-detected cancers n(%)	p-value[a]
Mass	Non-invasive	140	3(2)	137 (98)		393	110(28)	283 (72)	
Mass	Invasive	163	137(84)	26 (16)	0.0000	137	37(27)	110(73)	0.5123
Distortion	Non-invasive	23	0(0)	23 (100)		37	7(19)	30(81)	
Distortion	Invasive	11	9(82)	2 (18)	0.0000	13	4(31)	9(69)	0.3749
Calcification	Non-invasive	34	2(6)	32 (94)		151	26(17)	125(83)	
Calcification	Invasive	79	72 (91)	7 (8)	0.0000	73	17(23)	56(77)	0.2797
Asymmetric density	Non-invasive	96	0(0)	96 (100)		164	46(28)	118(72)	
Asymmetric density	Invasive	58	53(91)	5 (9)	0.0000	11	3(27)	8(73)	0.7805
Multiple features	Non-invasive	52	0(0)	52 (100)		130	47(36)	83(64)	
Multiple features	Invasive	100	84 (84)	16 (16)	0.0000	39	13(33)	26(67)	0.7468
Overall	Non-invasive	345	5 (1)	340 (99)		875	236(27)	639(73)	
	Invasive	411	353(86)	58(14)	0.0000	273	74(27)	199(73)	0.9651

Differences between the non-invasive assessment and invasive assessment groups regarding the method of detection during both periods of the follow-up after a FPR.

[a] Chi2 test

Table 4

Mammographic feature	Assessment group	Follow-up ≤ 2 years Ipsilateral n(%)	Follow-up ≤ 2 years Contralateral n(%)	Follow-up > 2 years Ipsilateral n(%)	Follow-up > 2 years Contralateral n(%)	p-value[a]
Mass	Non-invasive	83 (78)	23 (22)	172 (68)	82 (32)	0,04
	Invasive	60 (77)	18 (23)	54 (52)	49 (48)	<0.01
Distortion	Non-invasive	9 (64)	5 (36)	11 (55)	9 (45)	0,59
	Invasive	5 (71)	2 (29)	6 (75)	2 (25)	0,87
Calcification	Non-invasive	19 (73)	7 (27)	75 (71)	30 (29)	0,86
	Invasive	43 (91)	4 (9)	38 (66)	20 (34)	<0.01
Asymmetric density	Non-invasive	55 (74)	19 (26)	59 (67)	29 (33)	0,31
	Invasive	26 (87)	4 (13)	4 (50)	4 (50)	0,07
Multiple features	Non-invasive	36 (84)	7 (16)	57 (79)	15 (21)	0,54
	Invasive	45 (88)	6 (12)	17 (61)	11 (39)	<0.01
Overall	Non-invasive	202 (77)	61 (23)	374 (69)	165 (31)	0,03
	Invasive	179 (84)	34 (16)	119 (58)	86 (42)	<0.01

Percentage of breast cancers ipsilateral and contralateral to the mammographic features causing false positive recalls in both assessment groups, stratified by type of mammographic feature and period. We compared both periods of the follow-up.

[a] p-values using the two-sided Chi-square test when appropriate and the Yate's correction for continuity when 20 % of expected frequencies were less than 5.

Supplemental table 2:

The sensitivity analysis showed that differences regarding interval cancers when comparing Invasive and Non-invasive assessment groups and ≤ 2 or >2 years follow-up, were significant in each of three term periods of seven years each in which the 21 years period of the study was divided.

		Follow-up ≤ 2 years				Follow-up >2 years			
Term periods	Assessment group	Breast cancers n	Interval cancers n (%)	Screen-detected cancers n(%)	p-value[a]	Breast cancers n	Interval cancers n(%)	Screen-detected cancers n(%)	p-value[a]
Overall 1994-2001	Non-invasive	24	0 (0)	24 (100)		183	44 (24)	139 (76)	
	Invasive	33	21 (64)	12 (36)	0.0000	79	16 (20)	63 (80)	0.5028
Overall 2002-2008	Non-invasive	138	3 (2)	135 (98)		503	131 (26)	372 (74)	
	Invasive	124	109 (88)	15 (12)	0.0000	144	31 (22)	113 (78)	0.2701
Overall 2009-2015	Non-invasive	183	2 (1)	181 (99)		189	59 (31)	130 (69)	
	Invasive	254	225 (89)	29 (11)	0.0000	50	27 (54)	23 (46)	0.9651
Overall 1994-2015	Non-invasive	345	5 (1)	340 (99)		875	236(27)	639(73)	
	Invasive	411	353(86)	58(14)	0.0000	273	74(27)	199(73)	0.9643

Differences between the non-invasive assessment and invasive assessment groups regarding the method of detection during both periods of the follow-up after a FPR.

[a] Chi2 test