Evaluación del cribado mamográfico: la valoración de los resultados falsos positivos, los factores asociados, y su impacto sobre las mujeres cribadas

Marta Román Expósito

http://hdl.handle.net/10803/380550

Universitat de Lleida

Universitat de Lleida (UDL)

Departament de Ciències Mèdiques Bàsiques.

Doctorat en Salut

EVALUACIÓN DEL CRIBADO MAMOGRÁFICO: LA VALORACIÓN DE LOS RESULTADOS FALSOS POSITIVOS, LOS FACTORES ASOCIADOS, Y SU IMPACTO SOBRE LAS MUJERES CRIBADAS

Línia de Recerca en Anàlisi Econòmica i Salut

Tesis Doctoral

Doctoranda: Marta Román Expósito

Director: Xavier Castells Oliveres

co-Directora: Montserrat Rué Monné

Universitat de Lleida, 2015

AGRADECIMIENTOS

Aunque el protocolo pueda decir lo contrario, intentaré ser breve en esta sección de agradecimientos. Me cuesta encontrar las palabras. Son muchos años compartiendo experiencias. Experiencias que van mucho más allá de lo laboral y que se han transformando en acompañamiento en un viaje de largo recorrido.

En primer lugar y por encima de todo, quiero dar las gracias a Xavier Castells por su apoyo, su cuidado, y su dedicación. Además de ser el director de esta tesis, ha sido y es un formidable "jefe", compañero de trabajo, y consejero personal. La plena confianza depositada en mí desde el primer momento me ha hecho sentir cómoda y relajada, y me ha permitido crecer enormemente como persona y como profesional. Además, ha demostrado ser un excelente compañero de viaje en los momentos más complicados.

Me gustaría dar mi más sincero agradecimiento a todos los compañeros del Servei d'Epidemiologia i Avaluació del Hospital del Mar, por todo el tiempo que hemos compartido y por las experiencias que hemos vivido. Su apoyo ha sido esencial para llegar hasta aquí a flote: Maria Sala, Esther Martínez, Mercè Comas, Andrea Burón, Cristina Hernández, Jordi Blanch, Anabel Romero, Cristina González, Francesc Macià, Roser Terrades, Mercè Esturi, Eduard Molins, Priscila Giraldo, Miriam Caracuel, Angels Hospital, Isabel Torá-Rocamora, Milagros Herranz, Anna Collado, Immaculada Collet, …seguro que olvido a alguien. Perdón por ello.

Cómo no, agradecer a Montse Rué, co-directora de esta tesis, por su paciencia, su predisposición, y su asertividad. Las cuestiones más complejas se convierten en una tarea sumamente divertida a su lado.

Mi agradecimiento más sincero a todos los investigadores que de una manera u otra han estado involucrados en los proyectos que conforman esta tesis. Compartir con ellos todos estos años en el mismo equipo ha demostrado ser una experiencia altamente enriquecedora. Todos ellos han hecho que el trabajo fuese llevadero y sencillo: Raquel Zubizarreta, Nieves Ascunce, Lola Salas, Pepa Ibáñez, Marisa Baré, Jaume Galcerán, Carmen Natal, María Isabel González, Araceli Baroja, Raquel Almazán, Ana Barcos, Dolores Cuevas, Mariola De la Vega, Isabel Díez de la Lastra María Ederra, José María Velarde, Iosu Delfrade, Josep Maria Corominas, Ana Rodríguez,

Sònia Servitja, Mar Vernet, Núria Milà, Carmen Vidal, Maria Jesús Quintana, Mar Sánchez, Francina Saladie, Joana Ferrer, Josep Alfons Espinàs, Montse Martínez, Ester Vilaprinyo, Misericordia Carles.

Todo mi cariño para Laia Domingo, por mantenerme a salvo en los momentos más delicados y por estar siempre ahí. Por sus buenos consejos, y por las larguísimas conversaciones sobre “el sentido de la vida”.

Quiero agradecer también a Solveig Hofvind, responsable del Programa de Cribado de Cáncer de Mama de Noruega por confiar en mí y darme la oportunidad de expandir mi carrera profesional aquí en Noruega, donde resido actualmente.

Me gustaría dedicar este trabajo a mis padres, hermana, y demás familia. Por su apoyo incondicional y su cariño inagotable.

FINANCIACIÓN

La realización de esta tesis doctoral ha sido posible gracias a diferentes ayudas:

- Proyecto de investigación financiado por el Fondo de Investigaciones Sanitarias (FIS-ISCIII) en la convocatoria del año 2006, "Análisis de la probabilidad acumulada de al menos un falso positivo en el cribado mamográfico en España en la indicación de alguna exploración adicional invasiva (PAAF, core-biopsia o biopsia quirúrgica)" (PI06/1230)

- Proyecto de investigación financiado por el Fondo de Investigaciones Sanitarias (FIS-ISCIII) en la convocatoria del año 2009, "Estudio de la relación entre falso positivo, verdaderos positivos (Cánceres) y adherencia en los programas de detección precoz de cáncer de mama en España" (PI09/90251)

- Beca concedida por el CIBER de Epidemiología y Salud Pública (CIBERESP), "Ayuda para estancias breves en el extranjero con el fin de obtener la mención europea del doctorado" en la convocatoria del año 2012 para realizar una estancia de 3 meses en el Cancer Registry of Norway, Oslo, Noruega.

- Ayuda concedida por el Fondo de Investigaciones Sanitarias (FIS-ISCIII) "Contrato de apoyo del Sistema Nacional de Salud. Instituto Nacional Carlos III" para financiar el 50% de mi contrato de 2010 a 2012 (CA09/00428).

- Ayuda concedida por el Fondo de Investigación Sanitaria (FIS-ISCIII) "Contrato de apoyo del Sistema Nacional de Salud. Instituto Nacional Carlos III" para financiar el 50% de mi contrato de 2007 a 2009 (CA06/0087).

RESUMEN

Palabras clave (términos MeSH): Breast neoplasms, Mass screening, Mammography False positive reactions, Predictive value of tests, Observer variation, Early detection of cancer

Antecedentes: El cribado de cáncer de mama mediante mamografía reduce la mortalidad por esta enfermedad. Siguiendo las recomendaciones del Consejo Europeo, la mayoría de países europeos han puesto en marcha programas de carácter poblacional que ofrecen, en su mayoría mamografías bienales a mujeres entre 50 y 69 años. Los resultados sobre la efectividad del cribado poblacional son controvertidos en lo referente al balance entre reducción de mortalidad y efectos adversos del cribado, como son los resultados falsos positivos, el sobrediagnóstico, y los resultados falsos negativos. Los falsos positivos son el efecto adverso más frecuente del cribado mamográfico. Es importante profundizar en el conocimiento de los resultados falsos positivos y su impacto para poder mejorar la efectividad del cribado mamográfico.

Objetivos: En particular, en este trabajo de tesis se ha estudiado el riesgo acumulado de falsos positivos a lo largo de la participación secuencial de la mujer en el cribado mamográfico, valorando el impacto de diferentes factores del protocolo de lectura mamográfica y de las características de la mujer sobre el riesgo estimado. Posteriormente se ha evaluado el impacto que los resultados falsos positivos tienen sobre la adherencia al cribado en sucesivas convocatorias de cribado, así como su efecto sobre el riesgo de detección de cáncer de mama. Además, se valoró la tendencia temporal en las tasas de detección de carcinoma ductal in situ y carcinoma invasivo en el cribado poblacional desde la puesta en marcha de los programas de cribado, así como el impacto del uso de terapia hormonal sustitutiva sobre éstas tasas de detección. Por último, se ha aplicado la metodología para la estimación del riesgo acumulado de falsos positivos desarrollada para los trabajos de esta tesis con los datos del programa de cribado poblacional de cáncer de mama de Noruega.

Métodos: La población de estudio estaba comprendida por las mujeres participantes en los programas de cribado de 8 Comunidades Autónomas del territorio español, desde la puesta en marcha de los mismos en 1991 hasta diciembre de 2006. Las mamografías de cribado se llevaron a cabo en 74 unidades radiológicas diferentes. Se analizaron 4,739,498 mamografías de cribado de 1,565,364 mujeres de 45 a 69 años de edad. Para el estudio del programa de cribado de cáncer

de mama de Noruega se dispuso de información de todas las mamografías realizadas en el programa en el periodo 1996-2010. Se analizaron 715,311 mamografías de cribado de 231,310 mujeres cribadas por primera vez con 50-51 años. Para obtener las estimaciones de los estudios de esta tesis se utilizaron modelos lineales generalizados de regresión. En particular, se construyeron diversos modelos de regresión de riesgo con tiempo discreto y efectos aleatorios. Mediante estos modelos se estimó el riesgo de falsos positivos en cada participación de la mujer en el cribado (por cualquier tipo de prueba y por pruebas invasivas), así como el impacto sobre el riesgo de falsos positivos de las características del protocolo de lectura mamográfica y características de la mujer. Igualmente, se utilizaron estos modelos de regresión para estimar el riesgo de detección de cáncer de mama después de un resultado falso positivo, y para estimar la adherencia al cribado en convocatorias posteriores a un resultado falso positivo. Para el estudio de la tendencia temporal en las tasas de detección de carcinoma ductal in situ y carcinoma invasivo se utilizaron modelos lineales generalizados de Poisson.

Resultados: i) El riesgo acumulado de tener un resultado falso positivo a lo largo de 10 participaciones bienales en el cribado, para las mujeres que empezaron el cribado con 50-51 años, fue del 20.4% (IC95%: 20.02-20.76). El riesgo acumulado de tener un resultado falso positivo con pruebas invasivas fue del 1.8% (IC95%: 1.66-1.87). Los factores asociados con el riesgo de falso positivo, por cualquier tipo de procedimiento adicional y por procedimientos invasivos fueron la doble lectura de la mamografía (OR= 2.06; IC95%: 2.00-2.13 y OR=4.44; IC95%: 4.08-4.84, respectivamente), doble proyección mamográfica (OR= 0.77; IC95%: 0.76-0.79 y OR= 1.56; IC95%: 1.48-1.64), mamografía digital (OR= 0.83; IC95%: 0.72-0.96 para pruebas invasivas), pruebas invasivas previas (OR= 1.52; IC95%: 1.49-1.56 y OR=2.00; IC95%: 1.89-2.12), e historia familiar de cáncer de mama (OR= 1.18; IC95%: 1.15-1.20; y OR=1.21; IC95%: 1.13-1.30). ii) La adherencia a la segunda invitación al cribado en las mujeres con y sin falsos positivos previos fue 79.3 vs. 85.3%, respectivamente. En el cuarto y séptimo cribado, estos porcentajes fueron 86.3 vs. 89.9% y 94.6 vs. 96.0%, respectivamente. Las variables asociadas con una mayor probabilidad de no participar en sucesivas convocatorias de cribado fueron, grupo de edad mayor (OR= 8.48; IC95%: 8.31-8.65), y tener pruebas invasivas previas (OR= 1.09; IC95%I: 1.07-1.10). iii) El riesgo de detección de cáncer de mama en cribados sucesivos fue mayor en mujeres con falsos positivos con pruebas invasivas (OR= 2.69; IC95%: 2.28-3.16), y mujeres con falsos positivos con pruebas de imagen únicamente (OR= 1.81; IC95%: 1.70-1.94), en comparación a las mujeres sin resultados falsos

positivos previos. El riesgo de detección de cáncer aumentó de manera substancial en mujeres con falsos positivos con pruebas invasivas y antecedentes familiares de cáncer de mama de manera conjunta (OR= 4.64; IC95%: 3.23-6.66). iv) Las tasas de detección de cáncer de mama invasivo por 100,000 mujeres cribadas fue 394.0 en cribado inicial y 229.9 en cribados sucesivos. Las tasas de detección de carcinoma ductal in situ por 100,000 mujeres cribadas fueron 66.8 en cribado inicial y 43.9 en cribados sucesivos. No se encontró evidencia estadística de un cambio de tendencia en las tasas de carcinoma ductal in situ y cáncer invasivo a lo largo del periodo de los 16 años del periodo de estudio. Las tasas de detección de carcinoma ductal in situ aumentaron un 2.5% anual (IC95%: 1.3; 3.8), mientras que las tasas de cáncer invasivo fueron estables. v) En el contexto del cribado poblacional de cáncer de mama de Noruega, el riesgo acumulado de falso positivo a lo largo de 20 años de cribado bienal para las mujeres que empezaron el cribado con 50-51 años fue del 20.0% (IC95%: 19.7%-20.4%). El riesgo acumulado de falsos positivos con pruebas invasivas fue del 4.1% (IC95%: 3.9%-4.3%).

Conclusiones: i) El riesgo acumulado de falso positivo a lo largo de 10 participaciones secuenciales en el cribado varió ampliamente, en función de distintos factores del protocolo de lectura mamográfica y de las características de la mujer estudiados. ii) La adherencia en sucesivas convocatorias de cribado fue menor en mujeres con resultados falsos positivos previos en comparación a las mujeres con resultados negativos. Las diferencias en la adherencia disminuyeron con el número de participaciones en el cribado completadas. Estos resultados sugieren que los resultados falsos positivos en convocatorias iniciales tienen un mayor impacto sobre la adherencia al cribado. iii) Las mujeres con falsos positivos previos tenían un mayor riesgo de detección de cáncer, especialmente las mujeres con falsos positivos previos con pruebas invasivas. iv) A pesar del descenso observado en la incidencia de cáncer de mama en la población española, las tasas de detección de cáncer invasivo en el cribado fueron estables a lo largo de los 16 años del periodo de estudio. La proporción de carcinoma ductal in situ sobre el total de tumores de mama detectados en el cribado incrementó del 13% al 17% a lo largo del periodo de estudio. Las tasas de detección de cáncer invasivo y carcinoma ductal in situ no mostraron asociación con la tendencia decreciente de uso de terapia hormonal sustitutiva observada en las mujeres cribadas a partir del año 2002. v) En el programa de cribado de cáncer de mama de Noruega se estimó que una de cada 5 mujeres participantes sufrirán un resultado falso-positivo a

lo largo de 10 participaciones bienales en el cribado. El riesgo acumulado de falsos positivos con pruebas invasivas fue aproximadamente del 4%.

RESUM

Paraules clau (términos MeSH): Breast neoplasms, Mass screening, Mammography False positive reactions, Predictive value of tests, Observer variation, Early detection of cancer

Antecedents: El cribratge de càncer de mama mitjançant mamografia redueix la mortalitat per aquesta malaltia. Seguint les recomanacions del Consell Europeu, la majoria de països europeus han posat en marxa programes de caràcter poblacional que ofereixen, majoritàriament mamografies biennals a dones entre 50 i 69 anys. Els resultats sobre l'efectivitat del cribratge poblacional són controvertits pel que fa al balanç entre reducció de mortalitat i efectes adversos del cribratge, com són els resultats falsos positius, el sobrediagnòstic, i els resultats falsos negatius. Els falsos positius són l'efecte advers més freqüent del cribratge mamogràfic. És important aprofundir en el coneixement dels resultats falsos positius i el seu impacte per poder millorar l'efectivitat del cribratge mamogràfic.

Objectius: En particular, en aquest treball de tesi s'ha estudiat el risc acumulat de falsos positius al llarg de la participació seqüencial de la dona en el cribratge mamogràfic, valorant l'impacte de diferents factors del protocol de lectura mamogràfica i de les característiques de la dona sobre el risc estimat. Posteriorment s'ha avaluat l'impacte que els resultats falsos positius tenen sobre l'adherència al cribratge en successives convocatòries de cribratge, així com el seu efecte sobre el risc de detecció de càncer de mama. A més, es va valorar la tendència temporal en les taxes de detecció de carcinoma ductal in situ i carcinoma invasiu en el cribratge poblacional des de la posada en marxa dels programes de cribratge, així com l'impacte de l'ús de teràpia hormonal substitutiva sobre aquestes taxes de detecció. Finalment, s'ha aplicat la metodologia per a l'estimació del risc acumulat de falsos positius desenvolupada pels treballs d'aquesta tesi amb les dades del programa de cribratge poblacional de càncer de mama de Noruega.

Mètodes: La població d'estudi estava compresa per les dones participants en els programes de cribratge de 8 comunitats autònomes del territori espanyol, des de la posada en marxa dels mateixos en 1991 fins a desembre de 2006. Les mamografies de cribratge es van dur a terme en 74 unitats radiològiques diferents. Es van analitzar 4,739,498 mamografies de cribratge de 1,565,364 dones de 45 a 69 anys d'edat. Per a l'estudi del programa de cribratge de càncer de mama de Noruega es va disposar d'informació de totes les mamografies realitzades en el

programa en el període 1996-2010. Es van analitzar 715,311 mamografies de cribratge de 231,310 dones cribrades per primera vegada amb 50-51 anys. Per obtenir les estimacions dels estudis d'aquesta tesi es van utilitzar models lineals generalitzats de regressió. En particular, es van construir diversos models de regressió de risc amb temps discret i efectes aleatoris. Mitjançant aquests models es va estimar el risc de falsos positius en cada participació de la dona en el cribratge (per qualsevol tipus de prova i per proves invasives), així com l'impacte sobre el risc de falsos positius de les característiques del protocol de lectura mamogràfica i característiques de la dona. Igualment, es van utilitzar aquests models de regressió per estimar el risc de detecció de càncer de mama després d'un resultat fals positiu, i per estimar l'adherència al cribratge en convocatòries posteriors a un resultat fals positiu. Per a l'estudi de la tendència temporal en les taxes de detecció de carcinoma ductal in situ i carcinoma invasiu es van utilitzar models lineals generalitzats de Poisson.

Resultats: i) El risc acumulat de tenir un resultat fals positiu al llarg de 10 participacions biennals en el cribratge, per a les dones que van començar el cribratge amb 50-51 anys, va ser del 20.4% (IC95%: 20.02-20.76). El risc acumulat de tenir un resultat fals positiu amb proves invasives va ser del 1.8% (IC95%: 1.66-1.87). Els factors associats amb el risc de fals positiu, per qualsevol tipus de procediment addicional i per procediments invasius van ser la doble lectura de la mamografia (OR= 2.06; IC95%: 2.00-2.13 i OR= 4.44; IC95%: 4.08-4.84, respectivament), doble projecció mamogràfica (OR= 0.77; IC95%: 0,76-0,79 i OR= 1.56; IC95%: 1.48-1.64), proves invasives prèvies (OR= 1.52; IC95%: 1.49-1.56 i OR= 2.00; IC95%: 1.89-2.12), i història familiar de càncer de mama (OR= 1.18; IC95%: 1.15-1.20; i OR= 1.21; IC95 %: 1.13-1.30). ii) L'adherència a la segona invitació al cribratge en les dones amb i sense falsos positius previs va ser 79.3 vs. 85.3%, respectivament. En el quart i setè cribratge, aquests percentatges van ser 86.3 vs. 89.9% i 94.6 vs. 96.0%, respectivament. Les variables associades amb una major probabilitat de no participar en successives convocatòries de cribratge van ser, grup d'edat major (OR= 8.48; IC95%: 8.31-8.65), i tenir proves invasives prèvies (OR= 1.09; IC95%: 1.07 -1.10). iii) El risc de detecció de càncer de mama en cribratges successius va ser major en dones amb falsos positius amb proves invasives (OR= 2.69; IC95%: 2.28-3.16), i dones amb falsos positius amb proves d'imatge únicament (OR= 1.81; IC95%: 1.70-1.94), en comparació a les dones sense resultats falsos positius previs. El risc de detecció de càncer va augmentar de manera substancial en dones amb falsos positius amb proves invasives i antecedents familiars de càncer de mama de manera conjunta (OR= 4.64; IC95%: 3.23-

6.66). iv) Les taxes de detecció de càncer de mama invasiu per 100,000 dones cribrades va ser 394.0 en cribratge inicial i 229.9 en cribratges successius. Les taxes de detecció de carcinoma ductal in situ per 100,000 dones cribrades van ser 66.8 en cribratge inicial i 43.9 en cribratges successius. No s'ha trobat evidència estadística d'un canvi de tendència en les taxes de carcinoma ductal in situ i càncer invasiu al llarg dels 16 anys del període d'estudi. Les taxes de detecció de carcinoma ductal in situ van augmentar un 2.5% anual (IC95%: 1.3-3.8), mentre que les taxes de càncer invasiu van ser estables. v) En el context del cribratge poblacional de càncer de mama de Noruega, el risc acumulat de fals positiu al llarg de 20 anys de cribratge biennal per a les dones que van començar el cribratge amb 50-51 anys va ser del 20.0% (IC95%: 19.7%-20.4%). El risc acumulat de falsos positius amb proves invasives va ser del 4.1% (IC95%: 3.9%-4.3%).

Conclusions: i) El risc acumulat de fals positiu al llarg de 10 participacions seqüencials en el cribratge va variar àmpliament, en funció de diferents factors del protocol de lectura mamogràfica i de les característiques de la dona estudiats. ii) L'adherència en successives convocatòries de cribratge va ser menor en dones amb resultats falsos positius previs en comparació a les dones amb resultats negatius. Les diferències en l'adherència van disminuir amb el nombre de participacions en el cribratge completades. Aquests resultats suggereixen que els resultats falsos positius en convocatòries inicials tenen un major impacte sobre l'adherència al cribratge. iii) Les dones amb falsos positius previs tenien un major risc de detecció de càncer, especialment les dones amb falsos positius previs amb proves invasives. iv) Malgrat el descens observat en la incidència de càncer de mama en la població espanyola, les taxes de detecció de càncer invasiu en el cribratge van ser estables al llarg dels 16 anys del període d'estudi. La proporció de carcinoma ductal in situ sobre el total de tumors de mama detectats en el cribratge es va incrementar del 13% al 17% al llarg del període d'estudi. Les taxes de detecció de càncer invasiu i carcinoma ductal in situ no van mostrar associació amb la tendència decreixent d'ús de teràpia hormonal substitutiva observada en les dones cribrades a partir de l'any 2002. v) En el programa de cribratge de càncer de mama de Noruega es estimar que una de cada 5 dones participants patiran un resultat fals-positiu al llarg de 10 participacions biennals en el cribratge. El risc acumulat de falsos positius amb proves invasives va ser aproximadament del 4%.

SUMMARY

Keywords (MeSH terms): Breast neoplasms, Mass screening, Mammography False positive reactions, Predictive value of tests, observer variation, Early detection of cancer

Background: Mammographic screening has been shown to reduce mortality from this disease. Following the recommendations of the European Council, most European countries have started population-based screening programs that offer biennial mammograms to women between 50 and 69 years. The results on the effectiveness of population-based screening are controversial regarding the balance between mortality reduction and adverse effects, such as false positive results, overdiagnosis and false negative results. False positives are the most common adverse effect of breast screening. It is important to deepen the knowledge on false positive results and its impact in order to improve the effectiveness of mammographic screening.

Aims: In this thesis we have studied the cumulative risk of false positive results over 10 biennial participations in mammography screening age 50 to 69 years. We assessed the impact of different factors related to the mammographic reading protocol and women's characteristics on the estimated risk. Afterwards, we evaluated the impact of false positive results on re-attendance to subsequent screening invitations, as well as its effect on subsequent risk of breast cancer detection. In addition, we evaluated the time trends in the detection rates of ductal carcinoma in situ (DCIS) and invasive cancer in population based-screening. We also evaluated the impact of the reduction in hormonal therapy use on the rates of screen detected DCIS and invasive breast cancer. Finally, we applied the methodology to estimate the cumulative risk of false positive results developed for the studies in this thesis, to data from the Norwegian Breast Cancer Screening Program.

Methods: The study population included all women participating in mammographic screening in 8 different Regions of Spain, in the period 1991-2006. The screening mammograms were performed on 74 different radiology units. We analyzed 4,739,498 screening mammograms from 1,565,364 women aged 45-69 years. For the study based on the Norwegian Breast Cancer Screening Program, we gathered information on all screening mammograms performed in the period 1996-2010. We analyzed 715,311 screening tests from 231,310 women first screened at age 50-51. The estimates of the studies in this thesis, were computed using generalized linear regression models.

In particular, Discrete Time Hazard Models with random effects were used. The regression models were used to estimate the risk of false positive results (for any procedure and/or involving invasive procedures) in each participation of screened women in the program. These models were also used to estimate the impact of the reading protocol and women's characteristics on the risk of false positive results. Similarly, we assessed the risk of breast cancer detection after a false positive result, and the re-attendance to subsequent screening invitations in women with false positive results. To study the time trends in the rates of screen detected DCIS and invasive cancer Poisson regression models were used.

Results: i) The cumulative false-positive risk over 10 biennial participations in mammographic screening, for women who started screening at age 50-51 was 20.4% (95%CI: 20.02-20.76). The cumulative risk for false positives with invasive procedures was 1.8% (95%CI: 1.66-1.87). The factors associated with the false-positive risk, for any procedure and for invasive procedures were double mammogram reading (OR= 2.06; 95%CI: 2.00-2.13 and OR= 4.44; 95%CI: 4.08-4.84, respectively), two mammographic views (OR= 0.77; 95%CI: 0,76-0,79 and OR= 1.56; 95%CI: 1.48-1.64, respectively), previous invasive procedures (OR= 1.52; 95%CI: 1.49-1.56 and OR= 2.00; 95%CI: 1.89-2.12, respectively), and family history of breast cancer (OR= 1.18; 95%CI: 1.15-1.20; and OR= 1.21; 95%CI: 1.13-1.30, respectively). ii) At the second screening invitation re-attendance among women with and without a false-positive mammogram was 79.3 vs. 85.3%, respectively. At the fourth and seventh screenings, these percentages were 86.3 vs. 89.9% and 94.6 vs. 96.0%, respectively. The study variables associated with a higher risk of failing to participate in subsequent screenings were oldest age (OR= 8.48; 95%CI: 8.31-8.65), and having experienced previous invasive procedures (OR= 1.09; 95%CI: 1.07-1.10). iii) The risk of cancer detection was higher in women with false-positives involving an invasive procedure (OR= 2.69; 95%CI: 2.28-3.16), and women with false-positives involving additional imaging procedures alone (OR= 1.81; 95%CI: 1.70-1.94), compared with women without false positive results. The risk of cancer detection increased substantially if women with false positive results with invasive procedures had a familial history of breast cancer (OR= 4.64; 95%CI: 3.23-6.66). iv) The rates of screen detected invasive cancer per 100,000 screened women were 394.0 at first screening, and 229.9 at subsequent screen. The rates of screen detected DCIS per 100,000 screened women were 66.8 at first screen and 43.9 at subsequent screens. No evidence of a change point in trend in the rates of DCIS and invasive cancers over the study period were found. Screen detected DCIS increased at a steady 2.5% per year (95%CI: 1.3; 3.8), while screen detected invasive cancers were stable. v) The

cumulative false-positive risk after 10 biennial screening participations, for women who started screening at age 50-51 years was 20.0% (95%CI: 19.7%-20.4%). The cumulative risk of undergoing an invasive procedure with a benign outcome for the same group of women was 4.1% (95%CI: 3.9%-4.3%).

Conclusions: i) The cumulative risk of a false-positive result varied widely with factors related to the mammographic reading protocol and women's characteristics. ii) Re-attendance was lower in women with false positive results compared with those with negative results. The differences in re-attendance decreased with the number of completed screening participations, suggesting that abnormal results in earlier screenings more strongly influence behavior. iii) Women with a false-positive test had an increased risk of breast cancer detection in subsequent screening participations, especially those with a false-positive involving an invasive procedure with a benign outcome. iv) Despite the observed decrease in breast cancer incidence in the population, the rates of screen detected invasive cancer remained stable during the study period. The proportion of DCIS among screen detected breast malignancies increased from 13% to 17% throughout the study period. The rates of screen detected invasive cancer and DCIS were independent of the decreasing trend in hormone replacement therapy use observed among screened women after publication of the Women's Health Initiative trial in 2002. v) In the Norwegian Breast Cancer Screening Program it is estimated that one in every 5 women will be recalled for further assessment with a negative outcome if they attend biennial mammographic screening between ages 50 years to 69 years. The risk of an invasive procedure with a benign outcome was approximately 4%.

ÍNDICE

ÍNDICE DE TABLAS Y FIGURAS

Tablas

Figuras

PRESENTACIÓN

El cáncer de mama es el tumor más frecuente entre las mujeres a nivel mundial, y es además la primera causa de muerte relacionada con cáncer entre las mujeres europeas. Su amplio alcance hace que sea un problema de salud pública de primer nivel. En las últimas décadas se han dedicado numerosos esfuerzos para poder conocer los determinantes de este cáncer y promover estrategias que permitan mejorar el pronóstico de la enfermedad y así la reducir su mortalidad.

Diferentes ensayos clínicos realizados en la década de los 70 y 80, pusieron de manifiesto que la detección precoz del cáncer de mama mediante mamografía permitía detectar tumores en estadíos precoces, una mejora en el pronóstico y una disminución en la mortalidad. De este modo, a lo largo de las décadas de los 80 y 90, la mayoría de países europeos pusieron en marcha programas de cribado de cáncer de mama para fomentar la detección precoz de esta enfermedad. La gran mayoría de estos programas son de carácter poblacional y ofrecen mamografías bienales a las mujeres entre los 50 y 69 años.

Sin embargo, veinte años después del inicio de la implementación del cribado, en la actualidad se hace un análisis crítico de su efectividad. Tanto en lo que se refiere a los beneficios esperados, es decir la reducción de la mortalidad, como a los efectos adversos que comparta el diagnóstico precoz, especialmente el riesgo de falsos positivos y el tratamiento innecesario (sobrediagnóstico). Esta tesis doctoral se enmarca dentro del actual debate sobre el cribado de cáncer de mama, y la evaluación de sus efectos adversos. En concreto, se hace un énfasis especial en la evaluación de los resultados falsos positivos, que son el efecto adverso más frecuente del cribado, pero cuya magnitud e impacto a largo plazo no está evaluada de manera concluyente. Son escasos los estudios sobre el efecto acumulado de los resultados falsos positivos a lo largo de diferentes convocatorias de cribado. La mujer es invitada bienalmente a realizarse una mamografía de cribado durante un periodo de veinte años, pero la mayoría de análisis de los falsos positivos y de otras medidas de resultados del cribado se hacen desde una perspectiva transversal, generalmente basada en una única ronda de cribado.

Para profundizar en el estudio de los falsos positivos y su impacto, se puso en marcha un estudio de cohorte que permitiera el seguimiento de la mujer de manera secuencial a lo largo de su historial de cribado. La base de datos resultante contiene información de 10 programas de

cribado del contexto español, e incluye más de 4.500.000 mamografías de cribado de más de 1.500.000 mujeres, realizadas entre 1991 y 2006, con un promedio de 3 tres participaciones por mujer. Esta base de datos, diseñada específicamente para los estudios que aquí se presentan, es la base de datos más grande creada hasta la fecha para la evaluación de los efectos adversos del cribado.

En primer lugar, se han estudiado las características del protocolo de lectura mamográfica y características de la mujer asociados con los resultados falsos positivos. A continuación, se estudió el impacto que los falsos positivos tienen en sucesivas convocatorias de cribado, tanto en términos de reducción de la adherencia al cribado de estas mujeres, cómo su impacto en el riesgo de detección de cáncer de mama en convocatorias sucesivas. Posteriormente, aprovechando el potencial de la base de datos, se evaluó la tendencia temporal en las tasas de detección de cáncer de mama *in situ* e invasivo en el cribado mamográfico, desde la puesta en marcha de los programas en 1991 hasta el año 2006. Por último, y debido al interés suscitado por la metodología utilizada en estos análisis, se han actualizado los resultados sobre riesgo acumulado de falsos positivos en programa de cribado poblacional de Noruega, lo que ha servido de validación externa de la metodología.

Esta tesis se presenta como compendio de publicaciones, y está compuesta por cinco trabajos que pretenden dar respuestas concretas e inéditas sobre algunas de las cuestiones que se describen. Pero también, intenta poner de manifiesto la complejidad que entraña la evaluación de los efectos adversos del cribado en poblaciones dinámicas, así como la necesidad de utilizar una aproximación longitudinal para poder evaluar correctamente los efectos adversos.

Los trabajos que se presentan han sido realizados en el *Servei d'Epidemiologia i Avaluació* del *Hospital del Mar-IMIM*, bajo la dirección del Dr. Xavier Castells. Estas tareas recibieron financiación específica de dos proyectos del Fondo de investigaciones sanitarias, FIS-ISCIII (PI06/1230, PI09/90251). Personalmente he liderado el diseño, recogida de datos, validación, y análisis de la base de datos utilizada en estos proyectos. Por su parte, el estudio sobre el riesgo acumulado de resultados falsos positivos en el programa de Noruega ha sido posible gracias a una beca del *CIBERESP* (*CIBER de Epidemiología y Salud Pública*), para optar a la mención europea del doctorado. Con esta ayuda pude hacer una estancia de 3 meses en el *Cancer Registry of Norway*, en Oslo, Noruega.

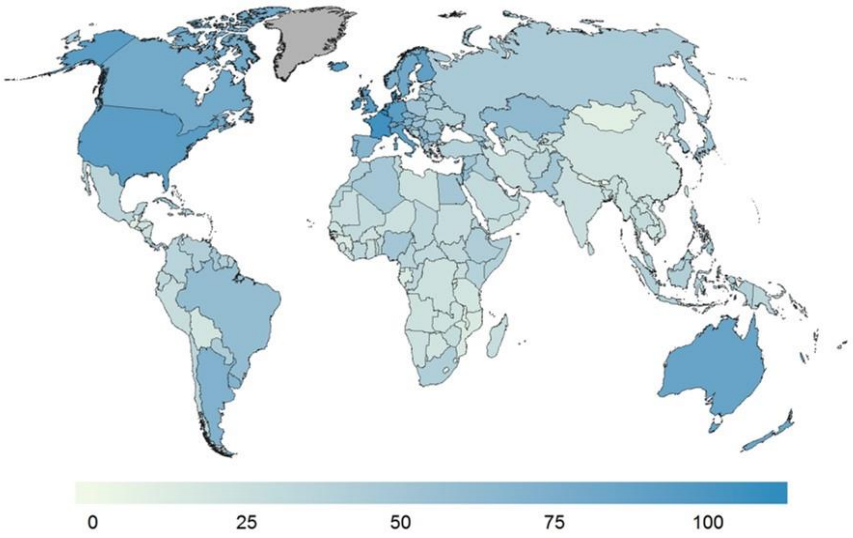
I. INTRODUCCIÓN

1. Epidemiología del cáncer de mama

El cáncer de mama es el segundo tumor más frecuente a nivel mundial, y de lejos, el más frecuente entre las mujeres a nivel mundial[1,2]. Se calcula que se diagnosticaron 1.67 millones de nuevos casos en el año 2012, lo que representa el 25.2% de todos los cánceres diagnosticados en mujeres. Además, es la segunda causa de muerte por cáncer en mujeres en los países desarrollados, con cerca de 198,000 muertes (el 15.4% de las muertes por cáncer)[1,2].

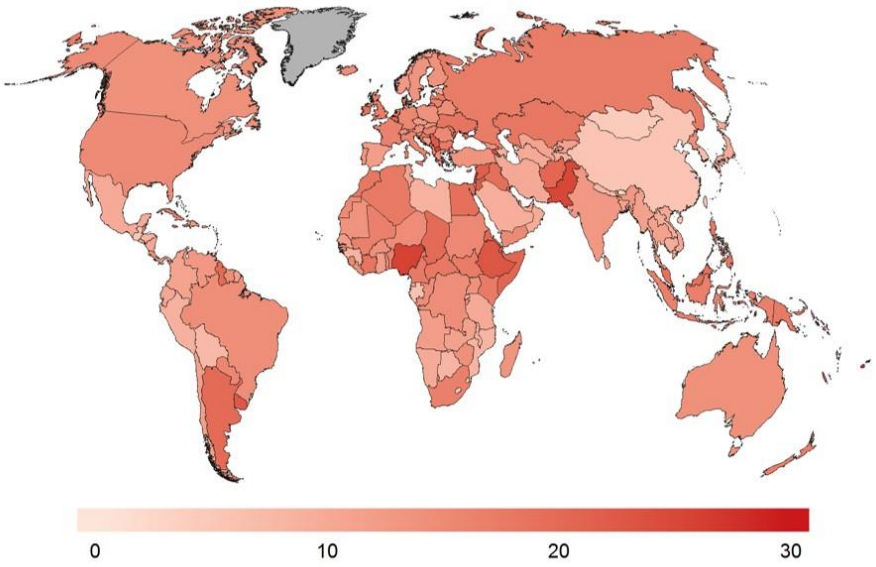
A nivel mundial, las tasas de incidencia varían hasta en casi un 400% entre países, con tasas que van desde 27 por 100,000 mujeres en África Central y el Este Asiático, hasta 96 por 100,000 en Europa Occidental[2,3]. El rango en las tasas de mortalidad a nivel mundial también presenta una gran variabilidad con tasas que varían desde un 6 por 100,000 mujeres en el Este Asiático, hasta un 20 por 100,000 en África Occidental[2,3]. Sin embargo debido a la supervivencia más favorable del cáncer de mama en los países desarrollados (que presentan una alta incidencia), el rango de mortalidad en las regiones desarrolladas es mucho más homogéneo, con valores que varían entre el 16.4 por 100,000 mujeres en el norte de Europa, y el 14.5 por 100,00 en Australia y Nueva Zelanda[1-4].

Figura 1.1: Incidencia mundial del cáncer de mama en 2012 (Tasas estandarizadas por 100,000 mujeres-año).



Fuente: GLOBOCAN 2012. http://globocan.iarc.fr/

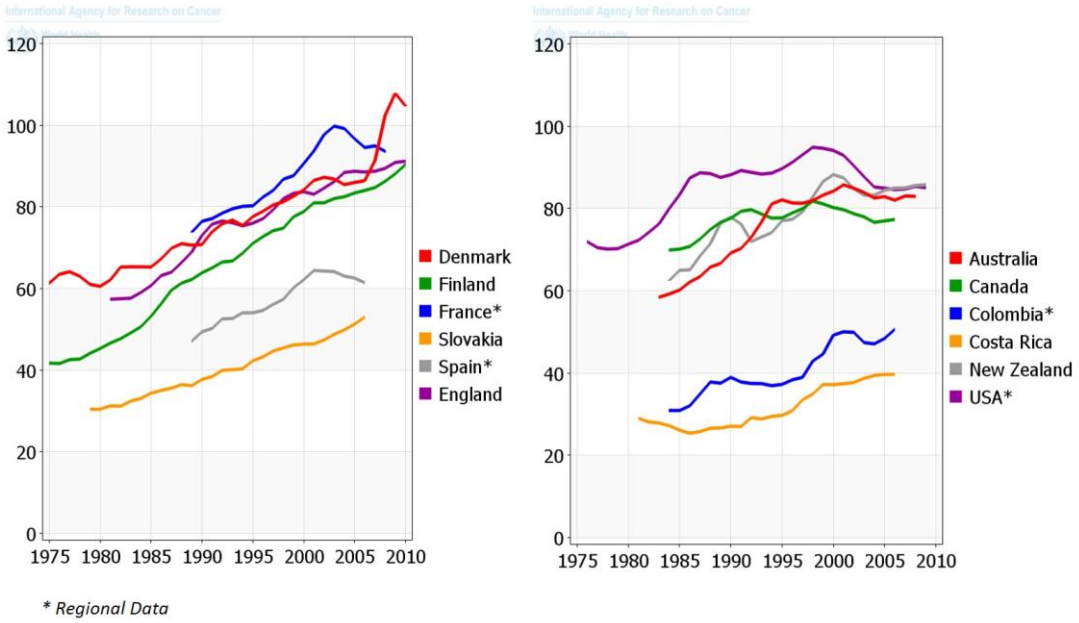
Figura 1.2: Mortalidad por cáncer de mama a nivel mundial en 2012 (Tasas estandarizadas por 100,000 mujeres-año).



Fuente: GLOBOCAN 2012. http://globocan.iarc.fr/

A pesar de la tendencia creciente en la incidencia de cáncer de mama a nivel mundial observada en las últimas 4 décadas, a partir del año 2000 en la mayoría de países con renta per cápita alta, entre ellos España, se ha comenzado a observar un descenso de dicha incidencia[5-14] (Figura 1.3).

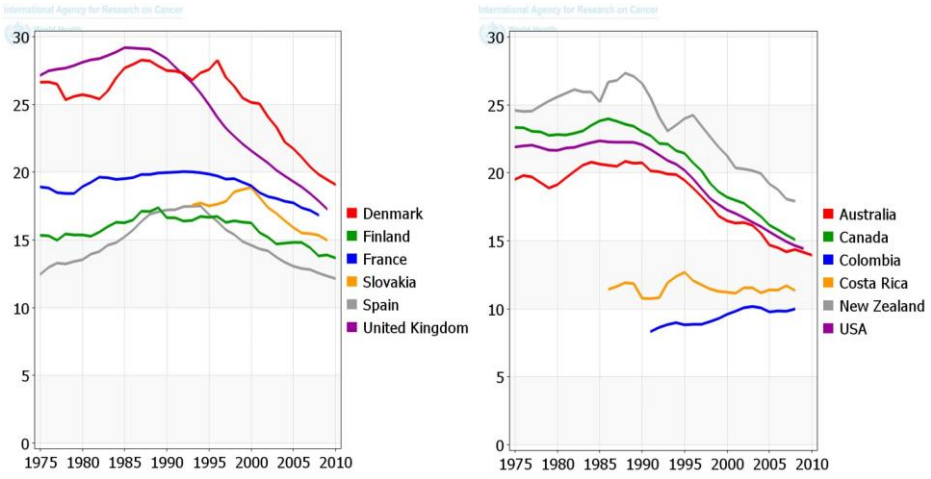
Figura 1.3: Tendencias temporales en la incidencia de cáncer de mama en una selección de países desde la década de los 70 hasta 2010 (Tasas estandarizadas por 100,000 mujeres-año).



Fuente: GLOBOCAN 2012. http://globocan.iarc.fr/

De manera análoga, a lo largo de las dos últimas décadas, la mortalidad por esta enfermedad ha disminuido sustancialmente, especialmente en los países desarrollados, gracias a una supervivencia más favorable, siendo actualmente una de las enfermedades oncológicas con mayor supervivencia, situándose por encima del 80% a los 5 años[15]. Esta mejora en la supervivencia se atribuye a la mejora en los tratamientos, a la introducción de las unidades funcionales y el trabajo multidisciplinar en los hospitales, y también a la implantación de las prácticas de detección precoz (Figura 1.4).

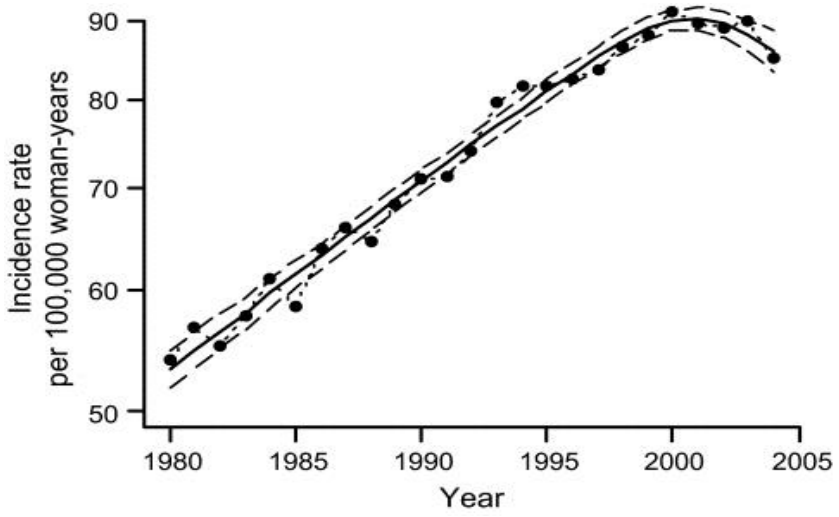
Figura 1.4: Tendencias temporales en la mortalidad por cáncer de mama en una selección de países desde la década de los 70 hasta 2010 (Tasas estandarizadas por 100,000 mujeres-año).



Fuente: GLOBOCAN 2012. http://globocan.iarc.fr/

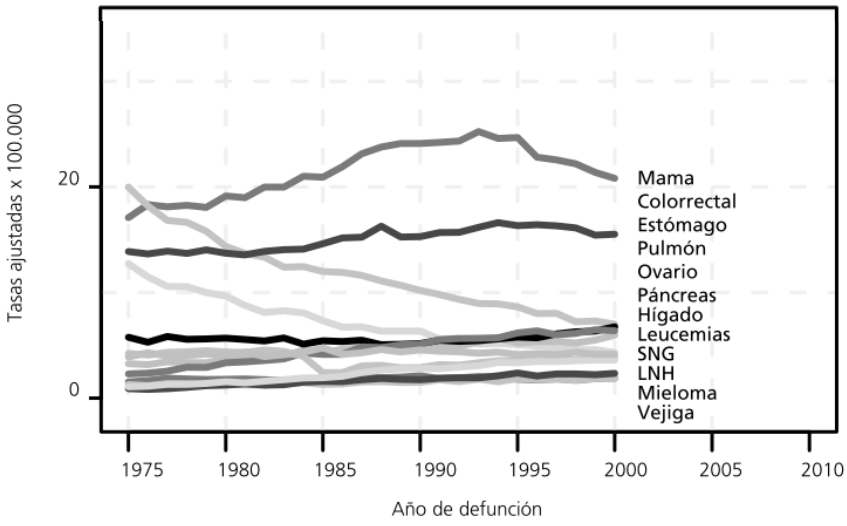
En el año 2012, se diagnosticaron en España más de 25,000 nuevos casos de cáncer de mama. La tasa de incidencia ajustada (utilizando como referencia la población mundial) fue de 67.3 nuevos casos por 100,000 mujeres, siendo la media de los países de la Unión Europea de 82.1 casos nuevos por 100,000 mujeres [2]. A pesar de la tendencia creciente observada en las últimas décadas (de manera análoga a la mayoría de países desarrollados), España sigue siendo uno de los países europeos con menor incidencia de cáncer de mama[2,4,5,16-19]. El cáncer de mama es la causa más frecuente de muerte por cáncer en las mujeres españolas, y se estima que se producen más de 6,000 muertes anuales por esta enfermedad. La tasa de mortalidad ajustada en el año 2012 (utilizando como referencia la población mundial) fue de 11.8 por 100,000 mujeres, siendo la media de los países de la Unión Europea de 15.5 por 100,000 mujeres y año[2].

Figura 1.5: Incidencia por cáncer de mama invasivo en España 1980–2004. Tasas estandarizadas por edad y sexo (mujer) de la población europea de referencia por 100,000 mujeres-año.



Fuente: Pollán et al. J Natl Cancer Inst. 2009; 101: 1584–1591.

Figura 1.6: Mortalidad por cáncer de mama en España 1975–2000 (línea superior). Tasas estandarizadas por edad y sexo de la población europea de referencia por 100,000 mujeres-año.



Fuente: La situación del cáncer en España; Instituto de Salud Carlos III

2. Cribado de cáncer de mama

La detección precoz del cáncer de mama mediante mamografía tiene como principal objetivo la reducción de la mortalidad global por esta enfermedad a partir de la detección de tumores en fases poco avanzadas. Para conseguir este objetivo de perspectiva poblacional, es imprescindible conseguir una amplia participación, lo que implica que un gran número de mujeres sanas se sometan periódicamente a exámenes mamográficos. Desde una óptica poblacional se espera que la implementación del cribado se traduzca a medio o largo término en una disminución de la mortalidad por la enfermedad, a pesar de que a nivel individual, no todas las mujeres participantes se beneficien directamente de la participación en los programas organizados.

2.1 Justificación para la aplicación del cribado mamográfico

A finales de los años 60 se describieron por primera vez cuales eran las condiciones necesarias para aplicar pruebas de detección precoz de una enfermedad[20]. En primer lugar, que la enfermedad represente un problema de salud importante, afectando de manera sustancial en la calidad y en la esperanza de vida. Segundo, que exista un tratamiento aceptado para dicha enfermedad. Tercero, deben de existir recursos para el diagnóstico y el tratamiento. Cuarto, ha de haber un periodo de latencia o con síntomas precoces detectables. Quinto, Debe de existir una prueba o exploración de cribado adecuada. Sexto, la prueba ha de ser ampliamente aceptada por médicos y pacientes. Séptimo, se ha de conocer suficientemente la historia natural de la enfermedad, incluyendo la progresión de la fase latente hacia la enfermedad declarada. Octavo, ha de haber una política clara sobre qué casos se tratan como pacientes. Noveno, que se tratase de una prueba de cribado con una buena relación coste-efectividad. Y finalmente, que la detección precoz de la enfermedad sea un proceso continuo y no que se haga una única vez.

Estos criterios se consideran clásicos y han sido citados en innumerables ocasiones. Con posterioridad han ido surgiendo revisiones, adaptaciones y actualizaciones de estos criterios.

En el caso particular del cáncer de mama, la efectividad del cribado se mide en términos de reducción de la mortalidad por esta enfermedad. La evidencia sobre la eficacia del cribado de cáncer de mama mediante mamografía se considera ampliamente aceptada[21-26]. En el cribado de cáncer de mama, se realiza una mamografía cada dos años a mujeres asintomáticas, con el

objetivo de detectar posibles tumores en fase preclínica. Aquellas mujeres en las que se confirma la enfermedad reciben tratamiento. En la actualidad, el Consejo Europeo recomienda la detección precoz del cáncer de mama. En España se ofrece el cribado de cáncer de mama con carácter poblacional, y forma parte de la estrategia del Sistema Nacional de Salud[27,28].

2.2 Evidencia e implementación del cribado de cáncer de mama en Europa y en España

Desde la década de los 80, distintos ensayos aleatorizados y controlados han estudiado el efecto del cribado de cáncer de mama mediante mamografía, valorando diferentes edades de cribado y periodicidad de la mamografía[29-34]. (Tabla 2.1).

Tabla 2.1: Principales características de los ensayos aleatorios sobre cribado de cáncer de mama

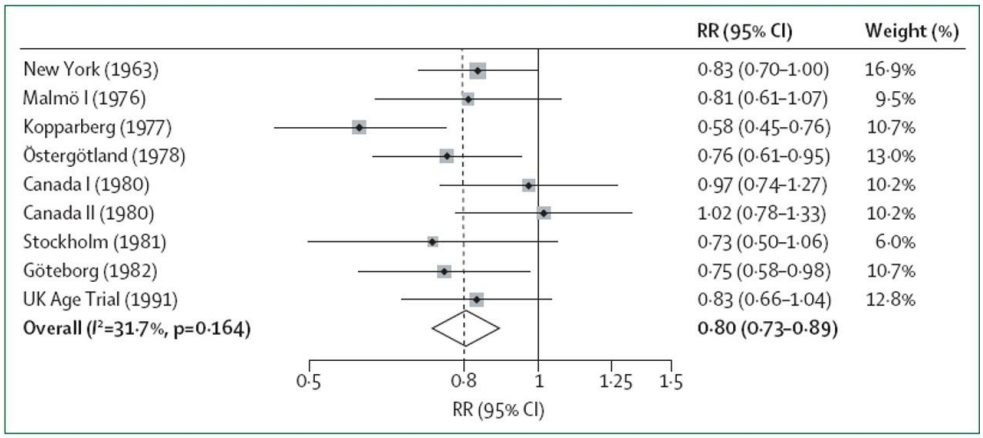
	New York HIP	Malmö I and II	Swedish two Country	Canada I and II	Stockholm	Göteborg	UK Age trial	Edinburgh
Start date	1963	1976	1977	1980	1981	1982	1991	1978
Randomisation method	Individual	Individual	Cluster	Individual	Day of birth	Day of birth	Individual	Cluster
Population of women								
Source	IC	P	P	Various	P	P	PC	PC
Number of women (clusters)	62 000	60 076	133 065 (45)	89 835	60 800	52 222	160 921	54 654 (87)
Age group (years)	40–64	45–69 and 43-49	38–75	40–49 and 50-59	39–65	39–59	39–41	45–64
Invited group intervention	M+PE	M	M+SE	M+PE+SE	M	M	M	M+PE
Mammography								
Number of views	2	2 then 1 or 2	1	2	1	2 then 1	2 then 1	2 then 1
Screening interval (months)	12	18–24	24–33	12	24–28	18	12	24
Number of screening rounds	4	6–8	2–4	4–5	2	4–5	8–10	2–4
Duration of screening (years)	3	12	7	5	4	7	8	6
Attendance	65 %	74 %	85 %	88 %	82 %	84 %	81 %	65 %
Control group intervention	None	None	None	PE+SE§	None	None	None	None
Follow-up								
Controls	Not known	Never	After 7 years	Never	After 4 years	After 7 years	After 10 years	After 10 years
Cause	L	IEC, NS	L, IEC, NS	IEC, NS	IEC, NS	NS	NS	NS

IC=insurance company register. P=population register. PC=primary care register. M=mammography. PE=physical examination. SE=self-examination. L=local. IEC=independent endpoint committee. NS=national statistics or register

Fuente: Adaptado de The Independent UK Panel on Breast Cancer Screening. Lancet 2012;380: 1778–86.

A pesar de que alguno de estos estudios ha sido cuestionado metodológicamente, debido principalmente a los criterios de aleatorización utilizados, en global los resultados muestran una reducción de la mortalidad en las mujeres cribadas. El cribado en mujeres de 50 a 69 años mostró una reducción estadísticamente significativa de la mortalidad por cáncer de mama alrededor del 20%[21] (Figura 2.1).

Figura 2.1: Resultados del meta-análisis sobre reducción de la mortalidad del cribado mamográfico en los ensayos aleatorios.



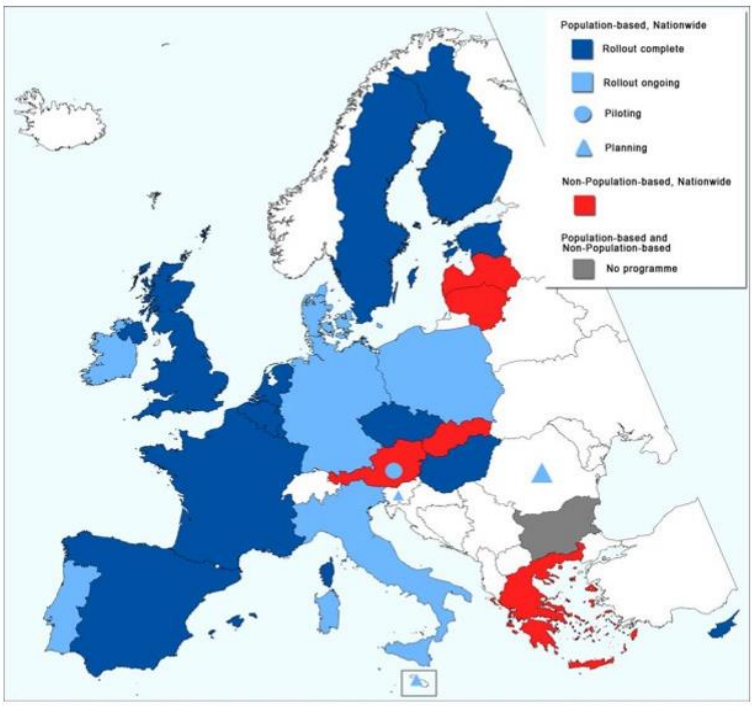
Fuente: The Independent UK Panel on Breast Cancer Screening. Lancet 2012;380: 1778–86.

A raíz de la publicación de los primeros ensayos aleatorizados que mostraban resultados favorables, en la década de los 90, la mayoría de comunidades científicas empezaron a recomendar el cribado mediante mamografía para la detección precoz del cáncer de mama. Progresivamente, numerosos países han ido poniendo en marcha programas de carácter poblacional. En la actualidad el Consejo Europeo recomienda el cribado poblacional del cáncer de mama a las mujeres entre 50-69 años. De igual manera se recomienda que los programas sigan los estándares establecidos en las "Guías europeas de Garantía de Calidad en el Cribado de Cáncer de Mama"[35,36].

En el año 2007, 26 de los 27 estados miembros de la Unión Europea habían puesto en marcha programas para la prevención de cáncer de mama, la mayoría de los cuales con carácter poblacional. Sin embargo, tan solo 11 de estos 26 países han desplegado el cribado a nivel de

todo el territorio. En conjunto, se calcula que en la Unión Europea cerca de 59 millones de mujeres se encuentran en la población diana según las recomendaciones de edad, y de estas el 41% viven en países con una cobertura total, y el 44% en países con una cobertura parcial[37].

Figura 2.2: distribución y alcance de los programas de cribado mamográfico en la Unión Europea en 2007.



Fuente: European Commission (DG SANCO, 2007); IARC (ECN and EUNICE projects, 2007).

Merece la pena mencionar, que a pesar de la existencia de las guías Europeas de Calidad, que tienen el objetivo de homogeneizar las prácticas de cribado, existen diferencias organizativas. El mayor consenso se encuentra en la periodicidad del cribado que es de 2 años en todos los países, excepto en el Reino Unido y en Malta que es de 3 años. Respecto a la edad de la población diana, si bien la mayoría cubren el rango recomendado (12 países, 50-69 años), existen algunos países donde el rango de edad es menor (50-59 años, 50-65 años) o incluso superior (45-69 años, 40-70 años). La mayoría de las diferencias entre programas, sin embargo, recaen en los criterios de elegibilidad y en el protocolo de las prácticas de cribado.

El primer programa de detección precoz de cáncer de mama en el territorio español se puso en marcha en Navarra en el año 1990. Con posterioridad y de manera gradual, se han ido iniciando programas en el resto de comunidades autónomas, hasta llegar a la cobertura total de la población diana en el año 2006. La organización de los programas recae en las autoridades sanitarias de las distintas comunidades autónomas. A pesar de la gestión independiente que tienen, los programas se coordinan a través de la Red de Programas de Cribado (http://www.programascancerdemama.org)[38]. Los responsables de los distintos programas se reúnen anualmente, y ponen en común los resultados de indicadores relacionados con la organización, los recursos y otros elementos destinados a garantizar su calidad a partir de la evaluación específica y conjunta de los programas.

Tabla 2.2: Año de inicio del programa, fuente de datos demográficos y grupo de edad diana en 2005.

Cominidad Autónoma	Año de inicio del programa	Grupo de edad diana	Fuente de datos demográficos
Navarra	1990	45-69	Padrón
Asturias	1991	50-69	Padrón y tarjeta sanitaria
Castilla-La Mancha	1992	45-68	Padrón y tarjeta sanitaria
Castilla y León	1992	45-69	Censo y tarjeta sanitaria
Cataluña	1992	50-69	Censo, padrón y tarjeta sanitaria
Comunidad Valencia	1992	45-69	Padrón y tarjeta sanitaria
Galicia	1992	50-65	Padrón y tarjeta sanitaria
La Rioja	1993	45-65	Padrón y tarjeta sanitaria
Murcia	1995	50-69	otro: PERSAN*
Andalucía	1995	50-65	Padrón y tarjeta sanitaria
País Vasco	1995	50-64	Padrón
Aragón	1997	50-64	Padrón y tarjeta sanitaria
Cantabria	1997	50-64	Padrón y tarjeta sanitaria
Baleares	1997	50-64	Padrón y tarjeta sanitaria
Extremadura	1998	50-65	Padrón y tarjeta sanitaria
Canarias	1999	50-69	Padrón y tarjeta sanitaria
Madrid	1999	50-64	Tarjeta sanitaria
Ceuta	2001	45-65	Tarjeta sanitaria

* Base de daot s de Salud Pública que se nutre, fundamentalmente, de tarjeta sanitaria

Fuente: Adaptado del Informe DESCRIC (AATRM Núm.2006/01).

Todos los programas del territorio español han adoptado las recomendaciones de las Guías Europeas para el cribado de cáncer de mama[36] y utilizan la mamografía bienal como prueba de

cribado. El grupo de edad de la población diana es de 50-64 años, pero la mayoría de programas han optado por ampliarlo hasta los 69 años. Además, en 6 comunidades autónomas también se incluye el grupo de 45 a 69 años. En la actualidad los programas cubren la práctica totalidad de la población diana que en el año 2005 incluía más de 4,300,000 mujeres.

Tabla 2.3: Población diana y cobertura de los programas del territorio español en 2005

Comunidad Autónoma	Mujeres del grupo de edad diana	Población cubierta	
		Número	%
Ceuta	6.767	3.382	50
La Rioja	31.584	31.584	100
Cantabria	49.850	49.850	100
Baleares*	70.453	69.351	98,4
Extremadura	74.908	74.908	100
Navarra	84.263	84.263	100
Aragón	111.742	107.719	96,4
Murcia	121.300	121.300	100
Asturias	134.279	134.279	100
Canarias	173.808	149.127	85,8
País Vasco	205.293	205.293	100
Castilla-La Mancha	230.098	230.098	100
Galicia	288.986	288.986	100
Castilla y León	366.127	366.127	100
Madrid	531.143	531.143	100
Cataluña	562.955	562.955	100
Comunidad Valenciana	581.047	581.047	100
Andalucía	689.000	689.000	100
TOTAL	4.313.603	4.280.412	99,39

* Mallorca e Ibiza

Fuente: Adaptado del Informe DESCRIC (AATRM Núm.2006/01).

Sin embargo, existen algunas diferencias importantes entre programas, especialmente en el método de lectura, simple o doble, y en el caso de doble lectura, si está se hace con consenso o arbitraje; y en el número de proyecciones, ya que algunos programas realizan una única proyección en cribados sucesivos, siendo el criterio unánime la doble proyección (cráneo-caudal y

contra-lateral) para cribado inicial[27]. Además, en lo referente al tipo de mamografía, desde el año 2000, algunos programas han optado por sustituir los mamógrafos analógicos por digitales. En el año 2012, nueve comunidades autónomas habían introducido total o parcialmente la mamografía digital, y se prevé que esta técnica se vaya extendiendo a todos los programas[38]. Todas estas características añaden heterogeneidad a la hora de evaluar y comparar los programas de cribado.

Tabla 2.4: Principales características de proceso y protocolo de lectura en los programas del territorio español en 2005.

Comunidad Autónoma	Prueba de cribado	Periodicidad	Nº de proyecciones	Sistema de lectura	Uso del Bi-Rads
Andalucía	Mamografía	2 años	2	Doble lectura sin consenso	Sí
Aragón	Mamografía	2 años	2	Lectura simple	Sí
Asturias	Mamografía	2 años	2 en primer cribado; 1 en cribados sucesivos	lectura simple (30% doble sin consenso)	Sí
Baleares	Mamografía	2 años	2	Doble lectura con consenso	Sí
Canarias	Mamografía	2 años	2	Doble lectura con consenso	Sí
Cantabria	Mamografía	2 años	2	Lectura simple	Sí
Castilla y León	Mamografía	2 años	2	Si lectura simple positiva, doble lectura con consenso (más el 25% de las lecturas simples negativas)	Sí
Castilla-La Mancha	Mamografía	2 años	2	lectura simple (doble con consenso en BI-RADS III)	Sí
Cataluña	Mamografía	2 años	2	Doble lectura con consenso	Sí
Ceuta	Mamografía	2 años	2	Lectura simple	No
Comunidad Valenciana	Mamografía	2 años	2 en primer cribado; 1 en cribados sucesivos	Doble lectura con consenso	Sí
Extremadura	Mamografía	2 años	2	Doble lectura con consenso	Sí
Galicia	Mamografía	2 años	2	Doble lectura sin consenso	Sí
La Rioja	Mamografía	2 años	2	Doble lectura con consenso	Sí
Madrid	Mamografía	2 años	2	Doble lectura con consenso	Sí
Murcia	Mamografía	2 años	2	Lectura simple	Sí
Navarra	Mamografía	2 años	2	Lectura simple	Sí
País Vasco	Mamografía	2 años	2	Lectura simple	Sí

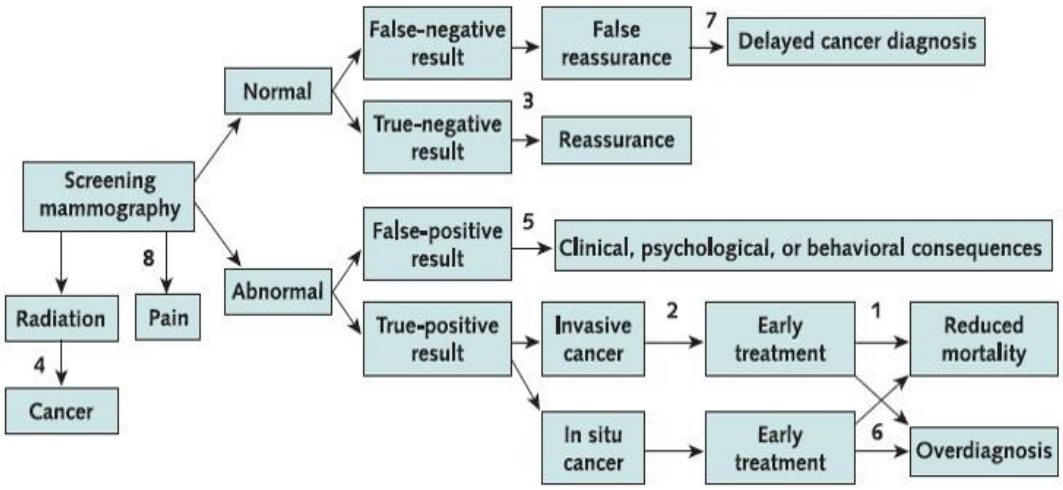
Fuente: Adaptado del Informe DESCRIC (AATRM Núm.2006/01).

2.3 Beneficios y efectos adversos del cribado

Los ensayos aleatorizados y controlados han demostrado un claro efecto del cribado sobre la reducción de la mortalidad por cáncer de mama bajo circunstancias relativamente controladas.

Sin embargo, los beneficios de la mamografía de cribado en el contexto poblacional no tienen por qué ser similares. Se espera que la efectividad del cribado haya mejorado desde la publicación de los primeros ensayos aleatorizados, ya que el control de calidad, la formación del personal especializado, y las técnicas mamográficas han mejorado con el tiempo[39], así como el tratamiento del cáncer de mama que ha mejorado gracias al uso extendido de tratamientos específicos[40].

Figura 2.3: Beneficios y efectos adversos del cribado mamográfico.



Fuente: Armstrong et al. *Ann Intern Med.* 2007;146:516-526.

En los últimos años se han publicado numerosos estudios observacionales que evalúan el impacto del cribado de cáncer de mama con resultados muy dispares. En parte, la disparidad en relación a los beneficios y efectos adversos del cribado, y el mantenimiento de esta controversia se deben a la diferente validez, a los diferentes diseños utilizados, y a las particulares exigencias metodológicas de los estudios del cribado.

Recientemente, el *EUROSCREEN Working Group* ha publicado una revisión sistemática de estudios poblacionales realizados en el contexto europeo en la que se evalúan los beneficios y los efectos adversos del cribado. En esta revisión se estima una reducción de la mortalidad de entre un 25% y un 31% [41]. Paralelamente, una revisión independiente sobre los beneficios y efectos adversos del cribado realizada por el, *The Independent UK Panel on Breast Cancer Screening* (IUKPBCS), estimó

que la reducción de la mortalidad es del 20% [21]. Más recientemente, se ha publicado un estudio que estimaría una reducción del 43% en la mortalidad atribuible al cribado[42]. A finales de 2014, un informe de la Organización Mundial de la Salud (OMS) que revisa la evidencia sobre beneficios y efectos adversos del cribado concluye que el cribado mamográfico poblacional en mujeres de 50 a 69 años reduce la mortalidad por esta enfermedad, y es coste-efectivo en países de renta media-alta[43].

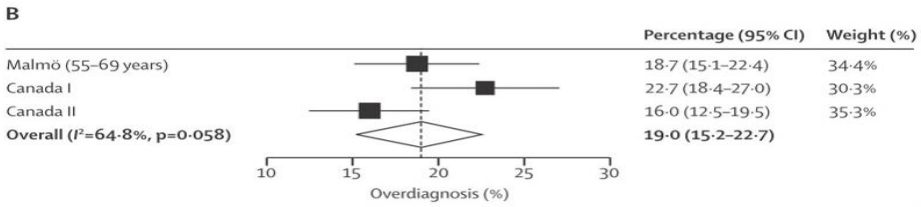
De manera inevitable, el cribado de cáncer de mama conlleva efectos adversos. La existencia de los mismos ha sido reconocida desde hace muchos años[44]. A nivel individual, no todas las mujeres participantes obtienen los mismos beneficios del cribado. Debido a que el cribado mamográfico se ofrece a una población sana y numerosa, sus efectos adversos deben mantenerse en el mínimo, manteniendo un nivel de beneficio aceptable. Los principales efectos adversos del cribado son los falsos positivos, los falsos negativos y el sobrediagnóstico. Cada uno de ellos con diferente alcance y consecuencias[45].

Los resultados falsos negativos son tumores que eran visibles en la mamografía de cribado pero que terminan diagnosticándose por otras vías fuera del cribado. Pueden ser debidos a errores en la interpretación radiológica, o a errores técnicos en la realización de la mamografía. Los falsos negativos generan retraso en el diagnóstico de la enfermedad, y están asociados a un mayor tamaño del tumor[46,47], más nodos positivos[46], y a una menor supervivencia[46-49], disminuyendo así la efectividad del cribado. Además, generan una falsa sensación de seguridad en la mujer que las hace estar menos alerta a los posibles signos de la enfermedad.

Por su parte, el sobrediagnóstico se puede definir como la detección de un cáncer de mama en el cribado, histológicamente confirmado, que nunca se habría diagnosticado clínicamente durante la vida de la mujer[50]. Por su impacto emocional y del propio tratamiento se considera que este es el efecto adverso más grave de la detección precoz, y es el que genera mayor controversia. A nivel individual, es imposible distinguir que tumores han sido sobrediagnosticados y cuales detectados precozmente, y ambos son tratados de igual manera. A nivel poblacional, sin embargo, el sobrediagnóstico implica un mayor número de mujeres diagnosticadas, lo que aumenta el coste diagnóstico y de tratamiento, así como el número de mujeres que sufrirá el estrés y la ansiedad debida al diagnóstico de la enfermedad.

Existen diversas aproximaciones matemáticas para la estimación del sobrediagnóstico sin que exista una metodología estándar para su cálculo, siendo el principal problema, disponer de una población de referencia comparable en ausencia del cribado. Las estimaciones publicadas hasta la fecha varían enormemente y son altamente susceptibles a interpretaciones sesgadas, variando entre el 1% y el 50% [51-53]. El monográfico del *EUROSCREEN Working Group* sitúa el sobrediagnóstico entre el 1% y el 10%, con un valor estimado del 6,5% [50], mientras que la revisión del IUKPBCS, estima que el 19% de tumores diagnosticados en el cribado estarían sobrediagnosticados[21].

Figura 2.4: Meta-análisis de los estimadores de sobrediagnóstico de los ensayos aleatorios sin final sistemático de cribado en el grupo control.



Fuente: The Independent UK Panel on Breast Cancer Screening. Lancet 2012;380: 1778–86.

Por último, los resultados falsos positivos, que son el eje central de esta tesis, se definen como la recomendación de realizar exploraciones adicionales (mamografía adicional, ecografía, resonancia magnética, punción con aguja fina, biopsia escisional, y/o biopsia quirúrgica) para detectar malignidad a mujeres en las que finalmente se descarta la presencia de un cáncer de mama. Los resultados falsos positivos generan preocupación y ansiedad en las mujeres afectadas[54-57], así como un mayor número de pruebas adicionales con un coste asociado[58]. Además, distintos estudios muestran que estas mujeres tienen una menor adherencia al programa en sucesivas convocatorias, y presentan un mayor riesgo de desarrollar cáncer de mama[59-62]. Sin embargo, los falsos positivos son una parte intrínseca e inevitable de los programas de cribado poblacional y se deben mantener en el nivel más bajo posible, manteniendo una tasa de detección adecuada. Las Guías Europeas recomiendan que el porcentaje de mujeres reconvocadas para exploraciones adicionales no supere el 7% en cribado inicial y el 5% en cribado sucesivo[36]. A nivel poblacional,

los falsos positivos son el efecto adverso con mayor impacto, tanto por el volumen de mujeres a las que afecta, como por los recursos y la carga asistencial que suponen. Se estima que el riesgo acumulado de presentar un resultado falso positivo a lo largo de la historia de cribado de una mujer en el contexto europeo es del 19.7% [56]. Con anterioridad a los estudios de esta tesis, en el territorio español, el riesgo acumulado de falso positivo había sido estimado únicamente en el contexto de un único programa de cribado de la ciudad de Barcelona, con un riesgo acumulado a lo largo de 10 convocatorias estimado en el 32.4% [63].

2.4 Controversias en la evaluación del cribado poblacional de cáncer de mama

A pesar de la evidencia existente sobre los beneficios del cribado de cáncer de mama, hay objeciones críticas al respecto. Algunas argumentan que los beneficios del cribado serían menores de lo asumido, mientras que los riesgos serían mayores.

En el año 2001, la *Cochrane Collaboration* publicó un meta-análisis de los resultados de los distintos ensayos aleatorizados y controlados, en el que se concluía que el cribado mamográfico no mejoraba la supervivencia, y que los efectos en la mortalidad por cáncer de mama no eran concluyentes[64,65]. La publicación de este estudio supuso un punto de inflexión a partir del cual se ha abierto un debate sobre la idoneidad del cribado de cáncer de mama con las actuales estrategias. Sus resultados fueron fuertemente rebatidos[25,66] y en el año 2006, la propia *Cochrane Collaboration* matizó sus conclusiones en una revisión del meta-análisis que mostraba una reducción de 15% de la mortalidad de cáncer de mama debida al cribado mamográfico[67].

Recientemente, el *EUROSCREEN Working Group* concluye que con la evidencia actual, el cribado de cáncer de mama continua siendo recomendable y que los beneficios del cribado justifican los efectos adversos. En este estudio se calcula que por cada 1000 mujeres cribadas bienalmente durante 20 años entre los 50 y los 69 años, y seguidas hasta los 79 años se diagnostican 71 cánceres de mama, se evitan entre 7 y 9 muertes, se sobrediagnostican 4 tumores, y 200 mujeres experimentan un resultado falso positivo [41]. Por su parte el informe del IUKPBCS concluye que por cada 1000 mujeres de 50-69 años cribadas durante 20 y seguidas hasta los 79 años se diagnostican 68 cánceres de mama, se evitan 4 muertes, y se sobrediagnostican 13 tumores[21].

Este informe subraya la necesidad de revisar la información que se da a las mujeres invitadas, para que sea lo más transparente posible y permita una toma de decisiones informada. En EEUU el debate se mantiene activo con la publicación de un estudio que encuentra que uno de cada tres cánceres diagnosticados mediante cribado estarían sobrediagnosticado, y que en el mejor de los casos el cribado solo tendría un pequeño efecto sobre la reducción de la mortalidad[68,69]. Por contra, una reciente publicación argumenta que la estimación de dicho estudio estaría fuertemente sobreestimada[70].

Las limitaciones metodológicas en la selección de los grupos de comparación, especialmente el grupo de mujeres no cribadas, y las distintas aproximaciones metodológicas utilizadas hacen que la evaluación de los beneficios del cribado poblacional de cáncer de mama sea una cuestión con respuestas heterogéneas e inacabadas. Actualmente el debate se centra en la mejora de la efectividad y comienza a focalizar su interés en la personalización del cribado [71]. Este enfoque pretende proponer estrategias de cribado personalizadas en función de distintos grupos de riesgo de la mujer en contraposición a la actual estrategia de “one size fits all” del cribado poblacional actual. De esta manera se pretendería maximizar los beneficios del cribado, y reducir los efectos adversos. Aunque se han hecho algunos avances en este aspecto, aún quedan cuestiones por resolver sobre el beneficio real del cribado mamográfico, y sobre los niveles aceptables de efectos adversos.

3. La necesidad de evaluar los resultados falsos positivos en el cribado mamográfico

Cómo ya se ha comentado anteriormente, la efectividad del cribado mamográfico está estrechamente ligada al balance entre beneficios y efectos adversos de esta práctica. Niveles altos de efectos adversos, reducen la efectividad y tienen un fuerte impacto negativo en las mujeres participantes. La valoración del impacto que los resultados falsos positivos tienen sobre las mujeres cribadas, y sobre la efectividad del cribado mamográfico es una cuestión clave para poder desarrollar programas de cribado efectivos.

3.1 Evidencia

El carácter poblacional del cribado hace que, en valores absolutos, los resultados falsos positivos sean el efecto adverso con mayor impacto, tanto por el volumen de mujeres a las que afecta, como por los recursos y la carga asistencial que suponen.

A pesar de su elevado impacto, un mínimo volumen de resultados falsos positivos es inevitable y es una parte intrínseca de los programas de cribado poblacional. Las dos grandes cuestiones de interés sobre los resultados falsos positivos en la actualidad son 1) ¿Cómo mantener en el nivel más bajo posible los resultados falsos positivos, manteniendo una tasa de detección adecuada?, y 2) ¿Cuál es el impacto que estos resultados falsos positivos tienen en las mujeres participantes a largo término?

3.1.1 Relación entre resultados falsos positivos y tasa de detección

Un buen test de cribado debe tener una buena sensibilidad, para no perder los casos que presentan la enfermedad, y una elevada especificidad, para reducir el número de personas con resultados falsos positivos que requieren exploraciones adicionales. Dado que la prevalencia del cáncer de mama es baja en términos absolutos, el valor predictivo positivo del cribado mamográfico será bajo, a pesar de que el test tenga una buena especificidad [72]. Es decir, deberemos de tener presente que pequeños cambios en la especificidad se amplifican cuando hay

una baja prevalencia. Desde esta perspectiva, deberemos de aceptar un mínimo volumen de falsos positivos para que el cribado mamográfico sea efectivo.

Las Guías Europeas establecen que el porcentaje de mujeres reconvocadas para exploraciones adicionales no supere el 7% en cribado inicial y el 5% en cribado sucesivo[36]. Si las tasas de detección en cribado inicial y sucesivo giran en torno al 0.45% en cribado inicial y 0.3 % en cribado sucesivo[27], implicaría que una proporción de resultados falsos positivos alrededor del 6.5% en cribado inicial y del 4.7% en cribado sucesivo se entendería como el máximo aceptable dada la baja prevalencia de la enfermedad.

Es conocida la existencia de variabilidad en la sensibilidad y en la especificidad del test de cribado entre diferentes contextos (inter-programa), e incluso dentro de un mismo programa de cribado (intra-programa). Estas diferencias se ven reflejadas en el valor predictivo positivo y valor predictivo negativo de la mamografía[73-78]. En términos generales, la variabilidad en el test dependerá; a) de las características de la mujer y b) diferencias en el observador.

La variabilidad en las características de la mujer no está relacionada con la mamografía de cribado en sí misma. La edad de la mujer, la densidad mamaría y los antecedentes familiares de cáncer de mama son algunas de las características que influyen sobre la sensibilidad y especificidad de la mamografía. La probabilidad de tener un resultado falso positivo disminuye con la edad de la mujer, mientras que la probabilidad de detección de cáncer aumenta. Por otro lado, la densidad mamaría y los antecedentes familiares aumentan el riesgo de presentar resultados falsos positivos y de presentar un cáncer de mama. La variabilidad en las características de la mujer cuestiona la adecuación de realizar cribado mamográfico bienal a todas las mujeres en la población diana, ya que la población cribada es heterogénea y habrá mujeres con poco beneficio esperado y un alto riesgo de efectos adversos. En la actualidad, se empiezan a debatir estrategias personalizadas de cribado que aumenten la efectividad en función de las características de la mujer. Por ejemplo, reducir la periodicidad del cribado en mujeres mayores de 60 años, sin antecedentes familiares y con baja densidad mamaría, o aumentar la periodicidad en mujeres jóvenes, con antecedentes familiares y alta densidad mamaria.

Por su parte, la variabilidad en el observador hace referencia a la realización del test de cribado. Características propias del protocolo de lectura mamográfica, como el uso de lectura simple o

doble, o el uso de una o dos proyecciones pueden tener efecto sobre la variabilidad, al igual que el uso de mamógrafo digital o analógico. Se considera que la variabilidad en la sensibilidad y especificidad de los radiólogos lectores es una de las principales fuentes de variación en la efectividad del cribado[74,75,77,78]. La experiencia de los radiólogos lectores y el número de mamografías leídas tienen impacto sobre el número de tumores detectados y el número de mujeres reconvocadas[75-78]. Incluso dentro de un contexto con un protocolo de lectura mamográfica común, donde los radiólogos lectores cumplen con las recomendaciones de las guías europeas se ha detectado una gran variabilidad en la tasa de resultados falsos positivos, y en menor medida en las tasas de detección[73].

Dado que algunos de estos factores, especialmente los relacionados con las características de la prueba, son potencialmente modificables, será importante conocer el impacto que estos tienen sobre el riesgo de tener un resultado falso positivo. Reducir los factores asociados con el desempeño del cribado mamográfico (experiencia de los radiólogos, método de lectura, etc.) deja espacio para mejorar la efectividad del mismo.

3.1.2 Impacto de los resultados falsos positivos a largo plazo

Como ya se ha comentado anteriormente, el cribado mamográfico recomienda una mamografía bienal a las mujeres entre 50 y 69 años de edad (10 cribados entre los 50 y los 69 años). El beneficio del cribado se mide en términos de reducción de la mortalidad como consecuencia de la participación de una población numerosa de mujeres a lo largo de sucesivas rondas de cribado. Para poder evaluar correctamente el balance beneficio-riesgo del cribado, es importante evaluar los efectos adversos de manera análoga, calculando el impacto de los mismos a lo largo de las sucesivas participaciones de la mujer. Las mujeres participantes son monitorizadas durante un amplio periodo de tiempo en que las observaciones no son independientes por dos razones; se observa a la misma mujer repetidas veces, y porque en el caso de los falsos positivos, el resultado está condicionado por los resultados previos. Por lo tanto, el impacto y la magnitud de los resultados falsos positivos deben analizarse como una prueba secuencial en el tiempo, es decir, como una cohorte.

Desde el punto de vista de la información proporcionada a las mujeres participantes, es importante analizar e informar a la mujer del riesgo para cada cribado (transversal) y para todo el periodo que se le propone controlarse (longitudinal, 10 cribados bienales). Por otra parte, desde una perspectiva clínica, la elección de realizar cribado bienal (test seriados) es de vital importancia. El valor predictivo del test es la característica más relevante cuando se interpretan los resultados del cribado[79]. Mediante la realización de pruebas de cribado secuenciales se maximiza la especificidad y el valor predictivo positivo, a pesar de que disminuirá la sensibilidad y el valor predictivo negativo. Dado que la mamografía de cribado no tiene una alta especificidad los test seriados son particularmente útiles[79].

En la actualidad la evaluación y el impacto de los resultados falsos positivos a largo plazo no están suficientemente estudiados y los resultados que los evalúan no son concluyentes.

3.2 Retos metodológicos en la evaluación de los falsos positivos

Para poder evaluar los falsos positivos será necesario disponer de información sobre la práctica del cribado, con el mayor nivel de detalle posible, en un gran número de mujeres participantes, seguidas durante un largo periodo de tiempo. El análisis detallado y en profundidad de los falsos positivos, más allá de la evaluación transversal propia de los programas, exigirá disponer de información individualizada de cada participación de la mujer en el cribado. Es deseable contar con información precisa sobre el resultado de la interpretación mamográfica, la recomendación de realizar exploraciones adicionales, el tipo de pruebas realizadas, el resultado de las mismas, y el diagnóstico o no de cáncer de mama. Además, será recomendable disponer de información referente a los factores de riesgo que puedan tener un impacto sobre los falsos positivos (antecedentes familiares, método de lectura mamográfica, etc.). En términos generales, esta información es específica de cada programa de cribado y no se recoge ni se codifica de manera uniforme.

Por su parte, la necesidad de un volumen elevado de mujeres para poder evaluar los resultados falsos positivos se justifica por la frecuencia del evento. A pesar de que un resultado falso positivo afectará aproximadamente a un 8.4 % de las mujeres participantes en cribado inicial, y un 3.3% en

cribados sucesivos[56], la frecuencia de los distintos factores de interés asociados a los resultados falsos positivos, como pueden ser la presencia de antecedentes familiares, el uso de terapia hormonal sustitutiva, la participación o el diagnóstico de cáncer en rondas posteriores, será aún menor dentro de este porcentaje de mujeres. A pesar de esto, su impacto a nivel poblacional será elevado, ya que millones de mujeres son invitadas a participar en el cribado poblacional cada año. Desde una perspectiva poblacional será especialmente relevante poder trabajar con poblaciones de mujeres cribadas y no con muestras.

A medida que los programas poblacionales se iban desplegando durante el último quinquenio de la década de los noventa y la década del 2000, han ido surgiendo diferentes estudios y aproximaciones para evaluar sus efectos adversos. La situación actual en la que la mayoría de programas llevan funcionando al menos una década presenta un escenario actualizado con posibilidad de disponer de datos longitudinales sobre la práctica del cribado en un largo periodo de tiempo. Este escenario abre nuevas cuestiones sobre como evaluar los falsos positivos a largo plazo.

Una aproximación adecuada al análisis de los falsos positivos a largo plazo deberá de tener en cuenta la correlación entre las múltiples observaciones de una mujer a lo largo del periodo de estudio (medidas repetidas). Además, determinadas características de estudio, como la presencia de antecedentes familiares, el método de lectura o la técnica (digital o analógica) pueden cambiar entre las distintas observaciones de una misma mujer, lo que hace deseable poder incluir variables cambiantes en el tiempo, en contraposición a asumir información fija a lo largo de todo el periodo de estudio. Por último, las mujeres en la población diana, son invitadas a realizarse una mamografía cada dos años, y por lo tanto las mujeres participantes solo están a riesgo de tener un resultado falso positivo cuando participen en el programa, siendo imposible tener un falso positivo si la mujer no participa o en el intervalo entre mamografías. Esta última peculiaridad hace que la aproximación al estudio de los falsos positivos se haga desde una perspectiva de tiempo discreto, en contraposición a las técnicas más habituales que asumen riesgo continuo en el tiempo, como por ejemplo los modelos de regresión de Cox.

En el año 1998 se publicó el primer estudio que estimaba el riesgo acumulado de experimentar un resultado falso positivo a lo largo de 10 cribados. El estudio se desarrolló en el contexto de cribado oportunista de EEUU, donde las mujeres participantes tenían entre 40 y 69 años, y la

periodicidad del cribado variaba, siendo frecuente el cribado anual[80]. Este estudio estimaba un riesgo acumulado de falsos positivos del 49.1%. Posteriormente, en 2004 se publicó la primera estimación del riesgo acumulado de falsos positivos a lo largo de 10 participaciones bienales entre los 50 y los 69 años en el cribado poblacional[81]. Este estudio se realizó en el programa de Noruega y estimaba un 20% de riesgo acumulado. El estudio ha sido citado en numerosas ocasiones y abrió la puerta a la evaluación de los falsos positivos y su impacto desde una perspectiva poblacional y no tanto de la evaluación interna y transversal del propio programa. Desde una perspectiva metodológica se trata de un estudio relativamente limitado, que ha sido criticado por sus deficiencias[82,83]. Posteriormente a la publicación de este estudio, únicamente 2 trabajos han estimado el riesgo acumulado de falsos positivos en el cribado poblacional, y ninguno de ellos incluía ajuste por factores de estudio adicionales a la edad[84,85].

Dentro del contenido de esta tesis hay una actualización de la estimación del riesgo acumulado de falsos positivos en el programa noruego, ampliando el periodo de estudio, y utilizando la metodología desarrollada para los trabajos de esta tesis.

3.3 Justificación del estudio sobre Riesgo Acumulado de Falsos Positivos (RAFP)

El territorio español representa un claro ejemplo de la complejidad existente para disponer de la información necesaria para la evaluación del cribado mamográfico, donde hay un programa de cribado en cada Comunidad Autónoma. En la totalidad de los programas se han adoptado las recomendaciones de las Guías Europeas para el cribado de cáncer de mama. Sin embargo, tal y como se ha comentado anteriormente, hay diferencias relevantes entre programas con relación al protocolo de puesta en práctica de la mamografía de cribado (método de lectura, número de proyecciones, o la edad de inicio del cribado). Tales diferencias, al igual que otras de orden más organizativo, conllevan variabilidad en entre los distintos programas de cribado. Esta situación ponía de manifiesto la necesidad de desarrollar un proyecto específico para estudiar los resultados falsos positivos en el cribado poblacional en España.

En el año 2007 se inicia un proyecto para evaluar, de forma conjunta, el riesgo de falsos positivos en el cribado mamográfico y su asociación con el protocolo de lectura mamográfica y las

características de la mujer en 10 programas poblacionales de cribado de cáncer de mama de 8 comunidades autónomas del territorio español. Este proyecto, fue financiado por el Instituto de Salud Carlos III a través del Fondo de Investigaciones Sanitarias (PI06/1230 y PI09/90251). En el diseño y desarrollo del estudio se contó con la participación de los responsables de cada uno de los programas participantes, lo que permitió precisar y matizar la información recogida y las definiciones utilizadas, de manera que fuesen representativas de la realidad del cribado desde una perspectiva clínica y epidemiológica.

El objetivo de este estudio era estimar la probabilidad acumulada de presentar un resultado falso positivo a lo largo de 10 participaciones bienales de la mujer en el cribado entre los 50 y los 69 años. Además el estudio tenía como objetivos: 1) Estimar la probabilidad acumulada de presentar un falso positivo con exploraciones adicionales invasivas; 2) Evaluar la asociación entre la probabilidad acumulada de falsos positivos y el protocolo de lectura mamográfica, y las características propias de la mujer; 3) Evaluar el impacto de los resultados falsos positivos en la participación de la mujer en sucesivas convocatorias de cribado; 4) Evaluar el impacto de los resultados falsos positivos sobre el riesgo de detección de cáncer de mama en sucesivas convocatorias de cribado.

La población de estudio estaba comprendida por las mujeres participantes en el cribado de 8 Comunidades Autónomas (Navarra, Valencia, Galicia, Canarias, Castilla y León, La Rioja, Asturias y Cataluña). Cada uno de los programas participantes aportó información individualizada de todas las mujeres participantes en el cribado al menos una vez desde el inicio de los mismos, hasta diciembre de 2006. El estudio cubría el 44% de la población diana española, con 4,739,498 mamografías de cribado de 1,565,364 mujeres. Las mamografías de cribado se llevaron a cabo en 74 unidades radiológicas diferentes.

4. Presentación de los trabajos que conforman la tesis

Esta tesis está compuesta por 5 estudios publicados como artículos científicos en revistas internacionales indexadas.

Artículo 1. Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening.

Román M, Sala M, Salas D, Ascunce N, Zubizarreta R, Castells X, and the Cumulative False Positive Risk Group.

Ann Oncol. 2012; 23: 104-111.

Artículo 2. Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening

Román M, Sala M, De La Vega M, Natal C, Galceran J, González-Román I, Baroja A, Zubizarreta R, Ascunce N, Salas D, Castells X, and the Cumulative False Positive Risk Group.

Breast Cancer Res Tr. 2011; 130(2): 543-552.

Artículo 3. Breast cancer detection risk in screening mammography after a false-positive result

Castells X, Román M, Romero A, Blanch J, Zubizarreta R, Ascunce N, Salas D, Burón A, Sala M, the Cumulative False Positive Risk Group.

Cancer Epidemiol. 2013; 37(1): 85-90.

Artículo 4. Trends in detection of invasive cancer and ductal carcinoma in situ at biennial screening mammography in Spain: A retrospective cohort study.

Román M, Rué M, Sala M, Ascunce N, Baré M, Baroja A, De la Vega M, Galcerán J, Natal C, Salas D, Sánchez-Jacob M, Zubizarreta R, Castells X, and the Cumulative False Positive Risk Group.

PLoS ONE. 8(12): e83121.

Artículo 5. The cumulative risk of false-positive results in the Norwegian Breast Cancer Screening Program: Updated results.

Román M, Hubbard RA, Sebuødegård S, Miglioretti DL, Castells X, Hofvind S.

Cancer. 2013; 119: 3952-3958

Artículos anexos

Anexo 1. The cumulative risk of false-positive screening results across screening centres in the Norwegian Breast Cancer Screening Program.

Román M, Skaane P, Hofvind S.

Eur J Radiol 2014; 83: 1639–1644

Anexo 2. Effect of start age of breast cancer screening mammography on the risk of false-positive results.

Salas D, Ibañez J, Román M, Cuevas D, Sala M, Ascunce N, Zubizarreta R, Castells X, and The Cumulative False Positive Risk group.

Prev Med 2011; 53(1-2): 76-81.

Anexo 3. Effect of radiologist experience on the risk of false-positive results in breast cancer screening programs.

Zubizarreta R, Fernández AB, Almazán R, Román M, Velarde J, Queiró T, Natal C, Ederra M, Salas D, Castells X, and the Cumulative False Positive Risk group.

Eur Radiol 2011; 21(10): 2083-2090.

5. Justificación de la unidad temática

El cribado de cáncer de mama es una de las intervenciones poblacionales más evaluadas. A pesar del debate actual sobre el balance entre beneficios y efectos adversos del cribado mamográfico las autoridades sanitarias a nivel estatal y europeo continúan recomendando la realización del cribado poblacional de cáncer de mama. El cuestionamiento actual sobre la efectividad del cribado pone de manifiesto la necesidad de avanzar en su evaluación, con la aportación de nuevos puntos de vista y nuevos datos que presenten los elementos necesarios para evaluar posibles alternativas en las actuales estrategias de detección precoz y mejorar así la eficiencia del cribado. Esta tesis se enmarca en la actual controversia del cribado, haciendo especial hincapié en la valoración de los efectos adversos, y de manera especial en la evaluación de los resultados falsos-positivos y su impacto.

La posibilidad de ampliar la información conocida hasta la fecha sobre los efectos adversos del cribado, y en particular sobre los resultados falsos positivos, mediante el análisis de grandes bases de datos poblacionales abre una ventana a una evaluación más precisa y fiable del cribado mamográfico. En particular, los trabajos de esta tesis están basados en un mismo proyecto, realizado con una base de datos común, diseñada expresamente para el estudio de los falsos positivos, y con una metodología específica que permite analizar las estructuras complejas de información que la evaluación de los falsos positivos requiere.

La estructura longitudinal y de carácter completo de la base de datos diseñada para los proyectos de esta tesis ha permitido dar respuestas a cuestiones de investigación que hasta ahora eran irrealizables. Además, debido a su potencial, se ha podido ampliar la utilidad de los datos dando respuesta a otras cuestiones. Como parte de esta tesis, se ha desarrollado un estudio específico sobre la evolución temporal en las tasas de detección de carcinoma ductal in situ y cáncer invasivo a lo largo de 16 años, desde la puesta en marcha de los programas en el territorio español (1991-2006). Así mismo, la metodología utilizada en esta tesis para la evaluación de los resultados falsos positivos es de especial interés, y ha podido ser validada externamente con su

aplicación a los datos del programa de cribado nacional de Noruega, lo que ha supuesto una excelente oportunidad para su aplicación y validación en otros contextos.

Como complemento a los trabajos de esta tesis y como parte del proyecto sobre Riesgo Acumulado de Falsos Positivos (RAFP) en el que se sustenta esta tesis, se han desarrollado otros dos trabajos, presentados en los anexos, y en los que he colaborado estrechamente con la aportación de datos y el análisis estadístico. Estos trabajos evalúan el impacto de la edad de inicio del cribado y la experiencia de los radiólogos lectores sobre el riesgo de falsos positivos (ver anexos). Además, y en relación a la variabilidad existente en la tasa de resultados falsos positivos y las tasas de detección dentro del contexto de un mismo programa de cribado, se añade a los anexos un trabajo sobre variabilidad desarrollado recientemente en el programa de cribado de cáncer de mama de Noruega.

II. HIPÓTESIS Y OBJETIVOS

1. Hipótesis

Hipótesis general

El cribado de cáncer de mama es una actividad de alcance poblacional. Por su diseño y organización, así como por las características de la prueba de cribado, se espera un cierto volumen de efectos adversos, además de los beneficios. El efecto adverso más frecuente del cribado mamográfico son los resultados falsos positivos. Éstos pueden ser cuantificados, y es posible valorar los factores asociados con los mismos y su impacto.

Hipótesis específicas

Sobre los resultados falsos positivos y sus factores asociados

1. El riesgo de tener un resultado falso positivo en el cribado mamográfico es diferente en función de las participaciones de la mujer en el cribado, siendo mayor en cribados prevalentes.
2. El riesgo de tener un resultado falso positivo está asociado con las características del protocolo de lectura mamográfica (método de lectura, número de proyecciones, uso de mamografía digital).
3. El riesgo de tener un resultado falso positivo está asociado con las características de las mujeres participantes (edad, uso de terapia hormonal sustitutiva, menopausia, antecedentes familiares de cáncer de mama, y antecedentes personales de pruebas invasivas).

Sobre el impacto de los resultados falsos positivos en el cribado de cáncer de mama

4. La presencia de un resultado falso positivo en el cribado mamográfico puede condicionar la adherencia en sucesivas convocatorias de cribado.
5. La presencia de un resultado falso positivo en el cribado mamográfico está asociado con una mayor probabilidad de detectar un cáncer en rondas de cribado sucesivas.
    5.1. Los resultados falsos positivos con pruebas invasivas presentan un riesgo aun mayor de detectar un cáncer en rondas sucesivas que los falsos positivos con pruebas de imagen únicamente.

Sobre las tendencias en las tasas de detección de cáncer de mama en el cribado poblacional, y el uso de terapia hormonal sustitutiva.

6. La tendencia en las tasas de detección de carcinoma ductal in situ y carcinoma invasivo en el cribado poblacional no han sufrido ningún cambio de tendencia desde la puesta en marcha de los programas hasta el final del periodo de estudio.

7. La disminución en el uso de terapia hormonal sustitutiva puede tener un impacto sobre las tasas de detección de carcinoma ductal in situ y carcinoma invasivo en el cribado poblacional.

2. Objetivos

El **objetivo general** de la tesis es profundizar en la evaluación del cribado poblacional del cáncer de mama. Específicamente en los siguientes aspectos:

- Cuantificar el riesgo de experimentar un resultado falso positivo y evaluar sus factores asociados
- Evaluar el impacto de los resultados falsos positivos en el cribado de cáncer de mama
- Evaluar las tendencias en las tastas de detección de cáncer de mama en el cribado poblacional, y el posible impacto de la terapia hormonal sustitutiva

Objetivos específicos

Sobre los resultados falsos positivos y sus factores asociados

1. Estimar el riesgo acumulado de tener un resultado falso positivo a lo largo de la participación secuencial de la mujer en el cribado bienal desde los 50 a los 69 años.
2. Estimar el riesgo acumulado de tener un resultado falso positivo con pruebas invasivas a lo largo de la participación secuencial de la mujer en el cribado bienal desde los 50 a los 69 años.
3. Estimar el impacto de las características del protocolo de lectura mamográfica en el riesgo de tener un resultado falso positivo (método de lectura, número de proyecciones, uso de mamografía digital).
4. Estimar el impacto de las características de las mujeres participantes en el riesgo de tener un resultado falso positivo (edad, uso de terapia hormonal sustitutiva, menopausia, antecedentes familiares de cáncer de mama, y antecedentes personales de pruebas invasivas).

Sobre el impacto de los resultados falsos positivos en el cribado de cáncer de mama

5. Evaluar la adherencia al cribado en convocatorias sucesivas en mujeres con y sin resultados falsos positivos previos.
6. Evaluar el riesgo de detectar un cáncer de mama en el cribado mamográfico en mujeres con y sin resultados falsos positivos previos.

    6.1. Evaluar el riesgo de presentar un cáncer de mama en el cribado mamográfico en mujeres con falsos positivos previos con pruebas invasivas.

Sobre las tendencias en las tasas de detección de cáncer de mama en el cribado poblacional, y el uso de terapia hormonal sustitutiva.

7. Evaluar la tendencia temporal en las tasas de detección de carcinoma ductal in situ y carcinoma invasivo en el cribado poblacional desde la puesta en marcha de los programas hasta el final del periodo de estudio.
8. Evaluar el efecto del uso de terapia hormonal sustitutiva sobre las tasas de detección de Carcinoma ductal in situ y carcinoma invasivo en el cribado poblacional.

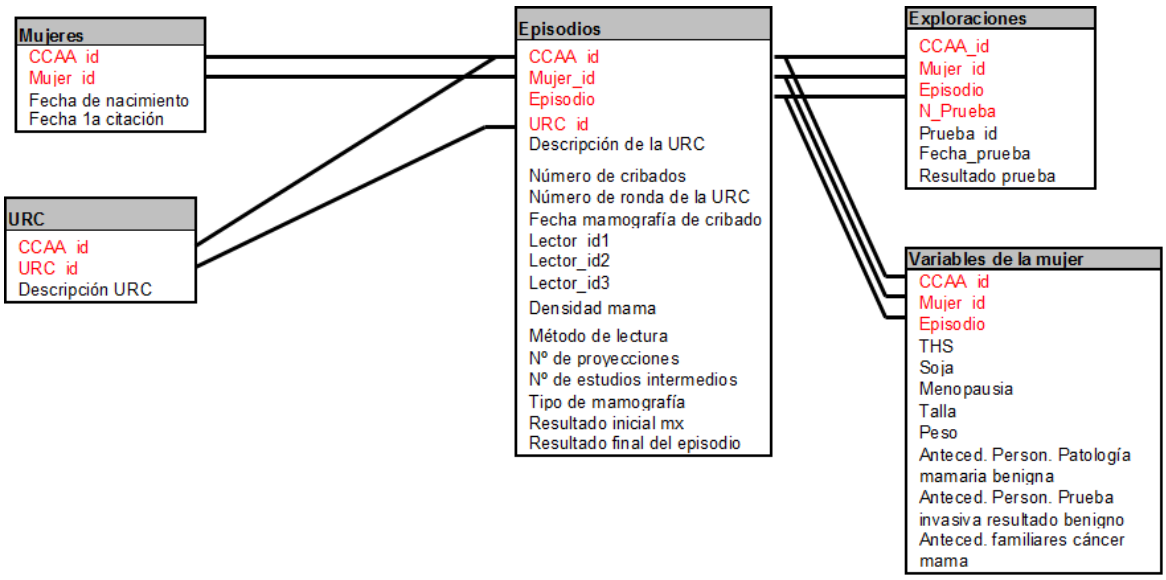
III. MÉTODOS Y RESULTADOS

1. Aproximación metodológica

1.1 Creación de la base de datos

Para la construcción de la base de datos utilizada en los estudios de esta tesis, se solicitó a cada uno de los programas participantes que aportasen la información completa de todas las mujeres participantes en el cribado al menos una vez desde la puesta en marcha de los mismos hasta diciembre de 2006. Los programas de cribado participantes tienen estructuras administrativas independientes y específicas de cada programa para recoger la información, por lo que se elaboró un detallado protocolo de definición de variables y validación de la información de los mismos (ver anexos). El protocolo de variables fue desarrollado y consensuado con la ayuda de los responsables de todos los programas participantes. De esta manera se consiguió evitar la ambigüedad en la interpretación de las definiciones, y se homogeneizó la codificación de las variables de interés. La información se recogió en una estructura de tablas multidimensional con diferentes niveles de información (figura III.1). Una vez recogidos los datos de cada programa, se validaron de manera individual, y en caso necesario se contactó con los programas para evitar posibles errores. Posteriormente se fusionaron las bases de datos de los distintos programas en una base de datos común, que es la utilizada para los distintos análisis.

Figura III.1: Estructura de información de la base de datos y variables de estudio



1.2 Población de estudio

La población de estudio estaba comprendida por las mujeres participantes en los programas de cribado de 8 Comunidades Autónomas: Navarra, Valencia, Galicia, Canarias, Castilla y León, La Rioja, Asturias y Cataluña. Los programas participantes tienen información estructurada de al menos tres rondas de cribado consecutivas, de manera que la información puede ser analizada como una cohorte (cada mujer debe ser identificada unívocamente mediante un único código en cada participación en el programa). Además, para cada participación de la mujer se dispone de información sobre el tipo de exploraciones adicionales realizadas y del resultado histopatológico, con el fin de verificar si el positivo es verdadero o falso. Se incluyen todas las mujeres participantes al menos una vez en cualquiera de los programas. Se excluye a las mujeres con historia de cáncer de mama o implante mamario anterior al primer cribado.

El periodo de estudio del proyecto abarcaba desde 1991, con la puesta en marcha del primer programa de cribado del territorio español en Navarra, hasta diciembre de 2006. El carácter retrospectivo del estudio implica la recogida de información de un gran volumen de mamografías de cribado. El estudio cubría el 44% de la población diana española y contenía información de 4,739,498 mamografías de cribado de 1,565,364 mujeres cribadas. Además, se incluía información de 378,060 exploraciones adicionales para confirmar o descartar malignidad, y 16,529 cánceres de mama detectados en el cribado. El 74% de las mujeres participantes tenían al menos dos mamografías de cribado (n=1,205,943), el 55% al menos 3 (n=867,160), y el 10% 6 o más (n=156,414). Las mamografías de cribado se llevaron a cabo en 74 unidades radiológicas diferentes con un promedio de 64,047 mamografías y 21,154 mujeres cribadas por unidad radiológica.

Tabla III.1: Descriptiva de la población de estudio

Mamografías	4 739 498
Mujeres cribadas	1 565 364
Exploraciones adicionales	378 060
Falsos positivos	264 801
Falsos positivos (invasivas)	24 436
Tumores detectados	16 529
Unidades radiológicas	74

Para el artículo del contexto noruego (artículo 5), se dispuso de información de todas las mamografías realizadas en el programa de cribado de cáncer de mama de Noruega, desde la puesta en marcha del mismo en 1996, hasta diciembre de 2010. Para este estudió se analizaron 715,311 mamografías de cribado de 231,310 mujeres cribadas por primera vez con 50-51 años.

1.3 Variabilidad entre las unidades radiológicas

Cada uno de los programas participantes tiene una o más unidades radiológicas de cribado en las que se realizan las mamografías de cribado, hasta un total de 74 unidades radiológicas, y máximo de 12 unidades en un mismo programa. Cada unidad radiológica se considera una fuente de variabilidad en la que sus observaciones están correlacionadas, ya que cada una de estas unidades utiliza su propio mamógrafo y por lo general tienen un equipo de radiólogos lectores propio.

1.4 Análisis estadístico

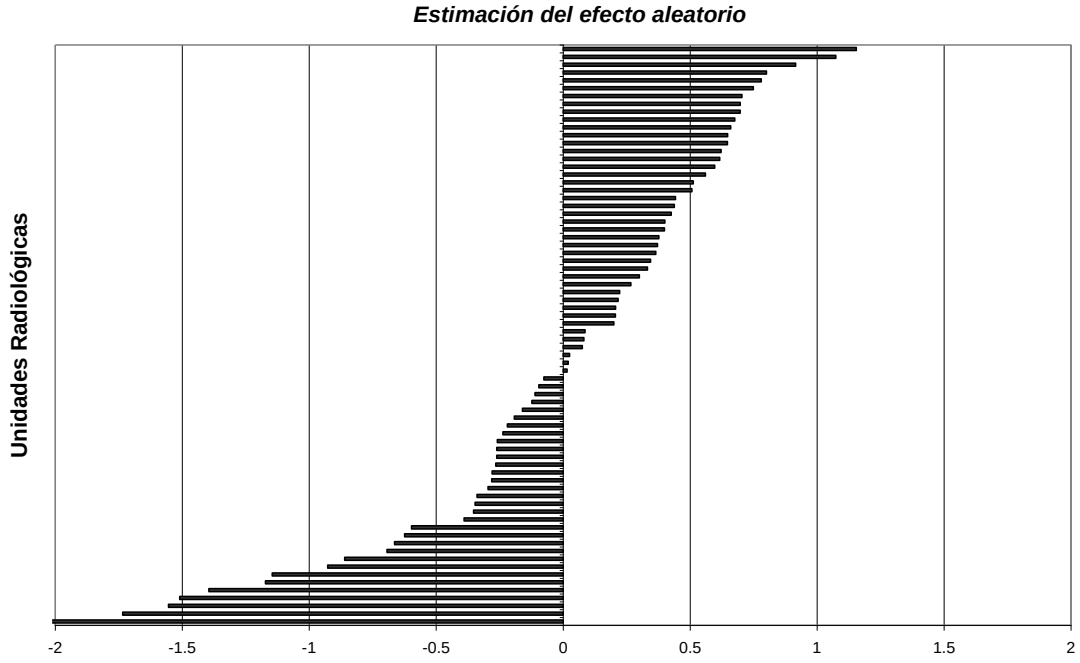
El análisis estadístico de la base de datos se hizo mediante Modelos de riesgo mixtos a tiempo discreto (Discrete Time Mixed Hazard Models). Se eligió este tipo de modelos porqué permitían incorporar las necesidades desarrolladas por el equipo investigador. El modelo debía tener en cuenta que; a) la variable respuesta (falso-positivo) es una variable dicotómica; b) el tiempo es discreto, y equivale a cada una de las participaciones en el cribado de la mujer; c) debe tener en cuenta la correlación entre las mamografías de una misma unidad radiológica (estructura jerárquica); d) debe tener en cuenta las repetidas observaciones de una misma mujer a lo largo del tiempo. e) debe proporcionar un estimador del riesgo de falso positivo en cada posible participación de la mujer; y f) Debe tener en cuenta las variables cambiantes en el tiempo (edad, método de lectura, etc.).

Estos modelos fueron desarrollados en profundidad por Judith D. Singer y John B. Willet en su libro “Applied Longitudinal Data Analysis” del año 2003. En ellos se modeliza el ‘hazard’ de presentar un determinado evento a partir de modelos de supervivencia para intervalos de tiempo discretos, en los que los individuos únicamente están a riesgo cuando son observados. Como

extensión a los modelos propuestos por Singer y Willet, en los modelos utilizados en el desarrollo de este trabajo se incorpora una componente aleatoria que tiene en cuenta la estructura jerárquica del muestreo de los datos, en el que mamografías realizadas en una misma unidad radiológica están correlacionadas. Estos modelos permiten una aproximación más precisa a la problemática comúnmente planteada para la estimación del riesgo de falos positivos.

Se muestra en la figura III.3 el estimador del efecto aleatorio para cada una de sus componentes (unidad radiológica). El conjunto de estos estimadores tiene media $\mu=0$ y una desviación estándar σ_u para cada una de las unidades. Se construyó el gráfico a partir de la estimación del efecto aleatorio en un modelo basal sin covariables.

Figura III.2: Estimadores del efecto aleatorio para cada una de sus componentes en el modelo basal sobre el riesgo de resultados falsos positivos



Por su parte, la categorización de variables utilizada en las variables explicativas se hizo en función de los criterios epidemiológicos propuestos por el grupo de investigadores, de manera que los resultados fuesen comparables con los de otros estudios.

2. Artículo 1

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

Autores: Román M, Sala M, Salas D, Ascunce N, Zubizarreta R, Castells X, Cumulative False Positive Risk Group.

Revista: Ann Oncol 2012; 23(1): 104-111.

Factor de impacto: 6.5 (Q1 Oncology)

DOI: 10.1093/annonc/mdr032

Abstract:

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, 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.

Annals of Oncology Advance Access published March 23, 2011

original article

Annals of Oncology
doi:10.1093/annonc/mdr032

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, 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

*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

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

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.

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 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).

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.

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

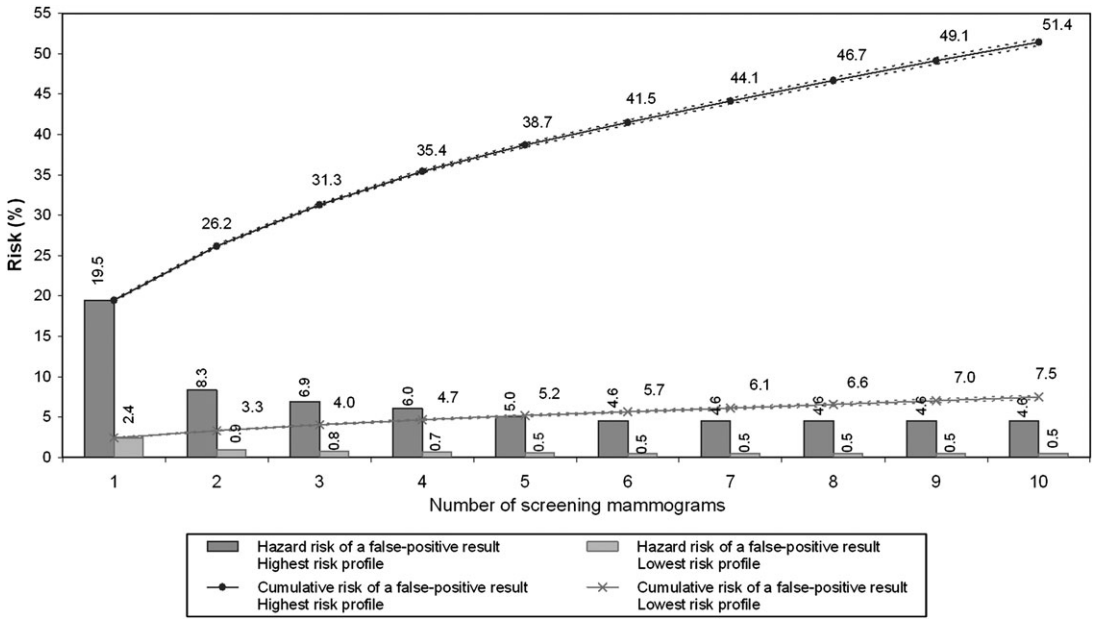


Figure 1. Cumulative risk and hazard risk of a false-positive result for any procedure for women starting screening at age 50–51 years. Highest risk (double reading, one view, film-screen mammography, premenopausal status, previous invasive procedures, and familial breast cancer) versus lowest risk profiles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

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 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].

Our results on the influence of women's characteristics are in agreement with those of previous studies. The risk of a false-

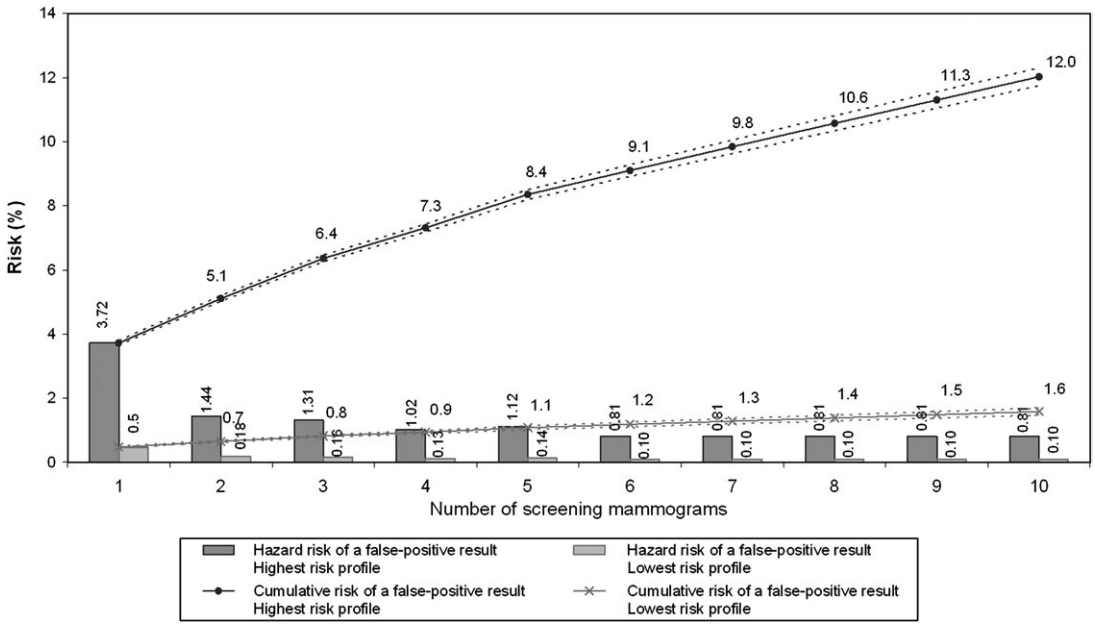


Figure 2. Cumulative risk and hazard risk of a false-positive result for invasive procedures for women starting screening at age 50–51 years. Highest risk (double reading, two views, not using HRT, premenopausal status, previous invasive procedures, and familial breast cancer) versus lowest risk profiles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

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.

3. Artículo 2

Título: Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening

Autores: Román M, Sala M, De La Vega M, Natal C, Galceran J, González-Román I, Baroja A, Zubizarreta R, Ascunce N, Salas D, Castells X, Cumulative False Positive Risk Group.

Revista: Breast Cancer Res Tr 2011; 130(2): 543-552.

Factor de impacto: 4.8 (Q1 Oncology)

DOI: 10.1007/s10549-011-1581-4

Abstract:

Background: False-positive results may influence adherence to mammography screening. The effectiveness of breast cancer screening is closely related to adequate adherence among the target population. The objective of this study was to evaluate how false-positives and women's characteristics affect the likelihood of reattendance at routine breast cancer screening in a sequence of routine screening invitations.

Methods: We performed a retrospective cohort study of 1,371,218 women aged 45-69 years, eligible for the next routine screening, who underwent 4,545,346 screening mammograms from 1990 to 2006. We estimated the likelihood of attendance at seven sequential screening mammograms. Multilevel discrete time hazard models were used to estimate the effect of false-positive results on reattendance, and the odds ratios (OR) of non-attendance for the women's personal characteristics studied.

Results: The overall reattendance rate at the second screening was 81.7% while at the seventh screening was 95.6%. At the second screening invitation reattendance among women with and without a false-positive mammogram was 79.3 vs. 85.3%, respectively. At the fourth and seventh screenings, these percentages were 86.3 vs. 89.9% and 94.6 vs. 96.0%, respectively. The study variables associated with a higher risk of failing to participate in subsequent screenings were oldest age (OR = 8.48; 95% CI: 8.31-8.65), not attending their first screening invitation (OR = 1.12; 95% CI: 1.11-1.14), and previous invasive procedures (OR = 1.09; 95% CI: 1.07-1.10). The risk of

non-attendance was lower in women with a familial history of breast cancer (OR = 0.97; 95% CI: 0.96-0.99), and those using hormone replacement therapy (OR = 0.96; 95% CI: 0.94-0.97).

Conclusion: Reattendance was lower in women with false-positive mammograms than in those with negative results, although this difference decreased with the number of completed screening participations, suggesting that abnormal results in earlier screenings more strongly influence behavior. These findings may be useful in providing women with accurate information and in improving the effectiveness of screening programs.

Breast Cancer Res Treat (2011) 130:543–552
DOI 10.1007/s10549-011-1581-4

EPIDEMIOLOGY

Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening

R. Román · M. Sala · M. De La Vega · C. Natal · J. Galceran · I. González-Román · A. Baroja · R. Zubizarreta · N. Ascunce · D. Salas · X. Castells

Received: 6 May 2011 / Accepted: 9 May 2011 / Published online: 27 May 2011

Abstract False-positive results may influence adherence to mammography screening. The effectiveness of breast cancer screening is closely related to adequate adherence among the target population. The objective of this study was to evaluate how false-positives and women's characteristics affect the likelihood of reattendance at routine breast cancer screening in a sequence of routine screening invitations. We performed a retrospective cohort study of 1,371,218 women aged 45–69 years, eligible for the next routine screening, who underwent 4,545,346 screening mammograms from 1990 to 2006. We estimated the likelihood of attendance at seven sequential screening mammograms. Multilevel discrete time hazard models were used to estimate the effect of false-positive results on reattendance, and the odds ratios (OR) of non-attendance for the women's personal characteristics studied. The overall reattendance rate at the second screening was 81.7% while at the seventh screening was 95.6%. At the second screening invitation reattendance among women with and without a false-positive mammogram was 79.3 vs. 85.3%, respectively. At the fourth and seventh screenings, these percentages were 86.3 vs. 89.9% and 94.6 vs. 96.0%, respectively. The study variables

The Members of the "The Cumulative False Positive Risk Group" are given in Appendix.

Electronic supplementary material The online version of this article (doi:10.1007/s10549-011-1581-4) contains supplementary material, which is available to authorized users.

R. Román · M. Sala · X. Castells (✉)
Department of Epidemiology and Evaluation, Institut Municipal d'Investigació Mèdica-Parc de Salut Mar., Mar Teaching Hospital, 25-29 Passeig Marítim, 08003 Barcelona, Spain
e-mail: xcastells@parcdesalutmar.cat

R. Román · M. Sala · N. Ascunce · X. Castells
Consortium for Biomedical Research in Epidemiology & Public Health (CIBER en Epidemiología y Salud Pública – CIBERESP), Barcelona, Spain

M. De La Vega
Dirección General de Programas Asistenciales, Consejería de Sanidad, Servicio Canario de Salud, Canary Islands, Spain

C. Natal
Program & Analysis Unit, Health Office, Asturias, Spain

J. Galceran
Foundation Society for Cancer Research and Prevention, Pere Virgili Health Research Institute, Reus, Tarragona, Spain

I. González-Román
Castilla-Leon Breast Cancer Screening Programme, Dirección General de Salud Pública ID e I, SACYL, Castilla y León, Spain

A. Baroja
La Rioja Breast Cancer Screening Programme, Fundacion Rioja Salud, Logroño, Spain

R. Zubizarreta
Galician Breast Cancer Screening Programme, Public Health & Planning Directorate, Health Office, Santiago de Compostela, Spain

N. Ascunce
Navarra Breast Cancer Screening Programme, Public Health Institute, Pamplona, Spain

D. Salas
General Directorate Public Health & Centre for Public Health Research, (CSISP), Valencia, Spain

X. Castells
Universitat Autònoma de Barcelona, Barcelona, Spain

associated with a higher risk of failing to participate in subsequent screenings were oldest age (OR = 8.48; 95% CI: 8.31–8.65), not attending their first screening invitation (OR = 1.12; 95% CI: 1.11–1.14), and previous invasive procedures (OR = 1.09; 95% CI: 1.07–1.10). The risk of non-attendance was lower in women with a familial history of breast cancer (OR = 0.97; 95% CI: 0.96–0.99), and those using hormone replacement therapy (OR = 0.96; 95% CI: 0.94–0.97). In conclusion, reattendance was lower in women with false-positive mammograms than in those with negative results, although this difference decreased with the number of completed screening participations, suggesting that abnormal results in earlier screenings more strongly influence behavior. These findings may be useful in providing women with accurate information and in improving the effectiveness of screening programs.

Keywords Breast cancer · False-positive · Screening · Reattendance

Introduction

False-positive results are a major concern of breast cancer screening. The negative effects of these results have been widely noted and include anxiety [1], additional physician visits, diagnostic tests, and excision biopsies [2, 3]; attitudes toward subsequent mammographic screening are also affected [4, 5].

The benefits of screening are measured by the reduction in mortality after participation in several screenings. False-positive results affecting participation in subsequent screenings may reduce the overall benefit of screening [6]. Study of reattendance at subsequent screenings after a false-positive result has provided contradictory results. While some studies carried out in Europe and Canada have associated false-positive results with lower reattendance [7–11], other European studies have found no association [12, 13], or even greater reattendance after a false-positive result has been shown in the USA [14–16]. These differences between the USA and Europe could be explained by differences in the screening setting (opportunistic or population-based with quality standards), a higher tolerance to missed breast cancer, and a higher reading volume among radiologists interpreting mammograms [5, 17]. In addition to false-positive results, other factors related to women's personal characteristics, such as age, use of hormone replacement therapy (HRT), age at menarche, a family history of breast cancer, and socio-economic status [16, 18], may also affect adherence to breast cancer screening.

As a result of the debate about the risk–benefit balance of breast cancer screening, several agencies have recommended that women be provided with reliable information on the adverse effects of screening programs [17, 19]. Non participating women do not benefit from the reduction in breast cancer mortality obtained from repeated screening. Besides, they might be more likely to fail to participate in further screening and, moreover, may be less likely to accept future screening invitations. Most studies on the impact of false-positives on reattendance evaluate the effect of these results on the next screening invitation, but lack information on the effects over a sequence of screening invitations, as a cohort. Widening knowledge of the factors affecting attendance at subsequent screenings, especially the impact of false-positives and their long-term effects, could improve the information given to women and improve the effectiveness of the screening programs.

We investigated adherence to population-based breast cancer screening in a cohort of screened women over a sequence of screening invitations, and examined how false-positives and women's personal variables modify the likelihood of reattendance.

Methods

Setting

The information analyzed was collected for a previous study conducted to evaluate the cumulative risk of a false-positive result in breast cancer screening, and is explained in detail elsewhere [20].

Briefly, all women residing in Spain aged 50–69 years are actively invited to participate in a population-based screening program, with screening intervals every 2 years. In Spain, breast cancer screening follows the European Guidelines for Quality Assurance in Mammographic Screening [17] and its results meet the required standards [19, 21]. Population-based breast cancer screening in Spain started in one region in 1990 and was implemented nationwide in 2006. Data from eight regions, covering 44% of the target population in 2005, were collected. Each region has one or several radiology units that perform screening. All regions participating in the study had completed at least three screening rounds by December 2006. Local application of the screening program can vary, the start age for screening being either 50–51 or 45–46 years.

The study included variables related to the presence of false-positive results in the screening process and to women's personal characteristics. At the first screening, information on previous invitations to the screening program was available. All information was gathered from each participant at each attendance. 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 post-menopausal), previous invasive procedures

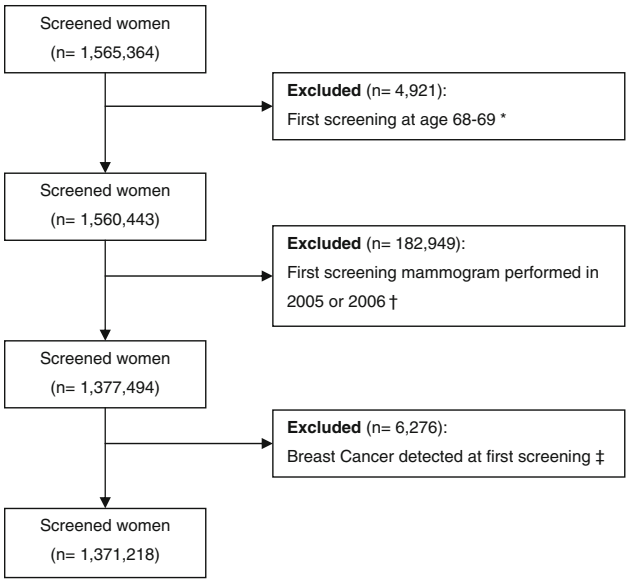


Fig. 1 Eligibility of women for study of reattendance at the following screening round, based on the initial study population. *Target population between 45 and 69 years. Women having a screening mammogram at 68–69 years are not invited again. †Women whose first screening mammograms were performed in 2005 or 2006 were not eligible for the study of attendance to the following screening invitation as the next invitation would be outside the study period (1990–December 2006). ‡Women in whom breast cancer is detected are not invited for screening

with a benign result, and the presence or absence of a first-degree familial history of breast cancer. Information on women's personal characteristics was not routinely gathered and was incomplete in some radiology units.

Study population

Information was available for 1,565,364 women who had undergone at least one screening mammogram in any of 74 radiology units of the eight participating regions between March 1990 and December 2006. Of these women, 6,276 (0.4%) were diagnosed with breast cancer in their first screening participation, and 4,921 (0.3%) were aged 68 or 69 years at their first screening mammogram and were thus ineligible for rescreening. We excluded 182,949 (11.7%) women who were first screened in 2005 or 2006 and could not be observed at 24 months in the following invitation. Finally, 1,371,218 women eligible for at least one screening invitation after the first participation were analyzed (Fig. 1). These women underwent a total of 4,545,346 mammographic screening tests.

Definition of reattendance

Attendance at breast cancer screening was defined as participation in a breast cancer screening program through a mammogram test following a routine screening invitation. Reattendance was measured as participation, in the next routine screening invitation (at 24 months) following a screening mammogram test. Eligibility for reattendance involved at least one screening mammogram previously performed (first screening).

Definition of a false-positive result

A result was considered a false-positive if, after recall for additional evaluation breast cancer was not diagnosed. Additional evaluation included both non-invasive (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 additional evaluation was carried out within a maximum of 2 months after screening. A definitive diagnosis of breast cancer was always histopathologically confirmed (invasive ductal carcinoma or carcinoma in situ). Women with a negative result (at mammographic reading or after additional evaluation) were recalled for a new screening mammography 24 months after the previous screen.

In addition to the definition of a "false-positive result" (involving a non-invasive and/or invasive additional evaluation), we defined a "false-positive for invasive procedures" as a false-positive involving at least one invasive procedure performed during the additional evaluation.

Statistical analysis

To calculate the risk of non-attendance at the following screening invitation, discrete time hazard models were used, as described in detail by Singer et al. [22, 23]. This methodology uses a logistic regression approach to compute these particular survival models with discrete time intervals. Two types of predictors are introduced in the model: "time indicators", given by the women's screening participation (acting as multiple intercepts), and "substantive predictors" for the effect of the study variables on the model. The event of interest was non-attendance at next routine screening invitation after a screening mammogram test. Analogously to standard survival analysis methods, women diagnosed with breast cancer, those who underwent a screening test in 2005 or 2006 or women aged 68 or 69 years were censored at their last screening participation. In addition, as repeated observations of non-attendance in the same woman would be correlated and could bias the estimates obtained, 255,243 (5.6%) rescreening tests performed after at least one missed screening invitation were censored to compute the regression model estimates.

Time changing variables were computed in the regression models as data were collected at each attendance. Besides,

interaction terms between the "time indicators" and specific covariates were included. Simple and multivariate discrete time hazard models were used to estimate the individual and simultaneous effect of all predictors. The multivariate models included the women's personal variables (age, HRT use, menopausal status, previous invasive procedures, familial history of breast cancer), whether or not the woman attended the first invitation to the screening program, the presence of false-positives and their interaction with the time indicators (women's screening participation). In addition, the multivariate models included a period effect (calendar years) as the start date of the radiology units differed, and both, simple and multivariate models, included a random effect component defined by the radiology units. The radiology unit was considered a random effect in the models because of the correlation among screening tests performed in the same radiology unit. The model including all the study variables was performed with the subset of screening mammograms for which information on women's personal variables was complete. This subset accounted for 2,660,155 (58.5%) screening mammograms from 45 radiology units. The GLIMMIX procedure in SAS 9.1 (SAS Institute, Cary, NC) was used. In all the models, residual pseudo-likelihood estimation was used.

Results

We analyzed information from 1,371,218 women who underwent 4,545,346 screening mammograms and were eligible for rescreening (mean (standard deviation) screening participations per woman: 3.31 (1.67)). Among the screened women 867,160 (63.2%) had three or more screening mammograms performed, while 67,609 (4.9%) had 7 or more screening mammograms. The flowchart of women eligible for the analysis is shown in Fig. 1.

Table 1 compares several characteristics of the sample analyzed with respect to reattendance at subsequent screening invitations. Non-attendees were older, had higher false-positive rates, had more previous invasive procedures and more frequently failed to attend their first screening invitation. In addition, lower percentages of premenopausal women, HRT use, and women with a familial history of breast cancer were observed among non-attendees.

Figure 2 shows the probability of reattendance up to the 7th screening participation. The probability increased with the number of completed screening participations. At the 1st screening the likelihood of returning for the following screening was 81.7% (95% CI: 81.63–81.76), that is, of the 1,371,218 women (N_1) who had a first screening mammogram and were eligible for a second screening 1,120,251 (N_2) attended the next screening invitation (81.7%). At the 3rd screening, the probability of returning for the 4th screening was 88.1% (95% CI: 87.98–88.12), whereas at 6th screening the probability of reattendance to the 7th screening rose to 95.6% (95% CI: 95.52–95.73).

Figure 3 shows the probability of reattendance at screening according to the number of missed screening invitations. The likelihood of reattendance at screening decreased as the number of missed screening invitations

Table 1 Characteristics of the screening tests performed with respect to reattendance at subsequent screenings ($N = 4{,}545{,}346$)

Characteristics	Number (%)		*P* value
	Re-attendance ($N = 3{,}940{,}283$)	Non-attenders ($N = 605{,}063$)	
Age at screening (years)			<0.001
44–49	601,898 (15.3)	83,642 (13.8)	
50–54	1,045,268 (26.5)	117,048 (19.3)	
55–59	1,077,387 (27.3)	103,966 (17.2)	
60–64	891,278 (22.6)	187,766 (31.0)	
65–69	324,452 (8.2)	112,641 (18.6)	
False-positive (any procedure)	205,533 (5.2)	44,325 (7.3)	<0.001
False-positive (invasive procedures)	16,488 (0.4)	6,501 (1.1)	<0.001
Failed to attend first invitation	403,662 (10.2)	79,918 (13.2)	<0.001
HRT[a]	271,202 (9.8)	26,422 (6.9)	<0.001
Premenopausal[a]	528,054 (20.9)	47,911 (15.3)	<0.001
Previous invasive procedure[a]	171,029 (6.9)	22,173 (7.2)	<0.001
Familial breast cancer[a]	188,387 (6.9)	24,275 (6.4)	<0.001

HRT use of hormone replacement therapy, *Premenopausal* pre/peri-menopausal status, *Previous invasive procedure* personal previous invasive procedure, *Familial breast cancer* first-degree familial history of breast cancer

[a] Due to missing data, numbers may vary for women related variables

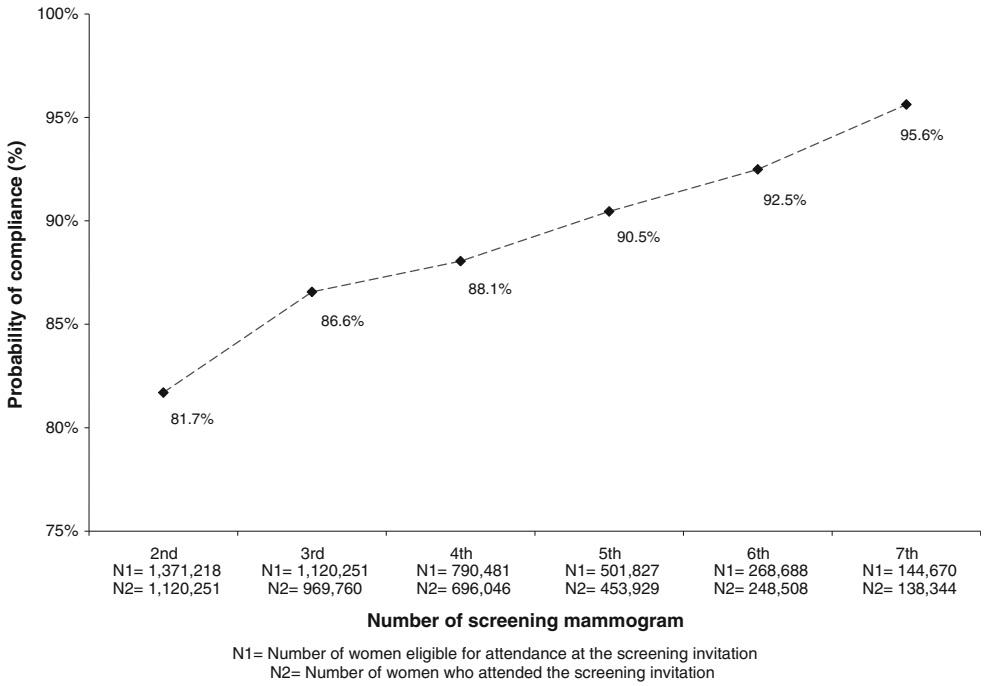


Fig. 2 Percentage of women attending screening from the 2nd to the 7th screening participations

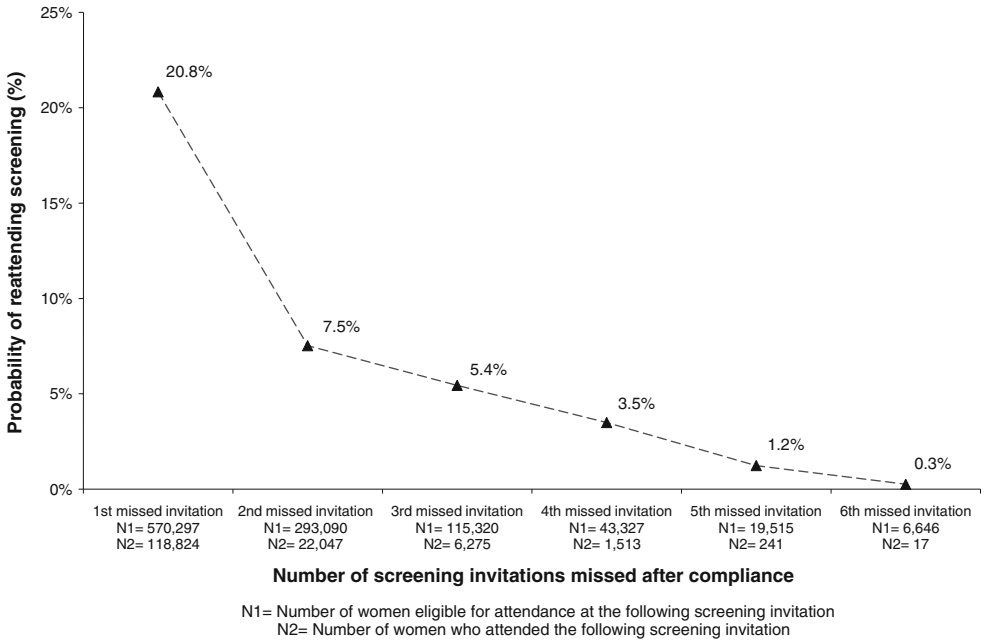


Fig. 3 Percentage of women reattending the following screening round after 1–6 missed screening invitations

increased. The probability of reattendance after the 1st missed invitation was 20.8% (95% CI: 20.73–20.94), that is, of the 570,297 women who failed to participate in a screening invitation, only 118,824 (20.8%) attended the next program's screening invitation. The probability of reattendance after three consecutive missed invitations was 5.4% (95% CI: 5.31–5.57); and was 0.3% (95% CI: 0.13–0.38) after six missed invitations.

Figure 4 shows the effect of false-positive mammograms on the probability of attendance up to the 7th screening participation, obtained from the regression model including the interaction between false-positives and the number of screening participation. Women not experiencing a false-positive result were more likely to return for the following screening invitation. The difference in the probability of reattendance among women with and without a false-positive decreased with the number of screening participations. At the first screening, the probability of attending the 2nd screening invitation was 79.3% (99% CI: 79.0–79.6) and 85.3% (99% CI: 85.2–85.4) for women with and without a false-positive result, respectively. At the 3rd screening participation, women experiencing a

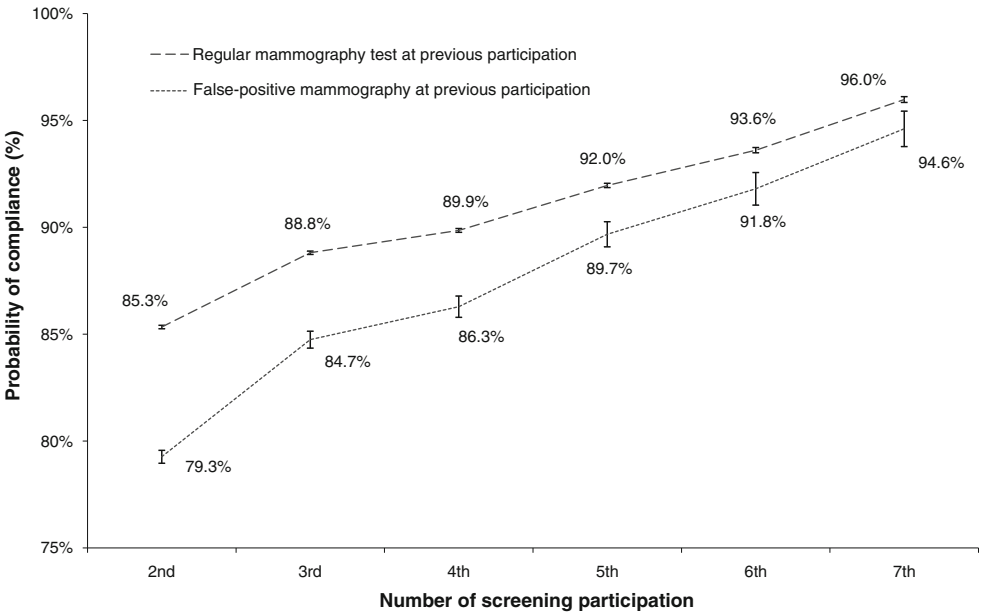


Fig. 4 Estimated probability of attendance at screening in the regression model, according to the presence of a false-positive result in the previous mammogram from the 2nd to the 7th screening participations

false-positive had a probability of attending to the 4th screening invitation of 86.3% (99% CI: 85.8–86.8), whereas women with a negative mammographic reading had a 90.0% (99% CI: 90.0–90.1) probability. At the 6th screening participation, the likelihood of attendance at the following screening was 94.6% (99% CI: 93.8–95.4) and 96.0% (99% CI: 95.8–96.1), for women with and without a false-positive result.

Adjusted odds ratios (OR) of the association between the women related variables and the risk of failing to participate in the following screening invitation obtained from the adjusted regression model are shown in Table 2. The risk of non-attendance was higher in the oldest women (OR = 8.48; 95% CI: 8.31–8.65), women not attending their first screening invitation (OR = 1.12; 95% CI: 1.11–1.14), and those with previous invasive procedures (OR = 1.09; 95% CI: 1.07–1.10), and was lower in women with a familial history of breast cancer (OR = 0.97; 95% CI: 0.96–0.99), and HRT users (OR = 0.96; 95% CI: 0.94–0.97). Menopausal status was not statistically significant.

False-positive results due to an invasive procedure had a stronger impact than false-positives for any procedure in the likelihood of reattending screening (data not shown). However, the decrease in the probability of reattendance remained proportional over the various screening invitations. At the first screening, the probability of attendance at the 2nd screening invitation was 72.6% (99% CI: 71.7–73.6) and 85.0% (99% CI: 84.9–85.1) for women with and without a false-positive for invasive procedures, respectively. The probability of attending to the 4th screening invitation was 78.2% (99% CI: 75.9–80.6) and 89.8% (99% CI: 89.7–89.9), respectively. At the 6th screening participation the probability of attendance to the 7th screening invitation was 86.9% (99% CI: 81.2–92.7) and 96.0% (99% CI: 95.8–96.1).

Discussion

The effectiveness of breast cancer screening is closely related to adequate adherence among the target population [6]. The aim of evaluating long-term reattendance in a cohort of screened women and the causes related to lower adherence is to improve the results of breast cancer screening, and provide the fullest information available to women invited to participate. Our results show that reattendance was lower in women with a false-positive result, and also that this effect was higher when the false-positive occurred in the firsts mammograms, showing that earlier experiences of abnormal mammograms had a greater impact on future behavior.

The attendance rate at the second screening invitation after the first screening participation was 81.7%. This result is consistent with findings in the UK [10], Norway [8] and Switzerland [18], where breast cancer screening is population-based and screening program organization is similar. We estimated that the re-attendance rate increased to 88.1% from the 3rd to the 4th screening participation and to 95.6% at the 7th. These results suggest that women's adherence to the screening program increases with the number of participations.

In agreement with the results of previous studies [7, 8, 10, 11], we found that a false-positive result was associated with lower screening reattendance. As suggested, women requiring further assessment after an abnormal mammogram may be discouraged from participating at the next screening, which may reflect the effects of the anxiety and

Table 2 Estimated odds ratios (OR) from the multiple regression model for the association (unadjusted and adjusted) between women's characteristics and reattendance at the following screening invitation ($N = 2{,}660{,}155$)

	Screening mammograms	Risk of leaving screening	
		Univariate analysis (unadjusted OR, 95% CI)[a]	Multivariate analysis (adjusted OR, 95% CI)[b]
Age at screening			
44–49	685,540	Ref.	Ref.
50–54	1,162,316	1.02 (1.01, 1.03)*	0.91 (0.89, 0.92)*
55–59	1,181,353	1.04 (1.03, 1.06)*	0.90 (0.88, 0.92)*
60–64	1,079,044	2.72 (2.69, 2.75)*	2.23 (2.19, 2.27)*
65–69	437,093	8.61 (8.50, 8.71)*	8.48 (8.31, 8.65)*
Attended first invitation			
Yes	4,061,766	Ref.	Ref.
No	483,580	1.19 (1.18, 1.20)*	1.12 (1.11, 1.14)*
HRT			
No	2,857,354	Ref.	Ref.
Yes	297,624	0.79 (0.78, 0.80)*	0.96 (0.94, 0.97)*
Menopause			
Menopausal	2,262,797	Ref.	Ref.
Premenopausal	575,965	0.62 (0.61, 0.62)*	1.01 (0.99, 1.02)
Previous invasive procedure			
No	2,595,815	Ref.	Ref.
Yes	193,202	1.09 (1.07, 1.11)*	1.09 (1.07, 1.10)*
Familial breast cancer			
No	2,891,655	Ref.	Ref.
Yes	212,662	1.00 (0.99, 1.02)	0.97 (0.96, 0.99)*

Due to missing data in the women's related variables the number expresses the maximum number of available information

HRT hormone replacement therapy, *Menopause* pre/peri-menopausal or menopausal status, *Previous invasive procedure* personal previous invasive procedure, *Familial breast cancer* first-degree familial history of breast cancer

* Significant at the 95% Confidence level. An OR > 1 indicates that women with that characteristic are more likely to fail to return to the following screening invitation

[a] Analysis adjusted by women's screening participation, screening period (years), and radiology unit (random effect)

[b] Multivariate analysis adjusted by women's screening participation, screening period (years), radiology unit (random effect), interaction between false-positives and the women's screening participation, and all other factors in the table

discomfort produced in women during the screening process generating the false-positive result, outweighing the perceived benefit of further screening [5, 11]. Besides, we observed that the reduction in reattendance was higher if the false-positive involved an invasive procedure, which may be a consequence of the higher anxiety generated in women requiring invasive procedures.

In a previous study, we observed that women experiencing a false-positive at first screening had lower reattendance rates to the second screening invitation [24]. However, other European studies found no association between the presence of false-positives and the reattendance rates [12, 13], while studies carried out in the USA showed higher reattendance rates in women experiencing a false-positive result [14–16, 25]. A meta-analysis performed in 2005 by Brewer et al. [4], and extended in 2010 [5], showed that reattendance was significantly higher in women in the USA with a false-positive result whereas no statistically significant differences where found in Europe, although the trend was toward a slightly lower rate of return for women who received a false-positive. These differences between the USA and Europe could be explained by differences in the screening setting. Breast cancer screening in Europe is population-based, offering a routine screening mammogram to almost 100% of the target population [26–29], whereas in the USA screening is not government-sponsored and organized, and women with a false-positive may be more likely to be closely followed-up by their practitioners with the recommendation to undergo screening or continued clinical follow-up. Other causes might be a higher tolerance to missed breast cancer, and a higher reading volume among radiologists interpreting mammograms [5, 17].

The false-positive risk has been shown to be higher at first screening [30–32], and we observed that the strongest reduction in reattendance in women experiencing a false-positive was at first screening and that the observed differences decreased with the number of screening participations, showing that earlier experiences of abnormal mammograms had a stronger influence on behavior. Furthermore, our results also showed that the probability of failing to attend a screening invitation increased with the number of missed invitations, suggesting an increasingly reluctant attitude in women missing a screening invitation. These findings should be carefully considered to evaluate long-term effects of experiencing a false-positive and emphasize the importance of providing adequate information to women at their first screening to ensure future attendance.

In addition to false-positives, several other factors have previously been described as influencing the attendance rate [14, 16], including age, a first-degree familial history of breast cancer, HRT use, and age at menarche. In accordance with previous studies we found a lower reattendance rate in the oldest women [7, 18]. Decreasing adherence with increasing age may be related to poorer health status, a lower perception of the benefits of screening, or to greater difficulty in travelling to screening centers [33, 34]. This finding is particularly important given that cancer detection rates are the highest in older women and consequently efforts should be made to keep these women in routine screening programs.

As previously described [18, 35, 36], women who failed to attend invitations to their first participation had lower reattendance rates, indicating that the initial attitude to recruitment to the program predicts future attendance. As for the menopausal status, a non statistically significant association between menopausal status and reattendance rates was found after adjustment for age, and other risk factors. An association between late age at menopause and an increased reattendance rate has been previously described [14, 16]. Use of HRT was associated with higher reattendance rates, as observed in a previous study [14]. The higher adherence could be explained by closer follow-up by the gynecologist or recommendation by the physician to undergo screening, and thus these women may be more conscious of the benefits of screening.

Women with previous invasive procedures had lower adherence rates. Possible explanations are that these women, similarly to women experiencing a false-positive result, are more likely to develop anxiety in the diagnostic process and hence might be reluctant to attend future screenings. Another explanation could be that these women might be already followed-up by their own gynaecologist; however, no information to confirm this hypothesis is available. Women with a first-degree familial history of breast cancer were more likely to attend subsequent screening, although the magnitude of this effect was small (OR: 0.97). Some authors have found a similar association [14], while others have found no statistically significant differences [16, 18],

This study has several major strengths: its retrospective cohort design enabled us to assess the probability of reattendance and its modifying factors, over a sequence of screening invitations. To date, this is the largest study performed that assesses the impact of false-positives on reattendance at breast cancer screening. However, this study also has some limitations. Firstly, the information on women's personal variables was not always available or complete in all the radiology units. To perform the multivariate regression models we analyzed a subsample with the maximum available information about women's characteristics ($N = 2{,}660{,}155$). A sensitivity analysis was performed to evaluate the overall reattendance rate, false-positive rate, and the age distribution among the initial study population and the subset with complete information for each personal variable, and no major differences were observed (see Online Appendix). Secondly, we were unable to determine whether women failing to participate had mammograms performed outside the screening program setting. Furthermore, no information was available on the socio-economic status social support, etc. of the screening participants, which could have been highly informative.

In conclusion, our study evaluated attendance at breast cancer screening at six subsequent invitations and the causes related to lower adherence. We found that false-positive results and women's personal characteristics affected attendance at subsequent screenings. Because repeated sequential screening is essential to reduce breast cancer mortality, understanding the factors modifying reattendance is important when the risks and benefits of screening are analyzed. This information could be useful to provide the best available information to women invited to participate and to improve adherence to subsequent screenings.

Acknowledgments This study was founded by grants from the Instituto de Salud Carlos III FEDER (PI09/90251). The funding sources had no role in the performance of the study or in the preparation of the manuscript. The authors would like to thank the entire *Cumulative False Positive Risk Group* for their support and dedication to this study.

Conflicts of interest The authors have declared no conflicts of interest.

Appendix

The Cumulative False Positive Risk Group (alphabetical order): Department of Epidemiology and Evaluation, Institut Municipal d'Investigació Mèdica-Parc de Salut

Mar, Barcelona: Xavier Castells, Francesc Macià, Cristiane Murta, Rubén Román, María Sala. 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, Iosu Delfrade, María Ederra, Nieves Erdozain, 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. Foundation Society for Cancer Research and Prevention. Pere Virgili 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. Jefa de la Sección de Promoción y Protección de la Salud. Cancer Screening and Epidemiology Department, UDIAT-CD. Corporació Parc Taulí-Institut Universitari Parc Taulı (UAB), Sabadell: Marisa Baré. Castilla-Leon Breast Cancer Screening Programme. Dirección General de Salud Pública ID e I. SACYL, Castilla y León: Isabel González-Román.

References

1. Brett J, Bankhead C, Henderson B, Watson E, Austoker J (2005) The psychological impact of mammographic screening. A systematic review. Psychooncology 14:917–938
2. Lerman C, Trock B, Rimer BK, Boyce A, Jepson C, Engstrom PF (1991) Psychological and behavioral implications of abnormal mammograms. Ann Intern Med 114:657–661
3. Bull AR, Campbell MJ (1991) Assessment of the psychological impact of a breast screening programme. Br J Radiol 64:510–515
4. Brewer NT, Salz T, Lillie SE (2007) Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med 146:502–510
5. Salz T, Defrank JT, Brewer NT (2011) False positive mammograms in Europe: do they affect reattendance? Breast Cancer Res Treat 127:229–230
6. Day NE, Williams DR, Khaw KT (1989) Breast cancer screening programmes: the development of a monitoring and evaluation system. Br J Cancer 59:954–958
7. Chiarelli AM, Moravan V, Halapy E, Majpruz V, Mai V, Tatla RK (2003) False-positive result and reattendance in the Ontario Breast Screening Program. J Med Screen 10:129–133
8. Hofvind S, Wang H, Thoresen S (2003) The Norwegian Breast Cancer Screening Program: re-attendance related to the women's experiences, intentions and previous screening result. Cancer Causes Control 14:391–398
9. Johnson MM, Hislop TG, Kan L, Coldman AJ, Lai A (1996) Compliance with the screening mammography program of British Columbia: will she return? Can J Public Health 87:176–180
10. McCann J, Stockton D, Godward S (2002) Impact of false-positive mammography on subsequent screening attendance and risk of cancer. Breast Cancer Res 4:R11
11. Brett J, Austoker J (2001) Women who are recalled for further investigation for breast screening: psychological consequences 3 years after recall and factors affecting re-attendance. J Public Health Med 23:292–300
12. Lampic C, Thurfjell E, Sjoden PO (2003) The influence of a false-positive mammogram on a woman's subsequent behaviour for detecting breast cancer. Eur J Cancer 39:1730–1737
13. O'Sullivan I, Sutton S, Dixon S, Perry N (2001) False positive results do not have a negative effect on reattendance for subsequent breast screening. J Med Screen 8:145–148
14. Pinckney RG, Geller BM, Burman M, Littenberg B (2003) Effect of false-positive mammograms on return for subsequent screening mammography. Am J Med 114:120–125
15. Lipkus IM, Halabi S, Strigo TS, Rimer BK (2000) The impact of abnormal mammograms on psychosocial outcomes and subsequent screening. Psychooncology 9:402–410
16. Burman ML, Taplin SH, Herta DF, Elmore JG (1999) Effect of false-positive mammograms on interval breast cancer screening in a health maintenance organization. Ann Intern Med 131:1–6
17. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L (2006) European guidelines for quality assurance in breast cancer screening and diagnosis. Office for Official Publications of the European Communities, Luxenburg
18. Bulliard JL, De Landtsheer JP, Levi F (2004) Reattendance in the Swiss mammography screening pilot programme. J Med Screen 11:59–64
19. Assembly of Members (2001) CONSTITUTION AND BYLAWS of the European Breast Cancer Network
20. Roman R, Sala M, Salas D, Ascunce N, Zubizarreta R, Castells X (2011) Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. Ann Oncol. doi:10.1093/annonc/mdr032
21. Ascunce N, Salas D, Zubizarreta R, Almazan R, Ibanez J, Ederra M (2010) Cancer screening in Spain. Ann Oncol 21(Suppl 3): iii43–iii51
22. Singer J, Willett JB (2003) Fitting basic discrete-time hazard models. Applied longitudinal data analysis: modelling change and event occurrence. Oxford University Press, New York, pp 357–467
23. Willett JB, Singer J (1995) It's deja vu all over again: using multiple-spell discrete-time survival analysis. J Educ Behav Stat 20(1):41–67
24. Álamo-Junquera D, Murta-Nascimento C, Macià F, Baré M, Galceran J, Ascunce N et al (2011) Effect of false-positive results on reattendance at breast cancer screening programmes in Spain. Eur J Public Health. doi:10.1093/eurpub/ckr057
25. Pisano ED, Earp J, Schell M, Vokaty K, Denham A (1998) Screening behavior of women after a false-positive mammogram. Radiology 208:245–249
26. Hofvind S, Vacek PM, Skelly J, Weaver DL, Geller BM (2008) Comparing screening mammography for early breast cancer detection in Vermont and Norway. J Natl Cancer Inst 100: 1082–1091
27. Smith-Bindman R, Ballard-Barbash R, Miglioretti DL, Patnick J, Kerlikowske K (2005) Comparing the performance of mammography screening in the USA and the UK. J Med Screen 12:50–54
28. Elmore JG, Nakano CY, Koepsell TD, Desnick LM, D'Orsi CJ, Ransohoff DF (2003) International variation in screening mammography interpretations in community-based programs. J Natl Cancer Inst 95:1384–1393
29. Yankaskas BC, Klabunde CN, Ancelle-Park R, Renner G, Wang H, Fracheboud J et al (2004) International comparison of performance measures for screening mammography: can it be done? J Med Screen 11:187–193
30. 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
31. 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
32. 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
33. Rodriguez C, Plasencia A, Schroeder DG (1995) Predictive factors of enrollment and adherence in a breast cancer screening program in Barcelona (Spain). Soc Sci Med 40:1155–1160
34. de Waard F, Collette HJ, Rombach JJ, Baanders-van Halewijn EA, Honing C (1984) The DOM project for the early detection of breast cancer, Utrecht, The Netherlands. J Chronic Dis 37:1–44
35. Cockburn J, Schofield P, White V, Hill D, Russell I (1997) Predictors of returning for second round screening at a population based mammographic screening programme in Melbourne, Australia. J Epidemiol Community Health 51:62–66
36. Sutton S, Bickler G, Sancho-Aldridge J, Saidi G (1994) Prospective study of predictors of attendance for breast screening in inner London. J Epidemiol Community Health 48:65–73

4. Artículo 3

Título: Breast cancer detection risk in screening mammography after a false-positive result

Autores: Castells X, Román M, Romero A, Blanch J, Zubizarreta R, Ascunce N, Salas D, Burón A, Sala M, Cumulative False Positive Risk Group.

Revista: Cancer Epidemiol. 2013; 37(1): 85-90.

Factor de impacto: 2.0 (Q2 Oncology)

DOI: 10.1016/j.canep.2012.10.004

Abstract:

Background: False-positives are a major concern in breast cancer screening. However, false-positives have been little evaluated as a prognostic factor for cancer detection. Our aim was to evaluate the association of false-positive results with the cancer detection risk in subsequent screening participations over a 17-year period.

Methods: This is a retrospective cohort study of 762,506 women aged 45-69 years, with at least two screening participations, who underwent 2,594,146 screening mammograms from 1990 to 2006. Multilevel discrete-time hazard models were used to estimate the adjusted odds ratios (OR) of breast cancer detection in subsequent screening participations in women with false-positive results.

Results: False-positives involving a fine-needle aspiration cytology or a biopsy had a higher cancer detection risk than those involving additional imaging procedures alone (OR = 2.69; 95%CI: 2.28-3.16 and OR = 1.81; 95%CI: 1.70-1.94, respectively). The risk of cancer detection increased substantially if women with cytology or biopsy had a familial history of breast cancer (OR = 4.64; 95%CI: 3.23-6.66). Other factors associated with an increased cancer detection risk were age 65-69 years (OR = 1.84; 95%CI: 1.67-2.03), non-attendance at the previous screening invitation (OR = 1.26; 95%CI: 1.11-1.43), and having undergone a previous benign biopsy outside the screening program (OR = 1.24; 95%CI: 1.13-1.35).

Conclusion: Women with a false-positive test have an increased risk of cancer detection in subsequent screening participations, especially those with a false-positive result involving

cytology or biopsy. Understanding the factors behind this association could provide valuable information to increase the effectiveness of breast cancer screening.

Cancer Epidemiology 37 (2013) 85–90

Contents lists available at SciVerse ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Breast cancer detection risk in screening mammography after a false-positive result

X. Castells [a,b,*], M. Román [a,b], A. Romero [a,b], J. Blanch [a], R. Zubizarreta [c], N. Ascunce [b,d], D. Salas [e], A. Burón [a], M. Sala [a,b], the Cumulative False Positive Risk Group

[a] Department of Epidemiology and Evaluation, Institut Municipal d'Investigació Mèdica-Parc de Salut Mar, Mar Teaching Hospital, 25-29 Passeig Marítim, 08003 Barcelona, Spain
[b] CIBERESP, CIBER Epidemiología y Salud Pública, Spain
[c] Galician Breast Cancer Screening Program, Public Health and Planning Directorate, Health Office, San Lázaro s/n, 15703 Santiago de Compostela, Galicia, Spain
[d] Navarra Breast Cancer Screening Program, Instituto de Salud Pública, Leyre 15, 31003 Pamplona, Spain
[e] General Directorate of Public Health and Center for Public Health Research (CSISP), Avda. Catalunya 21, Valencia, Spain

ARTICLE INFO

Article history:
Received 6 July 2012
Received in revised form 5 October 2012
Accepted 8 October 2012
Available online 9 November 2012

Keywords:
Breast cancer screening
Cancer detection
False-positive
Risk factors

ABSTRACT

Background: False-positives are a major concern in breast cancer screening. However, false-positives have been little evaluated as a prognostic factor for cancer detection. Our aim was to evaluate the association of false-positive results with the cancer detection risk in subsequent screening participations over a 17-year period. *Methods*: This is a retrospective cohort study of 762,506 women aged 45–69 years, with at least two screening participations, who underwent 2,594,146 screening mammograms from 1990 to 2006. Multilevel discrete-time hazard models were used to estimate the adjusted odds ratios (OR) of breast cancer detection in subsequent screening participations in women with false-positive results. *Results*: False-positives involving a fine-needle aspiration cytology or a biopsy had a higher cancer detection risk than those involving additional imaging procedures alone (OR = 2.69; 95%CI: 2.28–3.16 and OR = 1.81; 95%CI: 1.70–1.94, respectively). The risk of cancer detection increased substantially if women with cytology or biopsy had a familial history of breast cancer (OR = 4.64; 95%CI: 3.23–6.66). Other factors associated with an increased cancer detection risk were age 65–69 years (OR = 1.84; 95%CI: 1.67–2.03), non-attendance at the previous screening invitation (OR = 1.26; 95%CI: 1.11–1.43), and having undergone a previous benign biopsy outside the screening program (OR = 1.24; 95%CI: 1.13–1.35). *Conclusion*: Women with a false-positive test have an increased risk of cancer detection in subsequent screening participations, especially those with a false-positive result involving cytology or biopsy. Understanding the factors behind this association could provide valuable information to increase the effectiveness of breast cancer screening.

1. Introduction

One of the major concerns in breast cancer screening is the false-positive result. The negative effects of a positive mammographic reading in which cancer is excluded after additional evaluation include psychological [1] and behavioral consequences to the screened women [2], as well as additional physician visits, diagnostic tests, and excision biopsies [3,4].

The widespread adoption of breast cancer screening programs involves screening thousands of women periodically, of whom a large number will have a positive mammographic reading requiring additional evaluation. The estimated proportion of women with a false-positive result after ten screening participations ranges from 20% to 32% in Europe [5–7] and around 49% in the USA [8]. If the false-positive test involves cytology or a biopsy, variability in the estimations increases substantially, ranging from 1.7% to 5% in Europe [5,7], and 18.6% in the USA [8]. However, a negative result after additional evaluation does not necessarily indicate the absence of a benign lesion or a suspicious mammographic pattern.

The dissemination of screening mammography has increased the number of women with radiological abnormalities or benign breast lesions, although there is no general agreement for the follow-up of these women in the screening context. In most population-based screening programs women with a false-positive result follow the same screening recommendations as those with a negative mammographic reading [9]. However, benign breast lesions are a known risk factor for subsequent breast cancer [10,11], and women with benign breast surgery have lower sensitivity at screening [12]. Indeed, the presence of previous

* Corresponding author at: 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 address: xcastells@parcdesalutmar.cat (X. Castells).

1877-7821/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.canep.2012.10.004

benign breast lesions is a commonly included variable in the models assessing individual breast cancer risk, along with other factors such as the use of hormone replacement therapy (HRT) and a familial history of breast cancer [13–15].

Although several basic aspects of false positives and their effects have previously been studied, the association between false-positive results and detection of breast cancer in subsequent screening participations has been little studied [16–20]. Most of these studies had a small sample size and a short follow-up time, or had no information on whether the false-positive result involved a cytology examination or biopsy.

In the context of population-based screening programs, in which large cohorts of women are sequentially invited for a mammographic test over a time span of 20 years, the long-term follow-up of women with false-positive results could enhance the prediction of breast cancer risk [13,15]. This information might be useful to improve the effectiveness of breast cancer screening programs by encouraging women with false-positive results to return for further screening.

The aim of this study was to evaluate the association of a false-positive result with risk of breast cancer detection in a cohort of screened women over a sequence of routine screening participations.

2. Methods

2.1. Setting and study population

The study sample was drawn from a retrospective cohort study of screened women, conducted to evaluate the cumulative risk of a false-positive result over ten sequential screening participations [7]. Briefly, all women aged 45–69 resident in Spain are actively invited to participate in a population-based screening program every 2 years. Population-based breast cancer screening in Spain started in 1990 and became nationwide in 2006. Data from eight regions, covering 44% of the Spanish target population, were collected for this study. Each region has one or several radiology units that perform screening [21]. Breast cancer screening in Spain follows the European Guidelines for Quality Assurance in Mammographic Screening [9].

Information was obtained from 945,789 women who had undergone at least one screening mammogram between March 1990 and December 2006. These women underwent 2,777,429 screening mammograms in any of the 45 radiology units of the eight participating regions that routinely collected information on the women's personal characteristics. The study was approved by the Mar Teaching Hospital Research Ethics Committee.

2.2. False-positive results, cancer detection and women's personal characteristics

Women with a positive mammographic reading are recalled for additional evaluation to exclude malignancy. The diagnostic work-up took place within a maximum of 2 months after the screening test. Some women with a probably benign result at mammographic reading are referred for an intermediate mammogram at 6 or 12 months before the interval corresponding to the normal sequence (early recall) [22].

A positive result in the screening test was considered a false-positive result if, after additional evaluation, breast cancer was not diagnosed. Additional evaluation may include additional imaging procedures (additional mammography, magnetic resonance imaging, and ultrasonography), cytology (fine-needle aspiration cytology), or biopsy (core or open biopsy). A definitive diagnosis of breast cancer was always histopathologically confirmed (invasive carcinoma or carcinoma ductal in situ). If cancer was excluded after additional evaluation, women were routinely invited to participate in the screening program 2 years after the previous screening invitation. No information was available on cancers diagnosed as interval cancers or after women left the screening program.

Information on women's characteristics was obtained by a face-to-face interview performed by a trained health professional at the time of each screening mammogram. This information included the women's age, HRT use (present use or in the previous 6 months), menopausal status (pre- or postmenopausal), previous benign biopsy outside the screening program, and first-degree familial history of breast cancer.

2.3. Statistical analysis

The cancer detection rates were calculated as the number of breast cancers detected at screening divided by the number of screened women. The odds ratios (OR) and the 95% confidence intervals (95%CIs) for the association between false-positive results and the risk of cancer detection in subsequent screening participations were estimated with discrete time-hazard models. These models use a logistic regression approach to compute these particular survival models with discrete time intervals [23,24]. The event of interest was whether or not cancer was detected at a routine screening invitation. The probability of a cancer being detected at a routine screening invitation ($\pi(x)$) was expressed as $\ln(\pi(x)/1-\pi(x)) = \alpha_i D_i + \beta_j X_j$, where $\pi(x)$ is estimated by means of the logit function, like any other logistic regression model. D_i corresponds to the time indicators: one for each woman's screening participation (first screening, second screening, etc.). D_i equals 1 if the woman has performed her ith screening, and is 0 otherwise. The coefficients of the time indicators are expressed by α_i and are the intercepts in the model (multiple intercept model). As in any other regression model X_j is the jth study factor (i.e. first-degree familial history of breast cancer, attended previous screening invitation, etc.), and β_j is the estimated coefficient for the associated study factor. As cancers detected at first screening would not have a previous false-positive result in the screening setting, first screens were censored to compute the regression model estimates, as they would underestimate the risk.

Simple and multivariate models were used to estimate the individual and simultaneous effect of all predictors. The multivariate models included the women's personal variables (age, HRT use, menopausal status, previous benign biopsy outside the screening program, a first-degree family history of breast cancer), whether or not the woman attended her previous screening invitation, and the presence of a false-positive result in any previous screening participation. In addition, the multivariate models included a period effect (calendar years), as the start date of the radiology units differed, and a random effect component defined by the radiology units, because of the correlation among screening tests performed in the same radiology unit. Residual pseudo-likelihood estimation was used in all models by means of the GLIMMIX procedure in SAS 9.1.2 (SAS Institute, Cary, NC).

In further analyses, we tested for interactions between false-positive results and menopausal status, HRT use, family history of breast cancer, and a previous benign biopsy outside the screening program. For simplicity in the interpretation, we performed a stratified analysis for those women's characteristics showing a statistically significant interaction with false-positive results. Besides, to study whether the number of screening rounds since the false-positive test had an effect on the breast cancer risk, we analyzed whether the false-positive test occurred in the previous screening round (2 years) or two or more screenings in advance ($\geq$4 years).

Finally, we studied whether the cytologies and biopsies carried out to exclude malignancy were associated with a differential

cancer detection risk. A regression model was computed that included the additional imaging procedures, cytologies, and biopsies as independent categories.

3. Results

Of the 945,789 women who had undergone at least one screening mammogram, we excluded information from 183,283 women (19.4%) who had participated in only one screening round and could not be followed up over subsequent screening rounds. We analyzed information from 762,506 women who had at least two screening participations, who underwent 2,594,146 mammographic screening tests between 1990 and 2006. Average (standard deviation) screening participations per woman was 3.70 (1.60); 73% of women had undergone three or more screening mammograms, while 25.5% had at least five screenings.

Overall, the cancer detection rate in subsequent screenings observed was 2.89 cases per 1000 screening mammograms (Table 1). The cancer detection rate for women with a previous false positive involving an additional imaging procedure and those involving a cytology or biopsy was 4.53 and 7.09 cases per 1000 screening mammograms, respectively. Other factors associated with a higher detection rate were a first-degree family history of breast cancer, non-attendance at the previous screening invitation, having experienced a benign biopsy outside the screening program, older age, and post-menopausal status.

False positives showed an increased cancer detection risk in subsequent screening participations. False positives involving a cytology or biopsy were associated with a significantly higher risk of cancer detection than false positives leading to additional imaging procedures (OR = 2.69; 95%CI: 2.28–3.16 and OR = 1.81; 95%CI: 1.70–1.94, respectively) (Table 2). A higher cancer detection risk was also observed in the oldest women (OR = 1.84; 95%CI: 1.67–2.03), women with a first-degree familial history of breast cancer (OR = 1.65; 95%CI: 1.52–1.79), those not attending the previous screening invitation (OR = 1.26; 95%CI: 1.11–1.43), and those with a previous benign biopsy outside the screening program (OR = 1.24; 95%CI: 1.13–1.35). Of all the factors studied, a previous false-positive result, independently of the additional procedure involved (additional imaging, cytology or biopsy), showed the highest risk of cancer detection (OR = 1.89; 95%CI: 1.77–2.01) (data not shown).

The stratified analyses showed a stronger association of false positives involving a cytology or biopsy with the risk of cancer detection in women with a familial history of breast cancer compared with that in women without a familial history of breast cancer (OR = 4.64; 95%CI: 3.23–6.66, and OR = 2.41; 95%CI: 2.00–2.89, respectively) (Table 3). No differences among women with a familial history of breast cancer were observed for women with a false positive involving additional imaging procedures. None of the other women's characteristics tested for an interaction showed a statistically significant difference.

Fig. 1 shows that false positives after additional imaging procedures or after cytology or biopsy had an increased cancer

Table 1
Number of cancers detected and cancer detection rates in subsequent screens for the women's characteristics studied.

Variable	Subsequent screens (*N*)	Cancers (*N*)	Rate[a] (95%CI)
	1,963,225	5670	2.89 (2.81–2.96)
Previous false-positive[b]			
Never	1,663,403	4256	2.56 (2.48–2.64)
Additional imaging	278,081	1261	4.53 (4.28–4.78)
Cytology or biopsy	21,588	153	7.09 (5.97–8.21)
Attended previous screening invitation			
Yes	1,896,407	5410	2.85 (2.78–2.93)
No	66,818	260	3.89 (3.42–4.36)
Age (years)			
45–49	177,671	333	1.87 (1.67–2.08)
50–54	467,619	1036	2.22 (2.08–2.35)
55–59	558,354	1569	2.81 (2.67–2.95)
60–64	514,556	1762	3.42 (3.26–3.58)
65–70	245,025	970	3.96 (3.71–4.21)
HRT[c]			
No	1,743,323	5071	2.91 (2.83–2.99)
Yes	219,902	599	2.72 (2.51–2.94)
Menopausal status			
Menopausal	1,656,585	5025	3.03 (2.95–3.12)
Premenopausal	306,640	645	2.10 (1.94–2.27)
First-degree family history of breast cancer			
No	1,817,823	4989	2.74 (2.67–2.82)
Yes	145,402	681	4.68 (4.33–5.03)
Previous benign biopsy outside screening			
No	1,826,679	5139	2.81 (2.74–2.89)
Yes	136,546	531	3.89 (3.56–4.22)

95%CI = 95% confidence interval.
[a] Rate is presented as number of cancers per 1000 screening mammograms.
[b] Previous false-positive: at least one false-positive result in previous screening rounds.
– Never: women who had never experienced a false-positive result.
– Additional imaging: a false-positive involving only an additional mammogram, or a magnetic resonance imaging scan, or an ultrasound scan.
– Cytology or biopsy: a false positive involving fine-needle aspiration cytology, or core biopsy, or open biopsy. 95%CI = 95% confidence interval.
[c] HRT: hormone replacement therapy use at the time of the mammogram or in the previous 6 months.

Table 2
Estimated odds ratios (OR) from the multiple regression model for the association (non-adjusted and adjusted) between women's characteristics and the risk of cancer detection in subsequent screening participations.

Risk factor	Subsequent screens (*N*)	OR (95%CI) Non-adjusted[a]	OR (95%CI) Adjusted[b]
Previous false-positive[c]			
Never	1,663,403	Ref	Ref
Additional imaging	278,013	1.73 (1.62–1.85)	1.81 (1.70–1.94)
Cytology or biopsy	21,809	2.89 (2.48–3.37)	2.69 (2.28–3.16)
Attended previous screening invitation			
Yes	1,896,407	Ref	Ref
No	66,818	1.42 (1.25–1.61)	1.26 (1.11–1.43)
Age			
45–49	177,671	0.83 (0.73–0.94)	0.83 (0.73–0.95)
50–54	467,619	Ref	Ref
55–59	558,354	1.27 (1.18–1.38)	1.30 (1.20–1.42)
60–64	514,556	1.55 (1.43–1.68)	1.62 (1.49–1.77)
65–70	245,025	1.78 (1.63–1.95)	1.84 (1.67–2.03)
HRT[d]			
No	1,743,323	Ref	Ref
Yes	219,902	0.93 (0.86–1.02)	0.96 (0.88–1.04)
Menopausal status			
Menopausal	1,656,585	Ref	Ref
Premenopausal	306,640	0.71 (0.65–0.77)	0.92 (0.83–1.02)
First-degree family history of breast cancer			
No	1,817,823	Ref	Ref
Yes	145,402	1.69 (1.56–1.84)	1.65 (1.52–1.79)
Previous benign biopsy outside screening			
No	1,826,679	Ref	Ref
Yes	136,546	1.38 (1.26–1.51)	1.24 (1.13–1.35)

95%CI = 95% confidence interval.
[a] Analysis adjusted by women's screening participation.
[b] Multivariate analysis adjusted by women's screening participation, screening period (years), radiology unit (random effect), and all other factors in the table.
[c] Previous false-positive: at least one false-positive result in previous screening rounds.
– Never: women who had never experienced a false-positive result.
– Additional imaging: a false-positive involving only an additional mammogram, or a magnetic resonance imaging scan, or an ultrasound scan.
– Cytology or biopsy: a false positive involving fine-needle aspiration cytology, or core biopsy, or open biopsy.
[d] HRT: hormone replacement therapy use at the time of the mammogram or in the previous 6 months.

Table 3
Estimated odds ratios (OR) from the multiple regression model for the association between false-positive results and subsequent breast cancer detection risk by the presence or absence of a first-degree familial history of breast cancer.

Previous false-positive[a]	Women with a first-degree family history of breast cancer			Women without a first-degree family history of breast cancer		
	Subsequent screens (*N*)	Cancer (*N*)	OR (95%CI) Adjusted[b]	Subsequent screens (*N*)	Cancer (*N*)	OR (95%CI) Adjusted[b]
Never	119,782	478	Ref	1,543,621	3778	Ref
Additional imaging	23,859	170	1.82 (1.51–2.18)	254,154	1091	1.81 (1.69–1.95)
Cytology or biopsy	17,961	33	4.64 (3.23–6.66)	20,048	120	2.41 (2.00–2.89)

[a] Previous false-positive: at least one false-positive result in previous screening rounds.
– Never: women who had never experienced a false-positive result.
– Additional imaging: a false-positive involving only an additional mammogram, or a magnetic resonance imaging scan, or an ultrasound scan.
– Cytology or biopsy: a false-positive involving fine-needle aspiration cytology, or core biopsy, or open biopsy.
[b] Multivariate analysis adjusted by women's screening participation, screening period (years), radiology unit (random effect), whether or not the woman attended the previous screening invitation, age at screening, hormone replacement therapy use, menopausal status, and previous benign biopsy outside screening.

detection risk, independently of whether the false-positive test occurred in the previous screening round or two or more screenings in advance. False-positive tests experienced in the previous screening round were significantly associated with a higher cancer detection risk than those experiencing two or more screenings in advance ($P = 0.025$ and $P = 0.045$, for false-positive test after additional imaging procedures and after cytology or biopsy, respectively).

The association between the type of additional procedure carried out in the process leading to the false-positive test and the cancer detection risk is shown in Fig. 2. No differences were found in the cancer detection risk between false positives involving a cytology and those involving a biopsy (OR = 2.95; 95%CI: 2.34–3.71, and OR = 2.72; 95%CI: 2.11–3.52, respectively) ($P = 0.90$). False positives leading to additional imaging procedures had a significantly lower cancer detection risk (OR = 1.75; 95%CI: 1.63–1.88) than those involving cytology or a biopsy ($P < 0.001$ and $P = 0.005$, respectively).

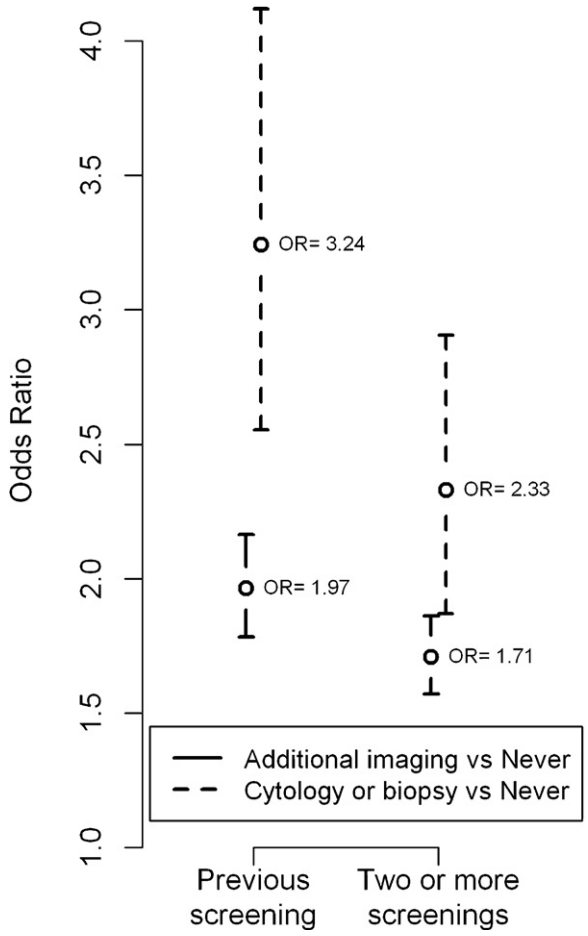


Fig. 1. Adjusted odds ratios (OR) for the cancer detection risk depending on whether the false-positive test occurred in the previous screening round or two or more screenings in advance.

4. Discussion

We observed an increased risk of breast cancer detection in women with a previous false-positive test in mammographic screening. Women with a false positive involving cytology or biopsy had a higher risk of cancer detection than those with a false positive involving only an additional imaging procedure. This risk remained significantly higher 4 years or more after the false-positive test. The cancer detection risk increased substantially if women with a cytology or biopsy had a familial history of breast cancer.

The increased cancer detection risk in women with a false-positive test observed in this study is in agreement with the results of previous studies. In a recent study, Euler-Chelpin et al. found an RR = 1.67 of breast cancer diagnosis after a false-positive test [16]. McCann et al. found an OR = 2.15 of cancer detection at the second screen in women with a false-positive test at the first screen [18].

A false-positive test in previous screening rounds is not in itself a risk factor for breast cancer. Some authors have reported false negatives in women undergoing additional evaluation after a

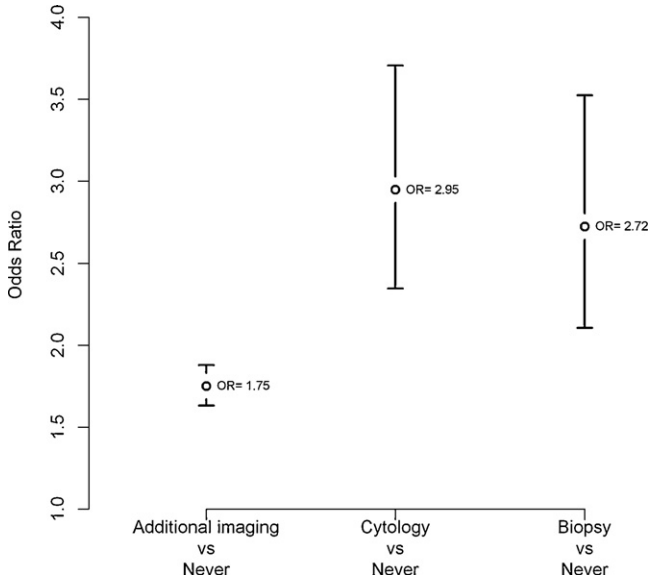


Fig. 2. Adjusted odds ratios (OR) for the cancer detection risk depending on the type of additional procedure leading to the false-positive test.

positive mammographic reading [18,25–27]. However, in agreement with the study of Euler-Chelpin et al., the cancer detection risk remained significantly higher 4 years or more after the false-positive test [16]. Besides, cancers missed at additional evaluation represent a small proportion of the whole [25], which could only partially explain the association between false-positive tests and the cancer detection risk in subsequent screening participations.

Women with a recommendation for additional evaluation are a specific subgroup of women with mammographic abnormalities. The absence of malignancy does not indicate the absence of benign abnormalities, especially in women recalled for a cytology examination or biopsy. A previous benign breast lesion is a known breast cancer risk factor [10,11,28] and is commonly included in models predicting breast cancer risk. However, few studies have assessed the impact of previous benign lesions in the context of breast cancer screening, in which non-symptomatic women are routinely evaluated. In our analyses, false positives involving a cytology examination or biopsy had an increased cancer detection risk (OR = 2.95 and OR = 2.72, respectively) compared with additional imaging procedures (OR = 1.75). This association was stronger than any other factor analyzed in the study, most of which are usually included in predictive models, such as a first-degree family history of breast cancer, older age, or a previous benign biopsy outside screening.

The risk of cancer detection after a false-positive test involving a cytology examination or biopsy was higher in women with a first-degree familial history of breast cancer (OR = 4.64). This differential effect could be partially explained by the presence of unknown genetic factors or malignant precursors in these women, as well as shared lifestyle and environment, which would involve prognostic factors for benign breast disease to develop into a malignant lesion [11]. In contrast with other studies [17], we found no significant differences in premenopausal women after adjusting for all the other study factors.

We analyzed information from a wide retrospective cohort over a 17-year period, which enabled us to ascertain the risk over a series of sequential screening participations. The wide spectrum of information analyzed – integrating information from several radiology units with different screening protocols – strengthens the consistency of the associations found, independently of possible differences in screening practice or the period analyzed. Moreover, the associations found were observed after adjustment was made for possible confounders, and in the stratified analysis. Nevertheless, our study also has some limitations. We performed specific analyses to outline possible causes for the association studied, which suggested some possible underlying reasons. Further studies are required to confirm the suggested hypothesis. No information was available on breast density, which could be associated with both an increased false-positive risk and an increased breast cancer risk. Previous studies have suggested that the association between previous false positives and cancer detection is independent of breast density [17].

The information provided in this study could be useful to increase the effectiveness of breast cancer screening programs if several surveillance strategies are rethought and defined taking into account personal factors related to breast cancer risk [29], including the results of the screening test. Women with a false-positive result should be encouraged to return for further screening as they have an increased cancer detection risk, and a decreased re-attendance probability [2]. Currently, the quality guidelines [9] define the target population for screening only by women's age and include women who may have very different breast cancer risks in the same target groups. In the actual debate about the effectiveness of breast cancer screening it seems straightforward to consider future screening strategies according to the breast cancer risk. Personalizing strategies would increase the positive and negative predictive values of mammographic screening, which in turn would enhance its effectiveness. Some studies have provided evidence in this regard [29].

In conclusion, our results showed a strong association between the presence of a false-positive test and the risk of cancer detection in subsequent screening participations. The association was stronger in false-positives involving a cytology examination or biopsy, and in women with a family history of breast cancer. Previous false-positive tests were a better predictor of cancer detection in subsequent screens than older age, a previous benign biopsy outside screening, or a family history of breast cancer alone. In the context of mammographic screening, in which large cohorts of women are assessed every 2 years, this personalized risk information could be useful to improve the effectiveness of breast cancer screening by emphasizing the need for return for further screening in women with false-positive results.

Role of the funding source

This work was supported by grants from the Instituto de Salud Carlos III-FEDER (PI09/90251). The funding sources had no role in the performance of the study or in the preparation of the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Appendix

The Cumulative False Positive Risk Group (alphabetical order): Department of Epidemiology and Evaluation, Mar Teaching Hospital, Barcelona: Jordi Blanch, Xavier Castells, Marta Román, Anabel Romero, Maria Sala. Galician Breast Cancer Screening Program. Public Health and Planning Directorate. Health Office, Galicia: Raquel Almazán, Ana Belén Fernández, María Teresa Queiro, Raquel Zubizarreta. Navarra Breast Cancer Screening Program. Public Health Institute, Pamplona: Nieves Ascunce, Iosu Delfrade, María Ederra, Nieves Erdozain, Juana Vidán. General Directorate Public Health and Centre for Public Health Research (CSISP). Valencian Health Agency and Center for Public Health Research (CSISP), Valencia: Dolores Cuevas, Josefa Ibáñez, Dolores Salas. Servicio Canario de la Salud, Canary Islands: María Obdulia De la Vega, Isabel Díez de la Lastra. Foundation Society for Cancer Research and Prevention. Pere Virgili Health Research Institute, Reus, Tarragona: Jaume Galceran. Program and Analysis Unit. Health Office, Asturias: Carmen Natal. La Rioja Breast Cancer Screening Program. Fundacion Rioja Salud, Logroño: Araceli Baroja. Cancer Screening and Epidemiology Department, UDIAT-CD. Corporació Parc Taulí-Institut Universitari Parc Taulı (UAB), Sabadell: Marisa Baré. Castilla-Leon Breast Cancer Screening Program. Dirección General de Salud Pública ID e I. SACYL, Castilla y León: Isabel González-Román.

References

[1] Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. Psychooncology 2005;14(November (11)):917–38.
[2] Roman R, Sala M, De L, Natal V, Galceran C, Gonzalez-Roman J, et al. Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening. Breast Cancer Res Treat 2011;543–52.
[3] Bull AR, Campbell MJ. Assessment of the psychological impact of a breast screening programme. Br J Radiol 1991;510–5.
[4] Lerman C, Trock B, Rimer BK, Boyce A, Jepson C, Engstrom PF. Psychological and behavioral implications of abnormal mammograms. Ann Intern Med 1991;657–61.

[5] Hofvind S, Thoresen S, Tretli S. The cumulative risk of a false-positive recall in the Norwegian Breast Cancer Screening Program. Cancer 2004;101(October (7)):1501–7.
[6] Njor SH, Olsen AH, Schwartz W, Vejborg I, Lynge E. Predicting the risk of a false-positive test for women following a mammography screening programme. J Med Screen 2007;14(2):94–7.
[7] Roman R, Sala M, Salas D, Ascunce N, Zubizarreta R, Castells X. Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. Ann Oncol 2011;March.
[8] Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med 1998;1089–96.
[9] Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Luxenburg: Office for Official Publications of the European Communities, 2006
[10] Ashbeck EL, Rosenberg RD, Stauber PM, Key CR. Benign breast biopsy diagnosis and subsequent risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2007;467–72.
[11] Hartmann LC, Sellers TA, Frost MH, Lingle WL, Degnim AC, Ghosh K, et al. Benign breast disease and the risk of breast cancer. N Engl J Med 2005;229–37.
[12] van BS, Duijm V, Voogd LE, Jansen AC, Louwman FHMW. Mammographic changes resulting from benign breast surgery impair breast cancer detection at screening mammography. Eur J Cancer 2012;2097–103.
[13] Cook NR, Rosner BA, Hankinson SE, Colditz GA. Mammographic screening and risk factors for breast cancer. Am J Epidemiol 2009;1422–32.
[14] Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;1879–86.
[15] Santen RJ, Boyd NF, Chlebowski RT, Cummings S, Cuzick J, Dowsett M, et al. Critical assessment of new risk factors for breast cancer: considerations for development of an improved risk prediction model. Endocr Relat Cancer 2007;169–87.
[16] Euler-Chelpin M, Risor LM, Thorsted BL, Vejborg I. Risk of breast cancer after false-positive test results in screening mammography. J Natl Cancer Inst 2012;682–9.
[17] Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer Inst 2006;1204–14.
[18] McCann J, Stockton D, Godward S. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. Breast Cancer Res 2002;4(5):R11 [Epub.2002.Jul.17. 2002].
[19] Groenendijk RP, Kochen MP, van Engelenburg KC, Boetes C, Strobbe LJ, Ruers TJ, et al. Detection of breast cancer after biopsy for false-positive screening mammography. An increased risk? Eur J Surg Oncol 2001;17–20.
[20] Peeters PH, Mravunac M, Hendriks JH, Verbeek AL, Holland R, Vooijs PG. Breast cancer risk for women with a false positive screening test. Br J Cancer 1988;211–2.
[21] Ascunce N, Salas D, Zubizarreta R, Almazan R, Ibanez J, Ederra M. Cancer screening in Spain. Ann Oncol 2010;21(May (Suppl. 3)):iii43–51.
[22] Ascunce N, Ederra M, Delfrade J, Baroja A, Erdozain N, Zubizarreta R, et al. Impact of intermediate mammography assessment on the likelihood of false-positive results in breast cancer screening programmes. Eur Radiol 2012; 331–40.
[23] 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. p. 325–56.
[24] 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. p. 357–467.
[25] Ciatto S, Houssami N, Ambrogetti D, Bonardi R, Collini G, Del Turco MR. Minority report – false negative breast assessment in women recalled for suspicious screening mammography: imaging and pathological features, and associated delay in diagnosis. Breast Cancer Res Treat 2007;37–43.
[26] Warren R, Allgood P, Hunnam G, Godward S, Duffy S. An audit of assessment procedures in women who develop breast cancer after a negative result. J Med Screen 2004;180–6.
[27] Utzon-Frank N, Vejborg I, Euler-Chelpin M, Lynge E. Balancing sensitivity and specificity: sixteen year's of experience from the mammography screening programme in Copenhagen, Denmark. Cancer Epidemiol 2011;393–8.
[28] Kabat GC, Jones JG, Olson N, Negassa A, Duggan C, Ginsberg M, et al. A multicenter prospective cohort study of benign breast disease and risk of subsequent breast cancer. Cancer Causes Control 2010;821–8.
[29] Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. Ann Intern Med 2011;10–20.

5. Artículo 4

Título: Trends in detection of invasive cancer and ductal carcinoma in situ at biennial screening mammography in Spain: A retrospective cohort study

Autores: Román M, Rué M, Sala M, Ascunce N, Baré M, Baroja A, De la Vega M, Galcerán J, Natal C, Salas D, Sánchez-Jacob M, Zubizarreta R, Castells X, Cumulative False Positive Risk Group.

Revista: PLoS ONE. 8(12): e83121.

Factor de impacto: 3.5 (Q1 Multidisciplinary Sciences)

DOI: 10.1371/journal.pone.0083121.

Abstract:

Background: Breast cancer incidence has decreased in the last decade, while the incidence of ductal carcinoma in situ (DCIS) has increased substantially in the western world. The phenomenon has been attributed to the widespread adaption of screening mammography. The aim of the study was to evaluate the temporal trends in the rates of screen detected invasive cancers and DCIS, and to compare the observed trends with respect to hormone replacement therapy (HRT) use along the same study period.

Methods: Retrospective cohort study of 1,564,080 women aged 45-69 years who underwent 4,705,681 screening mammograms from 1992 to 2006. Age-adjusted rates of screen detected invasive cancer, DCIS, and HRT use were calculated for first and subsequent screenings. Poisson regression was used to evaluate the existence of a change-point in trend, and to estimate the adjusted trends in screen detected invasive breast cancer and DCIS over the study period.

Results: The rates of screen detected invasive cancer per 100.000 screened women were 394.0 at first screening, and 229.9 at subsequent screen. The rates of screen detected DCIS per 100.000 screened women were 66.8 at first screen and 43.9 at subsequent screens. No evidence of a change point in trend in the rates of DCIS and invasive cancers over the study period were found. Screen detected DCIS increased at a steady 2.5% per year (95% CI: 1.3; 3.8), while invasive cancers were stable.

Conclusion: Despite the observed decrease in breast cancer incidence in the population, the rates of screen detected invasive cancer remained stable during the study period. The proportion of

DCIS among screen detected breast malignancies increased from 13% to 17% throughout the study period. The rates of screen detected invasive cancer and DCIS were independent of the decreasing trend in HRT use observed among screened women after 2002.

OPEN ACCESS Freely available online

PLOS ONE

Trends in Detection of Invasive Cancer and Ductal Carcinoma In Situ at Biennial Screening Mammography in Spain: A Retrospective Cohort Study

Marta Román[1,2], Montse Rué[2,3], Maria Sala[1,2,4], Nieves Ascunce[5], Marisa Baré[6], Araceli Baroja[7], Mariola De la Vega[8], Jaume Galcerán[9], Carmen Natal[10], Dolores Salas[11], Mercedes Sánchez-Jacob[12], Raquel Zubizarreta[13], Xavier Castells[1,2,4]*, the Cumulative False Positive Risk Group[¶]

1 Department of Epidemiology and Evaluation, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain, **2** Network for Research into Healthcare in Chronic Diseases (REDISECC), Madrid, Spain, **3** Basic Medical Sciences Department, Biomedical Research Institut of Lleida (IRBLLEIDA)-University of Lleida, Lleida, Spain, **4** Universitat Autònoma de Barcelona, Barcelona, Spain, **5** Navarra Breast Cancer Screening Programme, Public Health Institute, Pamplona, Navarra, Spain, **6** Oficina Tècnica de Cribratge, Corporació Sanitaria Parc Taulí-Institut Universitari Parc Taulí-UAB, Sabadell, Barcelona, Spain, **7** La Rioja Breast Cancer Screening Programme, Fundacion Rioja Salud, Logroño, La Rioja, Spain, **8** Dirección General de Programas Asistenciales, Consejería de Sanidad, Servicio Canario de Salud, Tenerife, Santa Cruz de Tenerife, Spain, **9** Foundation Society for Cancer Research and Prevention, Pere Virgili Health Research Institute, Reus, Tarragona, Spain, **10** Program & Analysis Unit, Health Office, Oviedo, Principado de Asturias, Spain, **11** General Directorate Public Health & Centre for Public Health Research, Valencia, Comunidad Valenciana, Spain, **12** Servicio de Promoción de la Salud y Programas Preventivos, Consejería de Sanidad, Valladolid, Castilla y León, Spain, **13** Galician Breast Cancer Screening Programme, Public Health & Planning Directorate, Health Office, Santiago de Compostela, La Coruña, Spain

Abstract

Background: Breast cancer incidence has decreased in the last decade, while the incidence of ductal carcinoma in situ (DCIS) has increased substantially in the western world. The phenomenon has been attributed to the widespread adaption of screening mammography. The aim of the study was to evaluate the temporal trends in the rates of screen detected invasive cancers and DCIS, and to compare the observed trends with respect to hormone replacement therapy (HRT) use along the same study period.

Methods: Retrospective cohort study of 1,564,080 women aged 45–69 years who underwent 4,705,681 screening mammograms from 1992 to 2006. Age-adjusted rates of screen detected invasive cancer, DCIS, and HRT use were calculated for first and subsequent screenings. Poisson regression was used to evaluate the existence of a change-point in trend, and to estimate the adjusted trends in screen detected invasive breast cancer and DCIS over the study period.

Results: The rates of screen detected invasive cancer per 100.000 screened women were 394.0 at first screening, and 229.9 at subsequent screen. The rates of screen detected DCIS per 100.000 screened women were 66.8 at first screen and 43.9 at subsequent screens. No evidence of a change point in trend in the rates of DCIS and invasive cancers over the study period were found. Screen detected DCIS increased at a steady 2.5% per year (95% CI: 1.3; 3.8), while invasive cancers were stable.

Conclusion: Despite the observed decrease in breast cancer incidence in the population, the rates of screen detected invasive cancer remained stable during the study period. The proportion of DCIS among screen detected breast malignancies increased from 13% to 17% throughout the study period. The rates of screen detected invasive cancer and DCIS were independent of the decreasing trend in HRT use observed among screened women after 2002.

Citation: Román M, Rué M, Sala M, Ascunce N, Baré M, et al. (2013) Trends in Detection of Invasive Cancer and Ductal Carcinoma In Situ at Biennial Screening Mammography in Spain: A Retrospective Cohort Study. PLoS ONE 8(12): e83121. doi:10.1371/journal.pone.0083121

Editor: Keitaro Matsuo, Kyushu University, Japan

Received June 10, 2013; **Accepted** October 31, 2013; **Published** December 23, 2013

Funding: This work was supported by grants from the Instituto de Salud Carlos III-FEDER (PI11/01296 and PI09/90251). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: xcastells@parcdesalutmar.cat

¶ Membership of the Cumulative False Positive Risk Group is provided in the Acknowledgments.

Introduction

Breast cancer is the most frequent tumour in women worldwide, and its incidence rates had risen steadily worldwide over these past decades [1]. However, since the early 2000's a downturn in its incidence rates have been reported in several developed countries [2–10]. The downturn has also been observed in Spain, more remarkably in women on the 45–69 age range [11,12]. The phenomenon has been attributed to the widespread adaption of screening mammography once screening saturation was nearly achieved [11–13], as well as to the reduction in the use of hormone replacement therapy (HRT) among post menopausal women after the publication of the results of the Women's Health Initiative trial in 2002 [14]. The prevalence of HRT use in Spain has always

been low compared to other countries [11,15–17]. Furthermore, the decline in breast cancer incidence due to the reduction in HRT use has not been studied in Spain.

Different trends have been observed in the incidence of invasive cancer compared to ductal carcinoma in situ (DCIS). While the incidence of invasive cancer has declined in the last decade, the incidence of DCIS of the breast has increased in several countries [18–22]. DCIS have substantially increased in the proportion of breast malignancies detected. The increase has been attributed to the implementation of breast cancer screening [21,23]. It is estimated that DCIS represents 20% of screen detected breast malignancies [21,24].

The availability of individual level data from a cohort of screened women in Spain, followed during 15 years provides the opportunity to analyze the screen detected rates of invasive breast cancer and DCIS over time. We wanted to evaluate the temporal trends in the rates of screen detected invasive cancers and DCIS, and to study the temporal trends with respect to the HRT use along the same study period.

Methods

Ethics Statement

The study was approved by the Mar Teaching Hospital Research Ethics Committee. The data was analyzed anonymously and therefore no additional informed consent was required.

Setting

The National Health System in Spain provides universal health coverage, including early detection of breast carcinoma. All women residing in Spain aged 50 to 69 years are actively invited to participate in population-based screening, with screening intervals every 2 years. However, some regions start inviting women at 45 years. Population-based breast cancer screening in Spain started in one region in 1990 and was implemented nationwide in 2005. Breast cancer screening in Spain follows the European Guidelines for Quality Assurance in Mammographic Screening [25] and its results meet the required standards [26]. Data from eight regions of Spain that perform population-based breast cancer screening were collected. The participating regions covered 44% of the Spanish target population for breast cancer screening in 2006. The participating women are provided with a unique personal identification number. Information about attendance, screening outcome, and diagnostic work-up was registered at an individual level in each screening region data base with the unique personal identification number.

Study Population and Data Collection

Information was collected from 1,564,080 women aged 45 to 69 years of age who had undergone at least one biennial screening examination between 1992 and December 2006. The women underwent 4,705,681 screening examinations during the study period. Due to the small sample size, information on screening examinations performed in 1990 and 1991 was not used for the study.

At the time of each screening examination information is routinely collected related to the mammographic interpretation, whether or not the woman was recalled for additional evaluation to rule out or confirm malignancy, and the specific additional evaluations performed, if any. Additional evaluation for breast cancer assessment included additional mammography, magnetic resonance imaging, ultrasonography, fine-needle aspiration cytology, core-needle biopsy and open surgical biopsy. The diagnostic work-up for additional evaluation was carried out within a maximum of 2 months after screening. A definitive diagnosis of breast cancer was always histopathologically confirmed. Information on histopathology classification was routinely collected in the screening regions for screen detected cancers using the ICD-10 classification codes. A case was considered as screen detected if the diagnosis was made on the basis of a screening examination with subsequent diagnosis work-up procedures. Cases were classified as DCIS or invasive breast cancer.

In addition, information on HRT use was obtained through a questionnaire administered face-to face by a trained health professional at each screening visit immediately before the screening examination. Women were considered to be users of hormone replacement therapy at a screening examination if they reported to be current users or to have used hormone therapy in the sixth months previous to that visit.

Statistical Analysis

Age-adjusted rates of screen detected invasive cancer and DCIS were calculated for first and subsequent screenings and 3-year period (1992–1994, 1995–1997, 1998–2000, 2001–2003, 2004–2006). The Age-adjusted rates of invasive breast cancer, DCIS, and HRT use among the screened women, were calculated for each calendar year. Age-specific incidence rates of invasive cancer and DCIS by 5 year age groups were computed standardized by first or subsequent screen. All age-standardizations were done using the direct method and the European standard population in 5-year age groups as reference.

Poisson regression analyses were used to estimate the trends in screen detected rates of invasive breast cancer and DCIS observed in the study population over the 15 year period. Calendar year, screening region, 5-year age groups and first/subsequent screen were used as explanatory variables. The estimated annual percentage change (APC) and 95% confidence intervals were obtained from the regression models. The APC was equal to 100 (e^m -1), where m is the coefficient of the variable of calendar year. Independent models were computed to evaluate the breast cancer trends of DCIS and invasive cancer separately, and to ascertain possible differences in the APC for first and subsequent screens.

In addition, changes in age- and region-adjusted detection rates of DCIS and invasive cancer over the study period were evaluated using transition change-point models [11,27]. These models assume a Poisson distribution for the number of cases in each stratum and afford a statistical test for the existence of a change-point in the overall trend, and where this is the case, estimate the year in which the change-point is located and the APC before and after the change point. Overall significance level was set at P-value$<$0.05. Statistical analyses were performed using SAS 9.1 (SAS Institute, Cary, NC, USA) and R (R Foundation for Statistical Computing, Vienna, Austria).

Results

Twenty nine percent of women were first screened at 45 to 49 years of age, and 30% at age 50 to 54 years (table 1). The crude number of screen detected cancers per 1,000 screening examinations increased with age, with and overall crude number of 2.73 per 1,000 screening examinations (table 1).

A total of 16,309 screen detected cancers were diagnosed in the 1992–2006 period analyzed. Of these cancers 78.8% (n = 12,851) were invasive cancers, 14.6% (n = 2,379) were DCIS, and 6.6% (n = 1,079) were unknown. At first screen 6,845 cancers were detected (14.6% DCIS, 76.7% invasive and 8.8% unknown) and 9,464 at subsequent screens (14.6% DCIS, 80.3% invasive and 5.1% unknown). Mean (standard deviation) age at detection of

Table 1. Number of women screened, screening examinations, screen detected ductal carcinoma in situ (DCIS) and invasive breast cancers by 5-year age groups.

	Women screened[a]	Screening examinations[b]	Screen detected DCIS[c]	Screen detected invasive cancer[c]
Age	**n**	**n**	**n (‰)**	**n (‰)**
45–49	464,434	764,069	421 (0.55)	1649 (2.16)
50–54	477,084	1,219,228	571 (0.47)	2831 (2.32)
55–59	300,245	1,191,627	572 (0.48)	3076 (2.58)
60–64	260,264	1,084,986	554 (0.51)	3600 (3.32)
65–69	62,053	445,771	261 (0.59)	1695 (3.80)
Overall	1,564,080	4,705,681	2379 (0.51)	12,851 (2.73)

[a]Number of women with that given age at first screening examination.
[b]Number of screening examinations performed in women at that given age.
[c]‰ calculated as number of cases per 1000 screening examinations.
doi:10.1371/journal.pone.0083121.t001

DCIS was 56.7 (6.39) and for invasive cancers was 57.8 (6.23) (P value$<$0.001).

Overall age-adjusted screen detected cancer rates were higher at first screening compared with subsequent screens for both, DCIS and invasive cancer (table 2). The screen detected rates of DCIS increased by 3-year period for first and subsequent screens. The highest screen detected rate of invasive cancer was observed in the 1998–2000 period for first screen, and in the 1992–1994 period for subsequent screens (table 2).

The overall age-specific rates of invasive cancer per 100,000 women-years increased with age. It was 215.8 for women aged 45–49 years, 232 at 50–54 years, 258 at 55–59 years, 332 at 60–64 years, and 380 at 65–69 years. The overall age-specific rates of DCIS per 100,000 women-years was 55 for women aged 45–49 years, 47 at 50–54 years, 48 at 55–59 years, 51 at 60–64 years, and 59 at 65–69 years.

After adjustment for age, screening region, and first or subsequent screen, the Poisson regression showed an absence of trend over the period studied for invasive cancers (p-value = 0.29), with a non-significant increase of 0.3% per year (APC 0.3, 95% CI: −0.2; 0.8), and a statistically significant increase of DCIS of 2.5% per year (APC 2.5, 95% CI: 1.3; 3.8). Figures 1a and 1b show the overall trends for first and subsequent screens for both, DCIS and invasive cancers. Fig. 1a shows that the incidence of screen detected invasive cancer was stable for first and subsequent screens with no significant trends over the period studied (p-value = 0.12 and 0.15 respectively for first and subsequent screens). As Fig. 1b depicts, the incidence of screen detected DCIS steadily increased along the study period for both, first and subsequent screens. The detection rates for DCIS increased by 2.9% per year for the first screen and 2.6% for the subsequent screens. There was no evidence of a change point in trend in the rates of DCIS and invasive cancers over the 17 year period studied (p-value for the existence of a change point = 0.3 for invasive cancer and p-value = 0.7 for DCIS).

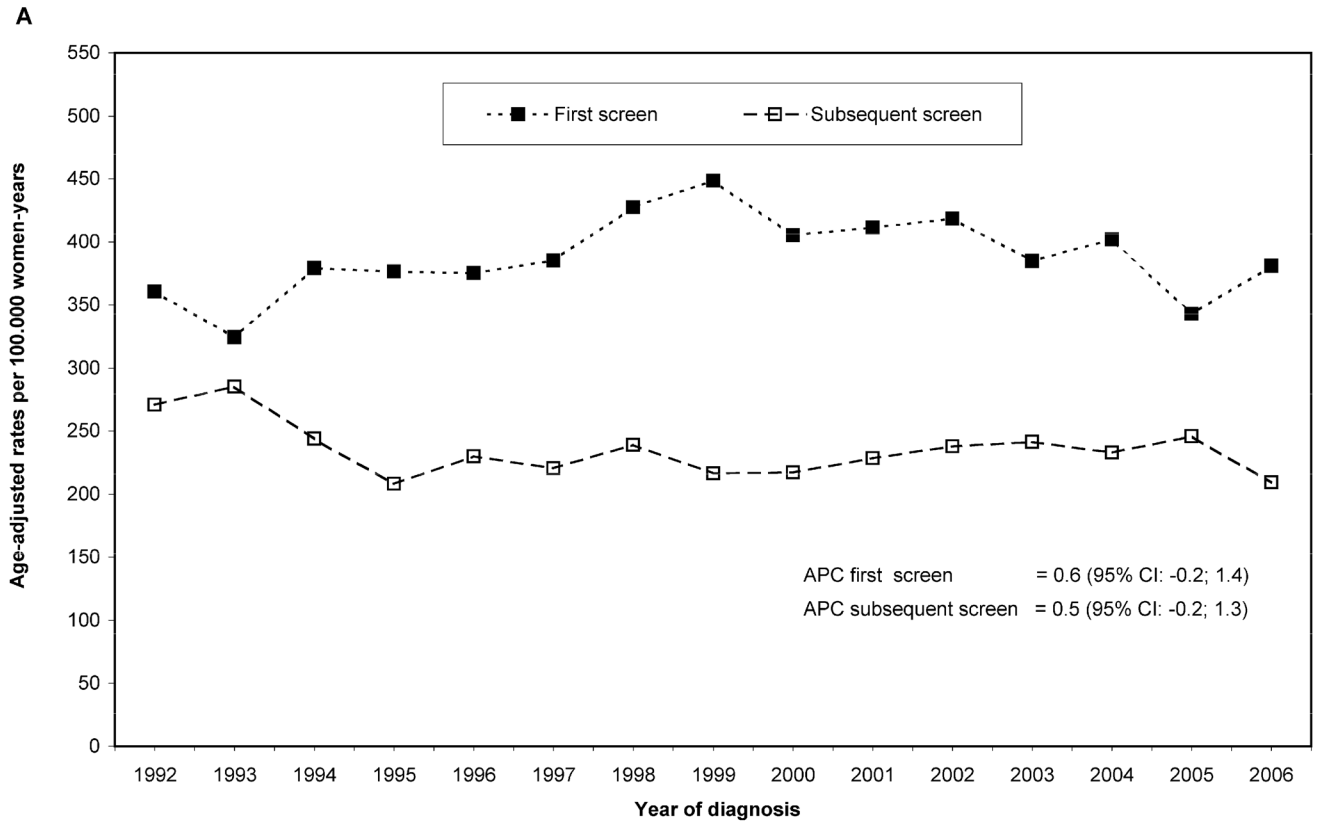
Table 3 shows time trends for the rates of screen detected DCIS and invasive cancer by 5-year age groups over the study period. The P values refer to the evaluation of the existence of a change-point in the overall trend. Estimates for the APC and 95% CI were obtained from the Poisson regression model for each 5-year age group, adjusted for screening region and first or subsequent screen. There was no evidence of a change point in trend among any of the 5-year age groups, for neither invasive cancers nor DCIS. No significant APC was found for any 5-year age group over the study period for screen detected invasive cancers. The APC of screen detected DCIS showed a significant increase over the study period for the 45–49, 50–54, and 55–59 years age groups (table 3).

We presented data on HRT use by the screened women in our data set, obtained from the administered questionnaire at the time of screening examination. Information on HRT use was available in 69.3% of screening examinations. The percentage of missing information on HRT use was stable in the study period. An

Table 2. Age-adjusted incidence rates of screen detected ductal carcinoma in situ (DCIS) and invasive breast cancer (per 100,000 European standard population) by period and first or subsequent screen.

	First screen				Subsequent screen			
	DCIS		Invasive breast cancer		DCIS		Invasive breast cancer	
Period (3-years)	**No. of cases**	**Rate**	**No. of cases**	**Rate**	**No. of cases**	**Rate**	**No. of cases**	**Rate**
1992–1994	116	60.5	681	357.8	24	39.7	162	266.4
1995–1997	185	65.4	1,037	378.8	127	43.7	666	221.2
1998–2000	257	64.7	1,544	426.1	236	40.7	1,365	222.9
2001–2003	235	69.9	1,140	409.0	414	43.6	2,377	236.6
2004–2006	203	70.3	846	372.3	582	45.8	3,033	228.7
Overall	996	66.8	5,248	394.0	1,383	43.9	7,603	229.9

doi:10.1371/journal.pone.0083121.t002


A
First screen
Subsequent screen
Age-adjusted rates per 100.000 women-years
APC first screen = 0.6 (95% CI: -0.2; 1.4)
APC subsequent screen = 0.5 (95% CI: -0.2; 1.3)
Year of diagnosis

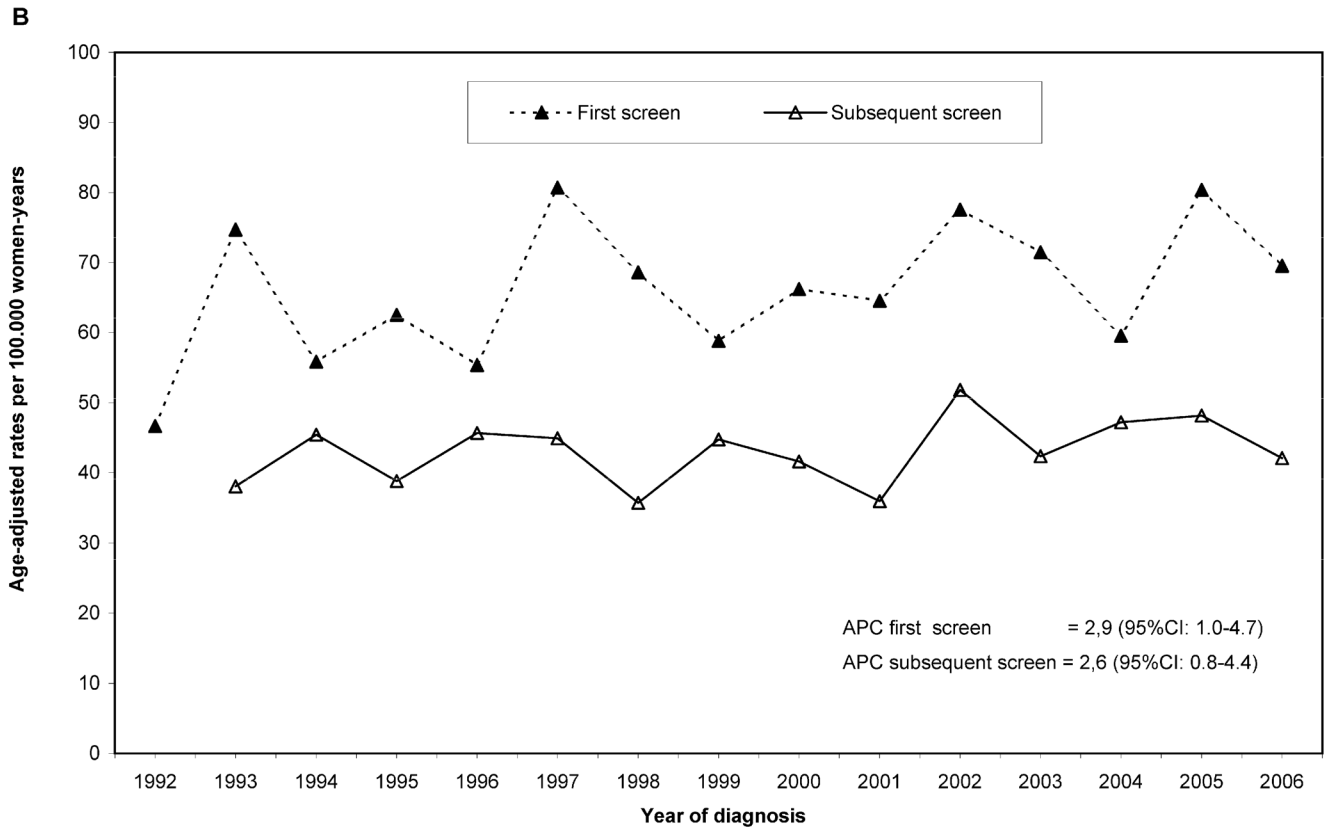

B
First screen
Subsequent screen
Age-adjusted rates per 100.000 women-years
APC first screen = 2,9 (95%CI: 1.0-4.7)
APC subsequent screen = 2,6 (95%CI: 0.8-4.4)
Year of diagnosis

Figure 1. Age-adjusted rates of screen detected breast cancer for first and subsequent screens, in the 1992–2006 period. Rates are given per 100,000 women-years and are standardized using the European standard population in 5-year age groups as reference. The annual percentage change (APC) and its 95% confidence intervals (95%CI) were estimated from the Poisson regression model adjusted by age and screening region. **A**) Age-adjusted rates of invasive cancer for first and subsequent screens, and estimated APC and 95%CI. **B**) Age-adjusted rates of screen detected ductal carcinoma in situ (DCIS) for first and subsequent screens, and estimated APC and 95%CI.
doi:10.1371/journal.pone.0083121.g001

increase in the prevalence of HRT use was observed from 1992 up to 2003. HRT use from 1992 to 1996 was relatively low with a prevalence of 2,749 HRT users per 100,000 women-years in 1996 (fig. 2). A large increase in HRT use was observed from 1997 to 2003 when the prevalence level peaked (13,303 per 100,000 women-years). A decrease was observed after 2003, with a prevalence of HRT users of 9,344 per 100,000 women-years in 2006. A stable incidence of screen detected invasive cancer in the study period is shown in figure 2, independently of the HRT use among screened women. Similarly, the steady increase of 2.5% per year in the incidence of screen detected DCIS showed to be independent of the HRT use among the screened women.

Discussion

Our results showed a steady increase of screen detected DCIS in Spain in the 1992–2006 period studied. The steady increase was observed for first and subsequent screens, and it was more markedly observed in screened women in the younger age groups. Despite the observed downturn in the population incidence of invasive cancer in women on the 45–69 age range the incidence of screen detected invasive cancers showed an absence of trend in the study period. The observed rates of screen detected DCIS and invasive cancer showed to be independent of HRT use among screened women.

The absence of trend in screen detected invasive cancers is in accordance with a previous study by Nederend et al. that reported an absence of trend in the rates of screen detected advanced cancers during a 12 year period [28]. With respect to DCIS, previous studies have shown an increase in the detection rates of DCIS. Van Steenbergen et al. found a ten-fold increase in the detection rate of DCIS between 1991 and 2000 in southern Netherlands, and a two-fold increase was found by Barchielli et al. in Italy [18,22]. The widespread adaption of screening mammography has often been used to explain the increase in the incidence of DCIS in the general population found in several studies [18,21,22,29,30]. However, our study is targeted exclusively to screening participants and our findings should be interpreted in the screening setting. A reason for the increase in screen detected DCIS could be the changes in the techniques and interpretation of screening mammograms over time, as well as the changes in the pathological classification of pre-malignant breast lesions. Population-based screening in Europe follows the recommendations of the European guidelines [25], but programs have progressively improved their quality indicators and efficiency over the years. On the other hand, the introduction of digital mammography has increased the sensitivity of screening mammography, more markedly in the detection of DCIS [31–34]. However, less than 1.5% of screening test were performed with digital mammography in this study.

The steady increase in screen detected DCIS over the study period while the screen detected invasive cancers remained stable has caused that DCIS have substantially increased in the proportion of breast malignancies detected in screening mammography. The proportion has increased from 13% in 1994 to 17% in 2006. The observed proportion of screen detected DCIS among all malignancies observed in the last part of the period (17%) was similar to what has been reported (18%) in other European countries [21].

The rates of screen detected invasive cancer by 5-years age groups showed no trend after adjustment for screening region and first or subsequent screen. The absence of trend in the 5-years age groups reinforced the idea of a steady, stable detection rate of invasive cancers along the study period. On the other hand, the rates of DCIS by 5-years age groups showed a statistically significant increase for the 45–49, 50–54, and 55–59 years age groups. The estimated increase in the rates of DCIS in the three youngest age groups showed a decreasing gradient with age that

Table 3. Trends in rates of screen detected ductal carcinoma in situ (DCIS) and invasive breast cancer in the 1992–2006 period by 5-year age groups.

	DCIS			Invasive breast cancer		
	Change-point p-value[a]	Annual percentage change[b]		Change-point p-value[a]	Annual percentage change[b]	
Age		Overall	95% CI		Overall	95% CI
45–49	0.48	3.9	1.2; 6.5[c]	0.22	0.5	−0.8; 1.8
50–54	1.00	3.0	0.4; 5.6[c]	0.28	−0.3	−1.5; 0.8
55–59	1.00	2.8	0.1; 5.6[c]	0.25	1.1	−0.1; 2.3
60–64	1.00	0.8	−1.9; 3.6	0.99	0.6	−0.4; 1.7
65–69	1.00	1.7	−2.2; 5.7	1.00	0.6	−0.9; 2.2
Overall	0.65	2.5	1.3; 3.8[c]	0.29	0.3	−0.2; 0.8

[a]P-value for the existence of a change point in trend obtained from the Poisson transition change-point model adjusted by screening region. Analyses performed for each 5-year age group and for the overall.
[b]Annual percentage change and 95% confidence intervals (95%CI) obtained from the Poisson regression model adjusted by screening region and participation status (first or subsequent screen). Analyses performed for each 5-year age group and for the overall.
[c]Significant trend at the 95% CI.
doi:10.1371/journal.pone.0083121.t003

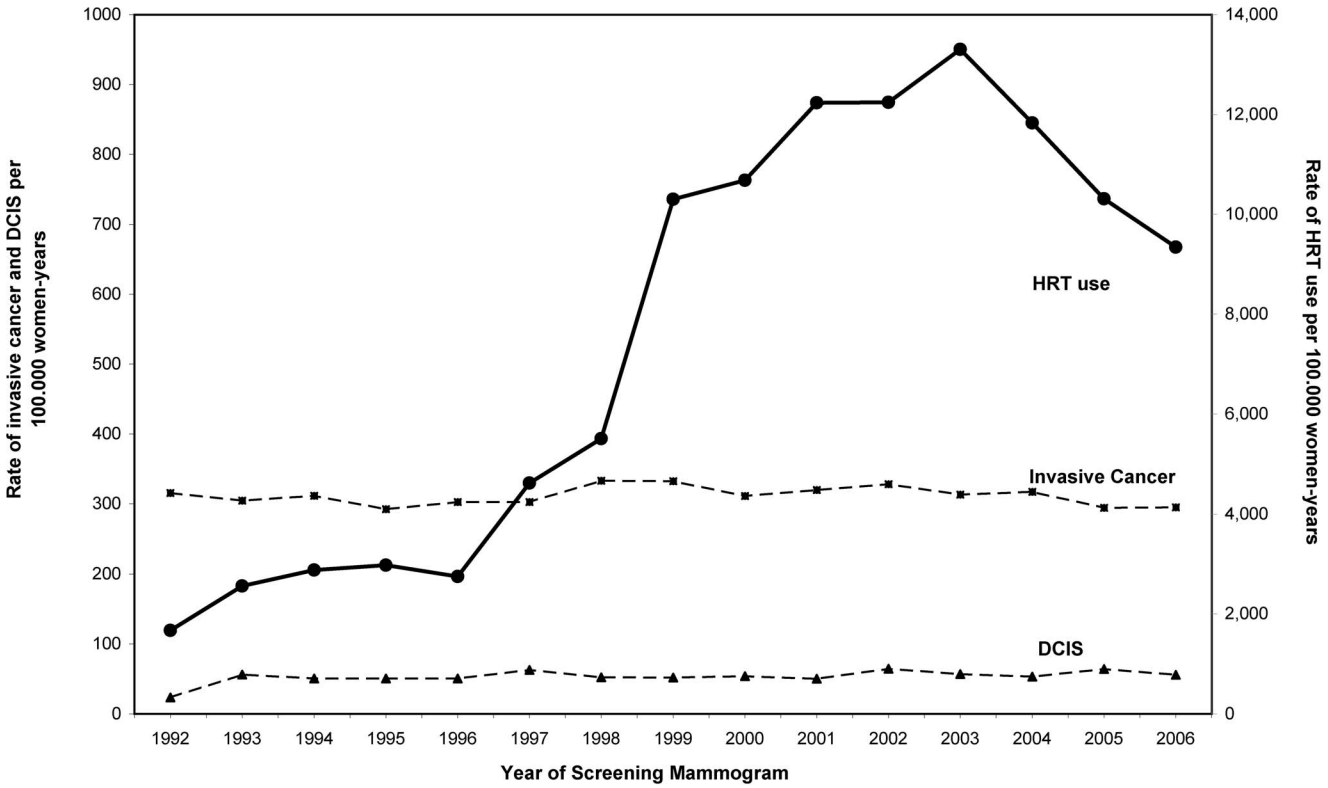


Figure 2. Rates of screen detected invasive cancer, screen detected ductal carcinoma in situ (DCIS), and hormone replacement therapy (HRT) use among screened women, in the 1992–2006 period. Rates are given per 100,000 women-years and are standardized to the age and first or subsequent screen using the European standard population in 5-year age groups as reference.
doi:10.1371/journal.pone.0083121.g002

ranged from 3.9% in the 45–49 years age group to 2.8% in the 55–59 years age group, and was not statistically significant for the 60–64, and 65–69 years age groups. Previous studies have also shown a highest proportion of DCIS among younger women [35].

The observed rates of screen detected DCIS and invasive cancer appeared to be independent of HRT use among screened women. An absence on change points in the overall trends was observed in all the analyses performed: screen detected DCIS and invasive breast cancer, first and subsequent screens, and 5-years age groups. If HRT use has had an effect in the screen detected rates of DCIS or invasive cancer we would expect to find a change point in the overall trends. The change in trend would be strongly expected after the year 2002 when the Women's Health Initiative trial was published [14], causing a reduction in the use of HRT among post menopausal women [17]. Figure 2 shows a reduction in the use of HRT starting in 2002, while the screen detected rates of DCIS and invasive cancer remain steady over the study period. Nevertheless, the time lag between the observed decreasing trend in HRT use and its impact in breast cancer incidence may be long. A reduction in screen detected breast cancer incidence may be observed in a longer term outside the end of our study period in 2006. However, we studied a four year offset from the reduction in the use of HRT in 2002 to the end of the study period in 2006. Several developed countries have reported data on population breast cancer incidence associated with a decrease in the use of HRT in shorter study periods, ranging from 2 to 5 years of offset [4–10]. On the other hand, longer duration of HRT use is known to increase women's breast cancer risk. However, an increased breast cancer risk is consistent for all estrogen plus progestin HRT users. The increased risk remains5-years or more after stop of HRT use [36]. In our study women were considered to be HRT users if they reported to be current users or to have used hormone therapy in the sixth months previous to the screening examination. The definition used ensures that HRT users had been users in a recent period (<6 months) avoiding misclassification of past users as current users.

We found that first and subsequent screens had similar trend patterns for both, invasive cancers and DCIS. By presenting the data for first and subsequent screens separately we avoided a potential confounding factor when analyzing long-term data for screen detected cancers. Higher screen detection rates were observed at first screens compared to subsequent screens, which was expected. However, the proportion of first and subsequent screens changes over time, with more first screenings performed as screening programmes are implemented during the study period, and in younger women who are first time invited. Not taking into consideration the participation when studying incidence trends in mammography screening may cause empirical estimators to be biased and confounded.

The widespread adaption of screening mammography once screening saturation was nearly achieve has been used to justify the observed downturn in the population incidence rates of invasive breast cancer reported since the early 2000 in women on the 45–69 age range [11,12]. During the 1990s screening programs were implemented in the corresponding populations, and screening mammography was widespread adapted. Most programs achieved full coverage of the target populations during the late 1990s and early 2000's [12,26,37]. The steady, stable detection rate of invasive cancers along the study period found in this study does not support the downturn in the incidence rates of invasive breast

cancer observed in the population since the early 2000 [4,5,7,8,10]. On the other hand, our findings could help to explain the increase in the population incidence of DCIS found in several studies [18–22]. The proportion of women in the population undergoing routine screening mammography will influence population-based estimates of breast cancer incidence [6]. The observed steady increase of DCIS in the proportion of screen detected breast malignancies from 13% to 17% is expected to influence the population incidence of DCIS. Previous studies have reported that over 67% of Spanish women in the 45–69 age range perform screening mammography in a publicly founded screening programme [26].

If the natural progression of invasive breast cancer is via DCIS, the detection of DCIS would help to prevent the development of breast carcinomas and consequently reduce breast cancer mortality [38]. However, the increasing number of screen detected DCIS, while the number of invasive cancers remains stable may present a clinical challenge if it implies an increase in the number of women overdiagnosed and overtreated [39].

Some limitations must be considered when interpreting our findings. Firstly, we did not have individual level data on non-participating women in the target population as we received anonymized data of screened women only from the participating regions. The attendance rate among invited women is reported to be 67% [26], and the re-attendance rate among participating women to be 91% [26]. The reported attendance and re-attendance rates are not dissimilar to other well established population-based screening programs in Europe [40]. It would have been desirable to have information on breast cancer risk factors among non-participating women. A previous study on usage of screening mammography previous to initiating a population-based breast cancer screening program in Spain showed that utilization of mammography was higher among younger women, women who had a higher education level, a family history of breast cancer, personal history of benign breast lesion, or had previous visits to a physician [41]. In addition, a substantial proportion of women in the 45–69 age range undergo opportunistic screening outside a screening programme [42]. Thus, the interpretations of the results in this study are related to detection in population-based screening, and its implication in the general population incidence should be carefully reviewed. However, a not dissimilar trend in screen detected DCIS and invasive cancer would be expected over time for population-based and opportunistic screening, as the changes in the interpretation of screening mammograms have occurred simultaneously. Besides, 6.6% of screen detected cancers in our study could not be classified as DCIS or invasive breast cancer because the histology classification was not available. The proportion of unknown histology of screen detected cancers decreased over time, as the screening programmes' databases achieved completeness and the established quality indicators were met. There were 9.5% unknown histology cancers cases in 1992, 6.2% in 1999, and 3.3% in 2006. To check whether the reduction in unknown histology cancer cases could have an effect in the observed increase in screen detected DCIS we performed a sensitivity analysis excluding the two screening regions with a highest proportion of unknown histology cancer cases at the beginning of the study period. No significant differences were observed compared to the analysis including all regions, therefore all cases were included in the analysis.

Conclusions

We studied the trends in screen detected DCIS and invasive breast cancer over a 15 year period, and found that the studied rates were independent of HRT use among screened women. Despite the observed downturn in the population incidence of invasive cancers, the screen detected rates of invasive cancers remained steady, stable over the study period, while the screen detected rates of DCIS steadily increased, causing an increase of DCIS in the proportion of screen detected breast malignancies. The increasing trend of screen detected DCIS was associated to younger ages, particularly women aged 45–60 years. The study provides substantial information to improve the knowledge about the impact of screening programmes over time. These results are particularly useful when the benefits and harms of screening mammography are evaluated in the long-term.

Acknowledgments

***The Cumulative False Positive Risk Group (alphabetical order)**: Department of Epidemiology and Evaluation, Mar, Barcelona: Jordi Blanch, Xavier Castells, Marta Román, Anabel Romero, Maria Sala. 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, Iosu Delfrade, María Ederra, Nieves Erdozain, Juana Vidán. 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. Foundation Society for Cancer Research and Prevention. Pere Virgili Health Research Institute, Reus, Tarragona: Jaume Galceran. Health Office, Asturias: Carmen Natal. La Rioja Breast Cancer Screening Programme. Fundacion Rioja Salud, Logroño: Araceli Baroja. Jefa de la Sección de Promoción y Protección de la Salud. Cancer Screening and Epidemiology Department, UDIAT-CD. Corporació Parc Taulí-Institut Universitari Parc Taulı (UAB), Sabadell Marisa Baré. Castilla-Leon Breast Cancer Screening Programme. Dirección General de Salud Pública ID e I. SACYL, Castilla y León: Isabel González-Román.*

Author Contributions

Conceived and designed the experiments: M. Román M. Rué MS XC NA MB AB MV JG CN DS MSJ RZ. Analyzed the data: M. Román M. Rue. Wrote the paper: M. Román M. Rué MS XC. Reviewed the manuscript: M. Román M. Rué MS XC NA MB AB MV JG CN DS MSJ RZ.

References

1. Ferlay J, Shin HR, Bray F (2010) Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10. Available: http://www.iarc.fr/en/publications/eresources/cancerbases/index.php. Accessed 13 November 2013.
2. Allemand H, Seradour B, Weill A, Ricordeau P (2008) Decline in breast cancer incidence in 2005 and 2006 in France: a paradoxical trend. Bull Cancer 95: 11–15.
3. Canfell K, Banks E, Moa AM, Beral V (2008) Decrease in breast cancer incidence following a rapid fall in use of hormone replacement therapy in Australia. Med J Aust 188: 641–644.
4. Glass AG, Lacey JV, Carreon JD, Hoover RN (2007) Breast cancer incidence, 1980–2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. J Natl Cancer Inst 99: 1152–1161.
5. Katalinic A, and Rawal R. (2008) Decline in breast cancer incidence after decrease in utilisation of hormone replacement therapy. Breast Cancer Res Treat 107: 427–430.
6. Kerlikowske K, Miglioretti DL, Buist DS, Walker R, Carney PA (2007) Declines in invasive breast cancer and use of postmenopausal hormone therapy in a screening mammography population. J Natl Cancer Inst 99: 1335–1339.
7. Kliewer EV, Demers AA, Nugent ZJ (2007) A decline in breast-cancer incidence. N Engl J Med 357: 509–510.
8. Kumle M (2008) Declining breast cancer incidence and decreased HRT use. Lancet 372: 608–610.
9. Ponti A, Rosso S, Zanetti R, Ricceri F, Tomatis M, et al. (2007) Breast cancer incidence, 1980–2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. J Natl Cancer Inst 99: 1817–1818.

10. Ravdin PM, Cronin KA, Howlader N, Berg CD, Chlebowski RT, et al. (2007) The decrease in breast-cancer incidence in 2003 in the United States. N Engl J Med 356: 1670–1674.
11. Pollan M, Pastor-Barriuso R, Ardanaz E, Arguelles M, Martos C, et al. (2009) Recent changes in breast cancer incidence in Spain, 1980–2004. J Natl Cancer Inst 101: 1584–1591
12. Pollan M, Michelena MJ, Ardanaz E, Izquierdo A, Sanchez-Perez MJ, et al. (2010) Breast cancer incidence in Spain before, during and after the implementation of screening programmes. Ann Oncol 21(Suppl 3): 97–102.
13. Li CI, Daling JR (2007) Changes in breast cancer incidence rates in the United States by histologic subtype and race/ethnicity, 1995 to 2004. Cancer Epidemiol Biomarkers Prev 16: 2773–2780.
14. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, et al. (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA 288: 321–333.
15. Benet M, Carvajal A, Garcia Del Pozo J, Álvarez A, Vega T (2002) [Hormonal replacement therapy in Spain] Tratamiento Hormonal Sustitutivo En Espana. Un Estudio De Utilizacion. Med Clin (Barc) 119: 4–8.
16. Schneider HP (1997) Cross-national study of women's use of hormone replacement therapy (HRT) in Europe. Int J Fertil Womens Med 42(Suppl 2): 365–375.
17. Barbaglia G, Macia F, Comas M, Sala M, del Mar V, et al. (2009) Trends in hormone therapy use before and after publication of the Women's Health Initiative trial: 10 years of follow-up. Menopause 16: 1061–1064.
18. Barchielli A, Federico M, De Lisi V, Bucchi L, Ferretti S, et al. (2005) In situ breast cancer: incidence trend and organised screening programmes in Italy. Eur J Cancer 41: 1045–1050.
19. Kricker A, Goumas C, Armstrong B (2004) Ductal carcinoma in situ of the breast, a population-based study of epidemiology and pathology. Br J Cancer 90: 1382–1385.
20. Li CI, Daling JR, Malone KE (2005) Age-specific incidence rates of in situ breast carcinomas by histologic type, 1980 to 2001. Cancer Epidemiol Biomarkers Prev 14: 1008–1011.
21. Sorum R, Hofvind S, Skaane P, Haldorsen T (2010) Trends in incidence of ductal carcinoma in situ: the effect of a population-based screening programme. Breast 19:499–505.
22. van Steenbergen LN, Voogd AC, Roukema JA, Louwman WJ, Duijm LE, et al. (2009) Screening caused rising incidence rates of ductal carcinoma in situ of the breast. Breast Cancer Res Treat 115: 181–183.
23. Puig-Vives M, Pollan M, Rue M, Osca-Gelis G, Saez M, et al. (2012) Rapid increase in incidence of breast ductal carcinoma in situ in Girona, Spain 1983–2007. Breast 21: 646–651.
24. Ernster VL, Ballard-Barbash R, Barlow WE, Zheng Y, Weaver DL, et al. (2002) Detection of ductal carcinoma in situ in women undergoing screening mammography. J Natl Cancer Inst 94: 1546–1554.
25. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L (2006) European guidelines for quality assurance in breast cancer screening and diagnosis. Luxenburg, Office for Official Publications of the European Communities.
26. Ascunce N, Salas D, Zubizarreta R, Almazan R, Ibanez J, et al. (2010) Cancer screening in Spain. Ann Oncol 21(Suppl 3): 43–51.
27. Pastor-Barriuso R, Guallar E, Coresh J (2003) Transition models for change-point estimation in logistic regression. Stat Med 22: 1141–1162.
28. Nederend J, Duijm LE, Voogd AC, Groenewoud JH, Jansen FH, et al. (2012) Trends in incidence and detection of advanced breast cancer at biennial screening mammography in The Netherlands: a population based study. Breast Cancer Res 14: R10.
29. Ernster VL, Barclay J, Kerlikowske K, Grady D, Henderson C (1996) Incidence of and treatment for ductal carcinoma in situ of the breast. JAMA 275: 913–918.
30. Virnig BA, Tuttle TM, Shamliyan T, Kane RL (2010) Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes. J Natl Cancer Inst 102:170–178.
31. De Gelder R, Fracheboud J, Heijnsdijk EA, Den Heeten G, Verbeek AL, et al. (2011) Digital mammography screening: weighing reduced mortality against increased overdiagnosis. Prev Med 53: 134–140.
32. Domingo L, Romero A, Belvis F, Sanchez M, Ferrer J, et al. (2011) Differences in radiological patterns, tumour characteristics and diagnostic precision between digital mammography and screen-film mammography in four breast cancer screening programmes in Spain. Eur Radiol 21: 2020–2028.
33. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, et al. (2005) Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 353: 1773–1783.
34. Bluekens AM, Holland R, Karssemeijer N, Broeders MJ, den Heeten GJ (2012) Comparison of digital screening mammography and screen-film mammography in the early detection of clinically relevant cancers: a multicenter study. Radiology 265: 707–714.
35. Evans WP, Starr AL, Bennos ES (1997) Comparison of the relative incidence of impalpable invasive breast carcinoma and ductal carcinoma in situ in cancers detected in patients older and younger than 50 years of age. Radiology 204: 489–491.
36. Chlebowski RT, Manson JE, Anderson GL, Cauley JA, Aragaki AK, et al. (2013) Estrogen plus progestin and breast cancer incidence and mortality in the Women's Health Initiative Observational Study. J Natl Cancer Inst 105:526–535.
37. Castells X, Sala M, Ascunce N, Salas D, Zubizarreta R, Casamitjana M (2007) Descripción del cribado del cáncer en España. Proyecto DESCRIC. Agència d'Avaluació de Tecnologia i Recerca Mèdiques de Cataluña. Informes de Evaluación de Tecnologías Sanitarias, AATRM.
38. Gunsoy NB, Garcia-Closas M, Moss SM (2012) Modelling the overdiagnosis of breast cancer due to mammography screening in women aged 40–49 in the United Kingdom. Breast Cancer Res 14: R152.
39. Paci E, Miccinesi G, Puliti D, Baldazzi P, De Lisi V, et al. (2006) Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening study in Italy. Breast Cancer Res 8: R68.
40. Broeders MJ, Scharpantgen A, Ascunce N, Gairard B, Olsen AH, et al. (2005) Comparison of early performance indicators for screening projects within the European Breast Cancer Network: 1989–2000. Eur J Cancer Prev 14: 107–116
41. Segura JM, Castells X, Casamitjana M, Macià F, Ferrer F (2000) Utilization of screening mammography as a preventive practice prior to initiating a population-based breast cancer screening program. J Clin Epidemiol 53: 595–603
42. Ministerio de Sanidad y Consumo and I.N.E (2006) Encuesta Nacional de Salud de España 2006. http://www.msc.es/estadEstudios/estadisticas/encuesta Nacional/encuesta2006.htm. Accessed 13 November 2013.

6. Artículo 5

Título: The cumulative risk of false-positive results in the Norwegian Breast Cancer Screening Program: Updated results

Autores: Román M, Hubbard RA, Sebuødegård S, Miglioretti DL, Castells X, Hofvind S.

Revista: Cancer. 2013; 119: 3952-3958.

Factor de impacto: 4.9 (Q1 Oncology)

DOI: 10.1002/cncr.28320

Abstract:

Background: Some false-positive results are inevitable in mammographic screening, but the impact of false-positive findings on the program and the participants is a disadvantage of screening. The objective of the current study was to estimate the cumulative risk of a false-positive result over 10 biennial screening examinations and the cumulative risk of undergoing an invasive procedure with a benign outcome in women screened between the ages of 50 years to 69 years.

Methods: A retrospective cohort study was performed in 231,310 women aged 50 years to 51 years at the time of first mammography screening who underwent 715,311 screening mammograms in the Norwegian Breast Cancer Screening Program from 1996 through 2010. Generalized linear mixed models were used to estimate the probability of a false-positive screening result and to compute the cumulative false-positive risk for up to 10 biennial screening examinations.

Results: The cumulative false-positive risk after 20 years of biennial screening for women who initiated screening aged 50 years to 51 years was 20.0% (95% confidence interval [95% CI], 19.7%-20.4%). The cumulative risk of undergoing an invasive procedure with a benign outcome for the same group of women was 4.1% (95% CI, 3.9%-4.3%). The cumulative risk of undergoing a fine-needle aspiration cytology, core needle biopsy, or open biopsy with a benign outcome was 1.4% (95% CI, 1.3%-1.5%), 2.0% (95% CI, 1.9%-2.1%), and 0.16% (95% CI, 0.13%-0.19%), respectively.

Conclusions: One in every 5 women will be recalled for further assessment with a negative outcome if they attend biennial mammographic screening between ages 50 years to 69 years. The

risk of an invasive procedure with a benign outcome is approximately 4%. It is important to communicate the existence and extent of this risk to the target group and to reduce to a minimum the waiting times between screening and further assessment.

Original Article

The Cumulative Risk of False-Positive Results in the Norwegian Breast Cancer Screening Program: Updated Results

Marta Roman, MSc[1,2]; Rebecca A. Hubbard, PhD[3,4]; Sofie Sebuodegard, BSc[5]; Diana L. Miglioretti, PhD[3,6]; Xavier Castells, MD, PhD[1,2]; and Solveig Hofvind, PhD[5,7]

BACKGROUND: Some false-positive results are inevitable in mammographic screening, but the impact of false-positive findings on the program and the participants is a disadvantage of screening. The objective of the current study was to estimate the cumulative risk of a false-positive result over 10 biennial screening examinations and the cumulative risk of undergoing an invasive procedure with a benign outcome in women screened between the ages of 50 years to 69 years. **METHODS:** A retrospective cohort study was performed in 231,310 women aged 50 years to 51 years at the time of first mammography screening who underwent 715,311 screening mammograms in the Norwegian Breast Cancer Screening Program from 1996 through 2010. Generalized linear mixed models were used to estimate the probability of a false-positive screening result and to compute the cumulative false-positive risk for up to 10 biennial screening examinations. **RESULTS:** The cumulative false-positive risk after 20 years of biennial screening for women who initiated screening aged 50 years to 51 years was 20.0% (95% confidence interval [95% CI], 19.7%-20.4%). The cumulative risk of undergoing an invasive procedure with a benign outcome for the same group of women was 4.1% (95% CI, 3.9%-4.3%). The cumulative risk of undergoing a fine-needle aspiration cytology, core needle biopsy, or open biopsy with a benign outcome was 1.4% (95% CI, 1.3%-1.5%), 2.0% (95% CI, 1.9%-2.1%), and 0.16% (95% CI, 0.13%-0.19%), respectively. **CONCLUSIONS:** One in every 5 women will be recalled for further assessment with a negative outcome if they attend biennial mammographic screening between ages 50 years to 69 years. The risk of an invasive procedure with a benign outcome is approximately 4%. It is important to communicate the existence and extent of this risk to the target group and to reduce to a minimum the waiting times between screening and further assessment. ***Cancer* 2013;119:3952-8**. *© 2013 American Cancer Society.*

KEYWORDS: breast neoplasms, mass screening, mammography, false-positive results, female.

INTRODUCTION

False-positive screening results are a concern in mammographic screening, although a certain rate is inevitable and must be accepted for adequate cancer detection. The negative effects of false-positive results have been widely noted and include the psychological harm of being recalled for further assessment, particularly in women who undergo a biopsy.[1] Furthermore, a false-positive screening result entails extra economic costs to the screening program[2,3] and may lead to decreased participation in future screenings.[4,5]

Most screening programs in Europe invite women aged 50 years to 69 years to mammographic screening every 2 years. A recent study based on results from European screening programs demonstrated an average recall rate of 4% at screening rounds after the first screen (range, 1%-11%).[6] The cumulative risk of a false-positive screening result is defined as the risk of experiencing at least 1 false-positive recall if a woman is screened biennially from ages 50 years to 69 years. In the European study, the pooled estimate of the cumulative risk of a false-positive recall after 10 rounds of screening was 20% and the cumulative risk of an invasive procedure with a benign outcome was 3%.[6] The recall rate and the cumulative risk of a false-positive screening result are reported to be substantially higher in the United States, ranging from 13% to 16% at first screen and 8% to 10% at subsequent screens. The cumulative risk of a false-positive result after 10 years of annual screening in the United States ranges from 42% to 61% for a recall and from 4.8% to 18.6% for a biopsy recommendation.[7-10]

Corresponding author: Solveig Hofvind, PhD, Cancer Registry of Norway, Oslo and Akershus University College of Applied Sciences, Majorstua 0403, Oslo, Norway; Fax: (011) 47 22 45 13 70; Solveig.Hofvind@kreftregisteret.no

[1]Department of Epidemiology and Evaluation, Hospital del Mar Medical Research Institute, Barcelona, Spain; [2]Network for Research into Healthcare in Chronic Diseases, Madrid, Spain; [3]Biostatistics Unit, Group Health Research Institute, Seattle, Washington; [4]Department of Biostatistics, University of Washington, Seattle, Washington; [5]Department of Research, Cancer Registry of Norway, Oslo, Norway; [6]Division of Biostatistics, Department of Public Health Sciences, University of California at Davis, Davis, California; [7]Faculty of Health, Oslo and Akershus University College of Applied Sciences, Oslo, Norway.

DOI: 10.1002/cncr.28320, **Received:** July 10, 2013; **Revised:** July 13, 2013; **Accepted:** July 19, 2013, **Published online** August 20, 2013 in Wiley Online Library (wileyonlinelibrary.com)

The first European study estimating the cumulative risk of a false-positive screening result used information from the first 3 screening rounds in only 4 of the 19 counties of Norway from 1996 through 2002.[11] The estimates were based on direct probability calculations and did not include adjustment for any factors such as the calendar year or the variability among the counties. In addition, the estimates assumed that each screening result was independent of prior screening results. The study was one of 3 included in the recently published review of false-positive screening results in European screening programs,[6] which identified only 2 prior studies estimating the cumulative risk of an invasive procedure with a benign outcome and only 1 study that had adjusted for confounding factors.

The availability of longer follow-up time, data from all 19 counties, and more appropriate estimation methods underscore the need for an update of the estimates of the cumulative risk of a false-positive screening result in the Norwegian Breast Cancer Screening Program (NBCSP). The goal of the current study was to update the estimates of the cumulative risk of a false-positive screening result (including additional assessment with mammography, ultrasound, and/or an invasive procedure) and the risk of a recall for further assessment including an invasive procedure (fine-needle aspiration cytology [FNAC], core needle biopsy [CNB], or open biopsy [OB]) with a benign outcome, using 15 years of individual-level data collected as a part of the NBCSP.

MATERIALS AND METHODS

The study population included all women with at least one screening examination performed in the NBCSP during the study period (1996-2010). The program invites women aged 50 years to 69 years to 2-view mammography every second year and is administered according to the European guidelines.[12] The screening program started as a pilot in 4 counties in 1996 and became nationwide in 2005.[13] The women are identified by a unique personal identification number given to all inhabitants of Norway. Information regarding attendance, screening outcome, and diagnostic workup was registered in the central nationwide database, with the personal identification number used as the unique identifier for each woman. We received an anonymized file with individual-level dates of invitations and attendances on all women targeted in the screening program. No ethical committee approval was necessary because we received anonymized data only.

The NBCSP uses independent double reading. An interpretation score ranging from 1 to 5 is given for both breasts and from both readers. A score of 1 indicates a negative screening examination whereas a score of 5 indicates a finding that is highly suspicious for malignancy. All mammograms with an interpretation score of ≥ 2 by one or both readers are discussed at a consensus/arbitration meeting to decide whether to recall the patient. Interpretation of the screening mammograms and the diagnostic workup take place at centralized breast clinics at university or county hospitals. Additional mammograms and ultrasound (noninvasive methods) are used to evaluate abnormal mammograms. If these methods are insufficient to rule out cancer, then an invasive procedure such as FNAC, CNB, or OB is performed. The diagnostic workup takes place 1 to 4 weeks after screening. If no malignancy is found, women are referred back to routine screening. Women who receive a diagnosis of breast cancer are referred for treatment. All malignancies (invasive carcinomas and ductal carcinoma in situ) are histologically verified.

Any recall for further assessments was considered a false-positive screening result if breast cancer was not diagnosed during the diagnostic workup (within 4 months), regardless of the procedures performed. We defined a false-positive screening result for a benign invasive procedure as any diagnostic workup including an FNAC, a CNB, or an OB with benign morphology. OB was defined as a diagnostic procedure including excision, incision, and marker biopsy. Women recalled due to insufficient technical quality or self-declared symptoms ($< 0.5\%$ for both together) were not included either in the nominator or denominator in the estimates of false-positive screening tests. We considered a woman as an irregular attendee if she missed her last screening invitation but attended after ≥ 4 years. Otherwise, she was considered a regular attendee.

Statistical Analysis

Our estimates are based on all screening examinations performed on women aged 50 years to 51 years at the time of first screening in the 19 Norwegian counties. The women contributed data from the time of their first invitation until the end of follow-up (December 31, 2010). Data regarding up to a maximum of 6 screening examinations performed during the study period were used for estimation. Screening examinations for the seventh and eighth screenings were not used because they represented $< 3\%$ of the overall screening examinations, and therefore the estimates of false-positive risk for these screening rounds were imprecise. The probability and 95% confidence intervals (95% CI) for the risk of a false-positive screening result at each screening examination were estimated using

TABLE 1. Number of Women Screened and Percentages of Women Recalled for Further Assessment With a Negative Outcome by Screening Round and 3-Year Time Period in the Norwegian Breast Cancer Screening Program, 1996 Through 2010

	First Screening		Second Screening		Third Screening		Fourth Screening		Fifth Screening		Sixth Screening	
3-Year Period	No. Screened	FP, % (95% CI)	No. Screened	FP, % (95% CI)	No. Screened	FP, % (95% CI)	No. Screened	FP, % (95%CI)	No. Screened	FP, % (95% CI)	No. Screened	FP, % (95% CI)
1996-1998	23,768	5.1 (4.8-5.4)	7557	2.9 (2.5-3.2)	1	—						
1999-2001	30,082	4.9 (4.6-5.1)	20,371	2.4 (2.2-2.6)	12,843	2.3 (2.1-2.6)	40	—				
2002-2004	57,517	5.3 (5.1-5.5)	38,175	2.5 (2.3-2.6)	22,507	2.2 (2.0-2.4)	18,682	2.3 (2.1-2.5)	6471	2.0 (1.7-2.4)		
2005-2007	58,009	6.0 (5.8-6.2)	53,023	2.6 (2.5-2.7)	41,283	2.3 (2.2-2.4)	24,460	2.1 (1.9-2.3)	16,663	2.2 (1.9-2.4)	10,623	2.3 (2.0-2.6)
2008-2010	61,934	6.9 (6.7-7.1)	57,881	2.5 (2.4-2.6)	54,505	2.0 (1.9-2.1)	47,384	1.8 (1.7-2.0)	32,078	1.7 (1.6-1.9)	19,454	1.8 (1.6-2.0)
Overall	231,310	5.8 (5.7-5.9)	177,007	2.5 (2.5-2.6)	131,139	2.2 (2.1-2.3)	90,566	2.0 (1.9-2.1)	55,212	1.9 (1.8-2.0)	30,077	2.0 (1.8-2.1)

Abbreviations: 95% CI, 95% confidence interval; FP, false positive screening results.

generalized linear mixed models. The regression model included adjustment for year of the screening examination, taking the last year (2010) as the reference category, and a random intercept for county to allow for variation across counties in false-positive risk. Women were included in analyses only up to the time of their first false-positive result. The probability of a false-positive result at the i^{th} examination (π_i) was expressed as $\ln(\pi_i/1-\pi_i) = \alpha_i D_i + \beta_1 X_i + \delta$, in which D_i is a vector of binary indicators denoting participating in the i^{th} screening round. D_i is equal to 1 if the woman participated in the i^{th} screening examination and equals 0 otherwise. X_i is a mammogram-level covariate indicating the year in which the screening examination was performed. δ is a county-specific random effect to account for the correlation among screening tests performed in the same county. We reported the results for the county using the median false-positive risk. The models are described in detail by Singer and Willett.[14,15]

Separate models were computed to estimate the probability of a false-positive screening result, the probability of any invasive procedure with a benign outcome, and the probability of a benign invasive procedure involving an FNAC, CNB, or OB, independently. We tested whether irregular attendees had a higher false-positive risk than regularly screened women by incorporating "irregular attendance" as an additional covariate in our regression model. The point estimates to calculate the cumulative risks of a false-positive screening result were performed assuming that the probability of experiencing a false-positive result in the 7th to the 10th screening examination was equal to that of the 6th examination. The cumulative risk of a false-positive result for each round up to the 10th screening examination was calculated by multiplying the probability of receiving a first false-positive test result at each round by the probability of receiving no false-positive test results at any previous round. Standard errors for the calculation of the 95% CIs for the cumulative risk probability were estimated using the Greenwood approximation.[14] This approximation is based on the estimated probabilities and the observed sample size in the current study population. Standard errors based on the Greenwood formula will be inflated relative to true standard errors. To assess the possibility of dependent censoring, we conducted a sensitivity analysis in which cumulative false-positive risk was also estimated, conditional on the number of screening examinations a woman was observed to receive.[16] Statistical significance was defined using a 2-sided α level of .05. Model parameters were estimated via residual pseudo-likelihood using the GLIMMIX procedure in SAS statistical software, version 9.1 (SAS Institute Inc, Cary, NC).

RESULTS

We analyzed information from 231,310 women aged 50 years to 51 years at the time of the initial screening examination in the NBCSP, contributing 715,311 screening examinations. A second screening examination was performed in 177,007 women (76.5%), 131,139 women (56.7%) underwent a third screening examination, and 30,077 women (13.0%) had a sixth screening examination (Table 1).

The percentage of women with a false-positive screening result was higher at the time of the initial compared with subsequent screening examinations (Table 1). The overall crude false-positive rates decreased from 5.8% (95%CI, 5.7%-5.9%) at initial screening to 2.5% (95% CI, 2.5%-2.6%) at the second screening (Table 1). The highest crude false-positive rate, 6.9% (95% CI, 6.7%-7.1%), was observed in women receiving their first screening mammogram between 2008 and 2010 (Table 1). The overall crude rates of a benign invasive procedure

TABLE 2. Number of Women Screened and Percentages of Women Recalled for Further Assessment Including An Invasive Procedure With a Benign Outcome by Screening Round and 3-Year Time Period in the Norwegian Breast Cancer Screening Program, 1996 Through 2010

3-Year Period	First Screening		Second Screening		Third Screening		Fourth Screening		Fifth Screening		Sixth Screening	
	No. Screened	Benign Invasive Procedure,[a] % (95% CI)	No. Screened	Benign Invasive Procedure, % (95% CI)	No. Screened	Benign Invasive Procedure, % (95% CI)	No. Screened	Benign Invasive Procedure, % (95% CI)	No. Screened	Benign Invasive Procedure, % (95% CI)	No. Screened	Benign Invasive Procedure, % (95% CI)
1996-1998	23,768	1.7 (1.6-1.9)	7557	0.6 (0.4-0.8)	1	—						
1999-2001	30,082	1.4 (1.3-1.6)	20,371	0.5 (0.4-0.6)	12,843	0.6 (0.5-0.7)	40	—				
2002-2004	57,517	1.4 (1.3-1.5)	38,175	0.5 (0.5-0.6)	22,507	0.6 (0.5-0.7)	18,682	0.6 (0.5-0.7)	6471	0.6 (0.4-0.7)		
2005-2007	58,009	1.7 (1.6-1.9)	53,023	0.6 (0.5-0.6)	41,283	0.6 (0.5-0.6)	24,460	0.5 (0.4-0.6)	16,663	0.5 (0.4-0.6)	10,623	0.6 (0.5-0.7)
2008-2010	61,934	1.9 (1.8-2.0)	57,881	0.5 (0.4-0.5)	54,505	0.4 (0.4-0.5)	47,384	0.3 (0.3-0.4)	32,078	0.4 (0.4-0.5)	19,454	0.4 (0.3-0.5)
Overall	231,310	1.7 (1.6-1.7)	177,007	0.5 (0.5-0.6)	131,139	0.5 (0.5-0.6)	90,566	0.4 (0.4-0.5)	55,212	0.5 (0.4-0.5)	30,077	0.5 (0.4-0.5)

Abbreviation: 95% CI, 95% confidence interval.
[a] Benign invasive procedure indicates a fine-needle aspiration cytology, core needle biopsy, or open biopsy with a benign outcome.

decreased from 1.7% (95% CI, 1.6%-1.7%) at initial screening to 0.5% (95% CI, 0.5%-0.6%) at the second screening (Table 2). The highest crude rate of a benign invasive procedure, 1.9% (95% CI, 1.8%-2.0%), was observed in women receiving their first screening mammogram between 2008 and 2010.

The estimated cumulative risk at 10 screening examinations for the cohort of women who initiated screening at ages 50 to 51 years was 20.0% (95% CI, 19.7%-20.4%) (Fig. 1). The cumulative risk of undergoing an invasive procedure with a benign outcome at 10 screening examinations for the same group of women was 4.1% (95% CI, 3.9%-4.3%).

A total of 6063 screened women (2.6%) underwent an invasive procedure with a benign outcome. FNAC constituted 2862 of the benign invasive procedures performed (47.2%), CNB represented 2498 (41.2%), and OB represented 703 of the benign invasive procedures performed (11.6%). The estimated cumulative risk of undergoing an FNAC, CNB, or OB with a benign outcome after 10 screening examinations for women initiating screening at ages 50 years to 51 years was 1.4% (95% CI, 1.3%-1.5%), 2.0% (95% CI, 1.9%-2.1%), and 0.16% (95% CI, 0.13%-0.19%), respectively.

We found that irregular screening attendees had a higher false-positive risk of a false-positive screening result (odds ratio, 1.12; 95% CI, 1.06-1.20), and a nonstatistically significantly higher risk of an invasive procedure with a benign outcome (odds ratio, 1.11; 95% CI, 0.98-1.26) compared with regularly screened women.

We evaluated the possible impact of dependent censoring on our cumulative false-positive risk estimates. The cumulative risk projecting the first 6 observed examinations up to 10 screening examinations was 20.5%, whereas the cumulative risk with the dependent censoring model was 19.9%. Furthermore, the cumulative risk of a benign invasive procedure was 5.4% based on the first 6 observed observations, and was 5.2% in the dependent censoring model.

DISCUSSION

We estimated that 1 in every 5 women who participates in the NBCSP will have a false-positive screening result over the course of 10 biennial screening examinations. Furthermore, we found that these women had a cumulative risk of undergoing an invasive procedure with a benign outcome of 4.1%. The results, which are based on nationwide data, confirm the results from a study published in 2004 for false-positive screening results, but are somewhat lower for an invasive procedure (4.1% vs 6.2%).[11]

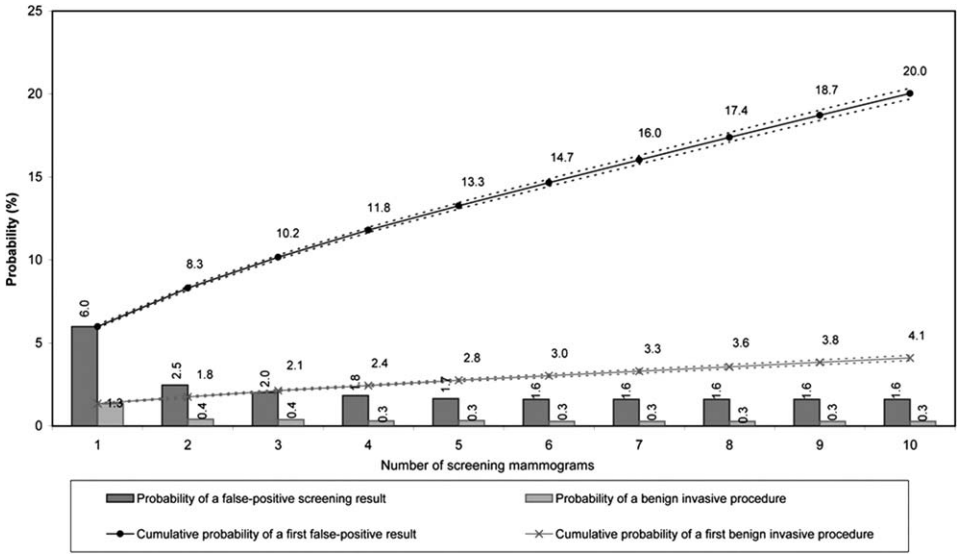


Figure 1. The risk of a false-positive recall for further assessment and risk of a recall for further assessment including an invasive procedure with a benign outcome at each screening round and cumulative risks are shown for women aged 50 years to 51 years at the time of first screening who underwent 1 to 10 biennial screening examinations in the Norwegian Breast Cancer Screening program.

These results are in agreement with other studies performed in European service screening programs based on biennial screening in women aged 50 years to 69 years.[17-19] The risk of a false-positive screening result was estimated to be 20.4% in a study from Spain, which used the same regression models as the current study.[18] For Copenhagen and Fyn, the cumulative risks were estimated to be 15.8% and 8.1%, respectively, in the study by Njor et al,[17] whereas a letter to the editor by Puliti et al gave a cumulative risk of 15.2% after 7 screening rounds in Italy.[19] The recent review by the Euroscreen Working Group, which included 4 countries, demonstrated a pooled estimate of 19.7%.[6] To the best of our knowledge, only 3 studies in Europe have estimated the cumulative risk of undergoing an invasive procedure with a benign outcome. The estimates ranged from 1.8% in Spain to 8.5% in the United Kingdom.[10,18]

False-positive risks estimates from the United States are substantially higher than those from Europe, ranging from 42% to 61% for false-positive results[7-9] and from 4.8% to 18.6% for a false-positive biopsy result after 10 screening examinations.[8-10] The differences have been attributed to the screening setting and practice environment. In Europe, breast cancer screening is population-based and all women aged 50 years to 69 years are invited every second year, whereas opportunistic screening of women aged ≥ 40 years using 1-year to 2-year screening intervals is the most common screening practice in the United States.[10,20-22] Furthermore, the recall rate might be influenced by different reading procedures (independent double reading with consensus in Europe and usually single reading in the United States) or different interpretive volumes, with a recommendation to read at least 5000 screening mammograms per year in Europe[13] compared with the requirement to read at least 960 mammograms every 2 years in the United States.[23] Approximately 40% of the radiologists reading screening mammograms in Norway reach the European volume standard[24] and are specialized in mammography, whereas most mammograms in the United States are read by general radiologists who interpret a wide range of imaging types. In addition, the recommended maximum level of recalls is 3% for subsequent screens under European guidelines[13] while it is 5% to 12% in the United States,[25,26] which may be due to medico-legal consequences in case of missed cancers at screening.[27] Furthermore, because screening in the United States tends to be opportunistic, US women may be more likely to attend screening at multiple facilities than their European counterparts. This could influence the false-positive rate if comparison films are not made available.[28] The underlying incidence cancer rate is the same in Europe (77 per 100,000) and the United States (76 per 100,000) and thus is unlikely to influence the false-positive rate.[29]

The estimated cumulative risk of undergoing an invasive procedure with a benign outcome decreased from

6.2% in the previous study from the Norwegian program[11] to 4.1% in the current study. The difference most likely is due to the performance of fewer FNACs of cysts during the last years compared with the time of the initiation of the program. In addition, in Norway, as in many other countries, CNB has replaced FNAC over the years due to its higher sensitivity for detecting breast cancer. Undergoing an invasive procedure is assumed to have a greater psychological impact than having additional mammography images and/or ultrasound[30]; thus, the rate should be kept as low as possible while maintaining adequate cancer detection. The increased breast cancer risk observed in women with a false-positive recall assessment, even > 6 years after the recall, underscores the importance of a complete assessment for any kind of breast abnormality.[31]

We also found a decrease in the cumulative risk of an OB with a benign outcome from 0.9% in the previous study[11] to 0.2% in the current study. This is likely due to the movement toward performing CNB instead of surgical biopsies as a first invasive diagnostic approach, with women undergoing surgical biopsy only if the CNB result is inconclusive. Surgical biopsies are assumed to have a high positive predictive value, but they also cause psychological stress for the women and more significant scarring than a CNB. A surgical biopsy could be performed for either the diagnosis or treatment of the breast malignancy.

The regression approach we used for estimation is appropriate for studying false-positive screening results, accounting for multiple adjustment variables and changes over time in the absence of dependent censoring. Risk studies from the United States have identified an association between the number of screening examinations and the risk of a false-positive test result.[9,32] Therefore, we computed the cumulative risk of a false-positive screening result, accounting for dependent censoring,[16] and found little difference in the false-positive risks estimates, suggesting that censoring was independent in our setting.

The availability of > 700,000 screening examinations from > 230,000 women aged 50 years to 51 years at the time of first screening and a study period of 15 years provide robust estimates for the cumulative risk of a false-positive screening examination, including recalls for different procedures. However, no women had the possibility of receiving 10 invitations during the study period, which led us to base our estimates on data from 6 rounds instead of 10. The current study is based on data from a population-based screening program with an attendance rate of 77% of the invited women, and in which 84% of eligible women had attended at least 1 screening examination during the study period. Estimating the cumulative probability of a false-positive result after 10 screening examinations is important for quantifying the potential harms of a screening program if a woman receives all recommended screens.

We estimated that approximately 1 in 5 women undergoing biennial mammography screening from ages 50 years to 69 years will have at least 1 false-positive screening result during that 20-year period, and < 5% will undergo an invasive procedure with a benign outcome. False-positive screening results are an unavoidable part of breast cancer screening and some risk of false-positive results must be accepted for adequate cancer detection. Undergoing an invasive procedure with a benign outcome does not mean that the biopsy was unnecessary, because some mammography findings require a biopsy to determine whether they are benign or malignant. The harm of false-positive recalls must be balanced against the goal of maintaining reasonable detection of early-stage cancers. There is a need for further knowledge regarding the recalls of patients with negative outcomes, and how to reduce their associated harms.

FUNDING SUPPORT

No specific funding was disclosed.

CONFLICT OF INTEREST DISCLOSURES

Dr. Hubbard has received grants from the National Institutes of Health and GE Healthcare. Dr. Hofvind and Ms. Sebuodegard are employed at the Cancer Registry of Norway, which administers the Norwegian Breast Cancer Screening Program.

REFERENCES

1. Brett J, Bankhead C, Henderson B, et al. The psychological impact of mammographic screening. A systematic review. *Psychooncology.* 2005;14:917-938.
2. Bull AR, Campbell MJ. Assessment of the psychological impact of a breast screening programme. *Br J Radiol.* 1991;64:510-515.
3. Lerman C, Trock B, Rimer BK, et al. Psychological and behavioral implications of abnormal mammograms. *Ann Intern Med.* 1991;114: 657-661.
4. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. *Ann Intern Med.* 2007;146: 502-510.
5. Roman R, Sala M, De la Vega M, et al. Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening. *Breast Cancer Res Treat.* 2011;130:543-552.
6. Hofvind S, Ponti A, Patnick J, et al. False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. *J Med Screen.* 2012; 19(suppl 1):57-66.
7. Christiansen CL, Wang F, Barton MB, et al. Predicting the cumulative risk of false-positive mammograms. *J Natl Cancer Inst.* 2000;92: 1657-1666.
8. 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.

9. Hubbard RA, Kerlikowske K, Flowers CI, et al. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. *Ann Intern Med.* 2011; 155:481-492.
10. 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.
11. 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.
12. 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.
13. Hofvind S, Geller BM, Skelly J, et al. Sensitivity and specificity of mammographic screening as practised in Vermont and Norway. *Br J Radiol.* 2012;85:e1226-e1232.
14. Singer JD, Willett JB. Fitting basic discrete-time hazard models. In: Applied Longitudinal Data Analysis: Modelling Change and Event Occurrence. New York: Oxford University Press; 2003:357-467.
15. Singer JD, 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.
16. Hubbard RA, Miglioretti DL, Smith RA. Modelling the cumulative risk of a false-positive screening test. *Stat Methods Med Res.* 2010;19: 429-449.
17. 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.
18. Roman R, Sala M, Salas D, et al. Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. *Ann Oncol.* 2012;23:104-111.
19. Puliti D, Miccinesi G, Zappa M. More on screening mammography. *N Engl J Med.* 2011;364:284-285.
20. 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.
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. Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology.* 2002;224:861-869.
24. Hofvind S, Geller BM, Rosenberg RD, et al. Screening-detected breast cancers: discordant independent double reading in a population-based screening program. *Radiology.* 2009;253:652-660.
25. Carney PA, Sickles EA, Monsees BS, et al. Identifying minimally acceptable interpretive performance criteria for screening mammography. *Radiology.* 2010;255:354-361.
26. Dick JF, Gallagher TH, Brenner RJ, et al. Predictors of radiologists' perceived risk of malpractice lawsuits in breast imaging. *AJR Am J Roentgenol.* 2009;192:327-333.
27. Elmore JG, Taplin SH, Barlow WE, et al. Does litigation influence medical practice? The influence of community radiologists' medical malpractice perceptions and experience on screening mammography. *Radiology.* 2005;236:37-46.
28. Thurfjell MG, Vitak B, Azavedo E, et al. Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. *Acta Radiol.* 2000;41:52-56.
29. Globocan 2008. Estimated Breast Cancer Incidence Worldwide in 2008. globocan.iarc.fr/factsheets/cancers/breast.asp. Accessed August 3, 2013.
30. Brett J, Austoker J. Women who are recalled for further investigation for breast screening: psychological consequences 3 years after recall and factors affecting re-attendance. *J Public Health Med.* 2001; 23:292-300.
31. von Euler-Chelpin M, Risor LM, Thorsted BL, et al. Risk of breast cancer after false-positive test results in screening mammography. *J Natl Cancer Inst.* 2012;104:682-689.
32. Xu JL, Fagerstrom RM, Prorok PC, et al. Estimating the cumulative risk of a false-positive test in a repeated screening program. *Biometrics.* 2004;60:651-660.

IV. DISCUSIÓN

1. Resultados principales

Los resultados de los cinco trabajos presentados en esta tesis dan soporte y profundizan en el conocimiento de los resultados falsos positivos, los factores asociados a los mismos y su impacto en las mujeres cribadas. Además, se hace hincapié en la evolución en las tasas de detección de cáncer en el cribado mamográfico, y su relación con el uso de terapia hormonal sustitutiva. Los resultados muestran la amplia variabilidad existente en el riesgo de tener un resultado falso positivo en función de distintas características propias de los programas de cribado y de las mujeres participantes, a la vez que reflejan el impacto que los falsos positivos tienen sobre las mujeres cribadas en términos de adherencia y riesgo de cáncer. Estos resultados ponen de manifiesto la posibilidad de reducir los efectos adversos del cribado mamográfico poblacional, lo que se traduciría en una mejora de la efectividad del mismo.

De manera resumida, los principales hallazgos de este trabajo de investigación son:

1) El riesgo acumulado de falsos positivos se ve fuertemente influenciado por las características del protocolo de lectura mamográfica, qué explican aproximadamente el 50% de la variabilidad observada, y las características personales de la mujer, que explicarían el 50% restante.
2) Los resultados falsos positivos reducen la adherencia de las mujeres en sucesivas convocatorias de cribado.
3) Las mujeres con resultados falsos positivos tienen un mayor riesgo de presentar un cáncer de mama en sucesivas convocatorias de cribado. Este riesgo es aún mayor en las mujeres con falsos positivos con pruebas invasivas.
4) Las tasas de detección de carcinoma ductal in situ ha aumentado de manera gradual y constante desde la puesta en marcha de los programas.
5) Las tasas de detección de cáncer de mama invasivo se han mantenido estables a lo largo del tiempo desde la puesta en marcha de los programas.
6) La disminución en la recomendación de terapia hormonal sustitutiva no tiene un impacto directo en las tasas de detección en el cribado.
7) La estimación del riesgo acumulado de falsos positivos en el programa de Noruega es altamente consistente con las estimaciones realizadas hace una década, y similar en valor absoluto al riesgo acumulado estimado en el contexto español.

2. Discusión conjunta de los artículos

La discusión conjunta de los artículos se articula en función de las preguntas de investigación en las que se centra esta tesis. La discusión específica de cada uno de los artículos se encuentra detallada en los mismos, donde también se encuentran las limitaciones y fortalezas de cada estudio.

Hemos observado que existe una amplia variabilidad en el riesgo acumulado de falsos positivos en función de distintas características del protocolo de lectura mamográfica, como el método de lectura, o el número de proyecciones, y en función de determinadas características de la mujer, como los antecedentes familiares de cáncer de mama, el uso de terapia hormonal sustitutiva, o la edad. El riesgo acumulado de falso positivo por cualquier tipo de prueba varió entre un 7.5% y 51.4%, en función de cuales de estas características estaban presentes. Lo que representa un ratio de variabilidad entre máximo y mínimo de 6.8. Las características del protocolo de lectura mamográfica explicaban más de la mitad de la variabilidad existente en el riesgo acumulado de falsos positivos entre el perfil de alto y bajo riesgo. Mientras que las características de la mujer explicaban el 46% restante. El grupo de alto riesgo corresponde a mujeres del grupo de edad más joven, con pruebas invasivas previas, historia familiar de cáncer de mama, estado pre-menopaúsico, doble lectura de la mamografía y una única proyección. Sin embargo, el riesgo acumulado de falsos positivos con pruebas invasivas varió entre el 12.0% y el 1.6%, lo que equivale a un ratio de variabilidad entre máximo y mínimo de 7.6. Es decir, la variabilidad es mayor que en la estimación del riesgo acumulado por cualquier tipo de prueba. Sin embargo, las características del protocolo de lectura mamográfica únicamente explicaban una cuarta parte de la variabilidad existente en el riesgo acumulado de falsos positivos con pruebas invasivas.

Desde un punto de vista organizativo, sería posible modificar las variables asociadas al protocolo de lectura mamográfica, mientras que las características de la mujer no se pueden modificar (antecedentes familiares, pruebas invasivas previas, etc.), o deben de ser consideradas desde una perspectiva más amplia para considerar su modificación (uso de terapia hormonal sustitutiva). Si queremos intervenir sobre el riesgo de falsos positivos para reducirlos y mejorar la efectividad del cribado, el esfuerzo debe centrarse en intervenir sobre los factores modificables (protocolo de lectura mamográfica).

A la vista de estos resultados surgen dos cuestiones clave para poner en perspectiva el impacto de los resultados falsos positivos en la mejora de la efectividad del cribado mamográfico. Por un lado, aunque el riesgo acumulado de falsos positivos con pruebas invasivas es menor en términos absolutos (rango: 1.6% a 12.0%), la variabilidad entre el perfil de alto y bajo riesgo es mayor que en el riesgo acumulado de falsos positivos por cualquier tipo de prueba (ratio máx-mín: 7.6 vs 6.8). Por otra parte, encontramos que a pesar de la mayor variabilidad existente en el riesgo acumulado de falsos positivos con pruebas invasivas, nuestra capacidad de intervenir para reducirlos es menor, ya que únicamente una cuarta parte de la variabilidad se explica con factores potencialmente modificables. Estos resultados deben de ser tenidos en cuenta si se quiere mejorar la efectividad del cribado mamográfico mediante la reducción de los efectos adversos y en particular de los resultados falsos positivos.

Otra parte sustancial en el conocimiento de los resultados falsos positivos es la evaluación de su impacto sobre las mujeres cribadas. Hemos podido comprobar que los resultados falsos positivos reducen la adherencia al cribado mamográfico en convocatorias sucesivas. Sin embargo, esta reducción de la adherencia fue más pronunciada si el falso positivo se experimentó en el primer cribado o cribados iniciales, mientras que las mujeres que experimentaban un falso positivo en cribados más avanzados (quinto, sexto, o séptimo), experimentaron una reducción de la adherencia mucho menor. Este hallazgo sugiere que las mujeres cribadas desarrollan una cierta fidelidad hacia los programas en los que participan, haciéndolas más tolerantes a posibles efectos adversos, y en particular a un potencial resultado falso positivo. Por otra parte, el aumento de riesgo de cáncer de mama observado en las mujeres con resultados falsos positivos previos identifica a un subgrupo específico de mujeres con una mayor probabilidad de presentar la enfermedad, siendo este riesgo aun estadísticamente significativo cuatro o más años después del resultado falso positivo. Además el riesgo era aún mayor si la mujer tuvo un resultado falso positivo con pruebas invasivas.

Dentro del actual debate sobre la mejora de la efectividad del cribado, se están replanteando las actuales estrategias de *'one size fits all'* por estrategias más personalizadas que permitan maximizar los beneficios en los grupos de mujeres más susceptibles de beneficiarse del cribado[71]. Las mujeres con resultados falsos positivos muestran una menor adherencia al cribado, y a su vez son mujeres con un mayor riesgo de detección de cáncer en convocatorias sucesivas, lo que las convierte en mujeres potencialmente más susceptibles de beneficiarse del cribado. Será

importante monitorizar a este grupo de mujeres para mantener el máximo nivel de participación en cribados sucesivos.

Muchos de los factores de riesgo asociados con los resultados falsos positivos son también factores de riesgo para el cáncer de mama (excepto la edad, que presenta una asociación inversa). Este hecho es de especial relevancia para diseñar posibles estrategias de personalización del cribado. Las mujeres con antecedentes familiares de cáncer de mama, pruebas invasivas previas, o elevada densidad mamaria han sido identificadas por algunos autores como población susceptible para la personalización del cribado, aplicando intervalos de cribado más cortos (cribado anual), comenzando el cribado a edad más temprana (45 años), o utilizando una prueba de cribado distinta a la mamografía (resonancia magnética). De esta manera, sería posible mejorar los beneficios del cribado en este grupo de mujeres, a la vez que sería posible reducir los efectos adversos (falsos positivos) en el grupo de mujeres sin estas características.

Las tendencias temporales en las tasas de detección de cáncer de mama en cribado poblacional están poco evaluadas en el contexto del territorio español. Diversos autores, dentro y fuera del territorio español han argumentado que el descenso en las tasas de incidencia de cáncer de mama invasivo en la población general observado en la última década (especialmente en el grupo de mujeres de 45 a 69 años) es debida, por un lado a la completa implantación de los programas de cribado poblacional que ya estarían cercanos a la saturación del cribado[5,18,86], y por otro, al descenso en la utilización de terapia hormonal sustitutiva tras la publicación del estudio del *Health Initiative Trial* en 2002 [87]. Sin embargo, los hallazgos de nuestro estudio muestran una tasa de detección de cáncer de mama invasivo estable a lo largo del periodo de estudio, que se mantiene después del año 2000. Estos resultados mantienen abierto el debate sobre las posibles causas de esta disminución en las tasas de incidencia poblacional, ya que sería esperable una reducción en las tasas de detección de cáncer de mama invasivo en el cribado para que la incidencia poblacional disminuyese debido a la saturación del cribado, y al mismo tiempo se esperaría una marcada reducción en las tasas de detección a partir del año 2002, en paralelo a la disminución observada en el uso de terapia hormonal.

El aumento gradual en la proporción de carcinoma ductal in situ detectado en el cribado desde la puesta en marcha de los programas puede ser parcialmente atribuible al sobrediagnóstico consecuencia del cribado mamográfico, pero también, a una mejora en la precisión de los

radiólogos lectores debida a la experiencia, y a posibles cambios en la clasificación histológica de las lesiones pre-malignas.

Cabe destacar que el seguimiento secuencial de las diversas participaciones de la mujer en el cribado y su análisis desde una perspectiva longitudinal supone una contribución novedosa en la metodología generalmente utilizada para evaluar el cribado poblacional de cáncer de mama. Este hecho se hace tangible con la utilización de esta metodología por otros autores[88], y con su aplicación en el contexto del programa poblacional de cribado de Noruega que valida y da consistencia a las estimaciones de riesgo acumulado de falsos positivos.

El gran volumen de información de esta base de datos con información de 10 programas de cribado del territorio español la convierte en la mayor base de datos para el estudio de los efectos adversos del cribado evaluada hasta la fecha. Son escasos los ejemplos de otros estudios internacionales que dispongan de bases de datos de estas características para la valoración de los efectos adversos del cribado poblacional. En el contexto de EEUU se puso en marcha en el año 1996 el *Breast Cancer Surveillance Consortium*[89]. Este proyecto aglutina baja un mismo criterio información procedente de diversas fuentes y tiene como finalidad promover la investigación en torno al cribado de cáncer de mama y garantizar una buena calidad de la información. La base de datos del *Breast Cancer Surveillance Consortiu*m no es completa del territorio de EEUU, y presenta deficiencias en algunos aspectos, pero tiene un gran valor en el impulso de la investigación sobre el cribado de cáncer de mama. En ese sentido, estudios como los expuestos en esta tesis, o los del *Breast Cancer Surveillance Consortiu*m son exponentes de un cambio de paradigma en los estudios centrados en la evaluación del cribado poblacional de cáncer de mama, donde es necesario disponer de grandes cohortes longitudinales en amplios marcos poblacionales.

3. Limitaciones

Las principales limitaciones de esta tesis son las propias de un estudio de cohortes retrospectivo. A pesar de que permiten disponer de información de un gran número de mujeres cribadas y amplios periodos temporales, queda limitada a la calidad de la información disponible en las fuentes de datos originales. En los 4 primeros artículos de esta tesis la información proviene de las bases de datos originales de los programas de cribado participantes en el estudio de Riesgo

Acumulado de Falsos Positivos (RAFP). Como se ha comentado anteriormente, se desarrolló un detallado protocolo de definiciones y recogida de variables que garantizaba la homogeneidad de la información recogida. Sin embargo, ciertas variables relacionadas con las características propias de la mujer, como el uso de terapia hormonal sustitutiva, antecedentes familiares, o la presencia de lesiones benignas previas, presentaban un volumen importante de valores desconocidos, que en algunos casos era del 40%. Para valorar el impacto de estas variables, se realizó un riguroso control de calidad sobre la procedencia y el método de información de las mismas, y se realizaron diversos análisis de sensibilidad para valorar el impacto de esta falta de información sobre las principales variables de estudio. Dado que el impacto de la falta de información sobre las variables de estudio, especialmente sobre los falsos positivos, se consideró mínimo o moderado, se realizaron sub-análisis específicos con el grupo de mujeres con información disponible sobre las características de la mujer. Por otro lado, variables inicialmente consideradas para su inclusión en el estudio como la densidad mamaria, no estaban disponibles en un número suficiente de programas, o la información recogida presentaba grandes limitaciones, hecho que no permitió incorporarla en los análisis o estudiar su efecto. Sin embargo, para responder a los objetivos planteados en estos estudios, no se requería específicamente el análisis de estas variables.

Para el tercer estudio sobre el riesgo de detección de cáncer en mujeres con falsos positivos previos, hubiese sido deseable disponer de información más detallada sobre la clasificación de los falsos positivos según la histología de las lesiones benignas diagnosticadas. En el estudio sobre la evolución temporal en las tasas de detección de cáncer de mama en el cribado mamográfico, presenta una limitación en cuanto a la información referente a los tumores, quedando restringida únicamente a la clasificación histológica de la invasividad. Además, la proporción de desconocidos en esta clasificación es del 6.6% sobre el total de tumores detectados en el periodo de estudio. En el caso del estudio noruego, los datos provienen de una única fuente de información de carácter nacional procedente del propio programa de cribado. Lo que garantiza la uniformidad de criterio en las definiciones, y la homogeneidad en la calidad y disponibilidad de información a nivel nacional. Además, los datos del programa de cribado están cruzados con el registro poblacional de cáncer de noruega, que cubre más del 99% de los casos de cáncer de mama diagnosticados en el país[90,91].

4. Fortalezas

El eje central de los artículos de estas tesis está basado en el análisis de la cohorte retrospectiva del estudio de Riesgo Acumulado de Falsos Positivos (RAFP). Este proyecto ha permitido disponer por primera vez en el territorio español de una base de datos conjunta de distintos programas de cribado, y que contiene información de un gran número de mujeres, seguidas de manera secuencial durante sus múltiples participaciones en el cribado. Cómo ya se ha comentado anteriormente, esta base de datos con información de 10 programas de cribado del territorio español es la mayor base de datos con información individualizada creada hasta la fecha para la valoración de los efectos adversos del cribado. Además, se dispone de información de un largo periodo temporal de 16 años, desde la puesta en marcha de los programas en 1991 hasta el año 2006. El intenso trabajo de redacción de protocolos, homogeneización de criterios, y validación de datos, garantiza un alto nivel de consistencia en la información analizada.

Por su parte, y como se comenta en la sección de métodos, la metodología desarrollada para realizar los estudios de esta tesis supone un avance en la evaluación del cribado poblacional desde una perspectiva longitudinal, y aporta una aproximación metodológica consistente, que permite ampliar la perspectiva sobre la valoración de los falsos positivos. Esta metodología, ha sido utilizada con posterioridad por otros autores para la evaluación de los falsos positivos en el contexto de otros países[88].

Todos los trabajos de esta tesis exploran aspectos controvertidos de la evaluación del cribado poblacional y aportan respuestas inéditas que se añaden al conocimiento sobre las prácticas de detección precoz del cáncer de mama, con particular énfasis en los resultados falsos positivos. A pesar de dejar algunos interrogantes abiertos, los artículos de esta tesis profundizan en un aspecto fundamental del cribado de cáncer de mama, como son los resultados falsos positivos, y extienden los resultados con la valoración de las tendencias en las tasas de detección del cribado mamográfico. Otros trabajos realizados con posterioridad o de manera simultánea corroboran o amplían los resultados de estos trabajos, y los toman como referencia, dando crédito a los hallazgos que se presentan[56,88].

Para el desarrollo de este proyecto se contó con la participación directa de personas con responsabilidades directivas y de gestión de los distintos programas participantes. Este hecho permite una mayor traslación de los resultados de la investigación a la práctica, y a la vez posibilita

poder plantear preguntas más relevantes en el contexto de la salud pública y los programas comunitarios.

5. Continuidad y futuras líneas de investigación

Como se ha comentado anteriormente, esta tesis se enmarca en la línea de investigación de cribado de cáncer de mama del Servei d'Epidemiologia i Avaluació de l'Hospital del Mar-IMIM y en la línea de evaluación de las intervenciones sanitarias de la Universitat de Lleida-IRBLleida. En el contexto del CIBER de Epidemiología y Salud Pública, y actualmente de la REDISSEC, se han propuesto y financiado diferentes iniciativas para profundizar en aspectos relacionados con el cribado mamográfico que pretenden dar continuidad a las cuestiones presentadas.

Como consecuencia lógica del estudio sobre riesgo de cáncer de mama en mujeres con resultados falsos positivos previos, se ha puesto en marcha un proyecto para profundizar en el conocimiento de las lesiones benignas y las sospechas radiológicas identificadas en el cribado y su posterior evolución a cáncer. Este proyecto, denominado BELLE Project, se puso en marcha en el 2012. Se trata de un proyecto de carácter estatal, financiado por el Fondo de Investigaciones Sanitarias, donde participan 8 programas de detección precoz del territorio español. He sido la persona responsable de diseñar la estructura de datos, la recogida y validación de información, y la creación de la base de datos de este proyecto. El proyecto BELLE ha dado sus primeros resultados con la reciente publicación de un artículo que muestra como el riesgo de cáncer de mama es superior en mujeres con lesiones benignas previas en comparación con mujeres sin lesiones previas, y como el riesgo de cáncer aumenta con el grado de proliferación de la lesión [92]. Se espera que los resultados de este proyecto aporten información relevante para la adecuación del cribado en función de los perfiles de riesgo de las mujeres participantes.

La experiencia adquirida a través de los proyectos RAFP y BELLE Project, juntamente con las colaboraciones iniciadas con grupos europeos y la necesidad de trabajar en un marco más amplio, se ha formalizado en diversas colaboraciones con grupos de investigación europeos y de EEUU. Por un lado, se ha puesto en marcha un ambicioso proyecto que incluye información completa sobre las mujeres cribadas de los programas de Noruega, Dinamarca, y de las 8 Comunidades Autónomas participantes en el estudio RAFP. La base de datos de este estudio incluye más de 6

millones de mamografías de casi 2 millones de mujeres participantes en el cribado. Esta base de datos ha sido creada y analizada por mí, y supone una nueva dimensión de complejidad metodológica. Existen estudios comparativos de indicadores entre países[93,94], o a partir de datos agregados[56], pero atendiendo a la evidencia disponible, esta es la primera vez que se analizan de manera conjunta datos individualizados procedentes de diferentes países en el contexto de cribado. Por otra parte, actualmente se está desarrollando un análisis colaborativo sobre indicadores de calidad del cribado mamográfico que incluye información de los programas de Noruega, España (proyecto RAFP), y EEUU (Breast Cancer Surveillance Consortium). Este análisis pretende mostrar y discutir la variabilidad en los indicadores de calidad existente entre los distintos programas, especialmente entre el contexto europeo y el de EEUU.

Desde agosto de 2013 desarrollo mi actividad profesional en el contexto del programa de cribado poblacional de Noruega, donde continúo profundizando en aspectos fundamentales de la evaluación de los programas de cribado poblacional. Una cuestión clave es el estudio de la variabilidad en los indicadores de los programas. Por un lado la variabilidad en las tasas crudas y riesgo acumulado de falsos positivos entre las distintas unidades radiológicas del programa[73], y por otro la variabilidad en el precisión de los radiólogos lectores en función del volumen de mamografías leídas, y la experiencia de los mismos[75,77,78].

Al igual que ocurre en otras ramas de la medicina, el futuro de la investigación sobre el cribado de cáncer de mama está encaminado hacia las estrategias de personalización del cribado, que permitan maximizar los beneficios en las mujeres más susceptibles de desarrollar la enfermedad, y minimizar los efectos adversos y los costes. La manera óptima de valorar la efectividad distintas estrategias de cribado sería mediante ensayos clínicos aleatorizados. Ante la imposibilidad de llevarlos a cabo en la mayoría de los casos, el uso de modelos matemáticos se está consolidando como una herramienta con mucho potencial. Dentro del grupo de investigación de Servei d'Epidemiologia i Avaluació de l'Hospital del Mar-IMIM se ha acumulado una amplia experiencia en la utilización de este tipo de modelos en el contexto de la evaluación de servicios sanitarios[95-98]. Los datos procedentes del estudio RAFP han sido utilizados, y se siguen utilizando, en combinación con datos sobre costes, para desarrollar diversos modelos matemáticos que evalúan el impacto de distintas estrategias de cribado sobre los beneficios, los efectos adversos y los costes de las distintas estrategias[58,99-101].

V. CONCLUSIONES E IMPLICACIONES

1. Conclusiones

1. Una de cada cinco mujeres participantes en el cribado bienal de cáncer de mama de manera secuencial entre los 50 y los 69 años de edad sufrirán un resultado falso positivo. Aproximadamente un 2% de las mujeres participantes sufrirán una prueba invasiva con resultado final negativo.

2. Existe una amplia variabilidad en el riesgo acumulado de falsos positivos por cualquier tipo de prueba y por pruebas invasivas en función de las características del protocolo de lectura mamográfica, y de las características de la mujer.

3. El protocolo de lectura mamográfica explica aproximadamente el 50% de la variabilidad observada en el riesgo acumulado de falsos positivos, y la cuarta parte de la variabilidad observada en el riesgo acumulado de falsos positivos con pruebas invasivas, mientras que las características de la mujer explican la parte restante.

4. Los resultados falsos positivos disminuyen la adherencia al cribado en convocatorias sucesivas. Esta disminución en la adherencia disminuye a medida que la mujer aumenta su participación en el cribado.

5. Los resultados falsos positivos aumentan el riesgo de detección de cáncer de mama en sucesivas convocatorias de cribado. El riesgo es mayor si el falso positivo conlleva pruebas invasivas.

6. La tendencia en las tasas de detección de cáncer de mama invasivo en el cribado poblacional son estables desde la puesta en marcha de los programas hasta el final del periodo de estudio.

7. Las tasas de detección de carcinoma ductal in situ en el cribado poblacional aumentaron de manera gradual y constante desde la puesta en marcha de los programas hasta el final del periodo de estudio.

8. En el programa de cribado de Noruega, una de cada cinco mujeres participantes bienalmente entre los 50 y los 69 años de edad sufrirán un resultado falso positivo; similar al contexto español. Aproximadamente un 4% de las mujeres participantes sufrirán una prueba invasiva con resultado final negativo.

2. Recomendaciones e implicaciones en Salud Pública

Los resultados de los trabajos que forman esta tesis contribuyen a mejorar la información existente para evaluar el balance entre riesgo y beneficios del cribado poblacional. La valoración de los efectos adversos, y en particular de los resultados falsos positivos, es una parte sustancial para poder mejorar la efectividad del cribado.

Los resultados de los diferentes trabajos son útiles para adecuar la información proporcionada a las mujeres invitadas a participar en los programas poblacionales de detección precoz del cáncer de mama. Como ya se ha comentado anteriormente, los resultados falsos positivos son el efecto adverso más frecuente del cribado mamográfico. Informar a las mujeres participantes sobre el riesgo de sufrir un resultado falso positivo puede contribuir a reducir la ansiedad de las mujeres ante una posible re-convocatoria para realizar exploraciones adicionales.

La amplia variabilidad observada en el riesgo acumulado de falsos positivos, además de ser potencialmente útil para mejorar la información proporcionada a las mujeres invitadas, puede ser especialmente útil para reducir los efectos adversos del cribado mamográfico y mejorar su efectividad. La identificación de los factores asociados con el riesgo de falsos positivos abre la puerta a la posibilidad de intervenir sobre los mismos, especialmente en aquellos que son potencialmente modificables, como los factores asociados al protocolo de lectura mamográfica.

La información sobre la disminución de la adherencia al cribado y el aumento de riesgo de detección de cáncer de mama en convocatorias sucesivas en las mujeres que presentan falsos positivos, puede ser útil para implementar futuras estrategias de optimización del cribado, revisando por ejemplo la periodicidad de la prueba de cribado, la edad de inicio, o la utilización de diferentes tecnologías para grupos específicos de mujeres de elevado riesgo. Además, esta información es útil para la organización estratégica de los programas, ya que será conveniente monitorizar a las mujeres con falsos positivos para promover su participación en futuras convocatorias de cribado.

La evaluación de la tendencia temporal en las tasas de detección de cáncer de mama en el cribado mamográfico aporta información muy relevante para valorar la evolución y el desarrollo de los programas de cribado. Esta información es útil para valorar el impacto que la detección de cáncer en el cribado mamográfico puede tener sobre la incidencia poblacional de cáncer de mama.

Asimismo, es útil para valorar el impacto que determinados factores, como la disminución en el uso de terapia hormonal sustitutiva, los cambios de técnica, o cambios en el criterio organizativo, pueden tener sobre las tasas de detección en el cribado.

La investigación de los beneficios y efectos adversos del cribado se ve limitada por la escasez de bases de datos poblacionales con información individualizada sobre el cribado mamográfico. Es importante continuar promoviendo la creación de grandes bases de datos poblacionales con perspectiva longitudinal. A pesar de la complejidad que representa la coordinación con las distintas fuentes de información y el trabajo de homogeneizar y validar la información, es necesario disponer de estas bases de datos para obtener una estimación precisa y conocer el impacto global de la detección precoz.

VI. CONCLUSIONS AND IMPLICATIONS

1. Conclusions

1. One in every five women participating in biennial mammographic screening between ages 50 to 69 years will experience a false positive result. Approximately, 2% of the participating women will experience an invasive procedure with benign outcome.

2. There is a wide variability in the cumulative risk of false positives results and false positive results with invasive procedures depending on the mammographic reading protocol, and women's characteristics.

3. The mammographic reading protocol explains about 50% of the observed variability in the cumulative risk of false positive results, and a quarter of the variability observed in the cumulative risk of false positives with invasive procedures, whereas the remaining part is explained by women's characteristics.

4. False positive results decrease reattendance to subsequent screening successive invitations. The decrease in reattendance is reduced as women increase the number of participations in the screening program.

5. False positive results increase the risk of breast cancer detection in subsequent screening tests. The risk of breast cancer detection is greater for false positive results with invasive procedures.

6. The time trend in the detection rates of invasive breast cancer in population-based screening are stable from the start of the screening programs until the end of the study period.

7. The detection rates of ductal carcinoma in situ in in population-based screening increased gradually and steadily from the start of the screening programs until the end of the study period.

8. In the Norwegian Breast Cancer Screening Program, one in every five women participating in biennial mammographic screening between ages 50 to 69 years will experience a false positive result; similar to the Spanish context. Approximately 4% of the participating women will experience an invasive procedure with benign outcome.

2. Recommendations and implications in Public Health

The results presented in this thesis improve the available information to evaluate the risks and benefits of mammographic screening. The assessment of the adverse effects, and in particular false positive results, is a substantial part to improve the effectiveness of population-based screening.

The results of the different studies are useful to enhance the information provided to women invited to participate in breast cancer screening programs. As previously mentioned, false positive results are the most common adverse effect of breast screening. Providing the women invited with information about the risk of a false positive result could help to reduce the anxiety of women in front of an eventual recall for further assessments.

The wide variability found in the cumulative risk of false positive results is potentially useful to improve the information provided to women invited to the screening programs. In addition, it is particularly useful to reduce the adverse effects of mammography screening and improve its effectiveness. The assessment of the risk factors associated with the risk of false positive results enhaces the possibility of intervening on them, especially those that are potentially modifiable, such as the factors associated with the mammographic reading protocol.

The results regarding the decreased reattendance to subsequent screening and the increased risk of breast cancer detection in women with previous false positive results, may be useful to implement future strategies for screening optimization. Optimization strategies may involve changing the time interval between screens, age at start of screening, or using a different screening test for specific groups of high-risk women (magnetic resonance, etc). In addition, this information is useful to improve the management of the screening programs, as it is advisable to monitor women with false positive results to promote their participation in subsequent screening invitations.

The assessment of the time trends in the rates of detection of breast cancer in mammographic screening provides very relevant information to assess the evolution and development of the screening programs. This information is useful in assessing the impact that screen detected breast cancer may have on the population incidence of the disease. It is also useful to assess the impact that specific factors, such as the decreased use of hormone replacement therapy, technical

changes, or changes in the organizational criterion, may have on the rates of screen detected breast cancer.

Research on the benefits and adverse effects of screening is limited by the scarcity of population-based databases with individualized information about mammographic screening. It is important to continue promoting the creation of large population-based databases with longitudinal perspective. Despite the complexity involved in the management and coordination of the various sources of information used, and the additional work required to standardize and validate the data, it is essential to make available these databases to produce accurate estimates and assess the overall impact of early detection.

VII. ANEXOS

1. Artículo anexo 1

Título: The cumulative risk of false-positive screening results across screening centres in the Norwegian Breast Cancer Screening Program

Autores: Román M, Skaane P, Hofvind S.

Revista: Eur J Radiol 2014; 83: 1639–1644.

Factor de impacto: 2.5 (Q2 Radiology)

DOI: 10.1016/j.ejrad.2014.05.038.

Abstract:

Background: Recall for assessment in mammographic screening entails an inevitable number of false-positive screening results. This study aimed to investigate the variation in the cumulative risk of a false positive screening result and the positive predictive value across the screening centres in the Norwegian Breast Cancer Screening Program.

Methods: We studied 618,636 women aged 50-69 years who underwent 2,090,575 screening exams (1996-2010). Recall rate, positive predictive value, rate of screen-detected cancer, and the cumulative risk of a false positive screening result, without and with invasive procedures across the screening centres were calculated. Generalized linear models were used to estimate the probability of a false positive screening result and to compute the cumulative false-positive risk for up to ten biennial screening examinations.

Results: The cumulative risk of a false-positive screening exam varied from 10.7% (95% CI: 9.4-12.0%) to 41.5% (95% CI: 34.1-48.9%) across screening centres, with a highest to lowest ratio of 3.9 (95% CI: 3.7-4.0). The highest to lowest ratio for the cumulative risk of undergoing an invasive procedure with a benign outcome was 4.3 (95% CI: 4.0-4.6). The positive predictive value of recall varied between 12.0% (95% CI: 11.0-12.9%) and 19.9% (95% CI: 18.3-21.5%), with a highest to lowest ratio of 1.7 (95% CI: 1.5-1.9).

Conclusions: A substantial variation in the performance measures across the screening centres in the Norwegian Breast Cancer Screening Program was identified, despite of similar administration, procedures, and quality assurance requirements. Differences in the readers' performance is

probably of influence for the variability. This results underscore the importance of continuous surveillance of the screening centres and the radiologists in order to sustain and improve the performance and effectiveness of screening programs.

European Journal of Radiology 83 (2014) 1639–1644

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

The cumulative risk of false-positive screening results across screening centres in the Norwegian Breast Cancer Screening Program

M. Roman [a,b], P. Skaane [c], S. Hofvind [a,d,*]

[a] Cancer Registry of Norway, Oslo, Norway
[b] Department of Women and Children's Health, Oslo University Hospital, Oslo, Norway
[c] Department of Radiology, Oslo University Hospital Ullevaal, University of Oslo, Oslo, Norway
[d] Oslo and Akershus University College of Applied Sciences, Faculty of Health Science, Oslo, Norway

ARTICLE INFO

Article history:
Received 6 March 2014
Received in revised form 20 May 2014
Accepted 26 May 2014

Keywords:
Breast neoplasms
Mass screening
Mammography
False positive reactions
Predictive value of tests

ABSTRACT

Background: Recall for assessment in mammographic screening entails an inevitable number of false-positive screening results. This study aimed to investigate the variation in the cumulative risk of a false positive screening result and the positive predictive value across the screening centres in the Norwegian Breast Cancer Screening Program.

Methods: We studied 618,636 women aged 50–69 years who underwent 2,090,575 screening exams (1996–2010. Recall rate, positive predictive value, rate of screen-detected cancer, and the cumulative risk of a false positive screening result, without and with invasive procedures across the screening centres were calculated. Generalized linear models were used to estimate the probability of a false positive screening result and to compute the cumulative false-positive risk for up to ten biennial screening examinations.

Results: The cumulative risk of a false-positive screening exam varied from 10.7% (95% CI: 9.4–12.0%) to 41.5% (95% CI: 34.1–48.9%) across screening centres, with a highest to lowest ratio of 3.9 (95% CI: 3.7–4.0). The highest to lowest ratio for the cumulative risk of undergoing an invasive procedure with a benign outcome was 4.3 (95% CI: 4.0–4.6). The positive predictive value of recall varied between 12.0% (95% CI: 11.0–12.9%) and 19.9% (95% CI: 18.3–21.5%), with a highest to lowest ratio of 1.7 (95% CI: 1.5–1.9).

Conclusions: A substantial variation in the performance measures across the screening centres in the Norwegian Breast Cancer Screening Program was identified, despite of similar administration, procedures, and quality assurance requirements. Differences in the readers' performance is probably of influence for the variability. This results underscore the importance of continuous surveillance of the screening centres and the radiologists in order to sustain and improve the performance and effectiveness of screening programs.

1. Introduction

Recall for assessments after screening mammography without detecting breast cancer is referred to as a 'false positive screening result', which is an inevitable adverse effect of mammographic screening. The negative effects of false positives have been widely discussed and include psychological harms [1,2], additional hospital visits and diagnostic tests [3], decreased participation in future screenings [4,5], and increased economic costs for the screening program.

The risk of a false-positive screening result is strongly associated with the recall rate, which is influenced by the screening procedures (e.g. prevalence or subsequent screen, screening interval, one or two views, single versus double reading, and use of screen film versus full field digital mammography), the radiologists (training and experience), and the characteristics of the women (e.g. age, screening history, use of hormone therapy, mammographic breast density, and previous invasive procedure) [6–10]. The reported recall rate was 9.3% for prevalent and 4.0% for subsequent screens in European screening programs during 2004–2007, ranging from 2.2% to 15.6% for prevalent, and from 1.2% to 10.5% for subsequent screens [7].

* Corresponding author at: Cancer Registry of Norway, Mammography Screening Program, P.O. 5313, Majorstua 0403, Oslo, Norway. Tel.: +47 23 333987.
E-mail addresses: Marta.Roman@kreftregisteret.no (M. Roman), PERSK@ous-hf.no (P. Skaane), Solveig.Hofvind@kreftregisteret.no (S. Hofvind).

http://dx.doi.org/10.1016/j.ejrad.2014.05.038
0720-048X/

However, the recall rate has to be evaluated together with the rate of screen-detected breast cancer. Positive predictive value (PPV) is thus an essential indicator in the evaluation of a screening program.

The cumulative risk of a false positive screening result could be defined as the risk of a false-positive recall for a woman, screened biennially from age 50–69. In Europe, the pooled cumulative risk of a false-positive recall was estimated to be 19.7%, and the risk of an invasive procedure with a benign outcome, 2.9% [7]. The cumulative risk for a false-positive screening result and of an invasive procedure with benign outcome is reported to be substantially higher in the U.S. compared to Europe, ranging from 41.6% to 63.3% and from 4.8% to 18.6%, respectively [11–13].

The recall rate and the related performance measures are a critical part in the evaluation of a screening program. It is important to keep the recall rate as low as possible without missing breast cancers. The Norwegian Breast Cancer Screening Program (NBCSP) includes 16 screening centres with a target population varying from 11,000 to 52,000 women. A substantial difference in early performance measures is observed in the program, despite of a common administration and guidelines for quality assurance [14]. However, these observations have not been systematized and no estimations related to false positive screening results have been documented previously. We wanted to take advantage of the data collected as a part of the quality assurance of the NBCSP since 1996 and investigate the variation in the PPV and the cumulative risk of a false positive screening result across screening centres in the program.

2. Materials and methods

The NBCSP is run according to the European guidelines for quality assurance in breast cancer screening and diagnosis [15]. The program started as a pilot in four counties in 1996 and was stepwise expanded becoming nationwide in 2005 [16]. Women aged 50–69 years are invited to a two-view mammography every second year. A unique personal identification number (PIN) given to all inhabitants of Norway at birth or immigration identifies the women. We received an anonymized file with individual level data. No ethical committee approval was thus necessary.

The NBCSP includes 19 counties and 16 screening centres. We defined 6 screening centres as representing a rural area, 6 as intermediate, and 4 as urban, based on the density of the population and accessibility to the screening centre. The central nationwide database, located at the Cancer Registry of Norway, gathers information on attendance, screening outcome, and diagnostic workup from all the screening centres. The program is described in detail elsewhere [16].

The study population includes all women with at least 1 screening exam performed in the NBCSP, 1996–2010. The screening mammograms are read by two independent radiologists. An interpretation score ranging from 1 to 5 is given for both breast, and from both readers. A score of 1 indicates a negative screening examination while a score of 5 indicates high susceptibility of malignancy. Screening exams with an interpretation score of 2 or higher by one or both readers are discussed at a consensus meeting. The consensus meeting decide whether to recall the women or not. Additional X-rays and ultrasound (non-invasive methods) are used to rule out findings on the screening mammograms. If further investigation is needed, fine-needle aspiration cytology (FNAC), core needle biopsy (CNB), and/or an open biopsy (OB) is performed. Vacuum assisted biopsies were included in the group of CNB. The diagnostic work-up takes place at the screening centre, located at university or county hospitals, usually 1–2 weeks after screening. If no malignancy is diagnosed, the women are referred back to screening. All women diagnosed with breast cancer are referred to treatment.

2.1. Measures and definitions

Recall rate was defined as the percentage of screening exams with positive mammographic findings and a call-back for assessment, among all screening exams. A false-positive screening result was defined as a recall where no breast cancer was diagnosed, regardless of the procedures performed (overall), while a false positive invasive procedure with a benign outcome was defined as a recall including a FNAC, CNB or OB with benign morphology. The cumulative risk of having a false-positive screening result was defined as the risk of being recalled with negative result during ten biennial screening exams from age 50–69. Likewise, the cumulative risk of undergoing an invasive procedure with a benign outcome was defined as the risk of having an invasive procedure with benign outcome during the same age and time span.

PPV of recall was defined as the number of screen detected breast cancers (invasive breast cancer and ductal carcinoma in situ, DCIS) divided by the number of recalls due to positive mammographic findings. PPV of invasive procedures was defined as the number of screen detected breast cancers divided by the number of recall exams with invasive diagnostic procedures (FNAC, CNB and/or OB). The rate of screen-detected cancer was defined as the number of breast cancer per 1000 screening exams. The interval cancer rate was defined as the rate of breast cancer diagnosed after a negative screening examination, with or without an invasive procedures, and before the next screening exam (within 730 days).

Variation across screening centres was illustrated by risk ratios, calculated as the highest to lowest ratio. Women recalled due to insufficient technical image quality or self-declared symptoms ($<$0.5% for both together) were not included either in the nominator or denominator in the estimates of false positive screening tests. The screening centres were anonymized in the tables and figures.

2.2. Statistical analysis

Our estimates of the cumulative risk of a false-positive screening result were based on all screening exams performed in the 16 screening centres of the NBCSP from 1996 to 2010. The women contributed data from their first screening examination, until end of follow up (December 31, 2010). Data up to a maximum of six screening examinations during the study period were used for the estimations. Generalized linear models were used to estimate the probability and 95% confidence intervals (CI) for the risk of a false positive screening result at each screening exam. The regression model included adjustment for year of the screening exam, taking the last year (2010) as the reference, and age at screening exam in three categories (50–54, 55–59 and $\geq$60), taking the youngest age as the reference. Women were included in analyses up to the time of their first false-positive test result. The probability of a false positive result at each screening exam was estimated as a function of time, given by the women's screening round (acting as multiple intercepts), and mammogram-level substantive predictors' given by the year in which the screening exam was performed and age at screening exam. The regression models are described in detail elsewhere [9,18]. Separate models were computed to estimate the probability of a false-positive screening result and the probability of an invasive procedure with benign outcome at each specific screening centre, and for the average of all screening centres.

The cumulative risk of a false-positive screening result for each round up to the 10th screening exam was calculated by multiplying the probability of receiving a first false-positive test result at each screening exam by the probability of receiving no false-positive test results at any previous exam. The point estimates of receiving a

Table 1
Number of screening exams, number and proportion of recall due to mammographic findings, recall including an invasive procedure, screen-detected breast cancer, interval cancer, and positive predictive value (PPV) of the screening test by screening centre in the Norwegian Breast Cancer Screening Program (1996–2010).

Screening centre	No. of screens	Recall due to mammographic findings, *n* (%)		Recall with invasive procedure *n* (%)		Breast cancer cases *n* (‰)		Interval cancer cases[a] *n* (‰)		PPV of recall[b] (%) (95% CI)[d]		PPV of invasive procedures[c](%) (95% CI)[d]	
A	38,539	1137	(3.0)	437	(1.1)	175	(4.5)	33	(1.2)	15.4	(13.3–17.5)	40.0	(35.5–44.6)
B	58,412	2611	(4.5)	918	(1.6)	341	(5.8)	73	(1.8)	13.1	(11.8–14.4)	37.1	(34.0–40.3)
C	69,519	1833	(2.6)	618	(0.9)	361	(5.2)	72	(1.3)	19.7	(17.9–21.5)	58.4	(54.5–62.3)
D	71,507	2288	(3.2)	948	(1.3)	456	(6.4)	110	(2.2)	19.9	(18.3–21.6)	48.1	(44.9–51.3)
E	78,077	2036	(2.6)	941	(1.2)	379	(4.9)	129	(2.2)	18.6	(16.9–20.3)	40.3	(37.1–43.4)
F	84,715	2324	(2.7)	975	(1.2)	431	(5.1)	129	(1.9)	18.5	(17.0–20.1)	44.2	(41.1–47.3)
G	99,997	2209	(2.2)	1000	(1.0)	423	(4.2)	129	(1.6)	19.1	(17.5–20.8)	42.3	(39.2–45.4)
H	100,017	2358	(2.4)	887	(0.9)	470	(4.7)	155	(2.0)	19.9	(18.3–21.5)	53.0	(49.7–56.3)
I	100,576	4738	(4.7)	1495	(1.5)	639	(6.4)	145	(1.8)	13.5	(12.5–14.5)	42.7	(40.2–45.3)
J	103,310	4562	(4.4)	1579	(1.5)	547	(5.3)	117	(1.5)	12.0	(11.0–12.9)	34.6	(32.3–37.0)
K	121,314	3015	(2.5)	1042	(0.9)	550	(4.5)	200	(2.0)	18.2	(16.9–19.6)	52.8	(49.8–55.8)
L	161,627	5071	(3.1)	1515	(0.9)	829	(5.1)	185	(1.5)	16.3	(15.3–17.4)	54.7	(52.2–57.2)
M	226,292	8106	(3.6)	2811	(1.2)	1383	(6.1)	495	(2.5)	17.1	(16.2–17.9)	49.2	(47.4–51.0)
N	226,420	7088	(3.1)	2826	(1.2)	1397	(6.2)	369	(1.9)	19.7	(18.8–20.6)	49.4	(47.6–51.3)
O	269,303	7468	(2.8)	3177	(1.2)	1467	(5.4)	401	(1.8)	19.6	(18.7–20.5)	46.2	(44.4–47.9)
P	280,950	12,478	(4.4)	5637	(2.0)	1578	(5.6)	505	(2.1)	12.6	(12.1–13.2)	28.0	(26.8–29.2)
Total	2,090,575	69,322	(3.3)	26,806	(1.3)	11,426	(5.5)	3247	(1.9)	16.5	(16.2–16.8)	42.6	(42.0–43.2)

[a] Interval cancer rate calculated as number of interval cancer cases divided by number of women screened 1996–2008, follow-up time until December 2010.
[b] PPV-1 calculated as number of screen detected breast cancers divided by number of recalls due to mammographic findings.
[c] PPV-2 calculated as number of screen detected breast cancers divided by number of recall exams including an invasive diagnostic procedure.
[d] 95% CI: 95% confidence interval.

first false-positive test result in the 7th to the 10th screening exam were assumed to be equal to that of the 6th exam. Standard errors for the calculation of the confidence intervals for the cumulative risk probability were estimated using Greenwood's approximation [19]. Statistical significance was defined using a two-sided α level of 0.05. Statistical analyses were performed in SAS 9.2 (SAS Institute, Cary, NC).

3. Results

We analysed data from 618,636 women aged 50–69 years who underwent 2,090,575 screening exams in the NBCSP, 1996–2010. A second screening exam was performed in 500,408 (80.9%) women, 385,936 (62.4%) had a third screening exam, and 81,976 (13.3%) a sixth exam. A total of 69,322 recalls due to mammographic findings, 26,806 recalls including an invasive procedure, and 11,426 screen detected breast cancers (DCIS and invasive) were identified (Table 1). The rate of screen detected breast cancer was 5.5‰, varying from 4.2‰ to 6.4‰ across screening centres. The highest to lowest ratio was 1.5 (95% CI: 1.3–1.7).

The rate of false positive screening results varied from 1.8% to 4.1% across screening centres, with a highest to lowest ratio of 2.3 (95% CI: 2.2–2.4) (Table 2). The rate of invasive procedures with benign outcome varied from 0.4% to 1.4%, with a highest to lowest ratio of 3.9 (95% CI: 3.4–4.4). The proportions of invasive procedures with benign outcome were below 1.0% in 13 of the 16 screening centres.

The cumulative risk of a false-positive screening result varied from 10.7% (95% CI: 9.4–12.0%) to 41.5% (95% CI: 34.1–48.9%) across screening centres, with 3.9 (95% CI: 3.7–4.0) as the highest to lowest ratio (Fig. 1). The overall cumulative risk of a false-positive result for all screening centres was 23.0% (95% CI: 22.8–23.2%),

Table 2
Area of residence, number of screening exams, and number and proportion of false-positive screening exams and invasive procedures with a benign outcome by screening centre in the Norwegian Breast Cancer Screening Program (1996–2010).

Screening centre	Area of residence[a]	No. of screens	False-positive screening results		Benign invasive procedures	
			N	Rate (%)	*N*	Rate (%)
A	Rural	38,539	965	2.5	262	0.7
B	Rural	58,412	2284	3.9	577	1.0
C	Rural	69,519	1474	2.1	257	0.4
D	Intermediate	71,507	1840	2.6	492	0.7
E	Rural	78,077	1661	2.1	562	0.7
F	Intermediate	84,715	1904	2.2	544	0.6
G	Rural	99,997	1795	1.8	577	0.6
H	Rural	100,017	1893	1.9	417	0.4
I	Intermediate	100,576	4121	4.1	856	0.9
J	Intermediate	103,310	4039	3.9	1032	1.0
K	Intermediate	121,314	2479	2.0	492	0.4
L	Intermediate	161,627	4258	2.6	686	0.4
M	Urban	226,292	6751	3.0	1428	0.6
N	Urban	226,420	5714	2.5	1429	0.6
O	Urban	269,303	6031	2.2	1710	0.6
P	Urban	280,950	10,929	3.9	4059	1.4
Total		2,090,575	58,138	2.8	15,380	0.7

[a] Screening centres were classified as representing a rural. intermediate or urban area based on the density of the population and accessibility to the screening centre.

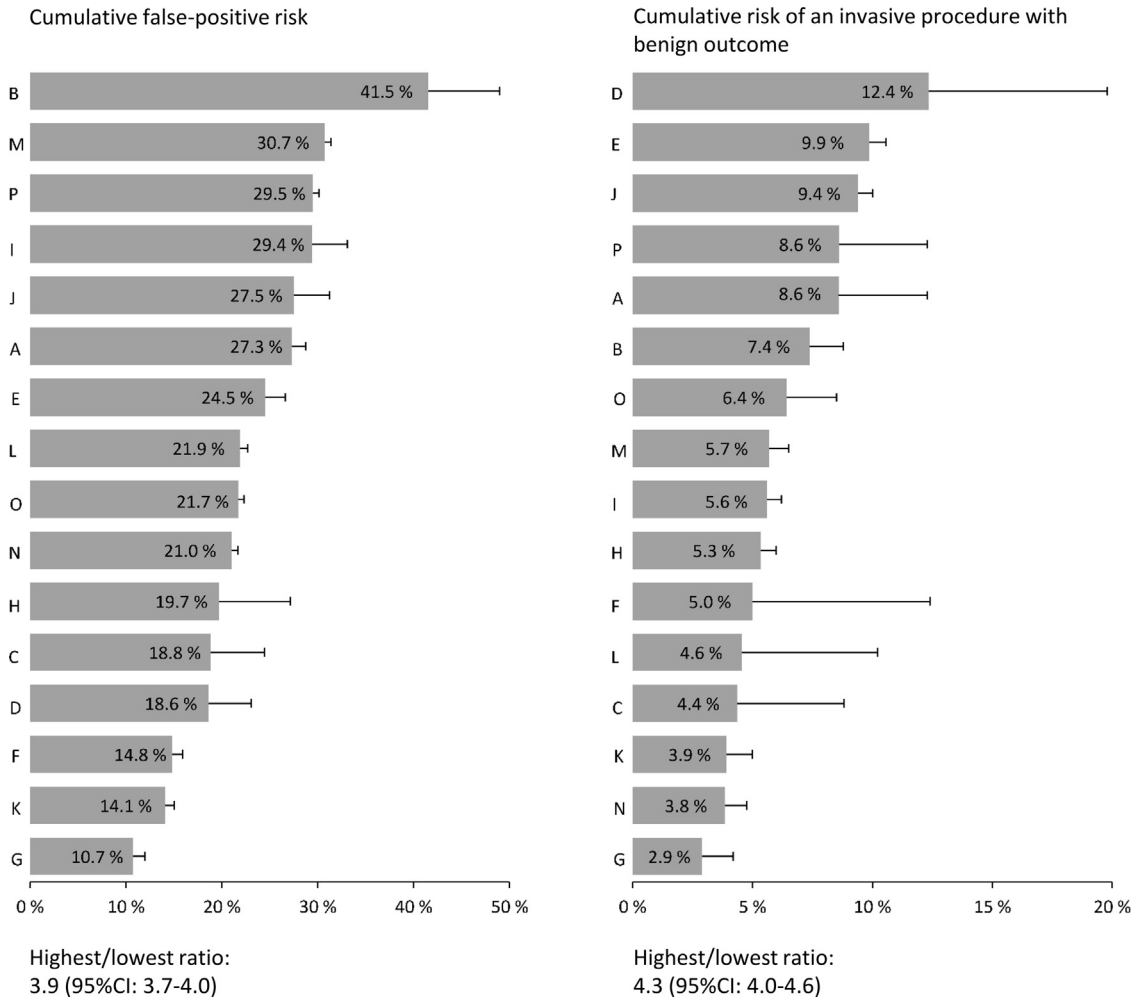


Fig. 1. The cumulative risk of a false-positive screening result and of undergoing an invasive procedure with a benign outcome by screening centre.

while the risk of an invasive procedure with a benign outcome varied from 2.9% (95% CI: 2.2–3.6%) to 12.4% (95% CI: 8.7–16.0%), with 4.3 (95% CI: 4.0–4.6) as the highest to lowest ratio. The overall risk was 5.3% (95% CI: 5.1–5.4%) for all the screening centres. Both the cumulative risk of a false-positive screening result and of an invasive procedure with a benign outcome decreased gradually from the highest to the lowest rate, but the sequences were not identical (Fig. 1). The Pearson's correlation coefficient between the cumulative risk of a false-positive screening result and of an invasive procedure with a benign outcome was 0.4. The screening centres D, E, F, and J had high cumulative risks of an invasive procedure with a benign outcome compared with their cumulative risk of a false-positive screening result, while centres B, I, L, N, and M had relatively low cumulative risk of an invasive procedure with a benign outcome compared with their cumulative risk of a false-positive screening result. No association was observed between the area of residence (rural, intermediate and urban) and the cumulative risk of a false-positive screening exam or the cumulative risk of an invasive procedure with a benign outcome.

PPV of recall varied between 12.0% (95% CI: 11.0–12.9%) and 19.9% (95% CI: 18.3–21.5%), with a highest to lowest ratio of 1.7 (95% CI: 1.5–1.9). PPV of invasive procedures varied between 28.0% (95% CI: 26.8–29.2%) and 58.4% (95% CI: 54.5–62.3%), with a highest to lowest ratio of 2.1 (95% CI: 1.9–2.3). The average PPV of recall and PPV of invasive procedures for all screening centres was 16.5% (95% CI: 16.2–16.8%) and 42.6% (95% CI: 42.2–43.2%), respectively (Table 1).

The rate of screen detected cancer was associated with the recall rate and the performance of an invasive procedure (Fig. 2). There was a wider variability in the recall rates (horizontal axis), than in the rates of screen-detected cancer (vertical axis) across the screening centres (Fig. 2). The size of the circles represent the number of screening mammograms performed at the different screening centres. The larger the circle, the more screening exams performed. No association was observed between the volume of screening exams performed at the screening centres and the PPV's. PPV's are presented as the recall rate (horizontal axis) plotted against the screen detected breast cancer rate (vertical axis) (left panel for PPV of recall and right panel for PPV of invasive procedures). A positive association was identified between the PPV of recall and PPV of invasive procedures (Pearson's correlation coefficient = 0.7), which means that screening centres with a high PPV of recall tended to have a high PPV of invasive procedures, and vice versa (Fig. 2). No association was observed between the rural, intermediate and urban areas and the PPV of recall and PPV of invasive procedures.

4. Discussion

We found a substantial variation in the cumulative risk of a false positive screening result with and without an invasive procedure, and in PPV of recall, and of invasive procedures, across the screening centres in the NBCSP. The variation was 3.9-fold for the cumulative risk of a false-positive screening result, and 4.3-fold for

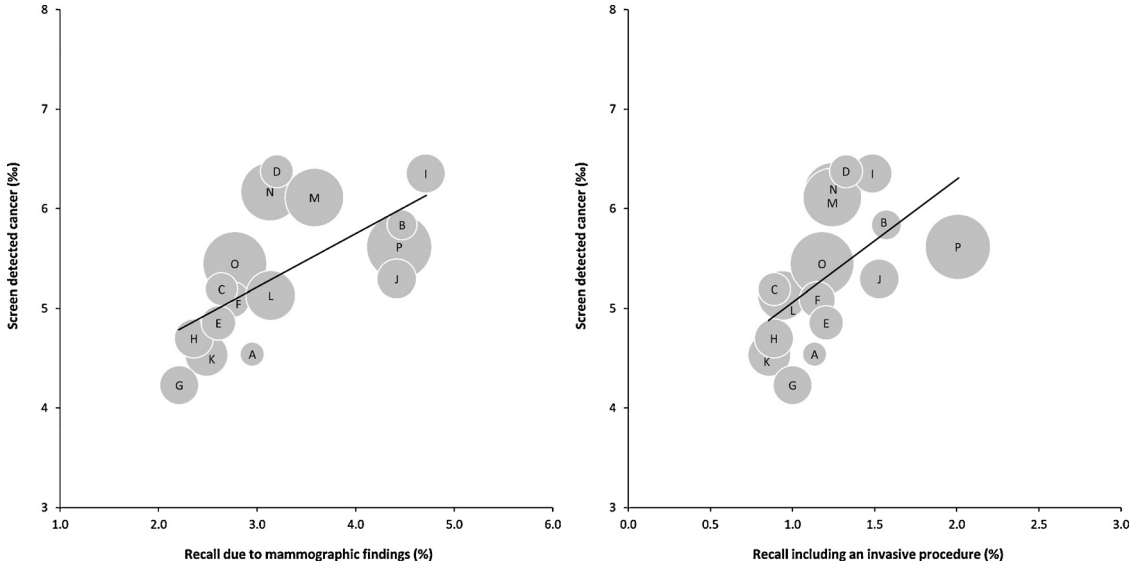


Fig. 2. Rate of screen-detected breast cancer per 1000 screening exams versus recall rate due to mammographic findings as a percentage of screening exams with positive mammographic findings among all exams (left), and rate of screen-detected breast cancer per 1000 screening exams versus recall rate including an invasive procedure as a percentage of screening exams with an invasive procedure (FNAC, CNB, OB) with a benign outcome among all exams (right).

an invasive procedure with benign outcome. PPV of recall varied a 1.7-fold and PPV of invasive procedures a 2.1-fold.

It is well known that recall rate and related performance measures vary between countries [7,10,17,18,20]. This is usually explained by differences in the screening algorithms and management policies, the screening interval, interpretation procedures, and legal consequences. Although the NBCSP is run according to the same management policy and guidelines, we found a remarkable variation across the screening centres.

The radiologist performance has been suggested to be one of the main sources of variation in the accuracy of screening mammography [21]. The diagnostic performance is related to the radiologist's age, years of experience, and number of yearly mammograms read [21–23]. Variation in the age and years of experience among radiologist is unavoidable in a national screening program, and might represent a benefit as well as harm. The European guidelines for quality assurance in breast cancer screening and diagnosis recommend that radiologist read at least 5000 screening mammograms annually [15]. A previous study has reported that 35% of the breast radiologists in the NBCSP reached that level in 2007 [24] and the average number of mammograms read each year to be 3600. However, the annual number of mammograms read varied from 275 to 13,395 [24]. Most radiologist had limited or no experience reading screening mammograms before they started to do screen reading in the NBCSP, but most of them were experienced in diagnostic mammography. Information about the radiologists reading volume and years of mammographic experience might be of influence for the outcome in this study. This possible association ought to be investigated in a separate study.

Training has been shown to improve the sensitivity of radiologists' performance [21]. The NBCSP requires training, shadow reading, and courses in mammography to start reading. Early performance measures and individual results for the radiologists are easily accessible through a software program available at all breast centres. In addition, epidemiological and radiological results are presented and discussed regularly at site visits and meetings. However, at some screening centres, radiologists are hired only to read screening mammograms. These radiologists do not necessarily take part in both, screening and diagnostic work-up, although it is recommended by the Norwegian and European guidelines [15]. Our results show that the recall rate is both among the highest and lowest at the screening centres with such practice.

Full field digital mammography (FFDM) has replaced screen film mammography (SFM) during the study period. FFDM is shown to affect the recall rate, and the rate of screen-detected DCIS compared with SFM [25–27], particularly in the transition phase [28]. Our study included data from the transition period between SFM to FFDM, which might overestimate the results. Further, CNB has replaced FNAC over the years due to its higher sensitivity for detecting breast cancer [29,30]. This factor might also have an impact on the results.

The correlation coefficient between the cumulative risk of a false-positive screening result and of an invasive procedure with benign outcome was 0.4. Although the variation in recall rate is influenced by the radiologist performance, this finding suggests that there are other reasons for this moderate correlation. A long distance from the women's residential community to the screening centre might have an impact on the decision to recall women. On the other hand, to avoid further call back of these women, the percentage of these women undergoing a needle biopsy might be higher compared to those residing close to the screening centre. However, low recall rates were shown for screening centres defined as rural, as well as centres defined as intermediate and urban.

A 3.9-fold variation was found in the cumulative risk of a false-positive screening result and a 4.3-fold for an invasive procedure with benign outcome, while the variation in the rate of screen-detected cancer was 1.5-fold, only. This results highlight that there is a potential to improve the effectiveness in some screening centres by reducing the rate of recalls and false positive screening results, while keeping acceptable detection rates. Such optimization would be beneficial also for the economic costs associated with recalled women, particularly for those including additional invasive procedures (FNAC, CNB, and OB). The procedures represent a disadvantage from the women's perspective.

The NBCSP define all cancers detected after a negative or a false positive screening result, with or without an invasive biopsy, as an interval cancer. The program does not accept short-term follow-up. The cases detected after a false positive screening result are thus defined as interval cancer. If a short-term follow-up was accepted,

most of these cases might be defined as screen-detected and true positive cancers, and thus resulted in a lower rate of false-positive screening result.

The estimated overall cumulative risks were slightly higher (23.0% and 5.3%) than previously reported from Norway (20.0% and 4.1%) [18]. This study included all women aged 50–69 with the age group 50–54 as the reference, while only women aged 50–51 at first screening were included in the previous paper [18]. Further, this study did not include the screening centres as a random source of variation in the overall estimation, because the variation across the screening centres was the main aim of this study, and the unadjusted estimate was considered the outcome of interest. As the screening program was implemented stepwise during ten years, the models included calendar year 2010 as the reference. The adjusted cumulative risk estimates should thus be carefully compared with the crude rates.

A major strength of this nationwide study is the number of women screened and screening exams included in the analyses. In addition, 15 years of follow-up represent seven complete screening rounds which provides a large sample size to study the variation in the PPV's and the cumulative risk of a false positive screening result.

5. Conclusion

A substantial variation of the early performance measures, including the cumulative risk of a false positive screening result and an invasive procedure with benign outcome was observed across the screening centres in the NBCSP, despite of the same administration, procedures, and quality assurance parameters. Differences in the readers' performance is probably of influence for the variability. The findings underscore the importance of continuous surveillance of the screening centres and the radiologists in order to sustain and improve the performance.

Conflict of interest

The authors declare that there are no conflict of interest.

Funding sources

This work had no specific funding source.

Acknowledgment

Thanks to all radiologists working in the Norwegian Breast Cancer Screening Program for their dedicated and enthusiastic work.

References

[1] Bond M, Pavey T, Welch K, et al. Systematic review of the psychological consequences of false-positive screening mammograms. Health Technol Assess 2013;17(13):1–170.
[2] Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. Psychooncology 2005;14(11):917–38.
[3] Lerman C, Trock B, Rimer BK, et al. Psychological and behavioral implications of abnormal mammograms. Ann Intern Med 1991;114:657–61.
[4] Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med 2007;146(7):502–10.
[5] Roman R, Sala M, De La Vega M, et al. Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening. Breast Cancer Res Treat 2011;130(2):543–52.
[6] Hofvind S, Moller B, Thoresen S, Ursin G. Use of hormone therapy and risk of breast cancer detected at screening and between mammographic screens. Int J Cancer 2006;118(12):3112–7.
[7] Hofvind S, Ponti A, Patnick J, et al. False-positive results in mammographic screening for breast in Europe: a literature review and survey of service screening programmes. J Med Screen 2012;19(Suppl. 1):57–66.
[8] Otten JD, Karssemeijer N, Hendriks JH, et al. Effect of recall rate on earlier screen detection of breast cancers based on the Dutch performance indicators. J Natl Cancer Inst 2005;97(10):748–54.
[9] Roman R, Sala M, Salas D, Ascunce N, Zubizarreta R, Castells X. Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. Ann Oncol 2012;23(1):104–11.
[10] Smith-Bindman R, Ballard-Barbash R, Miglioretti DL, Patnick J, Kerlikowske K. Comparing the performance of mammography screening in the USA and the UK. J Med Screen 2005;12(1):50–4.
[11] Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false-positive screening mammograms and clinical breast examinations. N Engl J Med 1998;338:1089–96.
[12] Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Ann Intern Med 2011;155(8):481–92.
[13] Hubbard RA, Miglioretti DL, Smith RA. Modelling the cumulative risk of a false-positive screening test. Stat Methods Med Res 2010;19(5):429–49.
[14] Quality manual of the Norwegian Breast Cancer Screening Program (Mammografiprogrammet: Kvalitetsmanual). Oslo, Norway: Kreftregisteret, Institute of Population-based Cancer Research; 2003.
[15] Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Luxemburg: Office for Official Publications of the European Communities; 2006.
[16] Hofvind S, Geller B, Vacek PM, Thoresen S, Skaane P. Using the European guidelines to evaluate the Norwegian Breast Cancer Screening Program. Eur J Epidemiol 2007;22(7):447–55.
[17] Hofvind S, Geller BM, Skelly J, Vacek PM. Sensitivity and specificity of mammographic screening as practised in Vermont and Norway. Br J Radiol 2012;85(1020):1226–32.
[18] Roman M, Hubbard RA, Sebuodegard S, Miglioretti DL, Castells X, Hofvind S. The cumulative risk of false-positive results in the Norwegian Breast Cancer Screening Program: updated results. Cancer 2013;119(22):3952–8.
[19] Singer J, Willett JB. Fitting basic discrete-time hazard models. Applied longitudinal data analysis: modelling change and event occurrence. New York: Oxford University Press; 2003. p. 357–467.
[20] Elmore JG, Nakano CY, Koepsell TD, Desnick LM, D'Orsi CJ, Ransohoff DF. International variation in screening mammography interpretations in community-based programs. J Natl Cancer Inst 2003;95(18):1384–93.
[21] Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. Radiology 2009;253(3):641–51.
[22] Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. J Natl Cancer Inst 1998;90(23):1801–9.
[23] Zubizarreta RZ, Llanes AB, Ortega RA, et al. Effect of radiologist experience on the risk of false-positive results in breast cancer screening program. Eur Radiol 2011;21(10):2083–90.
[24] Hofvind S, Geller BM, Rosenberg RD, Skaane P. Screening-detected breast cancers: discordant independent double reading in a population-based screening program. Radiology 2009;253(3):652–60.
[25] Bluekens AM, Holland R, Karssemeijer N, Broeders MJ, den Heeten GJ. Comparison of digital screening mammography and screen-film mammography in the early detection of clinically relevant cancers: a multicenter study. Radiology 2012;265(3):707–14.
[26] Bluekens AM, Karssemeijer N, Beijerinck D, et al. Consequences of digital mammography in population-based breast cancer screening: initial changes and long term impact on referral rates. Eur Radiol 2010;20(9):2067–73.
[27] Sala M, Salas D, Belvis F, et al. Reduction in false-positive results after introduction of digital mammography: analysis from four population-based breast screening programs in Spain. Radiology 2011;258(2):388–95.
[28] Hofvind S, Skaane P, Elmore JG, Sebuødegård S, Hoff SR, Lee CI. Mammographic performance in a population-based screening program: before, during, and after the transition from screen film to full field digital mammography. Radiology 2014;13:1502.
[29] Brancato B, Crocetti E, Bianchi S, et al. Accuracy of needle biopsy of breast lesions visible on ultrasound: audit of fine needle versus core needle biopsy in 3233 consecutive samplings with ascertained outcomes. Breast 2012;21(4):449–54.
[30] Vimpeli SM, Saarenmaa I, Huhtala H, Soimakallio S. Large-core needle biopsy versus fine-needle aspiration biopsy in solid breast lesions: comparison of costs and diagnostic value. Acta Radiol 2008;49(8):863–9.

2. Artículo anexo 2

Título: Effect of start age of breast cancer screening mammography on the risk of false-positive results

Autores: Salas D, Ibañez J, Román M, Cuevas D, Sala M, Ascunce N, Zubizarreta R, Castells X, Cumulative False Positive Risk Group.

Revista: Prev Med 2011; 53(1-2): 76-81.

Factor de impacto: 3.5 (Q1 Public Health)

DOI: 10.1016/j.ypmed.2011.04.013.

Abstract:

Background: To estimate the false-positive (FP) risk according to the start age of mammography screening (45-46 or 50-51 years).

Methods: Data from eight regions of the Spanish breast cancer screening programme from 1990 to 2006 were included (1,565,364 women). Discrete time-hazard models were used to ascertain the effect of age and time-related, programme-related and personal variables on FP leading to any further procedure and to invasive procedures (FPI). In a subset we estimated the differential FP risk of starting screening at 45-46 years (175,656 women) or 50-51 (251,275).

Results: A start age of 45-46 versus 50-51 years increased both FP (OR=1.20; 95%CI: 1.13-1.26) and FPI risks (OR=1.43 (95%CI: 1.18-1.73). Other factors increasing FP risk were premenopausal status (FP OR=1.26; 95%CI: 1.23-1.29 and FPI OR=1.22; 95%CI: 1.13-1.31), prior invasive procedures (FP OR=1.52; 95%CI: 1.47-1.57 and FPI (OR=2.08; 95%CI: 1.89-2.28) and family history (FP OR=1.16; 95%CI: 1.12-1.20 and FPI OR=1.26; 95%CI: 1.13-1.41). FP risk was increased by double reading (OR=1.36; 95%CI: 1.23-1.51) and FPI risk by double views (OR=1.34; 95%CI: 1.18-1.52). Both the cumulative FP and FPI risks were higher in women commencing screening at 45-46 years versus 50-51 years (33.30% versus 20.39% and 2.68% versus 1.76%).

Conclusions: Starting screening earlier increases the cumulative risk of FP and FPI.

Preventive Medicine 53 (2011) 76–81

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

Effect of start age of breast cancer screening mammography on the risk of false-positive results

Dolores Salas [a,*], Josefa Ibáñez [a,1], Rubén Román [b,2], Dolores Cuevas [c,3], Maria Sala [b,2], Nieves Ascunce [d,4], Raquel Zubizarreta [e,5], Xavier Castells [b,2] and The CFPR (Cumulative False Positive Risk) group [6]

[a] General Directorate Public Health and Centre for Public Health Research (CSISP), Avda. Catalunya, 21, 46020 Valencia, Spain
[b] Department of Epidemiology and Evaluation, IMIM-Parc de Salut Mar. CIBERESP, Passeig Marítim 25-29, 08003 Barcelona, Spain
[c] Valencian Health Agency and Centre for Public Health Research (CSISP), C/Micer Mascó 31, 33, 46010 Valencia, Spain
[d] Navarre Breast Cancer Screening Programme, Public Health Institute, CIBERESP, C/Francisco Bergamín, 2 bis. 31003-Pamplona, Spain
[e] Public Health and Planning Directorate, Health Office, Galicia, C/San Lázaro s/n 15781 Santiago, Spain

ARTICLE INFO

Available online 4 May 2011

Keywords:
Age factors
False positive
Breast cancer screening
Invasive procedures

ABSTRACT

Objective. To estimate the false-positive (FP) risk according to the start age of mammography screening (45–46 or 50–51 years).

Method. Data from eight regions of the Spanish breast cancer screening programme from 1990 to 2006 were included (1,565,364 women). Discrete time-hazard models were used to ascertain the effect of age and time-related, programme-related and personal variables on FP leading to any further procedure and to invasive procedures (FPI). In a subset we estimated the differential FP risk of starting screening at 45–46 years (175,656 women) or 50–51 (251,275).

Results. A start age of 45–46 versus 50–51 years increased both FP (OR = 1.20; 95%CI: 1.13–1.26) and FPI risks (OR = 1.43 (95%CI: 1.18–1.73).Other factors increasing FP risk were premenopausal status (FP OR = 1.26; 95%CI: 1.23–1.29 and FPI OR = 1.22; 95%CI: 1.13–1.31), prior invasive procedures (FP OR = 1.52; 95%CI: 1.47–1.57 and FPI (OR = 2.08; 95%CI: 1.89–2.28) and family history (FP OR = 1.16; 95% CI: 1.12–1.20 and FPI OR = 1.26; 95%CI: 1.13–1.41). FP risk was increased by double reading (OR = 1.36; 95% CI: 1.23–1.51) and FPI risk by double views (OR = 1.34; 95%CI: 1.18–1.52). Both the cumulative FP and FPI risks were higher in women commencing screening at 45–46 years versus 50–51 years (33.30% versus 20.39% and 2.68% versus 1.76%).

Conclusions. Starting screening earlier increases the cumulative risk of FP and FPI.

* Corresponding author at: Av/Catalunya, 21, 46020 Valencia, Spain. Fax: +34 96 1925832.
E-mail addresses: salas_dol@gva.es (D. Salas), ibanyez_jos@gva.es (J. Ibáñez), rroman@parcdesalutmar.cat (R. Román), cuevas_dol@gva.es (D. Cuevas), msalaserra@parcdesalutmar.cat (M. Sala), nascunce@cfnavarra.es (N. Ascunce), raquel.zubizarreta.alberdi@sergas.es (R. Zubizarreta), xcastells@parcdesalutmar.cat (X. Castells).
[1] Fax: +34 961 925832.
[2] Fax: +34 932 483496.
[3] Fax: +34 963 866607.
[4] Fax: +34 848 423474.
[5] Fax: +34 881 546571.
[6] The Cumulative False Positive Risk Group (alphabetical order):Department of Epidemiology and Evaluation, Institut Municipal d'Investigació Mèdica-Parc de Salut Mar, Barcelona: Castells X., Macià F., Murta C., Román R., Sala M.; Galician breast cancer screening programme. Public health & Planning Directorate. Health Office, Galicia: Almazán R., Fernández AB., Queiro MT, Zubizarreta R.; Navarra Breast Cancer Screening Programme Public Health Institute, Pamplona: Ascunce N., Delfrade I., Ederra M., Erdozain N., Vidan J.; General Directorate Public Health & Centre for Public Health Research (CSISP), Valencia: Ibáñez J., Salas D.; Valencian Health Agency & Centre for Public Health Research (CSISP), Valencia: Cuevas D.; Servicio Canario de la Salud, Canary Islands: De la Vega M, Díez de la Lastra I.; Foundation Society for Cancer Research and Prevention. Pere Virgili Health Research Institute, Reus, Tarragona: Galceran J.; Program & Analysis Unit. Health Office, Asturias: Natal C.; La Rioja Breast Cancer Screening Programme. Fundacion Rioja Salud, Logroño: Baroja A.; Cancer Screening and Epidemiology Department, UDIAT-CD. Corporació Parc Taulí-Institut Universitari Parc Taul* (UAB), Sabadell: Baré M; and Castilla-Leon Breast Cancer Screening Programme. Dirección General de Salud Pública ID e I. SACYL. Castilla y León: González-Román I.

0091-7435/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ypmed.2011.04.013

Introduction

Screening mammography is widely used to reduce mortality from breast cancer (Deck and Kakuma, 2006; Gotzsche and Nielsen, 2009; Nystrom et al., 2002; Tabar et al., 1985). One of the most important adverse effects of this practice is false-positive (FP) results. Some studies have estimated the cumulative risk of an FP result to be between 10% and 50% over 10 screening rounds (Elmore et al., 1998; Hofvind et al., 2004; Castells et al., 2006; Njor et al., 2007). Consensus is lacking on the optimal age to start screening (US Preventive Services Task Force, 2009; Woolf, 2010; Woloshin and Schwartz, 2010; Murphy, 2010; Berg, 2010; Yankaskas et al., 2010); some organisations recommend screening from 50 years (Council of the European Union, 2003) and others from 40 years (NCI Statement on Breast Cancer Screening, 2010). In 2009, the US Preventive Services Task Force revised their recommendations for screening and advocated a start age of 50 years (US Preventive Services Task Force, 2009). In Europe, 15 countries begin screening at 50 years, two at 40 years and two at 45 years (Karsa et al., 2008).

The effect of start age on FP risk has been little studied and few reports have included women who started screening at 40 years (Elmore et al., 1998; Armstrong et al., 2007). The risk of an FP result seems to be higher in younger women, which may be explained by their greater breast density and by other variables such as menopausal status and the use of hormone replacement therapy (HRT) (Carney et al., 2003), which has decreased in recent years (Barbaglia et al., 2009). In addition, the technical quality of mammography and reading protocols has improved and digital mammography has been widely adopted, although the impact of this type of mammography on FP results is currently under debate (Bick and Diekmann, 2007; del Turco et al., 2007; Skaane 2009; Sala et al., 2009).

The aims of the present study were to estimate the risk of an FP result leading to any type of procedure and to an invasive procedure (FPI) according to the start age of screening (50–51 or 45–46 years), to analyse the association between this risk and age, screening period and birth cohort, adjusted by personal and organisational variables, and to determine the cumulative risk of FP and FPI results after 10 screening rounds for each of the two start ages.

Methods

Design

Eight of the seventeen regional breast cancer screening programmes in Spain covering 44% of the target population participated in this study (Principality of Asturias, Canary Islands, Castile–León, Catalonia, Valencian region, Galicia, Navarre, and La Rioja). These programmes complied with the European guidelines for quality assurance in breast cancer screening (Perry et al., 2006). Four programmes start screening women aged 45 years and the remaining four start at 50 years (Ascunce et al., 2010). All programmes provided information on the mammograms performed in at least three consecutive rounds, the type of additional examinations carried out and the histopathological result confirming or excluding a diagnosis of cancer. The data were routinely collected and anonymised for the study.

Study population

The records for 1,565,364 women were included in this study. Age ranged from 45 to 69 years (157,656 started screening at 45–46 years and 251,275 at 50–51 years). There were 4,739,498 mammograms (428,710 mammograms corresponded to women starting screening at 45–46 years and 481,329 to women starting at 50–51 years) (Table 1). Women were invited to participate every 2 years, with a maximum of 10 or 12 screening rounds. The study period was from 1990 to 2006.

Definition of variables

FP results were divided into two types: (i) FP, defined as the recommendation for any additional procedure to exclude malignancy (additional views, ultrasound or other imaging tests and/or fine- or thick-needle aspiration biopsy, biopsy or other invasive tests) not leading to a definitive diagnosis of breast cancer, and (ii) FPI, defined as the recommendation for fine- or thick-needle biopsy, biopsy and/or other invasive tests to exclude malignancy without a final diagnosis of breast cancer.

The screening period was divided into four categories (1990–1994, 1995–1998, 1999–2002, and 2003–2006). Participating women were born between 1923 and 1962; the cohort was distributed into those born before 1940, from 1940 to 1949, and after 1949.

Programme-related variables consisted of the number of views (one or two), type of reading [a single radiologist (single reading) or two radiologists (double reading)] and mammographic technique (film-screen or digital). Participants' personal variables consisted of HRT use at screening or in the previous 6 months, the presence of a family history of breast cancer in first-degree relatives, menopausal or premenopausal status, and prior invasive tests with a benign result before the screening examination. The first mammogram in the programme was considered the first screen and the remainder were considered successive screens.

Statistical analysis

The FP, FPI, cancer detection rates and positive predictive values (PPV) in the study population were calculated for each age group at the first and successive screens. Differences between proportions were calculated using Wald's method.

To estimate the risk of an FP result, discrete time-hazard models were used (Singer and Willett, 2003a, 2003b). Two types of predictors were introduced: 'time indicators' given by the women's screening round (dummy variables acting as multiple intercepts) and 'substantive predictors' to evaluate the effect of covariates in the model. The event of interest was the occurrence of an FP result. To achieve independence among screens from the same participant, each participant had as many entries as mammograms until the first FP recall. Separate models were adjusted for FP and FPI. The models included a period effect variable.

Because of the correlation among FP results observed within single radiology units, the statistical models were adjusted by considering the radiology unit as a random effect. The GLIMMIX procedure in SAS 9.1 (SAS Institute, Cary, NC) was used. The models had two levels, in which screening mammograms (level 1) were nested within radiology units (level 2, random effect). Residual pseudo-likelihood estimation was used in the models.

Two different analyses were performed. A model with the whole database including start age, age at each screening, screening period, and birth cohort was computed to ascertain the effect of temporal variables on FP risk over the study period. A second analysis was performed with the subset of women who started screening at 45–46 years and at 50–51 years to ascertain the differential risk between the distinct start ages among the screening programmes, taking into account the programme-related and personal variables. To avoid confusion due to possible time-related changes, the analysis included age at screening and screening period. Validation of the analysis was based on deviance and Akaike's information criterion. Both analyses were performed for the two different response variables previously defined (FP and FPI).

The cumulative FP risk was projected forwards to 10 screening participations for women aged 45–46 and 50–51 years at their first screening round. The cumulative FP risk was projected forwards 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. Confidence intervals (CI) for the cumulative risk of an FP

Table 1
Study population. Eight Spanish breast cancer screening programs from 1990 to 2006.

	Age (years)	Screened women (N)	Mammograms (N)
Total study population	45–69	1,565,364	4,739,498
	Start age (years)	**Screened women (N)**	**Mammograms (N)**
Subset study population	45–46	175,656	428,710
	50–51	251,275	481,329
Total subset		426,931	910,039

Table 2
False positive results leading to any procedure and to invasive procedures, cancer detection and positive predictive values for the first and successive screens by age at screening, 1990–2006, Spain.

	First screen					Successive screens				
	Screens [a]	False positive (any procedure)	False positive (invasive procedures)	Cancer detection	Positive predictive value	Screens	False positive (any procedure)	False positive (invasive procedures)	Cancer detection	Positive predictive value
Age at screening (years)	n	n (%)	n (%)	n (%)	(%)	n	n (%)	n (%)	n (%)	(%)
45–49	467,910	54,699 (11.69)	5,475 (1.17)	1,523 (0.33)	2.71	304,079	23,993 (7.89)	1,324 (0.44)	725 (0.24)	2.93
50–54	477,177	37,601 (7.88)	5,002 (1.05)	1,790 (0.38)	4.54	749,285	36,838 (4.92)	2,439 (0.33)	1,795 (0.24)	4.65
55–59	300,901	21,157 (7.03)	2,716 (0.90)	1,405 (0.47)	6.23	899,299	30,639 (3.41)	2,070 (0.23)	2,484 (0.28)	7.50
60–64	260,223	17,339 (6.66)	2,214 (0.85)	1,724 (0.66)	9.04	833,028	25,914 (3.11)	1,777 (0.21)	2,817 (0.34)	9.80
65–69	59,153	3,961 (6.70)	487 (0.82)	623 (1.05)	13.59	388,443	12,660 (3.26)	932 (0.24)	1,643 (0.42)	11.49
Total	1,565,364	134,757 (8.61)	15,894 (1.02)	7,065 (0.45)	4.98	3,174,134	130,044 (4.10)	8,542 (0.27)	9,464 (0.30)	6.78

[a] At the first screen, the number of screening mammograms and the number of screened women were equivalent; N = 1,565,364 screened women; N = 4,739,498 mammograms.

result were calculated using Greenwood's approximation (Singer and Willett, 2003a, 2003b).

Ethical considerations

This study was performed in accordance with the principles of the declaration of Helsinki and the Spanish legal requirements of confidentiality.

Results

The FP and FPI rates in the first (11.69% and 7.88%, respectively) and successive screens (1.17% and 1.05%) were higher in women aged 45–49 than in those aged 50–54 years. The cancer detection rate was higher in women aged 50–54 at the first screening (0.38% versus 0.33%) and was equal to that in younger women at successive screens (0.24%). The PPV for additional procedures was higher in women aged 50–54 years at the first (4.54% versus 2.71%) or at successive screens (4.65% versus 2.93%). (Table 2).

The multiple regression model (Table 3) included the woman's age at first screening, the woman's current age at mammography screening, the screening period, and the woman's birth cohort. A statistically significant interaction was found between age at first screening and current age at screening, (Fig. 1). The FP risk was higher in screening mammograms performed in the 1995–1998 period (OR = 1.14; 95%CI: 1.12–1.16) and the FPI risk was higher in mammograms performed in the 1990–1994 period (OR = 1.50; 95% CI: 1.40–1.60) compared with the 2003–2006 period. FP and FPI risks were also higher in women in the oldest birth cohorts (OR = 1.21; 95% CI: 1.18–1.25 and OR = 1.17; 95%CI: 1.07–1.27, respectively) than in the youngest birth cohorts.

Fig. 1 shows the estimated OR for the interaction between age at first screening and current age at screening, obtained from the regression model (Table 3). For both FP (Fig. 1a) and FPI (Fig. 1b) the younger the start age of screening, the greater the FP risk, while the higher the age at screening, the lower the risk.

The analysis of the subset of women who started screening at 45–46 or 50–51 years is shown in Table 4. The univariate analysis is similar to the multivariate except for age at screening that was associated with an increased risk in the univariate model (data not shown). The variables conferring a higher FP risk were a start age of 45–46 (OR = 1.20; 95% CI: 1.13–1.26) versus 50–51 years, double reading (OR = 1.36; 95%CI: 1.23–1.51) versus single reading, pre-menopausal (OR = 1.26; 95%CI: 1.23–1.29) versus menopausal status, prior invasive procedures (OR = 1.52; 95%CI: 1.47–1.57) and a family history of breast cancer (OR = 1.16; 95%CI: 1.12–1.20) compared with the absence of these variables. The variables conferring a lower FP risk were two views (OR = 0.59; 95%CI: 0.56–0.61) versus a single view and digital (OR = 0.64; 95% CI: 0.56–0.73) versus screen-film mammography.

The variables increasing FPI risk were a start age of 45–46 years (OR = 1.43; 95%CI: 1.18–1.73), double views (OR = 1.34; 95%CI: 1.18–1.52), premenopausal status (OR = 1.22; 95%CI: 1.13–1.31), prior invasive procedures (OR = 2.08; 95%CI: 1.89–2.28) and a family history of breast cancer (OR = 1.26; 95%CI: 1.13–1.41).

Fig. 2 shows the cumulative FP and FPI risks up to the 10th screening mammogram for women starting screening at 45–46 and for those starting at 50–51 years. The cumulative FP risk was

Table 3
Estimated odds ratios (OR) for false-positive risk (unadjusted and adjusted) in the multiple regression model including age at first screening, current age at screening, period, cohort effects, and interaction term [‡], 1990 to 2006, Spain.

	Screening mammograms	False positive (any procedure)		False positive (invasive procedure)	
		Univariate analysis [unajusted OR, 95% C.I.] [±]	Multivariate analysis [adjusted OR, 95% C.I.] [†]	Univariate analysis [unadjusted OR, 95% C.I.] [±]	Multivariate analysis [adjusted OR, 95% C.I.] [†]
Period (at screening)					
1990–1994	301,213	1.05 (1.02, 1.07) *	1.03 (1.01, 1.05) *	1.52 (1.42, 1.63) *	1.50 (1.40, 1.60) *
1995–1998	856,833	1.16 (1.14, 1.18) *	1.14 (1.12, 1.16) *	1.31 (1.24, 1.37) *	1.28 (1.22, 1.35) *
1999–2002	1,564,396	1.11 (1.10, 1.13) *	1.10 (1.09, 1.12) *	1.27 (1.22, 1.32) *	1.26 (1.21, 1.31) *
2003–2006	2,017,056	Ref.	Ref.	Ref.	Ref.
Birth cohort					
<1940	1,190,456	1.07 (1.05, 1.10) *	1.21 (1.18, 1.25) *	0.99 (0.91, 1.07)	1.17 (1.07, 1.27) *
1940–1949	2,343,941	1.05 (1.03, 1.06) *	1.12 (1.10, 1.14) *	1.00 (0.96, 1.05)	1.10 (1.05, 1.15) *
≥1950	1,205,101	Ref.	Ref.	Ref.	Ref.

Age at first screening: women's age at the first screening; Age at screening: women's age when screening was performed; Period (at screening): period in which screening was performed; Birth cohort: woman's birth cohort. N = 1,565,364 screened women; N = 4,739,498 mammograms.

[‡] OR for the continuous variables 'age at first screening' and 'age at screening' and their interaction are shown graphically in Fig. 1.

* Significant at the 95% confidence level.

[±] Analysis adjusted by women's screen number and radiology unit (random effect).

[†] Multivariate analysis adjusted by women's screen number, radiology unit (random effect), and all other variables in the table.

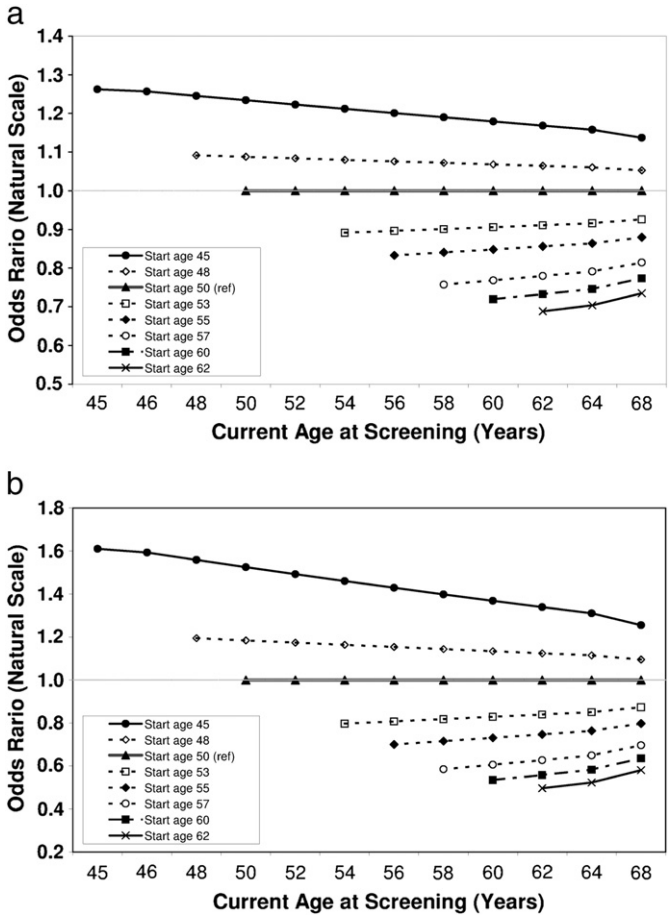


Fig. 1. Odds ratios from multivariate regression analysis showing the relationship between age at first screening mammography and false-positive results leading to any procedure (a) and to invasive procedures (b), by current age at screening. (1990–2006) Spain. Regression model adjusted by women's screening mammogram, screening period, birth cohort, and radiology unit in which the screening mammogram was performed (random effect); referent age, 50 years. See Table 3. N = 4,739,498 mammograms. Retrospective cohort study, 1990–2006.

estimated at 33.30% (95%CI: 32.91–33.70) for a start age of 45–46 and at 20.39% (95%CI: 20.02–20.76) for 50–51 years. The cumulative FPI risk was estimated as 2.68% (95%CI: 2.56–2.79) and 1.76% (95%CI: 1.66–1.87), respectively.

Table 4
Estimated odds ratios (OR) for the false-positive risk in the multiple regression model including women's personal variables and protocol-related characteristics for women starting screening at ages 45–46 versus 50–51 years, 1990 to 2006, Spain.

	Screening mammograms	Multivariate analysis [adjusted OR, 95% C.I.] †	
		False positive (any procedure)	False positive (invasive procedure)
Start age (years)			
45–46	481,329	1.20 (1.13, 1.26) *	1.43 (1.18, 1.73) *
50–51	428,710	Ref.	Ref.
Age at screening (years)			
44–49	302,998	0.89 (0.76, 1.04)	0.66 (0.39, 1.13)
50–54	388,622	0.88 (0.77, 0.99) *	0.79 (0.52, 1.21)
55–59	174,465	0.91 (0.81, 1.02)	0.79 (0.54, 1.15)
≥60	43,954	Ref.	Ref.
Period (at screening)			
1990–1994	39,720	1.18 (1.13, 1.23) *	1.67 (1.47, 1.89) *
1995–1998	95,383	1.20 (1.17, 1.24) *	1.44 (1.30, 1.59) *
1999–2002	278,325	1.11 (1.09, 1.14) *	1.18 (1.10, 1.26) *
2003–2006	496,611	Ref.	Ref.
Reading method			
Simple reading	171,083	Ref.	Ref.
Double reading	738,956	1.36 (1.23, 1.51) *	1.04 (0.67, 1.62)
Number of views			
One	430,110	Ref.	Ref.
Two	479,929	0.59 (0.56, 0.61) *	1.34 (1.18, 1.52) *
Mammography type			
Film-screen	904,061	Ref.	Ref.
Digital	5,978	0.64 (0.56, 0.73) *	0.70 (0.43, 1.14)
HRT			
No	811,931	Ref.	Ref.
Yes	98,108	0.99 (0.95, 1.02)	0.78 (0.69, 0.89)*
Menopausal status			
Menopausal	521,387	Ref.	Ref.
Premenopausal	388,652	1.26 (1.23, 1.29) *	1.22 (1.13, 1.31) *
Previous invasive procedure			
No	846,708	Ref.	Ref.
Yes	63,331	1.52 (1.47, 1.57) *	2.08 (1.89, 2.28) *
Previous familial breast cancer			
No	850,304	Ref.	Ref.
Yes	59,735	1.16 (1.12, 1.20) *	1.26 (1.13, 1.41) *

* Significant at the 95% confidence level.
† Multivariate analysis adjusted by women's screen number, radiology unit (random effect), and all other variables in the table; Start Age: age in which women started screening; Age at screening: women's age when screening was performed; Period (at screening): period in which screening was performed; Reading method: method used to read the screening mammogram (1 or 2 radiologists); Number of views: number of images performed to read the mammogram; Mammogram type: whether the mammogram was read analogically or using a digital monitor; HRT: hormone replacement therapy; Menopause: pre/peri-menopausal or menopausal status; Previous Invasive Procedure: personal previous invasive procedure; Previous Familial Breast Cancer: previously described; N = 426,931 screened women; N = 910,039 mammograms.

Discussion

The decision to commence screening mammography at a particular age should be determined by the balance of benefits and harms. A systematic review of screening mammography (Armstrong et al., 2007) concluded that the benefit–harms balance was favourable in most women with a start age of 50 or more but not in those who started screening at 40–49 years.

In our study, the cancer detection rate (an indicator of the possible benefits of screening) was significantly higher at the first screen in women aged 50–54 years than in those aged 45–49 years ($p<0.001$), while no differences were found for successive screens ($p=0.46$). Several studies have found that this rate increases with age (Elmore et al., 2005; Kerlikowske et al., 1993).

The FP rate is an indicator of adverse effects. We found differences in FP and FPI risks for the first and successive screens, depending on the age at screening. PPV was 70.98% higher (95%CI: 59.71–83.11%, $p<0.001$) at the first screen and was 53.54% higher (95%CI: 41.06–67.65, $p<0.001$) in successive screens in women aged 50–54 years at screening than in those aged 45–49 years.

We also studied the FP risk according to start age and found that this risk was higher in women who started screening at 45–46 years. After participating in 10 screening rounds, the cumulative risk in these women was 63% higher than that in women who started screening at 50–51 years (33.30% versus 20.39%).

The psychological impact of an FP result in women is much higher for invasive tests, which may also affect reattendance (Armstrong et al., 2007; Brett and Austoker, 2001). We found that the cumulative risk of an FPI was 50% higher in women with a start age of 45–46 versus 50–51 years (2.68% versus 1.76%).

Graphical representation of the interaction between start age and age at screening showed how the excess risk associated with initiating screening earlier gradually decreased as age at screening increased, tending to become equal, although the risk was always higher in younger age ranges. Screening period only influenced FPI risk, which was also higher in the oldest birth cohorts, a finding that could be due

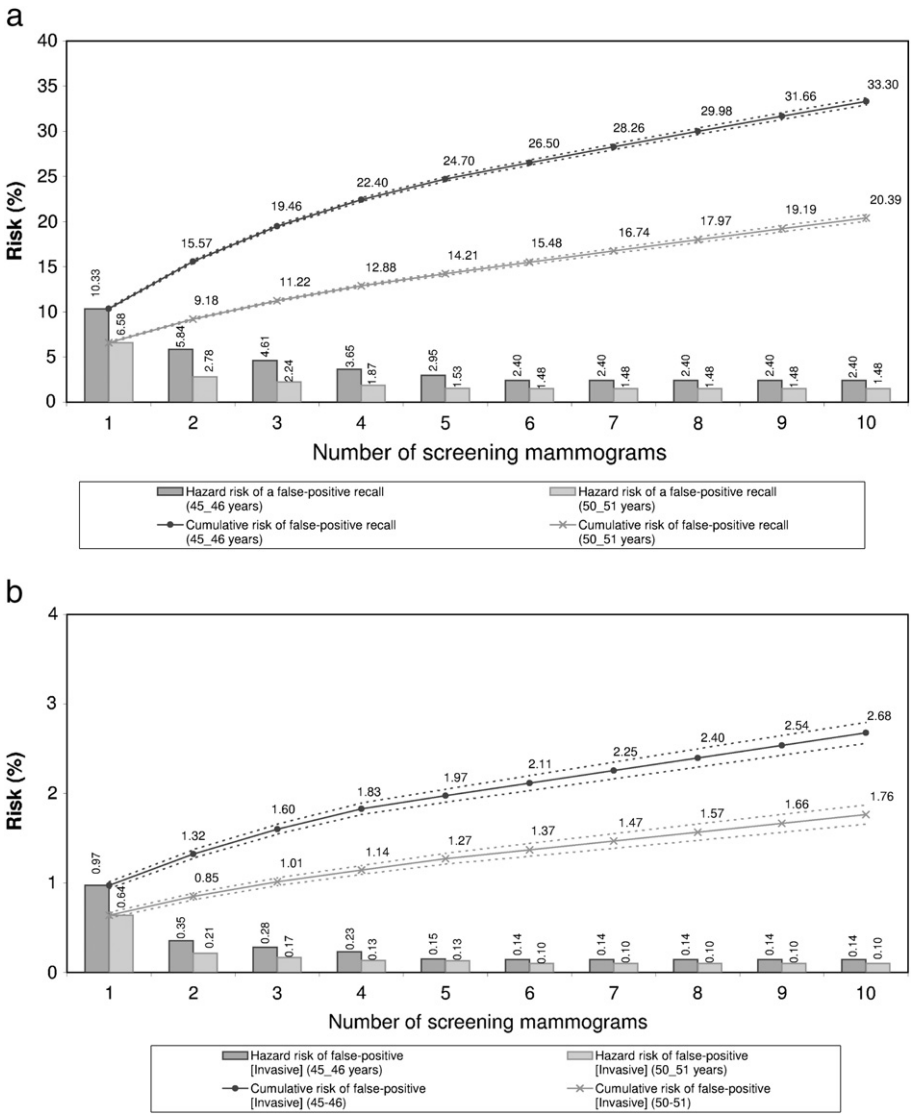


Fig. 2. Estimated cumulative risk and hazard risk of a false positive result leading to any procedures (a) and to invasive procedures (b) for women starting screening at ages 45–46 versus 50–51 years. (1990–2006) Spain. Women starting screening at age 45–46 (N = 175,656 screened women and N = 428,710 mammograms) versus 50–51 years (N = 251,275 screened women and N = 481,329 mammograms). Data were obtained from the regression models.

to more widespread use of HRT in this age group (Barbaglia et al., 2009).

In the subset of women with a start age of 45–46 or 50–51 years (Table 4), a period effect was observed independently of the other variables analyzed, especially in FPI risk, which was greater in the earlier periods (1990–1994). This difference in FP risk could be explained by the greater experience acquired over time among professionals, improved imaging hardware and changes in the criteria for clinical decisions (American College of Radiology, 2003).

One study found a higher FP risk in women aged 40–49 versus 50–59 years, with an OR = 1.1 at the first screen and an OR = 1.38 at successive screens. A family history of breast cancer at the first and successive screens, HRT use and prior breast lesions conferred a greater risk (Kavanagh et al., 2006). Another study of the combined effect of patients' risk profiles and radiologic characteristics on FP risk reported that the variables predictive of cumulative risk over eight rounds were lower age, a family history of breast cancer, prior biopsies and current oestrogen use (Christiansen et al., 2000).

In our study, the variables associated with a greater FP risk were a start age of 45–46 years, a family history of breast cancer, and prior invasive tests. However, unlike other studies (Christiansen et al., 2000) that found no association with premenopausal status, in our study this variable increased the risk of both FP and FPI (OR = 1.26 and OR = 1.22, respectively). Equally, while other studies found an association between HRT and oestrogen use and a greater FP risk (Kavanagh et al., 2006; Laya et al., 1996; Seradour et al., 1999; Thurfjell et al., 1997; Litherland et al., 1997) we did not observe this association. Although we were unable to study the combination of hormones used, the use of oestrogen plus progestin is lower in Spain than elsewhere, which may explain the discrepancies observed (Benet et al., 2002).

This study did not include mammograms performed outside organised screening programmes. Moreover, personal variables were not available in all the mammograms included, which may have influenced the results related to HRT use and other personal variables.

Comparison of our results with those of other studies is hampered by the fact that the start age in the screened population in Spain was 45 or 50 years, while other studies (Kavanagh et al., 2006; Christiansen et al., 2000) compared women aged 40 with those aged 50 years.

Few publications have reported cumulative FP and FPI risks and none have compared distinct start ages of screening. One study (Hofvind et al., 2004) estimated the cumulative FP risk after 10

completed rounds to be 20.8% for the cohort aged 50–51 years, a result similar to that of our study (20.39%) in the same age group. We found a cumulative FP risk of 33.30% in the group aged 45–46. The same study estimated a cumulative FPI risk of 6.2%, while in our study the same risk was 1.76% in the same age group and was 2.68% in the group aged 45–46.

Conclusions

Women who started screening at 45–46 years had greater FP and FPI risks than those starting at 50–51 years and this effect was maintained when adjustment was made for temporal, programme-related and personal variables. Some personal variables (premenopausal status, prior invasive procedures and a family history of breast cancer) can increase these risks. Programme-related variables such as double reading increase the FP risk and double views increase FPI risk. Screening period clearly influenced FPI.

Throughout participation in a screening programme, the cumulative risk of an FP and FPI is higher in women who start screening at an earlier age. Programmes should take the required organisational measures (type of reading and mammography technique) to reduce these effects.

Conflict of interest statement
No competing interests.

Acknowledgments

The authors thank the health professionals working in the breast cancer screening programmes for their contribution to early cancer detection and improving women's health.

This work was supported by grants from *Instituto de Salud Carlos III FEDER*, (PI061423).

References

American College of Radiology, 2003. American College of Radiology Breast Imaging Reporting and Data System (BI-RADS), fourth ed. American College of Radiology, Reston, Virginia.
Armstrong, K., Moye, E., Williams, S., et al., 2007. Screening mammography in women 40 to 49 years of age: a systematic review for the American College of Physicians. Ann. Intern. Med. 146, 516–526.
Ascunce, N., Salas, D., Zubizarreta, R., et al., 2010. Cancer screening in Spain. Ann. Oncol. 21, 43–51.
Barbaglia, G., Macià, F., Comas, M., et al., 2009. Trends in hormone therapy use before and after publication of the Women's Health Initiative trial: 10 years of follow-up. Menopause 16, 1061–1064.
Benet, R.M., Carvajal Garcia-Pando, A., Garcia, D.P., et al., 2002. Hormonal replacement therapy in Spain. Med Clin. 119 (1), 4–8 (Barc).
Berg, W.A., 2010. Benefits of screening mammography. JAMA 303, 168–169.
Bick, U., Diekmann, F., 2007. Digital mammography: what do we and what don't we know? Eur. Radiol. 17, 1931–1942.
Brett, J., Austoker, J., 2001. Women who are recalled for further investigation for breast screening: psychological consequences 3 years after recall and factors affecting re-attendance. J. Public Health Med. 23, 292–300.
Carney, P.A., Miglioretti, D.L., Yankaskas, B.C., 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.
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.
Christiansen, C.L., Wang, F., Barton, M.B., et al., 2000. Predicting the cumulative risk of false-positive mammograms. J. Natl. Cancer Inst. 92, 1657–1666.
Council of the European Union, 2003. Council recommendation of 2 December 2003 on cancer screening (2003/878/EC). Off. J. Eur. Union 327, 34–38.
Deck, W., Kakuma, R., 2006. Screening Mammography: a Reassessment Montréal (Canada): Agence d'Évaluation des Technologies et des Modes d'Intervention en Santé (AETMIS).
Del Turco, M.R., Mantellini, P., Ciatto, S., et al., 2007. Full-field digital versus screen-film mammography: comparative accuracy in concurrent screening cohorts. AJR. Am. J. Roentgenol. 189, 860–866.
Elmore, J., Barton, M., Moceri, V., et al., 1998. Ten-year risk of false positive screening mammograms and clinical breast examination. N. Engl. J. Med. 338, 1089–1096.
Elmore, J.G., Armstrong, K., Lehman, C.D., et al., 2005. Screening for breast cancer. JAMA 293, 1245–1256.
Gotzsche, P.C., Nielsen, M., 2009. Screening for breast cancer with mammography. Cochrane Database Syst. Rev., 4, p. CD001877.
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.
Karsa, L., Anttila, A., Ronco, G., et al., 2008. Cancer screening in the European Union. Report of the Implementation of the Council. Recommendation on Cancer Screening — First Report, Services of the European Commission, Luxembourg.
Kavanagh, A.M., Davidson, N., Jolley, D., et al., 2006. Determinants of false positive recall in an Australian mammographic screening program. Breast 15, 510–518.
Kerlikowske, K., Grady, D., Barclay, J., et al., 1993. Positive predictive value of screening mammography by age and family history of breast cancer. JAMA 270, 2444–2450.
Laya, M.B., Larson, E.B., Taplin, S.H., et al., 1996. Effect of estrogen replacement therapy on the specificity and sensitivity of screening mammography. J. Natl. Cancer Inst. 88, 643–649.
Litherland, J.C., Evans, A.J., Wilson, A.R.M., 1997. The effect of hormone replacement therapy on recall rate in the National Health Service Breast Screening Programme. Clin. Radiol. 52, 276–279.
Murphy, A.M., 2010. Mammography screening for breast cancer: a view from 2 worlds. JAMA 303, 166–167.
National Cancer Institute. NCI Statement on Breast Cancer Screening [Internet]. Bethesda: National Cancer Institute aat the National Institutes of Health (US); 2009 Nov 16 [updated 2010 Mar 3, cited 2010 Apr 27]. Available from: http://www.cancer.gov/newscenter/pressreleases/BreastScreen2009.
Njor, S.H., Olsen, A.H., Schwartz, W., et al., 2007. Predicting the risk of a false-positive test for women following a mammography screening programme. J. Med. Screen. 14, 94–97.
Nystrom, L., Andersson, I., Bjurstam, N., et al., 2002. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet 359, 909–919.
Perry, N., Broeders, M., de Wolf, C., et al. (Eds.). 2006. European guidelines for quality assurance in breast cancer screening and diagnosis. Office for Official Publications of the European Communities, Luxembourg.
Sala, M., Comas, M., Macià, F., et al., 2009. Implementation of digital mammography in a population-based breast cancer screening program: effect of screening round on recall rate and cancer detection. Radiology 252, 31–39.
Seradour, B., Esteve, J., Heid, P., et al., 1999. Hormone replacement therapy and screening mammography: analysis of the results of the du Rhône programme. J. Med. Screen. 6, 99–102.
Singer, J., Willett, J.B., 2003a. Describing discrete-time event occurrence data. Applied Longitudinal Data Analysis: Modelling Change and Event Occurrence. Oxford University Press, New York, pp. 325–356.
Singer, J., Willett, J.B., 2003b. Fitting basic discrete-time hazard models. Applied Longitudinal Data Analysis: Modelling Change and Event Occurrence. Oxford University Press, New York, pp. 357–467.
Skaane, P., 2009. Studies comparing screen-film mammography and full-field digital mammography in breast cancer screening: updated review. Acta Radiol. 50, 3–9.
Tabar, L., Gad, A., Holmberg, L., et al., 1985. Significant reduction in advanced breast cancer. Results of the first seven years of mammography screening in Kopparberg. Sweden. Diagn. Imag. Clin. Med. 54, 158–164.
Thurfjell, E.L., Holmberg, L.H., Persson, I.R., 1997. Screening mammography: sensitivity and specificity in relation to hormone replacement therapy. Radiology 203, 339–341.
US Preventive Services Task Force, 2009. Screening for breast cancer: US Preventive Services Task Force recommendation statement. Ann. Intern. Med. 151, 716–726.
US Preventive Services Task Force. Screening for Breast Cancer, 2009 [Internet]. Rockville: Agency for Healthcare Research and Quality; 2009 Dec [updated Dec 2009; cited 2010 Apr 27]. Available from http://www.uspreventiveservicestaskforce.org/uspstf/uspsbrca.htm.
Woloshin, S., Schwartz, L.M., 2010. The benefits and harms of mammography screening: understanding the trade-offs. JAMA 303, 164–165.
Woolf, S.H., 2010. The 2009 breast cancer screening recommendations of the US Preventive Services Task Force. JAMA 303, 162–163.
Yankaskas, B.C., Haneuse, S., Kapp, J.M., et al., 2010. Performance of first mammography examination in women younger than 40 years. J. Natl. Cancer Inst. 102, 692–701.

3. Artículo anexo 3

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

Autores: Zubizarreta R, Fernández AB, Almazán R, Román M, Velarde J, Queiró T, Natal C, Ederra M, Salas D, Castells X, Cumulative False Positive Risk Group.

Revista: Eur Radiol 2011; 21(10): 2083-2090.

Factor de impacto: 3.5 (Q1 Radiology)

DOI: 10.1007/s00330-011-2160-0.

Abstract:

Background: 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.

Conclusions: Radiologist experience is a determining factor in the risk of a false-positive result in breast cancer screening.

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.

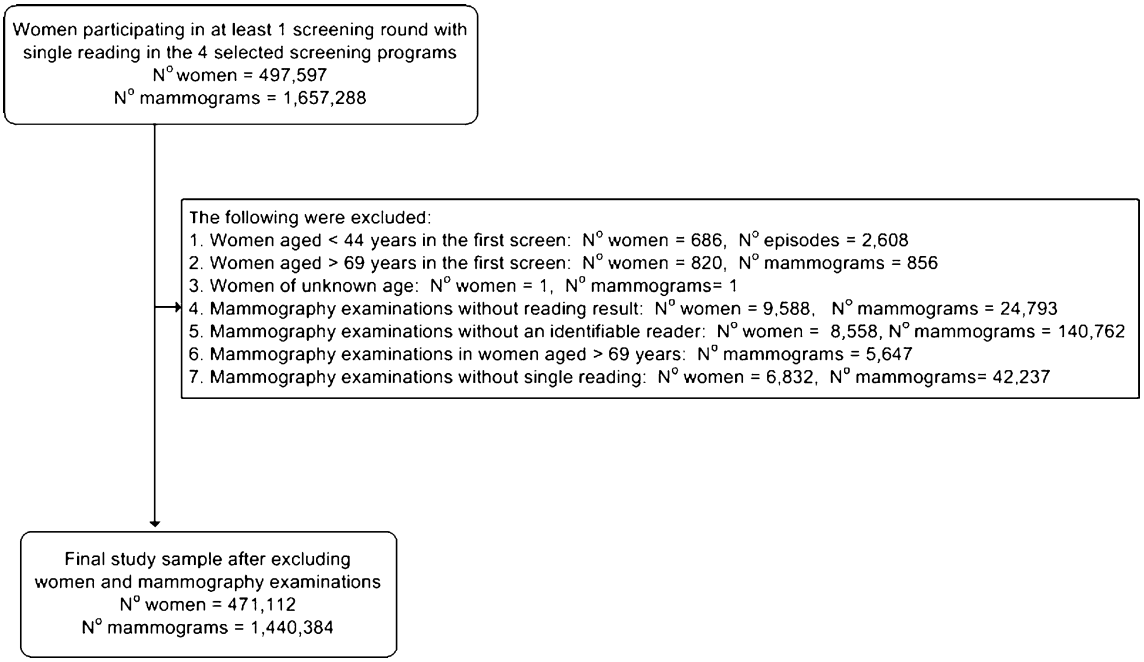


Fig. 1 Initial study population, final sample, number of excluded women and mammograms with the reasons for exclusion

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

4. Protocolo de variables de estudio del proyecto sobre riesgo acumulado de falsos positivos (Proyecto RAFP)

ANÁLISIS DE LA PROBABILIDAD ACUMULADA DE AL MENOS UN FALSO POSITIVO EN EL CRIBADO MAMOGRÁFICO EN ESPAÑA EN LA INDICACIÓN DE ALGUNA EXPLORACIÓN ADICIONAL INVASIVA (PAAF, CORE-BIOPSIA O BIOPSIA QUIRÚRGICA)

PROTOCOLO DE LAS VARIABLES DE ESTUDIO

31 de Enero de 2008

CRITERIOS DE INCLUSIÓN

1. Los programas disponen de información de al menos tres rondas de cribado consecutivas (e incluye la primera ronda).

2. Los programas disponen de un identificador único por mujer que permite enlazar (linkar) a la misma mujer a través de las distintas rondas de cribado (como si de una cohorte se tratase). En estas condiciones, para cada mujer y ronda, disponen de información sobre el resultado de la mamografía de cribado y se dispone de información respecto a las pruebas adicionales tanto invasivas como no invasivas.

3. Se incluirá en la base de datos a todas las mujeres incluidas en el programa (de cualquier edad en la fecha de exploración) con al menos 1 cribado (al menos una vez participante), pudiendo tener su primer cribado (cribado inicial) en cualquier ronda del programa.

4. Se incluirá a las mujeres con fecha de mamografía de cribado desde el inicio del programa hasta el 31 de Diciembre de 2006. Se registrarán las pruebas adicionales realizadas desde el inicio del programa hasta 30 de Junio de 2007

5. El estudio de los cánceres de intervalo / falsos negativos está fuera del marco de análisis de este estudio y no se contemplarán.

DESCRIPCIÓN DE LAS VARIABLES (VER CATEGORÍAS EN EL ANEXO 1)

Con el fin de mejorar la comprensión de las variables que vamos a recoger para el *análisis de la probabilidad acumulada de al menos un falso positivo en el cribado mamográfico en España en la indicación de alguna exploración adicional invasiva (PAAF, core-biopsia o biopsia quirúrgica)* proponemos el siguiente ejemplo:

IMAGINEMOS UN PROGRAMA DE CATALUNYA (CIUTAT VELLA) CON CINCO RONDAS (VUELTAS O CAMPAÑAS) Y UNA MUJER NACIDA 15/01/1948 Y CONVOCADA A PARTIR DE LA SEGUNDA (FECHA PRIMERA CONVOCATORIA 01/03/1998) Y HASTA LA QUINTA RONDA. SUPONGAMOS QUE LA MUJER HA PARTICIPADO EN LA SEGUNDA, CUARTA Y QUINTA RONDA DEL PROGRAMA.

SUPONGAMOS QUE SU TERCERA MAMOGRAFÍA DE CONTROL EN EL PROGRAMA (RONDA 5 DEL PROGRAMA) SE REALIZA EL 23/04/04. TRAS LA LECTURA DE LOS RADIÓLOGOS NO ES POSIBLE DESCARTAR MALIGNIDAD (MAMOGRAFÍA UNA PROYECCIÓN Y DOBLE LECTURA). EN FECHAS POSTERIORES A LA MAMOGRAFÍA SE LE REALIZAN CONSECUTIVAMENTE LAS SIGUIENTES PRUEBAS: ECOGRAFÍA, PAAF Y BIOPSIA QUIRÚRGICA. FINALMENTE NO SE DIAGNOSTICA CÁNCER DE MAMA PERO SE LE RECOMIENDA A LA MUJER LA REALIZACIÓN DE UNA MAMOGRAFÍA INTERMEDIA A LOS 6 MESES. PASADOS LOS 6 MESES (15/11/04), SE LE REALIZA LA MAMOGRAFÍA Y JUSTO DESPUÉS DE LA MAMOGRAFÍA (EN ESA MISMA FECHA), SE LE REALIZA UNA ECOGRAFÍA Y UNA BIOPSIA ESCISIONAL PERCUTÁNEA, CON UN RESULTADO DEFINITIVO HISTOLÓGICO DE CÁNCER DE MAMA.

1) CCAA (1 registro por CCAA)

• *CCAA_id:*

Código que se asignará a cada una de las Comunidades Autónomas participantes en el proyecto. Se van a enumerar sucesivamente.

• *Nombre de la CCAA:*

Descripción (etiqueta) de la Comunidad Autónoma.

EJEMPLO:

CATALUNYA

2) Unidad Radiológica de Cribado (1- N registros por CCAA / 1 registro por episodio)

• *URC_id:*

Código de cada una de las Unidades de Cribado donde se realizan las exploraciones. Se entenderá como URCs, unidades organizativas independientes dentro de una misma CCAA, sean estas fijas o móviles. Se van a enumerar las URC de forma sucesiva a partir del número 1, sin ningún orden específico. En el caso de CCAA sin estructura de URC bastará con poner 1 en esta variable y el nombre de la CCAA en la descripción de la URC.

• *Descripción de la URC:*

Nombre y/o descripción (etiqueta) de la Unidad de Cribado donde se realizan las exploraciones. Las CCAA sin estructura de URC la etiquetarán con el nombre de la propia CCAA. Aquellas CCAA que no identifiquen las URC, las etiquetaran con el nombre de la Comunidad seguido del número de URC_id que le hayan asignado (Comunidad1, Comunidad2,..., ComunidadN)

3) Mujeres (1 registro por mujer)

• *Mujer_id*:

Número identificador, interno de los programas, de la mujer. Para una CCAA no puede haber dos mujeres con el mismo número identificador. Una mujer debe mantener durante todo el tiempo del estudio el mismo número, aunque cambie de unidad de exploración a lo largo del periodo

• *Fecha de nacimiento:*

Fecha de nacimiento de la mujer.

EJEMPLO:

15/01/1948

• *Fecha de la 1ª citación en el programa:*

Fecha en la cual la mujer es invitada a participar por 1a vez en la URC (independientemente de si participa o no).

09/09/9999 si desconocido

EJEMPLO:

01/03/1998

4) Episodios (1 episodio por ronda del programa / 1-N registros por mujer)

• *Episodio:*

Indica el número ordinal de convocatoria (o invitación) de la mujer. El episodio incluye todo el proceso que transcurre desde la primera citación de una mujer para realizarse la mamografía rutinaria hasta el resultado final del episodio (cáncer, no cáncer, pendiente de exploraciones adicionales, pendiente mamografía intermedia, seguimiento incompleto, desconocido). Una mamografía intermedia no puede ser un episodio, aunque, dentro de un episodio, puede haber mamografías intermedias. **Solamente se van a registrar episodios de mujeres con mamografía realizada (participantes),** de manera que los episodios de mujeres no participantes (cribado externo) no van a quedar registrados en la base de datos. Por lo tanto, una mujer participante en una ronda del programa concreta genera exactamente un episodio, y para esta misma mujer, la base de datos contendrá tantos episodios como participaciones tenga.

Se excluirán los episodios derivados de la petición de una mamografía por parte de una mujer por síntomas.

Existen excepciones en que puede ocurrir que la mujer tenga un 'número de episodio' mayor al 'número de ronda del programa':

- *Mujeres con realización de una nueva mamografía de cribado por cambio de residencia (por ejemplo, dos mamografías realizadas en un mismo año en distintas zonas geográficas), se considerará dicha mamografía como un episodio nuevo.*
- *Mamografías de cribado realizadas a intervalos de un año.*
    - *Mujeres con antecedentes familiares*
    - *Edad en la exploración menor de 50 años*

EJEMPLO:

LA BASE DE DATOS DE EPISODIOS CONTENDRÍA TRES REGISTROS PARA ESTA MUJER QUE SERÍAN RELLENADOS CON LOS NÚMEROS 1,3 Y 4 CONSECUTIVAMENTE.

• *Número de cribados:*

Número ordinal que indica el número de exploraciones de cribado que una mujer lleva realizadas hasta el episodio (incluyendo la del episodio).

EJEMPLO:

EL CAMPO NÚMERO DE CRIBADOS SERÍA RELLENADO CON LOS NÚMEROS 1, 2 Y 3 CONSECUTIVAMENTE PARA CADA UNO DE LOS TRES EPISODIOS.

• *Número de ronda de la URC:*

Identifica el número de ronda (vuelta, campaña) del programa en cada episodio.

EJEMPLO:

EL CAMPO NÚMERO DE RONDA DE LA URC SERÍA RELLENADO CON LOS NÚMEROS 2, 4 Y 5 CONSECUTIVAMENTE PARA CADA UNO DE LOS TRES EPISODIOS.

• *Fecha mamografía de cribado:*

Fecha de realización de la exploración de cribado (test de cribado).

EJEMPLO:

23/04/04.(EN EL EPISODIO CODIFICADO COMO 4).

• *Lector_id1:*

Identificador del radiólogo lector número 1 de la mamografía de cribado.

• *Lector_id2:*

Identificador del radiólogo lector número 2 de la mamografía de cribado.

• *Lector_id3:*

Identificador del radiólogo lector número 3 de la mamografía de cribado (Ej: el de arbitraje).

• Densidad mama:

Se codificará el resultado global de ambas mamas.

• *Método de lectura:*

Indicación del tipo de lectura en el episodio para una mujer.

EJEMPLO:

DOBLE LECTURA (EN EL EPISODIO CODIFICADO COMO 4).

• *Nº de proyecciones:*

Indicación del número de proyecciones en el episodio para una mujer.

EJEMPLO:

UNA PROYECCIÓN (EN EL EPISODIO CODIFICADO COMO 4).

• *Número de estudios intermedios:*

Número total de estudios con mamografía, u otras pruebas, intermedias dentro de un episodio, realizados por indicación del programa independientemente de cuando y dónde se realicen. Si dentro del episodio no se ha realizado ningún estudio intermedio, el valor de la variable será 0 (cero).

EJEMPLO:

1 ESTUDIO INTERMEDIO (EN EL EPISODIO CODIFICADO COMO 4).

• *Tipo de mamografía:*

Permitirá recoger la información sobre si la mamografía realizada es convencional o digital (variable ligada al proyecto de la mamografía digital).

• *Resultado inicial de la mamografía de cribado*

Esta variable, juntamente con el resultado final del episodio, va a permitir evaluar el riesgo acumulado de falsos positivos de la mamografía como test de cribado.

Se indicará el resultado inicial de la mamografía de cribado para ese episodio, aunque posteriormente, y durante el mismo episodio, haya otros resultados de otros tests (mamografías intermedias, pruebas de imagen o exploraciones adicionales invasivas). En el caso de repetición técnica, se tomará como inicial la técnica.

Se codificará de la siguiente manera:

- *1: Negativo.*
- *2: Exploraciones adicionales: mujeres a las que se les recomienda la realización de alguna prueba o exploración adicional para descartar malignidad independientemente de cuando y dónde se realicen (estas pruebas pueden realizarse en la misma fecha de la exploración de cribado o en alguna fecha posterior previa reconvocatoria).*
- *3: Mamografía intermedia (Estudios intermedios): mujeres a las que a la vista del resultado de la mamografía de cribado se les recomienda la realización de una nueva mamografía*

(excepcionalmente otra prueba de imagen) antes de la que le correspondería de forma rutinaria (por ej. a los 3, 6 o 12 meses)

• *99: Desconocido*

EL CAMPO RESULTADO (INICIAL) DE LA MAMOGRAFÍA SERÍA RELLENADO CON LA ETIQUETA: EXPLORACIONES ADICIONALES (EN EL EPISODIO CODIFICADO COMO 4).

• *Resultado final del episodio:*

El resultado final del episodio es el diagnóstico definitivo del mismo, tenga o no tenga exploraciones adicionales, tenga o no tenga mamografías intermedias, sea cual sea la casuística durante el episodio. Se codificará de la siguiente manera:

• *1: Cáncer: el diagnóstico definitivo (histológico) del episodio es de cáncer de mama. No se incluirán los lobulares in situ.*

• *2: No cáncer: el resultado del episodio es de no cáncer de mama y se recomienda a la mujer un control rutinario.*

• *3: Pendiente de exploraciones adicionales: mujeres con episodios incompletos. Como se incluirá a las mujeres desde el inicio del programa hasta el 31 de Diciembre de 2006 es posible que durante el último período haya mujeres que están todavía pendientes de la realización (o de los resultados) de pruebas adicionales y no se pueda establecer un resultado final del episodio.*

• *4: Pendiente estudios intermedios: mujeres con episodios incompletos. Como se incluirá a las mujeres desde el inicio del programa hasta el 31 de Diciembre de 2006 es posible que durante el último período haya mujeres que están todavía pendientes de la realización de un estudio intermedio y no se pueda establecer un resultado final del episodio.*

• *5: Seguimiento incompleto: mujeres a las que no se puede completar las exploraciones recomendadas en el episodio (pruebas adicionales, exploraciones intermedias, etc) por ejemplo por rechazo a las mismas por parte de la mujer, fallecimiento antes de completar proceso diagnóstico etc.*

• *99: Desconocido: mujeres de las que no se conoce el resultado final del episodio (cáncer o no cáncer).*

En los programas de detección precoz con un circuito que no permita tener la información de todas la pruebas adicionales (por ejemplo, en el caso de un programa que tiene la información de algunas pruebas, pero la mujer puede ser derivada en algún momento a seguimiento hospitalario, momento a partir del cual la información sobre las pruebas que se realizan se desconocen), se procederá de la siguiente manera:

a. ***Si se conoce el resultado final del episodio*** *(Cáncer / No cáncer),*
   - *Se indicará dicho resultado.*
   - *Se informará en la base de datos de todas las pruebas adicionales que sean conocidas para este episodio si existen (invasivas y no invasivas).*
   - *Se considerará el proceso de seguimiento hospitalario como un estudio intermedio*
   - *Puesto que no todas las pruebas adicionales son conocidas (por derivación a seguimiento hospitalario):*
     i. *Si el resultado final del episodio es negativo (No cáncer): Se contabilizará una exploración adicional que se codificará como "prueba desconocida realizada durante el seguimiento hospitalario". Se reserva la codificación 88 para estos casos particulares. (Prueba_id = 88). El resultado de esta prueba desconocida será siempre 99 (desconocido).*
     ii. *Si el resultado final del episodio es positivo (Cáncer) se contabilizará una "prueba invasiva desconocida" (2.9) con resultado positivo.*

b. ***Si no se conoce el resultado final del episodio*** *éste será clasificado como 99 (Desconocido). En dicho caso, si existen, se informará en la base de datos de todas las pruebas adicionales conocidas.*

En el caso de mujeres pendientes de alguna exploración adicional derivada de una mamografía intermedia se codificará como "3: Pendiente de exploraciones adicionales".

EL CAMPO RESULTADO FINAL DEL EPISODIO NÚMERO 4 SERÍA RELLENADO CON LA ETIQUETA: CÁNCER.

5) Exploraciones adicionales (0-N registros por episodio)

• *N prueba:*

Es el contador del número de pruebas realizadas a una misma mujer dentro de cada episodio. Para cada prueba realizada a la mujer se le asignará el número de prueba que le corresponde.

• *Prueba_id:*

Se recogerán todas las pruebas realizadas en cada episodio aunque sean del mismo tipo (Ej: dos PAAF). En el caso de una mujer-episodio con pruebas en ambas mamas en la misma fecha, se registrará únicamente la más maligna.

Se codificará de la siguiente manera:

1. Pruebas no Invasivas

1.1. Otras proyecciones mamográficas

1.2. Ecografía

1.3. Resonancia magnética

1.4. Otras pruebas no invasivas

1.9. Prueba no invasiva desconocida

2. Pruebas Invasivas

2.1. PAAF

2.2. Biopsia aguja gruesa

2.3. Biopsia asistida por vacío

2.4. Biopsia escisional percutánea

2.5. Biopsia quirúrgica

2.6. Otras pruebas invasivas

2.9. Prueba invasiva desconocida

99. Desconocido

Además, se incluye la siguiente codificación para aquellos programas en que la mujer puede ser derivada a seguimiento hospitalario y se desconocen todas las pruebas realizadas.

88. Prueba desconocida realizada durante el seguimiento hospitalario

DENTRO DEL EPISODIO NÚMERO 4 (TERCER CRIBADO DE LA MUJER, RONDA 5 DEL PROGRAMA), EN LA BASE DE DATOS DE EXPLORACIONES ADICIONALES CONSTARÍAN LAS SIGUIENTES 5 PRUEBAS: ECOGRAFÍA, PAAF, BIOPSIA QUIRÚRGICA, ECOGRAFÍA Y BIOPSIA ESCISIONAL PERCUTÁNEA.

• *Fecha_prueba:*

09/09/9999 si desconocido.

• *Resultado prueba:*

Se incluirá el resultado de cada prueba.

La categoría no concluyente se reserva únicamente para pruebas invasivas no valorables y/o resultado no concluyente.

Se codificará de la siguiente manera:

1. Positiva, sospecha o certeza de malignidad

2. Negativa

3. No concluyente

4. Pendiente

99. Desconocido: sólo en el caso de que se confirme que la prueba se ha realizado pero no se conoce el resultado

SI SUPONEMOS QUE SE REALIZARON LAS 5 PRUEBAS ANTERIORES, ÉSTAS SERÍAN CLASIFICADAS CONSECUTIVAMENTE COMO: SOSPECHOSA, NEGATIVA (PODRIA SER NO CONCLUYENTE), NEGATIVA (PODRIA SER NO CONCLUYENTE), SOSPECHOSA, POSITIVA.

6) Variables de las mujeres (1 registros por episodio)

El objetivo es obtener información orientativa sobre el perfil de la mujer que se realiza la exploración de cribado. Puesto que algunas de las características de las mujeres pueden variar en cada ronda de cribado, las siguientes variables se recogerán para cada episodio:

• *THS*

Se evaluará el uso del THS en el momento de hacerse la exploración de cribado. Se considerará que la mujer usaba THS si lo tomaba en el momento de la mamografía de cribado o en los 6 meses anteriores.

• *Soja:*

Alimentación rica en Soja. Se considerará que la mujer usaba Soja si lo tomaba en el momento de la mamografía de cribado o en los 6 meses anteriores

• *Menopausia:*

Se considerarán dos niveles: posmenopáusica y, premenopáusica o perimenopáusica.

• *Talla*

Talla en centímetros.

• *Peso*

Peso en kilogramos.

• *Antecedentes personales de patología mamaria benigna (inespecífica):*

Se considerará que una mujer tiene antecedentes personales de patología mamaria benigna, cuando se tenga conocimiento de que ha sufrido alguna de las patologías mamarias benignas posibles, de forma previa a la exploración de cribado.

• *Antecedentes personales de prueba invasiva con resultado benigno:*

Cuando se tenga conocimiento de que la mujer ha sufrido alguna prueba invasiva con resultado benigno, previa a la exploración de cribado.

• *Antecedentes familiares de cáncer de mama:*

Se considerará que una mujer tiene antecedentes familiares de cáncer de mama cuando tiene algún familiar de primer grado afectado (madre, hermanas o hijas) con cáncer de mama.

ANEXO 1

CCAA (1 registro por CCAA)	**Codificación**
CCAA_id	Código asignado a cada CCAA
Nombre de la CCAA	Nombre de la CCAA
URC (1- N registros por CCAA, 1 registro por episodio)	**Codificación**
URC_id	Código asignado a cada Unidad Radiológica de Cribado donde se realizan las exploraciones
Descripción URC	Descripción identificativa de la Unidad Radiológica de Cribado
Mujeres (1 registros por mujer)	**Codificación**
CCAA_id	
Mujer_id	Número identificador interno del programa
Fecha de nacimiento	09/09/9999 Desconocido
Fecha 1a citación en el programa	09/09/9999 Desconocido
Episodios (1 registro por cribado de la mujer,	**Codificación**

1-N registros por mujer)	
CCAA_id	
Mujer_id	
Episodio	1,2,3,4,…
URC_id	Código asignado a cada Unidad Radiológica de Cribado donde se realizan las exploraciones
Descripción de la URC	Nombre de la URC
Número de cribados	1,2,3,4,…
Número de ronda de la URC	1,2,3,4…
Fecha mamografía de cribado	09/09/9999 Desconocido
Lector_id1	Identificador del radiólogo; 99: Desconocido
Lector_id2	Identificador del radiólogo; 99: Desconocido; [Vacío si no es pertinente]
Lector_id3	Identificador del radiólogo; 99: Desconocido;

	[Vacío si no es pertinente]
Densidad mama	1: Completamente grasa; 2: Densidad fibroglandular dispersa; 3: Densidad heterogénea; 4: Extremadamente densa; 99: Desconocido
Método de lectura	1: Lectura simple; 2: Doble lectura sin consenso (resultado más desfavorable); 3: Doble lectura con consenso (deciden ambos radiólogos); 4: Doble lectura con arbitraje; 5: Doble lectura con arbitraje a ciegas; 99: Desconocido
Nº de proyecciones	1,2
Nº de estudios intermedios	0,1,2,3,4,…
Tipo de mamografía	1: Mamografía convencional; 2: Digital indirecta / lectura placa; 3: Digital indirecta / lectura monitor; 4: Digital directa / lectura placa; 5: Digital directa / lectura monitor; 6: Otros; 99: Desconocido
Resultado inicial de la mamografía de cribado	1: Negativo; 2: Exploraciones adicionales; 3: Mamografía intermedia (Estudios

	intermedios) 99:Desconocido
Resultado final del episodio	1: Cáncer; 2: No cáncer; 3: Pendiente de exploraciones adicionales; 4: Pendiente estudios intermedios; 5: Seguimiento incompleto; 99: Desconocido
Exploraciones adicionales (0-N registros por episodio)	**Codificación**
CCAA_id	
Mujer_id	
Episodio	1,2,3,4,…
N_Prueba	1,2,3,4,…
Prueba_id	1.1. Otras proyecciones mamográficas; 1.2. Ecografía; 1.3. Resonancia magnética; 1.4. Otras pruebas no invasivas; 1.9. Prueba no invasiva desconocida; 2.1. PAAF; 2.2. Biopsia aguja gruesa; 2.3. Biopsia asistida por vacío; 2.4. Biopsia escisional percutánea; 2.5. Biopsia quirúrgica; 2.6. Otras pruebas invasivas 2.9. Prueba invasiva desconocida 99. Desconocida prueba

	88: Prueba desconocida seguimiento hospitalario
Fecha_prueba	09/09/9999 si desconocido
Resultado prueba	1:Sospecha o certeza de malignidad; 2: Negativo; 3: No concluyente; 4: Pendiente; 99: Desconocido
Variables de las mujeres (1 registro por episodio)	Codificación
CCAA_id	
Mujer_id	
Episodio	1,2,3,4,…
THS	1: Sí, en el momento del cribado o en los 6 meses anteriores; 2: No; 99: Desconocido
Soja	1: Sí, en el momento del cribado o en los 6 meses anteriores; 2: No; 99: Desconocido
Menopausia	1:Posmenopáusica; 2: Premenopáusica o

	perimenopáusica; 99: Desconocido
Talla	Talla en cm; 99: Desconocido
Peso	Peso en Kg; 999: Desconocido
Antecedentes personales de patología mamaria benigna (inespecífica)	1: Sí; 2: No; 99: Desconocido
Antecedentes personales de prueba invasiva con resultado benigno	1: Sí; 2: No; 99: Desconocido
Antecedentes familiares de cáncer de mama (madre, hermanas e hijas)	1: Sí; 2: No; 99: Desconocido

5. Protocolo de control de calidad del proyecto sobre riesgo acumulado de falsos positivos (Proyecto RAFP)

ANÁLISIS DE LA PROBABILIDAD ACUMULADA DE AL MENOS UN FALSO POSITIVO EN EL CRIBADO MAMOGRÁFICO EN ESPAÑA EN LA INDICACIÓN DE ALGUNA EXPLORACIÓN ADICIONAL INVASIVA (PAAF, CORE-BIOPSIA O BIOPSIA QUIRÚRGICA)

PROGRAMA DE CONTROL DE CALIDAD DE LA BASE DE DATOS

31 de Enero de 2008

RANGO DE VALORES

Se considera que un valor está fuera de rango, y por lo tanto es un valor imposible, cuando no está dentro del conjunto de valores predeterminados definidos para esa variable.

A continuación definimos el rango de valores para cada variable.

• *CCAA_id*

Campo definido como texto de longitud 2. Se codificará con un identificador del 1 hasta el 9 según la CCAA:

CCAA_id	Nombre CCAA
1	Asturias
2	Canarias
3	Castilla y León
4	Catalunya
5	Galicia
6	La Rioja
7	Región de Murcia
8	Navarra
9	Valencia

• *Nombre de la CCAA*

Campo definido como texto de longitud 25. Ver equivalencias en la definición de la anterior variable.

• *URC_id*

Campo definido como numérico de longitud 2. Se asignará un único valor 1 a está variable si no existe estructura de URCs (En este caso se asignará el nombre de la CCAA en la descripción de la URC).

• *Descripción de la URC*

Campo definido como texto de longitud 25. Puede ser introducido como un valor numérico, como un valor alfanumérico, o como una combinación de ambos. En el caso de que únicamente exista una URC_id para toda la CCAA, rellenar este campo con el nombre de la CCAA.

• *Mujer_id*

Campo definido como texto de longitud 15. Puede ser introducido como un valor numérico, un valor alfanumérico, o una combinación de ambos.

• *Fecha de nacimiento*

Campo definido como fecha de longitud 10, con formato 00/00/0000 correspondiente a día/mes/año (09/09/9999 si desconocido). No se añadirán restricciones de valor para esta variable, pero sí se incluirán valores poco probables.

• *Fecha de la 1ª citación en el programa*

Campo definido como fecha de longitud 10, con formato 00/00/0000 correspondiente a día/mes/año. Esta variable toma valores comprendidos entre 01/01/1989 y 31/12/2006 ó bien 09/09/9999 si es desconocido.

• *Episodio*

Campo definido como numérico de longitud 2. Será menor de 20

• *Número de cribados*

Campo definido como numérico de longitud 2. Será menor de 20

• *Número de ronda de la URC*

Campo definido como numérico de longitud 2. Esta variable toma valores en el rango 1 - 9 ó 99 si es desconocido.

• *Fecha mamografía de cribado*

Campo definido como fecha de longitud 10, con formato 00/00/0000 correspondiente a día/mes/año. Su rango de valores está comprendido entre 01/01/1989 y 31/12/2006 ó bien 09/09/9999 si es desconocido.

• *Lector_id1*

Campo definido como texto de longitud 50. Puede ser introducido con un valor numérico, con un valor alfanumérico, o con una combinación de ambos.

• *Lector_id2*

Campo definido como texto de longitud 50. Puede ser introducido con un valor numérico, con un valor alfanumérico, o con una combinación de ambos. Si no es pertinente, no se informará y se dejará su correspondiente campo 'vacío'.

• *Lector_id3*

Campo definido como texto de longitud 50. Puede ser introducido con un valor numérico, con un valor alfanumérico, o con una combinación de ambos. Si no es pertinente, no se informará y se dejará su correspondiente campo 'vacío'.

• *Densidad mama*

Campo definido como texto de longitud 2. Ver equivalencias:

1: Completamente grasa;
2: Densidad fibrogandular dispersa;
3: Densidad heterogénea;
4: Extremadamente densa;
99: Desconocido

• *Método de lectura*

Campo definido como texto de longitud 2. Ver equivalencias:

1: Lectura simple;
2: Doble lectura sin consenso (resultado más desfavorable);
3: Doble lectura con consenso (deciden ambos radiólogos);
4: Doble lectura con arbitraje;
5: Doble lectura con arbitraje a ciegas;
99: Desconocido

• *Nº de proyecciones*

Campo definido como numérico de longitud 2. Los únicos valores posibles son el 1, el 2, y el 99 si es desconocido.

• *Número de estudios intermedios*

Campo definido como numérico de longitud 2.

• *Tipo de mamografía*

Campo definido como texto de longitud 2. Ver equivalencias:

1: Mamografía convencional;
2: Digital indirecta / lectura placa;
3: Digital indirecta / lectura monitor;
4: Digital directa / lectura placa;
5: Digital directa / lectura monitor;
6: Otros;
99: Desconocido

• *Resultado inicial de la mamografía de cribado*

Campo definido como texto de longitud 2. Ver equivalencias:

1: Negativo;
2: Exploraciones adicionales;
3: Mamografía intermedia;
99: Desconocido

• *Resultado final del episodio*

Campo definido como texto de longitud 2. Ver equivalencias:

1: Cáncer;
2: No cáncer;
3: Pendiente de exploraciones adicionales;
4: Pendiente mamografía intermedia;
5: Seguimiento incompleto;
99: Desconocido

• *N prueba*

Campo definido como numérico de longitud 2.

• *Prueba_id*

Campo definido como texto de longitud 2. Ver equivalencias:

11: Otras proyecciones mamográficas;
12: Ecografía;
13: Resonancia magnética;
14: Otras pruebas no invasivas;
19: Prueba no invasiva desconocida;
21: PAAF;
22: Biopsia aguja gruesa;
23: Biopsia asistida por vacío;
24: Biopsia escisional percutánea;
25: Biopsia quirúrgica;
26: Otras pruebas invasivas;
29: Prueba invasiva desconocida;
99: Desconocido
88: Prueba desconocida realizada durante el seguimiento hospitalario

• *Fecha_prueba*

Campo definido como fecha de longitud 10, con formato 00/00/0000 correspondiente a día/mes/año (09/09/9999 si desconocido). Su rango de valores está comprendido entre 01/01/1989 y 30/06/2007 o bien 09/09/9999 si es desconocido.

> Para aquellas excepciones en que por un error en la información *la 'Fecha de prueba' sea anterior a la 'Fecha de mamografía de cribado'* la "Fecha de prueba" deberá tener el valor 09/09/9999 "Desconocido".

• *Resultado prueba*

Campo definido como texto de longitud 2. Ver equivalencias:

En el caso de realización de prueba invasiva, al menos una de las pruebas no invasivas del episodio debe estar clasificada como: positiva, sospecha de malignidad o certeza de malignidad.

1: Sospecha o certeza de malignidad;
2: Negativo;
3: No concluyente;
4: Pendiente;
99: Desconocido

• *THS*

Campo definido como texto de longitud 2. Ver equivalencias:

1: Sí, en el momento de la mamografía o en los 6 meses previos;
2: No;
99: Desconocido

• *Soja:*

Campo definido como texto de longitud 2. Ver equivalencias:

1: Sí, en el momento del cribado o en los 6 meses previos;
2: No;
99: Desconocido

• *Menopausia:*

Campo definido como texto de longitud 2. Ver equivalencias:

1: Posmenopáusica;
2: Premenopáusica o perimenopáusica;
99: Desconocido

• *Antecedentes personales de patología mamaria benigna (inespecífica).*

Campo definido como texto de longitud 2. Ver equivalencias:

1: Sí;
2: No;
99: Desconocido

• *Antecedentes personales de prueba invasiva con resultado benigno.*

Campo definido como texto de longitud 2. Ver equivalencias:

1: Sí;
2: No;
99: Desconocido

• *Antecedentes familiares de cáncer de mama*

Campo definido como texto de longitud 2. Ver equivalencias:

1: Sí;
2: No;
99: Desconocido

INCOHERENCIAS

1. Comprobar que los códigos para la variable “CCAA_id” se corresponden con los de la variable “Nombre de la CCAA” tal y como sigue:

1	Asturias
2	Canarias
3	Castilla y León
4	Catalunya
5	Galicia
6	La Rioja
7	Murcia
8	Navarra
9	Valencia

2. No pueden existir dos mujeres o más con el mismo código de “Mujer_id” para una determinada CCAA.

3. La variable “Fecha de nacimiento” tendrá formato 00/00/0000 en el orden día / mes / año y NO mes / día / año. Así mismo, el año se especificará con 4 dígitos.

4. La “Fecha de la 1ª citación en el programa” no puede ser un campo vacío. Ha de cumplir con la especificación correcta de fecha, tal y como se definió anteriormente, y además estará dentro del rango de valores predefinidos. En el caso de una mujer participante sin invitación previa (por los motivos que sea), este campo se rellenará con la fecha de realización de la primera mamografía.

5. La variable “Episodio” será un número natural estrictamente mayor o igual que 1. Es ordinal pero puede tener saltos (1,3,4…). Este campo no puede estar vacío para un registro existente. Comprobar que para una misma CCAA_id y Mujer_id no existan dos episodios iguales.

6. La variable “Número de cribados” es un contador ordinal, representado por un número natural. Empieza con el 1 y no tiene saltos (1,2,3,4…). Este campo no puede estar vació para un episodio existente. Es un número igual o menor al número de “Episodio”. Cuando se introduce un nuevo episodio con mamografía reliazada, la variable “Número de cribados” incrementa necesariamente en uno el contador.

7. El ”Número de ronda de la URC” es un número natural estrictamente mayor o igual que 1. En el caso de programas para el cual la ronda de cribado se asigna por municipios (no por Unidades de Exploración), se asignará el número de ronda del municipio en que reside la mujer.

8. La variable “Fecha mamografía de cribado” ha de ser igual o posterior a la “Fecha de 1ª citación en el programa” y además igual o inferior a “Fecha_prueba”. Tiene formato 00/00/0000 en el siguiente orden: día / mes / año. Así mismo, el año se especificará con 4 dígitos.

9. Para un mismo episodio, el “lector_id1” es distinto al “lector_id2”. Si para un mismo episodio, las variables “lector_id1” y “lector_id2” no son nulos, la variable “Método de lectura” no puede ser codificada con un 1 (lectura simple).

10. La variable “Tipo de mamografía” no puede estar vacía para un episodio concreto.

11. Si la variable “Resultado inicial de la mamografía de cribado” es igual a 1 (Negativo) para un episodio concreto, la variable “Prueba_id” quedará en blanco, es decir, no pueden existir pruebas para una mujer que tenga un resultado inicial de la mamografía negativo.

12. El “Resultado inicial de la mamografía de cribado” no puede ser un campo vacío para un episodio existente. Si el resultado inicial de la mamografía de cribado para un episodio concreto es igual a 1 (Negativo), la variable “Resultado final del episodio” será codificada necesariamente con el 2 (No cáncer).

13. Para un Mujer_id y episodio concreto, si la variable “Resultado final del episodio” se codifica como cáncer, no puede existir para esa misma mujer un episodio posterior. En el caso excepcional dónde se conozca el resultado final de cáncer una vez que se ha vuelto a invitar a la mujer y se ha realizado un nuevo cribado, esta mamografía se eliminará de la base de datos.

14. Si TODAS las pruebas no invasivas (códigos: 11, 12, 13, 14, o 19 de la variable Prueba_id) correspondientes a un episodio, tienen como valor en la variable “Resultado prueba” 2 (negativo), no debe haber ninguna prueba invasiva en el episodio.

> *Se puede dar el caso de episodios con ‘mamografía de cribado’ no negativa, resultado de TODAS las pruebas no invasivas 2 (negativo) y que posteriormente tengan una prueba invasiva. A pesar de que esta regla de validación se mantendrá como alerta de control para posibles errores de la base de datos queda registrado que esta casuística es factible*

15. Si la variable "Prueba_id" está codificada como una prueba de imagen (códigos: 11, 12, 13, 14, o 19) y la variable "Resultado prueba" es 1 (sospecha o certeza de malignidad), debe cumplirse al menos una de las condiciones siguientes:

a) Existe alguna prueba adicional invasiva en el episodio.

b) El número de controles avanzados en ese episodio es 1 o más.

c) El resultado final del episodio es 5 (Seguimiento incompleto) o 3 (Pendiente de exploraciones adicionales).

d) Existe alguna otra prueba adicional no invasiva (de cualquier tipo) en el episodio, con resultado negativo.

16. Si la variable “Resultado prueba” es codificada con un 3 (No concluyente) implica que la variable “Prueba_id” tiene que haber sido codificada necesariamente con un 21, ó 22, ó 23, ó 24, ó

25, ó 26, ó 29 (PAAF, BAG, BAV, BEP, Otra prueba invasiva). No puede haber un resultado no concluyente en una prueba no invasiva.

> *Dado que no se admite el resultado "no concluyente" en **pruebas no invasivas**, el resultado de las **pruebas no invasiva** en el caso en que remiten a la mujer a un control avanzado, y estuviese codificada como no concluyente, se codificará como **2. Negativo***

17. Si la variable "Resultado final del episodio" es codificada con un 3 (pendiente de exploraciones adicionales) implica que habrá por lo menos una prueba con "Resultado prueba" codificado como 4 (pendiente).

18. La variable "Menopausia" no puede ser codificada como 1 (posmenopáusica) en un episodio y en cualquier otro episodio posterior como 2 (premenopáusica o perimenopáusica).

19. La variable "Antecedentes personales de patología mamaria benigna (inespecífica)" si en un episodio se clasifica como 1 (Sí), en los subsiguientes no puede clasificarse como 2 (No). Debe seguir siendo clasificada como 1.

20. La variable "Antecedentes personales de prueba invasiva con resultado benigno" si en un episodio se clasifica como 1 (Sí), en los subsiguientes no puede clasificarse como 2 (No). Debe seguir siendo clasificada como 1.

21. La variable "Antecedentes familiares de cáncer de mama" si en un episodio se clasifica como 1 (Sí), en los subsiguientes no puede clasificarse como 2 (No). Debe seguir siendo clasificada como 1.

VALORES POCO PROBABLES

Los valores poco probables se definen para que la base de datos dé una señal de alerta indicando que probablemente existe una incoherencia.

1. Fecha de nacimiento

Se consideran valores poco probables de esta variable los que estén fuera del intervalo comprendido entre 01/01/1919 y 31/12/1962.

2. El número de cribados de la mujer no debería ser un número superior al número de ronda de la URC.

3. Número de estudios intermedios ≥ 4

4. N prueba ≥ 10 por episodio

5. La talla se recogerá en centímetros, así pues, es de esperar que sea un número que esté aproximadamente entre 120 y 200.

6. El peso se recogerá en kilogramos, así pues, es de esperar que sea un número que esté aproximadamente entre 40 y 120.

Para validar definitivamente la base de datos se sugiere la realización de “frecuencias” de cada variable. Los resultados de las frecuencias dependerán de los protocolos de cada CCAA, por lo que, de momento, no se propone una validación especifica.

VIII. BIBLIOGRAFÍA

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer. Journal international du cancer.* Sep 13 2014.

2. Ferlay J, Soerjomataram I, Ervik M, et al. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet].* 2013.

3. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA: a cancer journal for clinicians.* 2011;61(2):69-90.

4. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *European journal of cancer (Oxford, England : 1990).* 2013;49(6):1374-1403.

5. Pollan M, Pastor-Barriuso R, Ardanaz E, et al. Recent changes in breast cancer incidence in Spain, 1980-2004. *J Natl.Cancer Inst.* 2009;101:1584-1591.

6. Ravdin PM, Cronin KA, Howlader N, et al. The decrease in breast-cancer incidence in 2003 in the United States. *N.Engl.J.Med.* 2007;%19;356(16):1670-1674.

7. Parkin DM. Is the recent fall in incidence of post-menopausal breast cancer in UK related to changes in use of hormone replacement therapy? *European journal of cancer (Oxford, England : 1990).* 2009;45(9):1649-1653.

8. Canfell K, Banks E, Moa AM, Beral V. Decrease in breast cancer incidence following a rapid fall in use of hormone replacement therapy in Australia. *Med.J.Aust.* 2008;188(11):641-644.

9. Glass AG, Lacey JV, Jr., Carreon JD, Hoover RN. Breast cancer incidence, 1980-2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. *J.Natl.Cancer Inst.* 2007;99(15):1152-1161.

10. Katalinic A, Rawal R. Decline in breast cancer incidence after decrease in utilisation of hormone replacement therapy. *Breast Cancer Res.Treat.* 2008;107(3):427-430.

11. Kerlikowske K, Miglioretti DL, Buist DS, Walker R, Carney PA. Declines in invasive breast cancer and use of postmenopausal hormone therapy in a screening mammography population. *J.Natl.Cancer Inst.* 2007;99(17):1335-1339.

12. Kliewer EV, Demers AA, Nugent ZJ. A decline in breast-cancer incidence. *N.Engl.J.Med.* 2007;357(5):509-510.

13. Kumle M. Declining breast cancer incidence and decreased HRT use. *Lancet.* 2008;372:608-610.

14. Ponti A, Rosso S, Zanetti R, Ricceri F, Tomatis M, Segnan N. Re: Breast cancer incidence, 1980-2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. *J.Natl.Cancer Inst.* 2007;99(23):1817-1818.

15. Allemani C, Minicozzi P, Berrino F, et al. Predictions of survival up to 10 years after diagnosis for European women with breast cancer in 2000-2002. *International journal of cancer. Journal international du cancer.* 2013;132(10):2404-2412.

16. Forman D, Bray F, Brewster D, et al. *Cancer Incidence in Five Continents, Vol. X (electronic version).* 2013.

17. Parkin DM, Ferlay J, Curado MP, et al. Fifty years of cancer incidence: CI5 I-IX. *International journal of cancer. Journal international du cancer.* 2010;127(12):2918-2927.

18. Pollan M, Michelena MJ, Ardanaz E, Izquierdo A, Sanchez-Perez MJ, Torrella A. Breast cancer incidence in Spain before, during and after the implementation of screening programmes. *Ann.Oncol.* 2010;21 Suppl 3:iii97-102.

19. Botha JL, Bray F, Sankila R, Parkin DM. Breast cancer incidence and mortality trends in 16 European countries. *European journal of cancer (Oxford, England : 1990).* 2003;39(12):1718-1729.

20. Wilson J, Jungner F. *Principles and practice of screening for disease. Public Health Papers. no. 34.* 1968.

21. The benefits and harms of breast cancer screening: an independent review. *Lancet.* 2012;380(9855):1778-1786.

22. Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. *The Cochrane database of systematic reviews.* 2013;6:CD001877.

23. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann.Intern.Med.* 2002;137:347-360.

24. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Annals of internal medicine.* 2009;151(10):727-737, W237-742.

25. Vainio H, Soerjomataram I. *Breast cancer screening. IARC Handbooks of Cancer Prevention, volume 7.* 2002.

26. Tonelli M, Connor Gorber S, Joffres M, et al. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* 2011;183(17):1991-2001.

27. Castells X, Sala M, Ascunce N, Salas D, Zubizarreta R, Casamitjana M. *Descripción del cribado del cáncer en España. Proyecto DESCRIC.* Madrid: Plan de Calidad para el Sistema Nacional de Salud. Ministerio de Sanidad y Consumo. Agència d'Avaluació de Tecnologia i Recerca Mèiques de Cataluña; 2007.

28. Ascunce N, Salas D, Zubizarreta R, Almazan R, Ibanez J, Ederra M. Cancer screening in Spain. *Ann.Oncol.* 2010;21 Suppl 3:iii43-iii51.

29. Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. *BMJ (Clinical research ed.).* 1988;297(6654):943-948.

30. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. *Journal of the National Cancer Institute.* 2000;92(18):1490-1499.

31. Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet.* 2006;368(9552):2053-2060.

32. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet.* 2002;359(9310):909-919.

33. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. *Journal of the National Cancer Institute.* 1982;69(2):349-355.

34. Bjurstam N, Bjorneld L, Warwick J, et al. The Gothenburg Breast Screening Trial. *Cancer.* 2003;97(10):2387-2396.

35. Europe Against Cancer Programme of the European Community. European Code Against Cancer. Third version. In: Programme EAC, ed2003.

36. Perry N, Broeders M, de Wolf C, T"rnberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Luxenburg: Office for Official Publications of the European Communities 2006.

37. von Karsa L, Anttila A, Ronco G, Ponti A, Malila N, Arbyn M. Cancer Screening in the European Union. Report on the implementation of the Council Recommendation on cancer screening. International Agency for Research on Cancer (IARC); 2008.

38. Red de Programas de Cribado. 2014; http://www.programascancerdemama.org. Accessed 09/12/2014.

39. van Schoor G, Moss SM, Otten JD, et al. Increasingly strong reduction in breast cancer mortality due to screening. *British journal of cancer.* 2011;104(6):910-914.

40. Cronin KA, Feuer EJ, Clarke LD, Plevritis SK. Impact of adjuvant therapy and mammography on U.S. mortality from 1975 to 2000: comparison of mortality results from the cisnet breast cancer base case analysis. *Journal of the National Cancer Institute. Monographs.* 2006(36):112-121.

41. Paci E. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *Journal of medical screening.* 2012;19 Suppl 1:5-13.

42. Hofvind S, Ursin G, Tretli S, Sebuodegard S, Moller B. Breast cancer mortality in participants of the Norwegian Breast Cancer Screening Program. *Cancer.* 2013;119(17):3106-3112.

43. *WHO position paper on mammography screening.* Geneva, Switzerland: World Health Organization; 2014.

44. Morrison A. *Screening in chronic diseases.* 1985.

45. Armstrong K, Moye E, Williams S, Berlin JA, Reynolds EE. Screening mammography in women 40 to 49 years of age: a systematic review for the American College of Physicians. *Annals of internal medicine.* 2007;146(7):516-526.

46. Afzelius P, Zedeler K, Sommer H, Mouridsen HT, Blichert-Toft M. Patient's and doctor's delay in primary breast cancer. Prognostic implications. *Acta oncologica (Stockholm, Sweden).* 1994;33(4):345-351.

47. Burgess CC, Ramirez AJ, Richards MA, Love SB. Who and what influences delayed presentation in breast cancer? *British journal of cancer.* 1998;77(8):1343-1348.

48. Domingo L, Blanch J, Servitja S, et al. Aggressiveness features and outcomes of true interval cancers: comparison between screen-detected and symptom-detected cancers. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP).* 2013;22(1):21-28.

49. Richards MA, Smith P, Ramirez AJ, Fentiman IS, Rubens RD. The influence on survival of delay in the presentation and treatment of symptomatic breast cancer. *British journal of cancer.* 1999;79(5-6):858-864.

50. Puliti D, Duffy SW, Miccinesi G, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *Journal of medical screening.* 2012;19 Suppl 1:42-56.

51. Jorgensen KJ, Gotzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ (Clinical research ed.).* 2009;339:b2587.

52. Paci E, Miccinesi G, Puliti D, et al. Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening study in Italy. *Breast Cancer Res.2006.;8.(6.):R68.* 2006;8:R68.

53. Zahl PH, Strand BH, Maehlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. *BMJ (Clinical research ed.).* 2004;328(7445):921-924.

54. Brett J, Austoker J. Women who are recalled for further investigation for breast screening: psychological consequences 3 years after recall and factors affecting re-attendance. *J Public Health Med.* 2001;23:292-300.

55. Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. *Psychooncology.* 2005;14:917-938.

56. Hofvind S, Ponti A, Patnick J, et al. False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. *Journal of medical screening.* 2012;19 Suppl 1:57-66.

57. Bull AR, Campbell MJ. Assessment of the psychological impact of a breast screening programme. *Br.J.Radiol.* 1991;64:510-515.

58. Carles M, Vilaprinyo E, Cots F, et al. Cost-effectiveness of early detection of breast cancer in Catalonia (Spain). *BMC cancer.* 2011;11:192.

59. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. *Ann.Intern.Med.* 2007;1*46.(7.):502.-10.* 2007;146:502-510.

60. Salz T, Defrank JT, Brewer NT. False positive mammograms in Europe: do they affect reattendance? *Breast Cancer Res.Treat.2010.Nov.4.* 2010.

61. von Euler-Chelpin M, Risor LM, Thorsted BL, Vejborg I. - Risk of breast cancer after false-positive test results in screening mammography. - *J Natl Cancer Inst.* 2012(9):9.

62. Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *Journal of the National Cancer Institute.* 2006;98(17):1204-1214.

63. 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.(4):316.-21.* 2006;60:316-321.

64. Olsen O, Gotzsche PC. *Screening for breast cancer with mammography.* 2001.

65. Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet.* 2000;355(9198):129-134.

66. de Koning HJ. Mammographic screening: evidence from randomised controlled trials. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO.* 2003;14(8):1185-1189.

67. Gotzsche PC, Nielsen M. *Screening for breast cancer with mammography.* 2006.

68. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *The New England journal of medicine.* 2012;367(21):1998-2005.

69. Welch HG, Passow HJ. Quantifying the benefits and harms of screening mammography. *JAMA internal medicine.* 2014;174(3):448-454.

70. Etzioni R, Xia J, Hubbard R, Weiss NS, Gulati R. A reality check for overdiagnosis estimates associated with breast cancer screening. *Journal of the National Cancer Institute.* 2014;106(12).

71. Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Annals of internal medicine.* 2011;155(1):10-20.

72. Fletcher RW, Fletcher SW. Prevention (Chapter 8). *Clinical Epidemiology: The essentials* LWW; Fourth edition (March 7, 2005); 2005:165-185.

73. Roman M, Skaane P, Hofvind S. The cumulative risk of false-positive screening results across screening centres in the Norwegian breast cancer screening program. *European Journal of Radiology.* 2014.

74. Redondo A, Comas M, Macia F, et al. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. *The British journal of radiology.* 2012;85(1019):1465-1470.

75. Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology.* 2009;253(3):641-651.

76. Elmore JG, Miglioretti DL, Reisch LM, et al. Screening mammograms by community radiologists: variability in false-positive rates. *Journal of the National Cancer Institute.* 18 2002;94(18):1373-1380.

77. Moss SM, Blanks RG, Bennett RL. Is radiologists' volume of mammography reading related to accuracy? A critical review of the literature. *Clinical radiology.* 2005;60(6):623-626.

78. Theberge I, Chang SL, Vandal N, et al. Radiologist interpretive volume and breast cancer screening accuracy in a Canadian organized screening program. *Journal of the National Cancer Institute.* 2014;106(3):djt461.

79. Fletcher RW, Fletcher SW. Diagnosis (chapter 3). *Clinical Epidemiology: The essentials* LWW; Fourth edition (March 7, 2005); 2005:43-74.

80. Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N.Engl.J Med.* 1998;338:1089-1096.

81. 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.

82. Hubbard RA, Miglioretti DL, Smith RA. Modelling the cumulative risk of a false-positive screening test. *Stat.Methods Med.Res.* 2010;19(5):429-449.

83. Xu JL, Fagerstrom RM, Prorok PC, Kramer BS. Estimating the cumulative risk of a false-positive test in a repeated screening program. *Biometrics. 2004;60.(3):651.-60.* 2004;60:651-660.

84. Njor SH, Olsen AH, Schwartz W, Vejborg I, Lynge E. Predicting the risk of a false-positive test for women following a mammography screening programme. *J Med.Screen. 2007;14(2):94.-7.* 2007;14:94-97.

85. Puliti D, Miccinesi G, Zappa M. More on screening mammography. *N.Engl.J.Med. 2011;364.(3):284.-5*.

86. Li CI, Daling JR. Changes in breast cancer incidence rates in the United States by histologic subtype and race/ethnicity, 1995 to 2004. *Cancer Epidemiol.Biomarkers Prev.* 2007;16(12):2773-2780.

87. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA. 2002;288.(3):321.-33.* 2002;288:321-333.

88. Otten JD, Fracheboud J, den Heeten GJ, et al. Likelihood of early detection of breast cancer in relation to false-positive risk in life-time mammographic screening: population-based cohort study. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO.* 2013;24(10):2501-2506.

89. Breast Cancer Surveillance Consortium: Evaluating Screening Performance in Practice. In: National Cancer Institute NIoH, U.S. Department of Health and Human Services, ed. US: National Cancer Institute; 2004.

90. Cancer Registry of Norway. Regulation on the collection and processing of personal health data in the Cancer Registry of Norway (Cancer Registry Regulations). In: Norway CRo, ed. Oslo, Norway: Cancer Registry of Norway; 2013.

91. Larsen IK, Smastuen M, Johannesen TB, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. *European journal of cancer (Oxford, England : 1990).* 2009;45(7):1218-1231.

92. Castells X, Domingo L, Corominas JM, et al. 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 research and treatment.* 2015;149(1):237-244.

93. Hofvind S, Geller BM, Skelly J, Vacek PM. Sensitivity and specificity of mammographic screening as practised in Vermont and Norway. *The British journal of radiology.* 2012;85(1020):e1226-1232.

94. Smith-Bindman R, Ballard-Barbash R, Miglioretti DL, Patnick J, Kerlikowske K. Comparing the performance of mammography screening in the USA and the UK. *J Med.Screen. 2005.;12.(1):50.-4.* 2005;12:50-54.

95. Comas M, Castells X, Hoffmeister L, et al. Discrete-event simulation applied to analysis of waiting lists. Evaluation of a prioritization system for cataract surgery. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2008;11(7):1203-1213.

96. Comas M, Roman R, Cots F, et al. Unmet needs for cataract surgery in Spain according to indication criteria. Evaluation through a simulation model. *The British journal of ophthalmology.* 2008;92(7):888-892.

97. Comas M, Roman R, Quintana JM, Castells X. Unmet needs and waiting list prioritization for knee arthroplasty. *Clinical orthopaedics and related research.* 2010;468(3):789-797.

98. Roman R, Comas M, Mar J, et al. Geographical variations in the benefit of applying a prioritization system for cataract surgery in different regions of Spain. *BMC health services research.* 2008;8:32.

99. Comas M, Arrospide A, Mar J, et al. Budget impact analysis of switching to digital mammography in a population-based breast cancer screening program: a discrete event simulation model. *PloS one.* 2014;9(5):e97459.

100. Vilaprinyo E, Forne C, Carles M, et al. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. *PloS one.* 2014;9(2):e86858.

101. Rue M, Vilaprinyo E, Lee S, et al. Effectiveness of early detection on breast cancer mortality reduction in Catalonia (Spain). *BMC cancer.* 2009;9:326.