Universitat Autònoma de Barcelona

Avaluació de la taxa de detecció de càncer i del càncer d'interval en programes de detecció precoç del càncer de mama utilitzant mètodes longitudinals

Jordi Blanch i Font

Universitat Autònoma de Barcelona (UAB)
Facultat de Medicina
Doctorat en Salut Pública i Metodologia de la Recerca Biomèdica

Avaluació de la taxa de detecció de càncer i del càncer d'interval en programes de detecció precoç del càncer de mama utilitzant mètodes longitudinals

Tesi Doctoral

Memòria presentada per Jordi Blanch i Font per optar al grau de Doctor per la Universitat Autònoma de Barcelona.

Signatura de la directora de tesi	Signatura de la directora de tesi	Signatura del doctorand
Dra. Maria Sala Serra	Dra. Montserrat Rué Monné	Jordi Blanch i Font

Programa de Pediatria, Obstetrícia i Ginecologia, Medicina Preventiva i Salut Pública
Universitat Autònoma de Barcelona, 2017

Aquesta tesi doctoral s'ha realitzat al Servei d'Epidemiologia i Avaluació de l'Hospital del Mar-IMIM, a Barcelona, sota la direcció de la Dra. Maria Sala Serra i la Dra. Montserrat Rué Monné. Es presenta com a compendi de publicacions.

Barcelona, novembre del 2017.

Agraïments

Primer de tot, vull agrair a la Maria Sala i a la Montse Rué que acceptessin dirigir-me la tesi. Els vull agrair els seus consells, el suport rebut i la seva paciència durant aquests últims anys. A la Maria, per obrir-me les portes del Servei d'Epidemiologia de l'Hospital del Mar. Amb ella, he après molt sobre els estudis epidemiològics i com encarar els problemes que van sorgint al llarg dels projectes. A la Montse, per totes les trucades i per la seva infinita paciència. Moltes gràcies a les dues.

A tot el Servei d'Epidemiologia i Avaluació de l'Hospital del Mar, em van fer sentir com a casa tot el temps que hi vaig ser. En recordaré sempre les converses entre cafès i dinars. Vull dóna les gràcies a en Xavier Castells per l'oportunitat de formar part del Servei. A les companyes de despatx -la Laia, la Marta i l'Anabel- per compartit aquest viatge. A la Laia, per totes les estones que vam fer créixer el projecte i els articles. A la Marta, per ensenyar-me a superar qualsevol dificultat. A l'Anabel, per estar sempre de bon humor. No voldria oblidar la resta del Servei que està a l'hospital. A l'Esther, per animar-nos i ajudar-nos en tot el que podia. A la Mercè, per les converses sobre temes estadístics i friquis. Va ser un plaer treballar amb vosaltres.

A la Unitat de Suport a la Recerca de Girona de l'IDIAP Jordi Gol que m'ha donat l'oportunitat de poder continuar fent recerca des de la meva ciutat. En especial, a la Maria i en Rafael per fer possible tenir un gran equip. A en Josep Maria, per fer-nos riure cada dia. A en Marc, per tots els consells estadístics. A la Ruth, l'Ani i la Lia, per totes les converses on arreglem el món, proposem nous reptes i ens qüestionem les nostres certeses. És un plaer treballar amb tots vosaltres.

Finalment, vull agrair tot l'esforç que han fet els meus pares. Gràcies a ells he pogut fer el que volia i sempre m'han animat a anar més lluny. No m'oblido de l'Aida, gràcies a ella he descobert què és cuidar i deixar-se cuidar, discutir, tenir milions de

plans de futur i, sobretot, tenir un projecte en comú. A en Bernat, que m'ensenya on són els meus límits i ens ajudem a superar-los.

Moltes gràcies a tots.

Finançament

La realització d'aquesta tesi doctoral ha estat possible gràcies a diferents ajuts:

- Projecte de recerca finançat pel "Fondo de Investigación Sanitaria" (FIS-ISCIII), en la convocatòria de l'any 2006, pel projecte "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).
- Projecte de recerca finançat pel Fondo de Investigación Sanitaria (FIS-ISCIII), en la convocatòria de l'any 2009, pel projecte "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).
- Projecte de recerca finançat pel Fondo de Investigación Sanitaria (FIS-ISCIII), en la convocatòria de l'any 2009, pel projecte "Evaluación de la tasa del cáncer de intervalo y sus determinantes en los programas de detección precoz del cáncer de mama" (PS 09/01153).

Resum

Introducció Durant la dècada dels 80, es va demostrar que la detecció precoç del càncer de mama (CM) era la millor estratègia per disminuir la seva mortalitat a llarg termini i millorar-ne el pronòstic. La majoria de països europeus van iniciar programes poblacionals de cribratge mamogràfic biennal destinats a les dones de 50 a 69 anys. A partir de l'any 2000 es van començar a presentar diferents resultats sobre l'efectivitat d'aquesta estratègia. Van mostrant que la reducció de la mortalitat podria ser menor a l'atribuïda i, en canvi, podia produir efectes adversos poc estudiats fins a aquell moment. Aquesta tesi s'emmarca en la línia d'investigació per avaluar els determinants de l'aparició de CM en dones participants en el cribratge i els seus efectes adversos, especialment els resultats mamogràfics falsos positius (FP) i falsos negatius (FN). A més, s'estudia les característiques biològiques dels tumors detectats per a identificar subgrups de més risc i millorar la efectivitat del cribratge. El plantejament longitudinal suposa reptes metodològics que s'aborden en aquesta tesi.

Objectius L'objectiu general és aprofundir en l'avaluació del cribratge poblacional del CM, concretament 1) quantificar el risc acumulat de detectar un càncer dins el cribratge durant 10 anys de participació i avaluar els seus factors associats; 2) quantificar el risc de tenir un càncer d'interval (CI) i avaluar els seus factors associats; 3) comparar diferents mètodes per estimar l'efecte d'un factor de risc en el risc de presentar un CM en les participants, i 4) comparar les característiques biològiques dels CI amb les dels tumors detectats en les proves de cribratge.

Població i mètodes Per a cada objectiu s'ha utilitzat una població d'estudi i una metodologia diferent. Per al primer objectiu, s'estudia la taxa de detecció de càncer dins el cribratge, utilitza les dades de la cohort retrospectiva del projecte RAFP. S'estima la relació entre la taxa de detecció i els seus factors associats segons

les característiques del protocol per a la realització i lectura de les mamografies i les característiques de la dona. S'utilitzen mètodes de supervivència en temps discret. Per al segon objectiu, s'estudia la taxa de CI amb les dades de la cohort retrospectiva del projecte INCA, que conté informació dels subtipus de CI i de les característiques biològiques dels tumors. S'estima l'associació entre la taxa de CI amb els factors de risc relacionats amb les característiques del protocol i les propies de la dona. També, s'estima la relació d'aquests factors de risc segons la via diagnòstica. Hem utilitzat un model de Cox amb riscos proporcionals i risc competitiu específic per causa. Per al tercer objectiu, s'han comparat els models utilitzats en els dos primers objectius amb un tercer de diferent, un model multi-estat basat en un procès de Markov. Per a comparar les tres metodologies, s'ha emprat la simulació. A més, s'ha aplicat els tres models a la població catalana de la cohort del projecte INCA. Per al quart objectiu, s'ha utilitzat un disseny de cas-control amb els càncers detectats en el cribratge i els d'interval del projecte INCA. Per estudiar les diferències entre els càncers s'ha utilitzat un model de regressió logística. A més, els CI s'han classificat radiològicament (verdaders intervals, FN, tumors ocults, signes mínims). Per estudiar les diferències entre els diferents tipus de CI i els detectats en el cribratge, s'ha emprat un model de regressió multinomial. Tant la gestió de les bases de dades com l'anàlisi estadística, es realitzaren amb els softwares SPSS (versió 12.0 i 18.0), SAS (versió 9.2) i R (versió 2.14 i 3.1). En tots els treballs s'assumeix un nivell de significació del 5%, considerant així, valors $p < 0.05$ estadísticament significatius.

Resultats Els principals resultats per a cada objectiu han estat: 1) El risc acumulat de detectar un càncer en el cribratge al cap de 10 anys de participar-hi és d'entre 11,11 i 16,71 per cada 1.000 participants segons les variables del protocol. La utilització de la doble lectura i dues projeccions és l'estratègia que detecta més tumors, tan invasius com in situ. 2) Tenir un resultat FP previ és el principal factor de risc per al CI, especialment per al FN. La història familiar de CM està associada als veritables CI. A més, les característiques de les dones són factors de risc tant per a la detecció de càncer en el cribratge com per als CI amb el mateix ordre de magnitud. 3) Els models multi-estat i el model de Cox estimen correctament l'efecte del FP sobre l'aparició del CM. Els models multi-estat permeten estimar les transicions de no tenir càncer a estadi preclínic i de preclínic a clínic. 4) Les variables relacionades amb el protocol mamogràfic estan associades amb la detecció de càncer en el cribratge i no amb els CI. Els veritables intervals i els de signes mínims tenen una distribució dels fenotips similars entre ells amb una major proporció de triples negatius que els càncers de cribratge. La distribució dels fenotips per als FN

i els tumors ocults és més semblant a la dels tumors detectats en el cribratge. Les mames denses estan associades especialment als tumors ocults, però també amb els veritables intervals i els FN.

Conclusions 1) Un millor coneixement dels efectes adversos del cribratge de CM, com ara la variabilitat en les estratègies de protocol i de les característiques de la dona, ha de permetre millorar el cribratge i oferir a les dones estratègies millors i fent que estiguin més ben informades. 2) Els factors de risc del CI són els mateixos que els dels CM detectats en el cribratge. La relació entre les característiques personals i organitzatives amb el risc de CI permet identificar subgrups de dones amb diferent risc de desenvolupar CM. 3) Els models multi-estat permeten modelar la història natural del CM i incloure les participacions de les dones en el cribratge dins el model per avaluar millor l'efecte de les variables en el risc de desenvolupar CM. 4) Gairebé la meitat dels CI van ser veritables intervals, entre els quals hi ha un alt percentatge de tumors amb fenotip associat a un mal pronòstic. La distribució dels fenotips per als FN i els tumors ocults és semblant a la distribució en els CM detectats en el cribratge. Els tumors ocults estan fortament associats a les mames denses.

Continuïtat de la recerca Els resultats d'aquesta tesi han donat lloc a l'inici de varis projectes per a aprofundir en diferents aspectes del cribratge. L'any 2012 s'inicià el projecte BELE que té per objectiu aprofundir en el coneixement de les lesions benignes. L'any 2013 s'inicià el projecte CAMISS que té per objectiu estudiar l'evolució clínica dels tumors segons la via diagnòstica. L'any 2015 s'inicià el projecte InforMa per a millorar el procés d'informar amb l'objectiu de dissenyar material informatiu i avaluar el seu efecte en la presa de decisions informades i en la participació en el cribratge.

Índex

Índex de figures

Índex de taules

Capítol 1

Introducció

1.1 Epidemiologia del càncer de mama

1.1.1 Incidència i mortalitat

El càncer de mama (CM) és el diagnòstic de càncer més freqüent en dones a escala mundial. 1,67 milions de dones van ser diagnosticades de CM l'any 2012, el 25% del total de càncers. El CM és la neoplàsia més freqüent en les dones mortes per càncer amb 522.000 defuncions l'any 2012, el 15% de les defuncions per càncer en dones[1].

Durant l'última dècada, el CM ha esdevingut la principal causa de mort per càncer entre les dones en els països desenvolupats. Suposa un canvi respecte a la dècada passada, on el càncer de cèrvix era la principal causa[1].

Existeix una alta variabilitat entre països (figura 1.1). La incidència més gran es localitza al nord d'Europa, Austràlia, Nova Zelanda i Amèrica del nord. La menor es localitza a l'Àfrica subsahariana i l'Àsia[1]. Aquesta gran variabilitat a escala mundial és degut a factors reproductius (història menstrual més llarga, nul·liparitat, etc.), hormonals (tractament hormonal substitutiu (THS), anticonceptius orals, etc.), la detecció precoç[2, 3] i altres causes de mort, sobretot a l'Àfrica i Àsia.

A finals dels anys 80 i principi dels 90, la incidència (figura 1.2) als països occidentals va augmentar a causa de canvis dels factors reproductius i a la implementació dels programes de cribratge[4]. A partir de l'any 2000, la incidència ha disminuït, una de les possibles causes és la disminució de l'ús de THS[5–8].

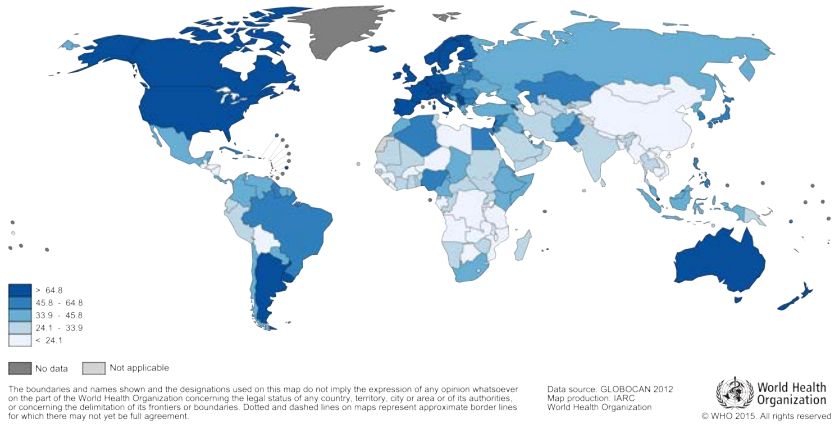


Figura 1.1: Incidència estimada del càncer de mama a nivell mundial a l'any 2012.

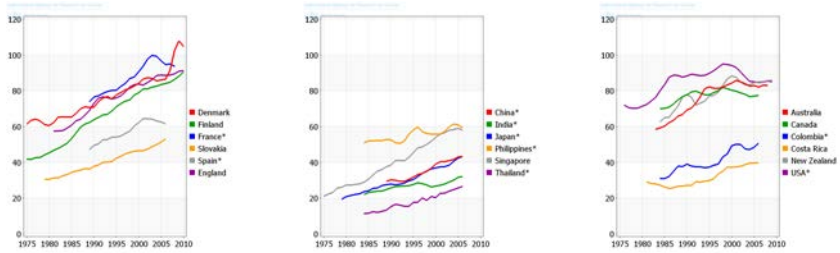


Figura 1.2: Tendència de la incidència del càncer de mama en diferents països: raons estandarditzades per edat per 100.000 habitants.[1]

La taxa de mortalitat (figura 1.3) ha disminuït durant els últims 25 anys com a resultat de la detecció precoç i les millores en el tractament[2, 4, 9]. També existeix una alta variabilitat entre països (figura 1.4). La mortalitat més gran es localitza a l'Àfrica occidental. La mortalitat més baixa es troba a l'est d'Àsia. Els països

[1]Regional data. Fonts: NORDCAN (www.ancr.nu), ECO (eco.iarc.fr), England (www.ons.gov.uk), CI5.iarc.fr, Australia (www.aihw.gov.au), New Zealand (www.health.govt.nz), USA (seer.cancer.gov).

desenvolupats tenen una mortalitat homogènia, perquè tenen una major supervivència deguda a poder accedir a millors serveis sanitaris.

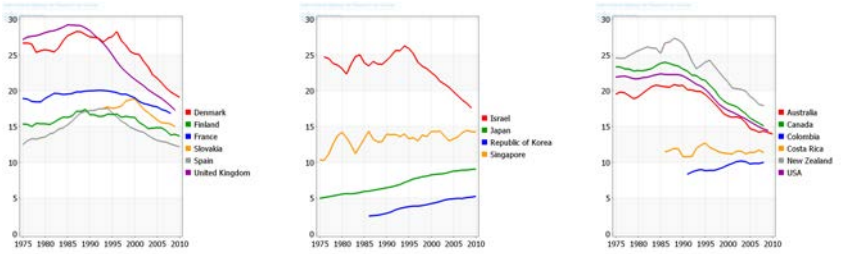


Figura 1.3: Tendència de la mortalitat per càncer de mama en diferents països: raons estandarditzades per edat per 100.000 habitants.[2]

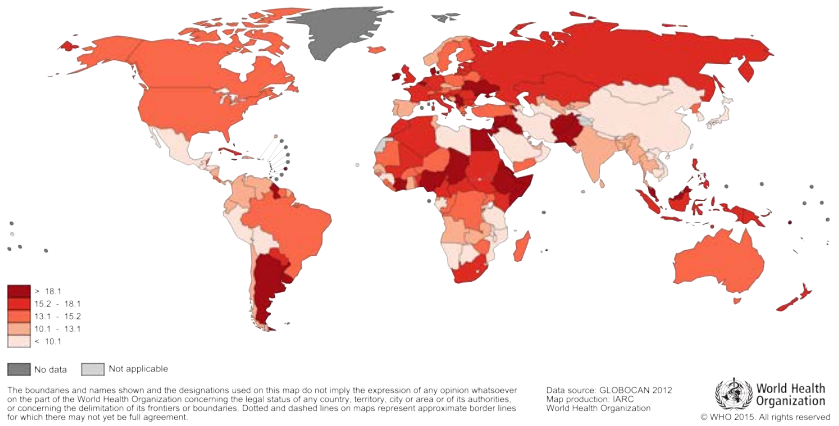


Figura 1.4: Mortalitat estimada del càncer de mama a nivell mundial a l'any 2012.

A Espanya, es moren 6.000 dones per CM i es diagnostiquen 26.000 nous casos anualment[10]. Té una incidència menor respecte EU-28 (Raó estandarditzada per l'edat és de 67,3 per 100.000 dones enfront de 82,1), l'Europa del nord (89,4) o

[2]Font: WHO (www.who.int/healthinfo/en/)

l'Europa occidental (96,0)[1]. Com en els països occidentals, la incidència ha disminuït els últims anys (figura 1.5); però no existeix una raó clara que expliqui aquesta disminució[11]. A Catalunya, en el període 2003-2007 es van diagnosticar 3.907 CM[12]. A Espanya, la taxa de mortalitat per CM a l'any 2012 va ser una de les més baixes d'Europa (11,9 per 100.000 dones ajustada per la població europea)[1]. A Catalunya, dades dels registres de càncer poblacional de Girona i Tarragona, mostren una mortalitat de 9,8 a Girona a l'any 2009[13], de 13,1 a Tarragona a l'any 2013 i de 10,5 a les terres de l'Ebre a l'any 2013[14]. La supervivència als 5 anys va ser del 82,8% a Espanya en el període 1999-2007 lleugerament superior a la mitjana europea (81,8%)[15]. A Girona i Tarragona, 84,6% i 82,8% en el període 2000-2004, respectivament[13, 14].

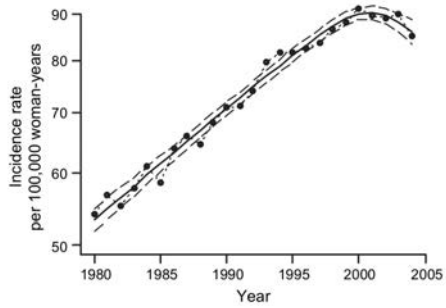


Figure 1. Age- and registry-adjusted incidence rates of invasive breast cancer over the period 1980–2004 among women aged 25 years or older included in all Spanish registries. The observed rates (**dotted line**) were obtained by using nominal categories for each single calendar year of diagnosis, and the estimated temporal trend (**solid line**) and its 95% confidence interval (**dashed lines**) were obtained from fitting a change-point model (see "Methods").

Figura 1.5: Taxes d'incidència de càncer de mama ajustades per edat.[3]

1.1.2 Factors de risc clàssics

Els diferents factors de risc del CM poden ser factors hormonals, reproductius, de dieta i d'estils de vida[16–20]. La taula 1.1 resumeix els principals factors de risc clàssics.

[3]Extreta de Pollán et al[11].

Factors hormonals i reproductius	
Increment de risc	Edat del primer naixement > 30 anys vs. < 20 anys
	Menopausa > 54 anys vs. < 45 anys
	Menarquia < 12 anys vs. > 14 anys
	Tractament hormonal substitutiu
	Anticonceptius orals
Disminució de risc	Alta paritat
	Lactància
	Edat jove en tenir el 1r fill
Dieta i estils de vida	
Increment de risc	Consum d'alcohol (≥ 1 beguda al dia)
	Sedentarisme
	Elevat índex de massa corporal (post menopausa)
Disminució de risc	Activitat física
	Alt consum de vegetals
Altres factors de risc	
Increment de risc	Mutació dels gens BRCA1 o BRCA2
	Història familiar de càncer de mama en familiars de primer grau
	Lesions benignes
	Elevada densitat mamària
	Elevats nivells endògens d'estrogen
	Exposició a radiacions
	Raça blanca

Taula 1.1: Factors de risc del càncer de mama.

Els factors de risc es poden utilitzar per crear funcions de risc de tenir un CM. Les funcions de risc són models estadístics que prediuen el risc que un individu amb un conjunt de factors de risc experimenti un determinat esdeveniment. Els actuals models de risc individual no són prou acurats per a identificar quines dones patiran un CM, ja que entre el 70 i 80% dels CM apareixen en dones sense factors de risc coneguts[21]. Només entre el 5 i 10% tenen un origen genètic degut a les mutacions pròpies de la dona, sobretot a causa del gen BRCA1 o BRCA2[21]. Actualment, a Catalunya els models de risc de tenir un CM no s'utilitzen en la pràctica mèdica ni durant el cribratge per modificar la periodicitat de les mamografies de cribratge.

La densitat mamària s'està estudiant com a principal factor per a personalitzar el cribratge, juntament amb altres factors com l'edat, la història de biòpsies mamàries i antecedents familiars[22]. Diversos treballs han identificat que les dones amb mames denses tenen més risc de patir un CM[23–25].

1.1.3 Història natural del càncer de mama

No es coneix perfectament la història natural del CM. La hipòtesi més estesa és el model lineal[26, 27]. Les cèl·lules sanes de l'epiteli progressen fins al carcinoma invasiu en diferents estats (figura 1.6). La progressió pot durar anys i requereix l'acumulació de diferents alteracions genètiques. Tot i que els carcinomes in situ són possibles precursors dels carcinomes invasius, sembla que no tots els carcinomes invasius provenen d'un in situ[28].

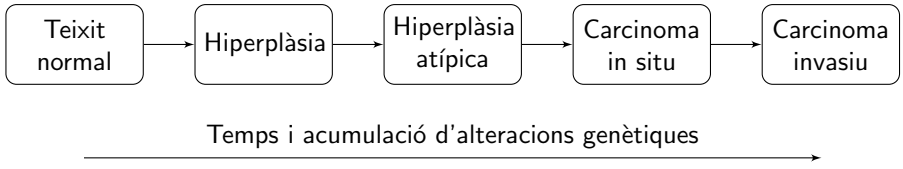


Figura 1.6: Model lineal del càncer de mama.[4]

Actualment, el model lineal està sent qüestionat. Farabegoli et al i Esserman et al afirmen que els carcinomes in situ poden ser precursors dels invasiu, però no un pas obligatori [28, 31]. Ja que alguns subtipus poden desenvolupar-se a partir de cèl·lules progenitores afectades per diferents mutacions[28]. Als carcinomes in situ de baix risc se'ls hauria de fer un seguiment actiu per identificar quins evolucionen cap a un invasius[31].

1.1.4 Factors pronòstics

Els factors pronòstics són utilitzats per a la presa de decisions sobre el maneig i el tractament de la malaltia. Es determinen en el moment del diagnòstic a partir de les mostres obtingudes de les proves invasives. En el CM, són la histologia, el grau histològic, l'afectació dels ganglis limfàtics, l'estadiatge i els receptors tumorals.

- La **histologia** és l'estudi de l'anatomia microscòpica de les cèl·lules. En el cas del CM s'estudien les cèl·lules de l'epiteli provinents dels ductes i lòbuls mamaris. Identifica si el carcinoma és ductal o lobular i si és invasiu o in situ. Es classifiquen seguint la Classificació Internacional de Malalties per Oncologia. Els carcinomes invasius tenen un pitjor pronòstic comparat amb els lobulars o els in situ[16, 17].

[4] Adaptat de Allred et al cite i Burstein et al[29, 30].

- El **grau histològic** informa com es diferencien del teixit normal comparant l'aparença de les cèl·lules cancerígenes enfront de teixit normal. S'utilitza la classificació de Scarff-Bloom-Richardson[32] agrupen els tumors en molt diferenciats (I), moderadament diferenciats (II) o poc diferenciats (III). Els tumors molt diferenciats tendeixen a mostrar un bon pronòstic. En canvi, els poc diferenciats tenen un alt risc de recurrència o de mort[17, 33]

- L'**afectació dels ganglis limfàtics** és el factor pronòstic més important[16, 17]. El nombre de ganglis limfàtics amb metàstasis està associat amb un mal pronòstic[34]. Per determinar l'estat ganglionar s'utilitza la tècnica del "gangli sentinella"[35]. El gangli sentinella és el primer nòdul limfàtic on les cèl·lules tumorals arribarien en cas de disseminació. Per conèixer la seva afectació s'utilitza un radioisòtop.

- L'**estadiatge** és la classificació de la "Union for International Cancer Control" segons la mida del tumor, l'afectació dels ganglis limfàtics i les metàstasis en el moment del diagnòstic. Es pot classificar amb la informació de la història clínica o amb l'informe patològic del tumor. Una mida gran, afectació ganglionar i presència de metàstasis tindran estadis majors i pitjor pronòstic. L'estadi 0 està format per formes pre-invasives o in situ. En els estadis 1, 2 i 3, el tumor es troba localitzat a la mama o als ganglis limfàtics regionals. En l'estadi 4 el càncer ja ha format metàstasi. La taxa de supervivència és major en càncers localitzats a la mama en comparació als estesos més enllà de la mama[36].

- Els **receptors tumorals** es determinen amb reaccions immunoenzimàtiques a la superfície de les cèl·lules tumorals. Rutinàriament, es determinen la presència de receptors d'estrogen (RE), receptors de progesterona (RP) i el factor de creixement HER-2. Segons la combinació de l'expressió d'aquests receptors es classifiquen els càncers en diferents fenotips[37], taula 1.2.

Fenotip	RE	RP	HER-2
Luminal A	+	+	-
Luminal B	+	-	-
	+	+/-	+
HER-2	-	-	+
Triple Negatiu	-	-	-

Taula 1.2: Classificació dels càncer de mama segons els diferents receptors tumorals.

Els càncers luminals són els de millor pronòstic i els triple negatius els de pitjor. Els HER-2 tenen un pitjor pronòstic que els luminals, però tenen una bona taxa de resposta al tractament amb l'anticòs monoclonal trastuzumab[38].

1.2 Cribratge del càncer de mama

1.2.1 Base científica

Frame i Carlson van establir quines condicions s'han de satisfer per a poder aplicar la detecció precoç en una malaltia concreta[39–41]:

1. Problema de salut important: afecta la qualitat i l'esperança de vida de manera important.

2. Etapa preclínica llarga i una història natural ben establerta.

3. Tractament eficaç en estadis inicials i acceptat per la població.

4. Existeix una prova de cribratge ràpida, segura, fàcil de realitzar i ben acceptada tant per metges com pacients. Aquesta ha de tenir una elevada sensibilitat, especificitat i alt valor predictiu positiu.

5. La prova de cribratge ha de tenir una bona relació cost-efectivitat.

6. El tractament en estadis preclínics redueix la mortalitat i les futures complicacions.

Durant les dècades dels 70 i 80, es van realitzar diversos assaigs clínics aleatoritzats i controlats per a estudiar l'eficàcia del cribratge de CM amb mamografia. La taula 1.5 resumeix les característiques i resultats dels principals assaigs clínics. Tots els assaigs clínics, excepte un, estimen una reducció del risc de morir de CM en les dones participants. Tot i que només la meitat ha conclòs que la reducció era estadísticament significativa, amb RR que van de 0.58 a 0.83.

D'aquests assaigs clínics s'han fet diferents revisions sistemàtiques amb metanàlisis. Tots conclouen que el cribratge de CM amb mamografia redueix la mortalitat per CM (taula 1.3). Aquesta reducció és aproximadament del 20% i estadísticament significativa. Totes les metanàlisis analitzen els estudis reportats a la taula 1.5, però algunes tenen diferents subanàlisis.

Les associacions de professionals sanitaris recomanen el cribratge de CM, però amb diferents edats i periodicitats. La taula 1.4 conté un resum de les principals organitzacions i de les seves recomanacions. La majoria recomanen una periodicitat biennal, excepte al Regne Unit amb una periodicitat triennal; en dones d'entre 50 i 69 anys, o fins i tot als 74 anys. Però la major diferència entre guies es troba en els criteris d'elegibilitat i en el protocol de les pràctiques de cribratge.

Metanàlisis	Subanàlisis	RR (IC95%)
Lancet 2012		0.80 (0.73 - 0.89)
Cochrane	Tots	0.81 (0.74 - 0.87)
	Excloent < 50a	0.77 (0.69 - 0.86)
	Ben aleatoritzats	0.90 (0.79 - 1.02)
	Aleatorietat subòptima	0.75 (0.67 - 0.83)
US Task Force		0.81*
	50-59 anys	0.86 (0.75 - 0.99)
	60-69 anys	0.68 (0.54 -0.87)
Canadian Task Force		0.79 (0.68 - 0.90)
Duffy 2012		0.79 (0.73 - 0.86)

Taula 1.3: Metanàlisis dels diferents assaigs clínics.[5,6]

Organització	Any	Interval d'edat recomanat	Periodicitat
IARC	2002	50-69	Biennal
European Comission	2006	50-69	Biennal
US Preventive Services Task Force	2009	< 50	No
		50-74	Biennal
Canadian Task force on Preventive Health Care	2011	50-74	1-2 anys
National Health Service UK	2011	47-73	Triennal
American Cancer Society (ACS)	2015	40-44	Oportunístic
Guidelines for Breast Cancer Screening in		45-54	Anual
Women at Average Risk		≤ 55	Biennal

Taula 1.4: Recomanacions per al cribratge del CM.[5]

[5] Adaptat de la metanàlisis Marmot[42].
[6] *: Mitjana ponderada per l'invers de la variància dels dos grups d'edat.
[7] Succ.=Successius.

New York HIP	Malmö I i II	Swedish Two counties	Edinburgh	Canada I i II	Stockholm	Göteborg	UK Age trial
Data d'inici							
1963	1976	1977	1978	1980	1981	1982	1991
Nombre de dones							
62.000	60.076	133.065 45 clústers	54.654 87 clústers	89.835	60.800	52.222	160.921
Edat							
40-64	45-66 i 43-49	38-75	45-64	40-49 i 50-59	39-65	39-59	39-41
Nombre de projeccions de la mamografia							
2	Inicial: 2 Succ.: 1/2	1	Inicial: 2 Succ.: 1	2	1	Inicial: 2 Succ.: 1	Inicial: 2 Succ.: 1
Interval entre proves (mesos)							
12	18-24	24-33	24	12	24-28	18	12
Nombre de rondes realitzades							
4	6-8	2-4	2-4	4-5	2	4-5	8-10
Duració (anys)							
3	12	7	6	5	4	7	8
Participació							
65%	74%	85%	65%	88%	82%	84%	81%
RR (CI95%)							
0,83 (0,70-1,00)	0,81 (0,61-1,07)	0,58 (0,45-0,76) 0,76 (0,61-0,95)		0,97 (0,74-1,27) 1,02 (0,78-1,33)	0,73 (0,50-1,06)	0,75 (0,58-0,98)	0,83 (0,73-0,89)

Taula 1.5: Assaigs clínics sobre l'eficàcia del cribratge de càncer de mama per reduir la mortalitat per càncer de mama inclosos en les diferents metanàlisis. [5,7]

1.2.2 Descripció dels programes espanyols

El primer programa espanyol de cribratge de CM es va iniciar a Navarra l'any 1990. El cribratge es va desplegar a la resta de comunitats autònomes (CCAA) gradualment fins a l'any 2001. Es va arribar a la cobertura total de la població diana l'any 2006[43]. L'hospital del Mar va iniciar el primer programa català l'any 1995.

La coordinació dels programes de cada CCAA la duu a terme la *Red de Programas de Cribado* [www.programascancerdemama.org]. Anualment, s'avaluen els diferents indicadors relacionats amb el cribratge per CCAA i conjuntament, es redacten guies de bones pràctiques, i altres temes.

A Espanya, la població diana dels programes de cribratge de CM són les dones de 50 a 69 anys. Encara que, inicialment, eren les dones de 50 a 64 anys. A més, sis CCAA també conviden les dones de 45 a 49 anys. La majoria de programes capten la població diana a través d'almenys dues fonts diferents com la targeta sanitària, el cens o el padró. La població diana l'any 2015 era de 5.449.445 dones amb una cobertura del 100% en aquell any[43].

Tots els programes utilitzen com a test de cribratge la mamografia amb una periodicitat biennal. La majoria de programes realitzen, en cada participació, dues projeccions de cada mama (crània-caudal i obliqua); però una minoria utilitza dues projeccions en la participació inicial i una en les successives (obliqua). En el mètode de lectura de les mamografies és on hi ha més variabilitat entre programes. Els mètodes més usats són: la simple (lectura amb un sol radiòleg), la doble amb consens (lectura amb dos radiòlegs on si hi ha discrepància es revisa conjuntament) o la doble sense consens (s'agafa el pitjor resultat). La majoria de programes utilitzen la classificació BI-RADS[43].

A l'informe DESCRIC es van avaluar els principals indicadors de qualitat del cribratge de CM. La participació global l'any 2005 va ser del 65,02% amb un rang de 24,87% a 88,04% segons CCAA. L'adherència calculada en 9 programes va ser del 87,79% (rang: 69,35% - 97,06%). La taxa de detecció global va ser de 3,76 càncers per cada 1.000 participacions, en la participació inicial va variar entre 2,60 i 8,93 i en les successives entre 2,03 i 3,61. Per tipus de tumors, els invasius van representar el 80,10% (rang: 65,24% - 90,33%) del total de càncers, els intraductals el 15,12% (rang: 8,95% - 27,83%) i els desconeguts el 4,78% (rang: 0,10% - 19,76%)[43].

La variabilitat en la taxa de detecció entre els diferents programes es pot explicar

per les pròpies característiques de les dones de cada regió[44], la incidència basal i les característiques dels programes[43]. Les diferències en el tipus de mamògraf, el nombre de projeccions, el mètode de lectura, els radiòlegs (nombre mínim de mamografies anuals, experiència, ...) comporten que els diferents programes tinguin taxes de detecció diferents.

1.3 Beneficis i efectes adversos

Les participants en els programes de detecció precoç obtenen una reducció de la mortalitat per CM, però també poden patir efectes adversos. Els beneficis i els efectes adversos estan resumits en la figura 1.7.

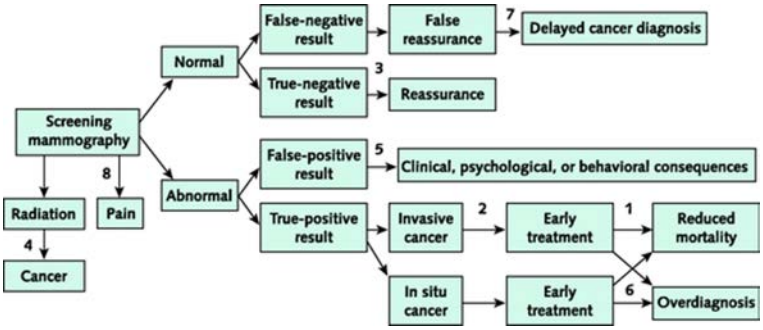


Figura 1.7: Beneficis i efectes adversos associats al cribratge del càncer de mama utilitzant la mamografia.[8]

De l'article de Paci et al.[46] i de la metanàlisis Marmot[42], es poden expressar els beneficis i els efectes adversos en valors absoluts. Suposem que tenim 1.000 dones d'entre 50 i 51 anys cribrades biennalment fins als 69 anys i seguides fins als 79 anys. D'aquestes, 71 se'ls hi diagnosticarà un CM en el cribratge durant els 20 anys (68 a Marmot), un cas per cada 14 participants. Entre 7 i 9 dones no moriran de CM (4.3 a Marmot), per salvar-ne una es necessiten entre 111 i 143 participants (235 convidades i 180 participants a Marmot). S'espera que 4 CM diagnosticats seran sobrediagnosticats (12.9 a Marmot), una sobrediagnosticada per cada 250 participants (per cada mort per CM n'hi haurà 3 de sobrediagnosticades

[8]Extret d'Armstrong et al[45].

a Marmot). A més, 200 dones patiran un fals-positiu (FP) de les quals 30 seran invasius, un FP per cada 6 participants -33 pels FP invasius[46].

1.3.1 Tipus d'efectes adversos

Els efectes adversos de la detecció precoç del CM es poden classificar en 5 grups: radiació associada a la mamografia, dolor causat per la prova, resultats FP, resultats falsos negatiu (FN) i el sobrediagnòstic[45].

La radiació associada a la mamografia és l'efecte advers més freqüent, però no s'ha pogut relacionar amb l'augment del risc de CM. En canvi, sí que s'han publicat estudis observacionals que conclouen que l'augment de risc és proporcional a la dosi de radiació i al nombre d'exposicions[45].

Quan es reporta el dolor causat per la prova, s'obté una gran variabilitat, ja que entre el 28% i el 77% de les dones reporta molèsties. Aquesta gran variabilitat és a causa de la subjectivitat de la mesura i està relacionat amb el cicle menstrual o l'ansietat. En canvi, no s'ha trobat cap relació entre el dolor i la no-adherència als programes de detecció precoç[45].

El FP es defineix com les participacions en què les mamografies de cribratge mostren sospita de malignitat i les proves complementàries no ho confirmen. Les guies europees recomanen que no se superi el 7% de dones amb FP en la participació inicial i del 5% de les participacions en les successives[47]. El 19,7% de les dones europees participants tindran un FP durant les 10 participacions[48]. A Espanya, aquest risc s'estima del 20,4%[49]. Les conseqüències més importants del FP són l'estrès i l'angoixa provocats. En diversos estudis de l'efecte sobre l'adherència, s'ha trobat que disminueix en el context espanyol[50], no varia a Copenhaguen[51] o augmenta a EUA[52].

Un altre efecte advers és el FN -càncers diagnosticats després d'una prova negativa i que ja eren presents a la mamografia de cribratge. Una conseqüència de tenir FN és la disminució del nombre de CM detectats pel cribratge provocant una disminució de l'efectivitat dels programes en la reducció de la mortalitat del CM[45]. Per conèixer els FN, necessitem identificar tots els càncers d'interval (CI), que presentarem més endavant.

Finalment, el sobrediagnòstic es defineix com els tumors detectats en el cribratge que al llarg de la vida de la dona no donarien mai símptomes[53]. Aquestes dones

poden tenir tumors diagnosticats que sempre serien asimptomàtics o que es moriran per altres causes. No es beneficien del cribratge i només pateixen els seus efectes adversos, per culpa de l'impacte emocional del diagnòstic i els tractaments. Actualment, no és possible saber quins casos són sobrediagnosticats. Puliti et al va estimar que el sobrediagnòstic en els programes europeus està entre l'1% i el 10%[53]. En canvi, l'informe Marmot estima que el 19%[42] dels càncers diagnosticats per la mamografia de cribratge ho són. Puliti et al utilitza només els articles publicats que tenen en compte la incidència basal del CM esperada en absència dels programes de detecció. En canvi, l'informe Marmot només utilitza assaigs clínics (taula 1.5).

Càncer d'interval

Els CI es defineixen com a "CM primaris sorgits després d'un episodi de cribratge amb resultat normal i abans de la següent invitació o en el termini de 24 mesos per les dones que han arribat a l'edat límit del cribratge"[47]. Aquesta definició depèn de la periodicitat de les proves i té diverses limitacions, però és útil com a indicador de l'efectivitat del cribratge. Els CI són identificats creuant l'identificador de les participants amb les dades provinents dels registres de càncer poblacionals. En les regions on no hi ha un registre poblacional, s'utilitzen els registres de tumors hospitalaris, el conjunt mínim de dades hospitalàries (CMBD-H) o fent seguiment actiu de les participants -contacte telefònic de les participants que no acudeixen a les invitacions del programa.

En el context europeu, els CI representen entre el 20% i el 30% dels casos diagnosticats [54]. Les guies europees estableixen que la taxa de CI no hauria de superar el 30% de la taxa esperada en absència de cribratge durant el primer any i el 50% en el segon any[47].

Per tal de conèixer quins CI són FN, cal classificar radiològicament tots els CI. La classificació radiològica ha de ser realitzada per un grup de radiòlegs experts en CM i cribratge. La classificació consta de dues parts. En la primera, es revisa la mamografia de cribratge sense la de diagnòstic i es classifica com a verdader interval (mamografia de cribratge normal), signes mínims (possible anormalitat), FN (anormalitat clarament visible). En la segona, es torna a revisar la mamografia de cribratge juntament amb la de diagnòstic i es torna a classificar i pot ser diferent de la primera. Així, els CI es divideixen en cinc tipus[47], resumits a la taula 1.6. Les guies europees recomanen que el FN no superi el 20% del total de CI.

Els verdaders intervals són càncers que en la última mamografia no eren detectables.

Tipus	Subtipus	Mamografia de cribratge	Mamografia diagnòstica
Verdader Interval		Negativa	Positiva
Ocult		Negativa	Negativa
Signes mínims		Signes mínims	Signes mínims o positiva
Fals negatiu	Error lectura	Positiva	Positiva
	Error tècnic	Negativa per error tècnic	Positiva
Inclassificable		Qualsevol	No trobada

Taula 1.6: Classificació dels càncer d'intervals segons la mamografia de cribratge i de diagnòstic.

La majoria de verdaders intervals són càncers més agressius i de creixement més ràpid que els detectats en el cribratge. Els FN són falles del programa, ja que són càncers detectables en la última mamografia, però que no es van diagnosticar[55]. Conèixer les característiques de cada subtipus de càncer d'interval, permetria minimitzar el risc de patir un CI.

1.4 Metodologies per estimar els factors de risc

Per estimar l'efecte d'un factor de risc sobre el CM podem utilitzar diversos enfocaments. En aquesta secció explicarem els mètodes transversals, utilitzant una sola mesura per dona; els mètodes longitudinals, utilitzant mesures repetides de la mateixa dona; i les simulacions, per avaluar el cribratge o els diferents mètodes. Per exemple, aquests mètodes ens permeten avaluar la qualitat del programa de detecció precoç o millorar la vigilància de les dones amb un risc elevat de tenir un CI.

1.4.1 Mètodes transversals

Els mètodes transversals utilitzen indicadors i taxes per cada participació. S'utilitzen els de les Guies Europees[47] centrats en dos aspectes: la qualitat dels programes i l'impacte sobre la mortalitat. L'impacte sobre la mortalitat està fora del context de la nostra tesi.

Les Guies Europees demanen estimar els indicadors de qualitat dels programes per ronda de cribratge. Mantenir uns estàndards de qualitat són necessaris, ja que s'està duent a terme una intervenció poblacional. Els indicadors proposats són:

- **Cobertura**: raó entre el nombre d'invitacions de la ronda i nombre de dones elegibles.

- **Taxa de participació**: nombre de dones participants dividit pel nombre de dones convidades.
- **Taxa de FP**: nombre de dones amb una prova complementària sense càncer dividit pel total de les participants. També es calcula segons si la prova és invasiva o no.
- **Taxa de detecció**: nombre de dones amb càncer detectat en una ronda de cribratge per cada 1,000 participants. També es calcula segons si el càncer detectat és in situ o invasiu.
- **Taxa de CI**: nombre de CI diagnosticats per cada 10,000 participants amb resultat negatiu. També es pot expressar com la proporció entre la taxa observada i la incidència esperada de CM en absència del cribratge.
- **Especificitat del programa**: proporció de participants amb un resultat negatiu sobre les participants realment lliures de càncer.

$$\frac{veritables\ negatius}{veritables\ negatius + FP}$$

- **Sensibilitat del programa**: proporció de participants amb un càncer detectat en el cribratge sobre les participants amb càncer.

$$\frac{veritables\ positius}{veritables\ positius + FN}$$

1.4.2 Mètodes longitudinals

En el programa de cribratge se segueixen les dones durant un període de temps i es registra la informació (càncer, FP, ...) de cada examen mamogràfic. Es disposa de tota la història de la dona dins el cribratge. Per tant, es poden estudiar els canvis que es produeixen en les dones i veure l'efecte de determinades variables sobre l'aparició d'un CM (en el cribratge o d'interval) o en altres aspectes del cribratge com l'aparició d'un FP. L'anàlisi de supervivència pot ser útil a l'hora d'analitzar aquest tipus de dades.

Models de supervivència a temps discret

Tal com expliquen els treballs d'Allison[56] i Singer and Willet[57], quan estudiem el cribratge, és adequat agafar el temps-discret, ja que la informació és recollida a

cada prova i els esdeveniments (detecció del càncer en el cribratge o FP) només poden ocórrer quan la dona participa en el cribratge. Per exemple, en el cas dels resultats FP només poden succeir si la dona participa en el programa. En el cas del càncer detectat en el cribratge, el càncer asimptomàtic pot aparèixer en qualsevol moment del temps, però només el detectarà el cribratge si la dona hi participa.

Els avantatges dels mètodes amb temps discret són que les variables canviants amb el temps són senzilles d'incorporar; o que les variables categòriques són estimades amb mètodes per a taules de contingència.

Els models a temps discret suposen que el temps només pot prendre valors naturals i que els n individus són observats tots a $t = 0$. Cada persona és observada fins a t_i, temps on ocorre l'esdeveniment, o aquest és censurat. Per cada persona i observació el model inclou les covariables, x_{ti}. Es defineix la taxa instantània de risc a temps discret (en anglès, «discrete-time hazard rate») com

$$P_{it} = P[T_i = t | T_i \geq t, x_{it}]$$

Aquesta taxa és la probabilitat condicionada de què l'esdeveniment succeeixi al temps t donat que no ha succeït abans. Aquesta probabilitat condicionada pot modelar-se usant la funció logística

$$\log\left(\frac{P_{it}}{1 - P_{it}}\right) = \alpha_t + \beta \cdot x_{it}$$

on α_t és el conjunt de funcions indicadores segons el temps-discret. La utilització del model logístic permet assegurar que per qualsevol combinació de α, β i x la probabilitat que es produeixi l'esdeveniment sempre estarà entre 0 i 1. El principal inconvenient és que els paràmetres del model, i per tant la probabilitat estimada, depèn de la durada de l'interval de temps, però poden ser estimats mitjançant qualsevol program estadístic.

Model de riscos competitius amb funcions de risc específiques per causa

En l'anàlisi de supervivència clàssic, es defineixen dues mesures de risc, la taxa de risc instantània i la incidència acumulada. La taxa de risc instantània, també anomenada hazard rate, és la fracció d'individus que tenen l'esdeveniment en un instant de temps sobre tots els individus a risc en aquell instant de temps. La incidència acumulada és la fracció d'individus que han tingut l'esdeveniment en un període de temps sobre tots els individus a risc al principi del període. La majoria

de models i mètodes es basen en la funció de risc instantani o hazard. El hazard a l'instant t es defineix com

$$h(t) = lim_{\Delta t \to 0} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t}.$$

La *incidència acumulada* es defineix com $F(t) = P(T \leq t)$.

En el nostre cas d'estudi, quan detectem un càncer en el cribratge no el podem observar com a interval i a l'inrevés. Quan tenim dos esdeveniments i només pot observar-se'n un, el model ha d'incloure aquest risc competitiu. En presència de riscos competitius es poden utilitzar tres metodologies diferents: específica per causa, hazard subdistribuït o combinat.

En aquest treball s'ha utilitzat el model amb una funció de risc específica per causa on només s'està interessat en un sol tipus d'esdeveniment. Aquest model estima el *hazard* i la *incidència acumulada* que s'observaria si no existís el risc competitiu. La incidència acumulada s'estima tenint en compte que existeix el risc competitiu, que fa disminuir la funció de supervivència global. Si no existís el risc competitiu hi hauria més supervivents i això provocaria més casos del risc d'interès. S'aplica l'anàlisi de supervivència clàssic per cada tipus d'esdeveniment i quan succeeix l'esdeveniment competitiu es censura. El $hazard$ per l'esdeveniment k i censurant la resta de possibles esdeveniments, es defineix com

$$\lambda_k(t) = lim_{\Delta t \to 0} \frac{P[t \leq T < t + \Delta t, E = k | T \geq t]}{\Delta t}.$$

La suma dels hazards de cada esdeveniment és el hazard global, $h(t) = \sum_{e=1}^{K} \lambda_e(t)$. La incidència acumulada s'obté combinant el hazard específic per causa i la probabilitat d'estar lliure de qualsevol esdeveniment.

Per estimar l'efecte de les covariables sobre el hazard causa-específic es pot utilitzar un model de Cox clàssic per cada esdeveniment,

$$h_k(t) = h_{k,0}(t) \exp(\beta \cdot x)$$

on $h_{k,0}(t)$ és el hazard específic per causa basal per l'esdeveniment k.

Per a una explicació més detallada es pot consultar Kleinbaum i Klein[58] o Geskus[59].

1.4.3 Simulació

Els models de simulació s'han emprat per avaluar diferents aspectes dels programes de detecció de CM. Per exemple, quina és la millor estratègia de cribratge[60], la utilització de recursos[61], o conèixer el sobrediagnostic[62], entre altres. En el nostre cas, hem utilitzat les simulacions per comparar els models multi-estat basats en un procès de Markov, els models de Cox i els de temps discret per a estimar l'efecte del FP sobre l'aparició d'un CM.

Model teòric del cribratge

Assumim un model de tres estats

$$S_0 \xrightarrow[x]{} S_p \xrightarrow[\tau]{\tau - x} S_c$$

on S_0 indica l'absència de CM, S_p l'estat preclínic -on el CM és asimptomàtic però pot ser detectat amb un prova diagnòstica (p.e., mamografia)- i S_c l'estat clínic o la presència de símptomes[63, 64]. Les edats a S_p i S_c són x i τ, respectivament. $\tau - x$ és el temps de sojorn en l'estat S_p.

Escala temporal Per l'entrada a preclínic, podem triar entre dues escales temporals, temps des de l'entrada al programa de cribratge o l'edat. L'edat és l'escala temporal més rellevant, ja que està relacionat directament amb el procés biològic. La incidència del càncer preclínic des de l'entrada depèn del disseny de l'estudi, és a dir, a quina edat les dones entren a l'estudi. Pel temps d'entrada a clínic, podem triar la mateixa escala o el temps de sojorn -el temps des de S_p a S_c.

Censura En el model, la transició $S_0 \to S_p$ té una censura d'interval, ja que no es pot conèixer exactament quan una dona entra a l'estat Sp. Podem assumir que la transició $S_p \to S_c$ s'observa exactament o és censurada per la dreta si acaba l'estudi o la dona es perd o mor sense ser diagnosticada de CM.

Les dones que se'ls detecta un càncer al cribratge són tractades. Com a conseqüència l'entrada a clínic no succeeix. Implica que la transició $S_p \to S_c$ mai serà observada en aquestes dones. És una censura informativa per a l'estimació de la distribució del temps de sojorn. Les dones amb un temps de sojorn curt tenen més probabilitats de ser diagnosticades amb un càncer en estadi clínic. Si la mamografia tingués

una sensibilitat del 100%, les dones amb un temps de sojorn superior als dos anys sempre estarien censurades a la següent visita al programa.

Entrada retardada En l'estudi INCA, les dones comencen a participar a partir dels 50 anys. Les dones de més de 50 anys quan s'inicia el programa només poden participar si no tenen símptomes de CM. Les dones amb càncer abans d'entrar al programa no es consideren. Això pot donar lloc a un truncament per l'esquerra.

Model multi-estat basat en un procès de Markov El model multi-estat basats en un procès de Markov descriu com els individus es mouen entre una sèrie d'estats en temps continu[65, 66]. En els models de Markov, el nou estat i el temps en que succeeix només depèn de l'estat present[67]. El canvi d'estat s'anomena transició o esdeveniment. La intensitat de la transició pot dependre del temps t o també d'un conjunt de covariables $z(t)$, representa el risc instantani de passar de l'estat r a l'estat $s \neq r$

$$q_{rs}(t; z(t)) = \lim_{\delta t \to 0} \frac{P(S(t+\delta t) = s | S(t) = r, z(t))}{\delta t}.$$

on $S(t)$ indica l'estat d'un individu en el temps t.

Per modelar el cribratge de CM com un model multi-estat hem assumit que

1. La intensitat de la transició $S_0 \to S_p$ s'incrementa amb l'edat, com ho fa la incidència de CM.
2. El temps d'entrada a S_p té una censura d'interval entre el temps de l'última mamografia negativa i la primera positiva.
3. El temps d'entrada a clínic, S_c, és exacte per a les dones amb CI i està censurat per la dreta per a les dones sense un càncer diagnosticat dins el programa.
4. Tenim estats mal classificats ja que la sensibilitat de la mamografia és inferior al 100%.
5. FP és una variable temps-depenent en forma esglaonada. Abans del primer FP és 0 i a partir del primer FP és 1.

Amb aquest model, hem estimat l'efecte del FP sobre el temps d'entrada a preclínic. Hem utilitzat el paquet `msm`[65] per analitzar les dades generades.

Creació de la simulació

En els estudis amb simulacions es bo seguir un protocol[68]. Els diferents punts que s'han de tenir en compte són:

1. **Protocol detallat de tots els aspectes de la simulació.**
   (a) Justificació de totes les decisions presses.
2. **Definir la finalitat i els objectius:** Ens permet centrar l'estudi i evita repeticions no necessàries i pèrdues de temps.
3. **Procediment de la simulació:**
   (a) Nivell de dependència entre les bases de dades simulades.
   (b) Què es farà amb les simulacions que fallin.
   (c) Programari utilitzat.
   (d) Especificació de la llavor inicial.
4. **Mètode per generar la base de dades:** Els mètodes per obtenir les base de dades simulades han de ser descrites tant en el protocol com en els articles publicats. Les simulacions necessiten una distribució assumida per a les variables i l'especificació de tots els paràmetres requerits.
5. **Escenaris que volem investigar:** El nombre d'escenaris a estudiar i els mètodes per avaluar han d'estar determinats i justificats en el protocol. Els escenaris estudiats han de reflectir les circumstàncies més comunes i cobrir el rang de valors plausibles.
6. **Descripció dels mètodes estadístics que volem avaluar**.
7. **Estimacions obtingudes per cada simulació:** És essencial pensar com es guardaran les estimacions per cada simulació, ja que ens permetrà comprovar la consistència del rendiment i ens permetrà identificar qualsevol error en les dades.
8. **Nombre de simulacions que es faran:** El nombre de simulacions a fer ha de permetre obtenir una bona precisió en les estimacions.
9. **Criteri per avaluar el comportament dels mètodes estadístics pels diferents escenaris**.

(a) Càlcul del biaix, $\delta = \hat{\beta} - \beta$.

(b) Càlcul de la precisió, $(\hat{\beta} - \beta)^2 + (SE(\hat{\beta}))^2$ on SE és l'error estàndard.

(c) Càlcul de la cobertura: proporció de cops que l'interval de confiança conté el paràmetre teòric.

10. **Presentació dels resultats de la simulació:** Els estudis simulats poden generar una gran quantitat de resultats que s'han de resumir i mostrar de manera clara i concisa per a facilitar la comprensió de les conclusions.

Capítol 2

Hipòtesis i objectius

2.1 Hipòtesi general

En l'estimació de la taxa de detecció de càncers en el cribratge i CI hi influeixen els protocols dels programes de cribratge, els factors intrínsecs de la dona i el mètode d'avaluació.

2.2 Hipòtesis específiques

Sobre els protocols dels programes i els factors intrínsecs de la dona:

H1. Les característiques dels programes (mètode de lectura, nombre de projeccions, tipus de mamògraf) i de la dona (edat, antecedents familiars, de biòpsia prèvia, ús de tractament hormonal, menopausa) influeixen en la probabilitat acumulada de detectar un CM tant de cribratge com d'interval.

Sobre el mètode d'avaluació:

H2. Els models estadístics basats en la història natural del CM presenten millors estimacions de l'efecte de determinades variables en el risc de presentar un CM en comparació als models que separen l'efecte segons el mètode de detecció.

Sobre les característiques biològiques:

H3. Les característiques dels CI (estadi clínic, grau histològic, marcadors biològics) són diferents de les dels tumors detectats en el cribratge.

2.3 Objectius

L'objectiu general de la tesi és aprofundir en l'avaluació del cribratge poblacional del CM i entendre millor els determinants de la detecció i dels efectes adversos. Ens centrarem en els següents aspectes:

O1. Quantificar el risc acumulat de detectar un càncer de cribratge en el període de 10 anys de participació en un programa de detecció precoç de CM i avaluar els seus factors associats.

O2. Quantificar el risc de tenir un CI i avaluar els seus factors associats.

O3. Comparar diferents mètodes per estimar l'efecte de determinades variables en el risc de presentar un CM en dones que participen regularment en un programa de detecció precoç.

2.3.1 Objectiu secundari

OS1. Comparar les característiques biològiques dels CI amb les dels tumors detectats en les proves de cribratge.

Capítol 3

Mètodes i resultats

La tesi està dividida en quatre treballs. Tres publicats en revistes científiques com a articles de recerca original i el darrer en procés de revisió. Tots s'han realitzat en el context del cribratge poblacional, però utilitzant dades de dos projectes amb finançament competitiu.

La taula 3.1 resumeix els diferents objectius específics de la tesi i les referències dels articles originals que hi donen resposta, juntament amb el projecte associat.

3.1 Disseny i població d'estudi

El disseny bàsic sobre el que es dóna resposta als tres objectius principals és el de cohort retrospectiva. El disseny per a respondre l'objectiu secundari és de cas-control niuat dins una cohort.

Les cohort estan formades amb dones participants en programes de detecció precoç de diferents CCAA d'Espanya d'entre 50 i 69 anys que han format part del estudi "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)" (RAFP, FIS PI06/1230 i PI09/90251) i l'estudi "Estudio de la relación entre falso positivo, verdaderos positivos (Cánceres) y adherencia en los programas de detección precoz de cáncer de mama en España" (INCA, FIS PS09/01153). Les característiques generals de les dues poblacions estan

[9] A més, s'indica el projecte d'on provenen les dades.

Objectiu	Article	Estudi
Sobre els protocols dels programes i els factors intrínsecs de la dona		
O1. Estimar el risc acumulat de detectar un càncer en el període de 10 anys i els seus determinants	Blanch et al. Breast cancer research and treatment 138 (3), 869-877	RAFP
O2. Estimar el risc de càncer d'interval	Blanch et al. PLoS ONE 9(10): e110207.	INCA
Sobre el mètode d'avaluació		
O3. Comparar diferents mètodes per estimar l'efecte de determinades variables en el risc de presentar un CM en dones que participen regularment en un programa de detecció precoç.	Blanch et al. 1a versió	INCA
Sobre les característiques biològiques		
OS1. Comparar les característiques biològiques dels càncers d'interval amb els tumors detectats en les proves de cribratge	Domingo et al. Breast Cancer Res. 2014;16(1):R3.	INCA

Taula 3.1: Relació entre els diferents objectius i els treballs presentats.[9]

resumides a la taula 3.2.

El projecte RAFP conté la informació de les dones participants dels programes de detecció precoç de 8 CCAA (Astúries, Canàries, Castella i Lleó, Catalunya, Galícia, La Rioja, Navarra i València) des del 1991 fins al desembre del 2006. Els programes almenys tenen la informació de tres rondes de cribratge consecutiu i cada dona està identificada amb un codi únic. Per cada participació es disposa de la informació de la mamografia de cribratge, el resultat (càncer o no), i, en el cas d'una mamografia positiva, el tipus d'exploració addicional i el seu resultat histopatològic. Es van incloure totes les dones que van participar almenys un cop en algun dels programa de cribratge participants. Es van excloure les que havien patit un CM o tenien un implant mamari anterior a la primera participació. En aquest estudi es va aconseguir tenir informació del 44% de la població diana espanyola de l'any 2005.

El projecte INCA conté la informació de les dones participants en programes de detecció precoç de 5 CCAA (Canàries, Catalunya, Galícia, País Basc i València) des del gener del 2000 fins al desembre del 2006. Es va fer una cerca activa per identificar tots els CI entre el gener del 2000 i el juny del 2009 utilitzant la informació dels registres poblacionals de tumors, el registres d'ingrès hospitalaris i els registres de càncer hospitalaris. Per cada participació, es va recollir la mateixa informació

que en el projecte RAFP. Per a les dones amb càncer, es va obtenir la informació clínica i biològica del tumor. Per als CI i una mostra de càncers de cribratge, es va obtenir informació radiològica per a poder classificar el tipus de càncer i mesurar la densitat mamaria. Aquesta mostra es va dissenyar per tenir dos càncers detectats en el cribratge per cada CI.

	RAFP	INCA
Dones	1.565.364	645.764
Més de dues participacions	1.205.943	515.030
Més de tres participacions	867.160	382.174
Mamografies	4.739.498	1.508.584
Falsos positius	264.801	62.330
Falsos positius invasius	24.436	7.352
Tumors detectats en el cribratge	16.529	5.309
Càncers d'intervals		1.653
Unitats radiològiques	74	32
Mitjana de dones per unitat	21.154	21.806
Mitjana de mamografies per unitat	64.047	47.143

Taula 3.2: Característiques generals de les poblacions d'estudi.

3.2 Creació de les bases de dades

Per construir les bases de dades dels dos projectes es va utilitzar la mateixa metodologia. Al començar la tesi, la base de dades del projecte RAFP ja estava creada. El primer pas de la tesi va ser crear la base de dades del projecte INCA.

Per al projecte INCA, es va demanar a cada programa de detecció precoç la informació completa de totes les dones que havien participat en el període 2000-2006. Es va elaborar un protocol de definició de variables i validació de la informació, ja que els diferents programes tenen estructures administratives independents i específiques. El protocol de variables es va basar en el del projecte RAFP i es va consensuar amb l'ajuda dels responsables de tots els programes participants. Així, vam poder homogeneïtzar la informació aportada per cada programa evitant ambigüitats. Les dades provinents de cada programa es van validar de forma individual i contactàrem amb els programes per resoldre dubtes quan va ser necessari. Al final, vam fusionar les diferents base de dades per cada programa en una base de dades comuna.

3.3 1r article

Breast Cancer Res Treat
DOI 10.1007/s10549-013-2458-5

EPIDEMIOLOGY

Cumulative risk of cancer detection in breast cancer screening by protocol strategy

J. Blanch · M. Sala · M. Román · M. Ederra · D. Salas · R. Zubizarreta · M. Sanchez · M. Rué · X. Castells · CFPR group

Received: 18 January 2013 / Accepted: 18 February 2013

Abstract *Background* There is little information on the individual risk of screen-detected cancer in women over successive participations. This study aimed to estimate the 10-year cumulative breast cancer detection risk (ductal carcinoma in situ and invasive carcinoma) in a population-based breast cancer screening program according to distinct protocol strategies. A further aim was to determine which strategies maximized the cancer detection risk and how this risk was affected by the radiologic protocol variables. *Methods* Data were drawn from a retrospective cohort of women from nine population-based screening programs in Spain from 1990 to 2006. We used logistic regression with discrete intervals to estimate the cumulative detection risk at 10 years of follow-up according to radiologic variables and protocol strategies. *Results* In women starting screening at the age of 45–59 years, the cumulative risk of screen-detected cancer at 10 years ranged from 11.11 to 16.71 per 1,000 participants according to the protocol strategy. The cumulative detection risk for overall cancer and invasive cancer was the highest with strategies using digital mammography, double reading, and two projections (16.71 and 12.07 ‰, respectively). For ductal carcinoma in situ, cumulative detection risk was the highest with strategies using screen-film, double reading, and two projections (2.32 ‰). The risk was the lowest with strategies using screen-film mammography, single reading, and two projections. *Conclusions* This study found that at least eleven cancers are detected per 1,000 women

Electronic supplementary material The online version of this article (doi:10.1007/s10549-013-2458-5) contains supplementary material, which is available to authorized users.

J. Blanch · M. Sala (✉) · M. Román · X. Castells
Epidemiology and Evaluation Department, Hospital del Mar-IMIM, Passeig Marítim, 25-29, 08003 Barcelona, Spain
e-mail: MSalaSerra@parcdesalutmar.cat

J. Blanch · M. Sala · X. Castells
EHEA Doctoral Program in Public Health. Department of Pediatrics, Obstetrics and Gynecology, Preventive Medicine and Public Health, Universitat Autònoma de Barcelona (UAB), Bellaterra, Barcelona, Spain

J. Blanch · M. Sala · M. Román · X. Castells
Red de investigacion en servicios de salud en enfermedades crónicas (REDISSEC), Barcelona, Spain

M. Ederra
Navarra Breast Cancer Screening Program, Instituto de Salud Pública, Leyre 15, 31003 Pamplona, Spain

M. Ederra
CIBER de Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain

D. Salas
General Directorate of Public Health and Center for Public Health Research (CSISP-FISABIO), Avda. Catalunya 21, Valencia, Spain

R. Zubizarreta
Galician Breast Cancer Screening Program, Directorate for innovation and management of public health. Health Office, San Lázaro s/n, 15703 Santiago de Compostela, Galicia, Spain

M. Sanchez
Programas de Salud de la Mujer, Dirección General de Salud Pública, Consejería de Sanidad y Servicios Sociales Gobierno de Cantabria, Santander, Spain

M. Rué
Basic Medical Sciences Department, Biomedical Research Institut of Lleida (IRBLLEIDA)-University of Lleida, Lleida, Spain

Published online: 08 March 2013

screened in the first 10 years of follow-up. Enhanced knowledge of the variability in cumulative risk of screen-detected cancer could improve protocol strategies.

Keywords Breast cancer · Screening · Cumulative risk · Detection risk · Protocol strategies

Abbreviations

DCIS	Ductal carcinoma in situ
SFM	Screen-film mammogram
DM	Digital mammogram
CI	Confidence interval
OR	Odds ratio
SDT	Screen-film mammography, double reading and two-views
SDTO	Screen-film mammography, double reading and two views in the initial screening and one view in successive screenings
SST	Screen-film mammography, single reading and two views
DDT	Digital mammography, double reading and two-views

Introduction

There is a need for continuous and complete evaluation of breast cancer screening programs to better understand their risk–benefit ratio. Cancer screening is based on the rationale that detecting and treating cancer at early and asymptomatic stages will prevent deaths and allow less aggressive treatments [1].

Screening focuses on the early detection of invasive cancer. The introduction of population-based breast cancer screening, however, increased detection of ductal carcinoma in situ (DCIS) [2–4] and raised concern about its potential over-diagnosis [5].

Population-based screening has been implemented in most European countries, including Spain. These programs follow the recommendations of the European Union guideline but differ in their use of screen-film mammography (SFM) or digital mammography (DM), single or double reading, and the number of projections in initial or successive screening mammograms; all these factors have an effect on screening accuracy [6]. Independent double reading of mammograms and two-view mammography at prevalent and incident screenings are associated with improved quality indicators [7–9]. Nevertheless, the specific effect of these protocol-related variables on DCIS and invasive cancer detection has been less thoroughly evaluated, and the introduction of DM is particularly controversial. The cancer detection rate, especially of DCIS, has been reported to be higher with DM [6, 10, 11]. However, most studies evaluating performance indicators have used cross-sectional estimates. Because participants are periodically screened for at least 20 years, the cumulative effect of these factors on quality indicators should also be evaluated to allow their impact to be estimated over the entire participation period.

The aim of this study was to estimate the cumulative risk of DCIS and invasive cancer detection in women participating in a population-based breast cancer screening program and to determine the effect of distinct radiologic protocol-related characteristics and protocol strategies on this risk.

Methods

Setting and study population

The study sample was drawn from a retrospective cohort study of screened women which aimed to evaluate the cumulative risk of a false-positive result and the impact of the introduction of DM. Details of these studies have been reported elsewhere [10, 12, 13].

The population-based breast cancer screening program in Spain began in 1990 and went nationwide in 2001. All women aged 50 (or 45 in one region) to 69 years receive a letter every 2 years inviting them to undergo mammography. We included information on all women participating in at least one screening round from nine Spanish regions with 76 radiology units. Each radiology unit had information from the beginning of each program to December 2006. The programs in each region satisfy the European Guidelines for Quality Assurance in Mammographic Screening [14], but differ in their protocol characteristics.

By December 2006, 12 radiology units had finished the third round, 23 were at the fourth round and 41 were at the fifth round or more. Single reading was used in 18 radiology units and double reading (with or without consensus or arbitration) in 55. Three radiology units used both types of reading in the study period. In 49 radiology units, two views (mediolateral oblique and craniocaudal) were always used. The remaining 27 radiology units used two views in the first participation and one view (craniocaudal) in successive participations. All radiology units began using SFM. Only five radiology units switched to DM between September 2004 (two radiology units) and January 2005 (three radiology units). The database contains information on 1,264,340 women with 4,015,431 mammograms.

Variables

Cancer detection

Women with negative mammography were recalled for a new mammogram at 2 years. Otherwise, women with a

positive result were recalled for immediate further assessments (invasive and/or non-invasive tests) to exclude malignancy.

Screen-detected cancers were defined as those histopathologically confirmed as DCIS or invasive cancer after further assessments. False-positive results were defined as those not confirmed as cancer after further evaluation. If cancer was excluded, the woman was invited to undergo a new mammogram at 2 years.

Study variables

At each participation, information was collected on the women's characteristics and protocol-related variables. Women's characteristics included age, the number of participations and, for successive participations, the interval between the current and previous round to distinguish between regular (every 2 years) and irregular participations (more than 30 months since last participation). The screening protocol variables were the type of mammogram (SFM or DM), reading method (single or double) and the number of views (one or two).

Protocol strategies

We studied four protocol strategies. There were three protocol strategies using SFM and one using DM. The three strategies using SFM were double reading and two views (SDT strategy); double reading and two views in the initial screening and one view in successive screenings (SDTO strategy); and single reading and two views (SST strategy). The fourth strategy, using DM, employed double reading and two views (DDT strategy).

Statistical analysis

The overall of screen-detected cancer, DCIS detection and invasive detection rates were calculated as number of breast cancer, DCIS cancers, and invasive cancers detected at screening divided by number of participations, respectively. The recall rate was calculated as the number of participations requiring at least one further assessment divided by number of participations. We calculated the 95 % confidence intervals (95 %CI) of these rates using the binomial approximation. The relative change in detection between any two protocol strategies was defined as the fraction of the absolute difference between rates with respect to the lower rate.

The probability of a cancer being detected at screening, and the 95 %CI, were estimated with generalized linear mixed models [15, 16]. This probability was conditioned on the distinct study variables and the random variation in the cancer detection rates among the different radiology units [12]. Two independent models were estimated for the probability of screen-detected cancer for radiological variables (model 1) and for protocol strategies (model 2). Also, two independent models were estimated for the probability of DCIS and invasive cancer for radiological variables (model 3) and for protocol strategies (model 4).

Cumulative risk of cancer detection

We estimated the cumulative cancer detection risk at 10 years in women who began screening at the age of 45–59 years. The cumulative risk was calculated from the model with screening strategies as the cancer detection risk at each participation multiplied by the proportion of cancer-free women up to that participation. The general equation is

$$\begin{aligned}\text{Cumulative\_risk}(i) &= 1 - S(i) \\ &= 1 - (S(i-1) \cdot (1 - h(i)))\end{aligned}$$

where i is the number of participation, $S(i)$ is the cancer-free function, and $h(i)$ is the probability estimated by the model. The cumulative detection risk was calculated for each strategy. The 95 %CI of the cumulative risk was calculated using Greenwood's formula for survival probabilities [17].

The statistical analysis was performed using SAS (SAS system for Windows, version 9.2). All p-values were bilateral, and p-values of less than 0.05 were considered statistically significant.

Results

Detection rates

During the study period, there were a total of 12,570 screen-detected cancers (3.13 cancers per 1,000 mammograms), of which 2,005 (0.50 ‰) were DCIS and 9,995 (2.49 ‰) were invasive carcinomas and 570 had no histological information (Table 1). The DCIS and invasive cancer detection rates were higher in the initial participation than in successive participations (Table 1).

The cancer detection rate was highest with the DDT strategy for both DCIS (0.84 ‰) and invasive cancer (3.04 ‰). No statistically significant differences were observed in detection rates between the SDT (3.36 ‰) and the SDTO (3.16 ‰) strategies, although the former had a lower recall rate (4.09 % versus 7.24 %, $p < 0.001$). The cancer detection rate was lowest with the SST strategy (2.63 ‰). The highest positive predictive value was found with the SDT strategy (8.22 %). In all screening strategies, the positive predictive value was highest in successive screenings than in the initial screening, the greatest

Table 1 Information on recalled women, cancer detection (DCIS or invasive carcinoma), and positive predictive value by screening strategy

	Protocol strategy				Total[a]
	SDT	SDTO	SST	DDT	
Screening test (*n*)	1,244,172	1,202,181	1,004,640	29,633	4,015,431
Initial (*n* (%))	438,208 (35.2)	377,078 (31.4)	325,983 (32.4)	8,212 (27.7)	1,264,340 (31.5)
Recall (*n* (rate %))	50,889 (4.1)	87,056 (7.2)	57,833 (5.8)	1,893 (6.4)	24,3394 (6.1)
Initial	31,057 (7.1)	43,363 (11.5)	28,396 (8.7)	1,080 (13.2)	120,660 (9.5)
Successive	19,832 (2.5)	43,693 (5.3)	29,437 (4.3)	813 (3.8)	123,334 (4.5)
Cancers (*n* (rate ‰))	4,183 (3.4)	3,799 (3.2)	2,644 (2.6)	131 (4.4)	12,570 (3.1)
Initial	1,734 (4.0)	1,532 (4.1)	1,034 (3.2)	39 (4.7)	4,826 (3.8)
Successive	2,449 (3.0)	2,267 (2.7)	1,610 (2.4)	92 (4.3)	7,744 (2.8)
DCIS (*n* (rate ‰))[b]	661 (0.5)	593 (0.5)	392 (0.4)	25 (0.8)	2,005 (0.5)
Initial	288 (0.7)	264 (0.7)	164 (0.5)	10 (1.2)	812 (0.6)
Successive	373 (0.5)	329 (0.4)	228 (0.3)	15 (0.7)	1,193 (0.4)
Invasive (*n* (rate ‰))[b]	3,394 (2.7)	2,982 (2.5)	2,059 (2.0)	90 (3.0)	9,995 (2.5)
Initial	1,392 (3.2)	1,148 (3.0)	786 (2.4)	27 (3.3)	3,752 (3.0)
Successive	2,002 (2.5)	1,834 (2.2)	1,273 (1.9)	63 (2.9)	6,243 (2.3)
PPV (%)	8.22	4.36	4.57	6.92	5.15
Initial	5.58	3.53	3.64	3.61	4.00
Successive	12.35	5.19	5.47	11.32	6.28
No. of radiology units[c]	49	23	14	4	76

Strategy: *SDT* screen-film mammography; double reading and two views, *SDTO* screen-film mammography; double reading and two views in the initial participation and one view in successive participations, *SST* screen-film mammography; single reading and two views, *DDT* digital mammography; double reading and two views, *DCIS* ductal carcinoma in situ, *PPV* positive predictive value

[a] There were 534,865 screening tests with other protocol strategies

[b] There were 570 cancers with unknown histology

[c] Some radiology units changed their strategy during the study period

difference being found in the DDT strategy (initial: 3.6 % and successive: 11.3 %).

Table 2 shows the cancer detection rates, and specifically for DCIS and invasive cancers, by women's characteristics and radiologic protocol-related variables. Overall and invasive cancer detection increased with age (from 2.92 ‰ to 3.71 ‰). No statistically significant differences were observed in DCIS. Cancer detection was higher in women not attending the previous invitation than in those screened regularly, with double reading than with single reading, and with two views than with one view. DM detected more invasive and DCIS cancers than did SFM. The increase in the relative change between DM and SFM and between two views and one view was higher for DCIS than for invasive cancer (84.8 % vs 23.3 % and 14.2 % and 6.6 %, respectively), while that between double and single reading was lower in DCIS than in invasive cancers (9.2 % and 15.7 %).

Risk detection model

Table 3 shows the adjusted effect of radiologic protocol-related variables (model 1 and 3) and protocol strategies (model 2 and 4) on risk detection. Model 3 showed that double reading increased DCIS detection (OR = 1.44 95 %CI: 1.17–1.79) and invasive cancer (OR = 1.27 95 %CI: 1.14–1.42). DM increased detection of DCIS (OR = 1.40 95 %CI: 1.06–1.86) and invasive cancers (OR = 1.21 95 %CI: 1.04–1.40). Risk was higher in older women than in those aged 50-54 years for both types of cancer (DCIS: OR = 1.59 95 %CI: 1.22–2.07; invasive carcinoma: OR = 1.58 95 %CI: 1.41–1.79). This risk was unchanged by the number of projections. Model 2 and 4 showed that the SST strategy decreased the cancer detection respect the SDT strategy (OR = 0.61 95 %CI: 0.48–0.79 for DCIS and OR = 0.82 95 %CI: 0.72–0.94 for invasive cancers). No statistically significant differences were observed in cancer detection risk between DDT and SDT strategies.

Cumulative risk of detection by protocol strategies

The cumulative risk of cancer detection was 13.25 ‰ (95 %CI: 13.04–13.46), 10.61 ‰ (95 %CI: 10.43–10.80) for invasive cancer, and 2.11 ‰ (95 %CI: 2.02–2.19) for DCIS. The highest overall cancer detection risk was found

Table 2 Cancer detection (overall, DCIS and invasive carcinoma) and their rates separated by study variables

	Mammography	Overall Cancers		DCIS		Invasive	
	N	*N*	Rate [‰ (95 %CI)]	*N*	Rate [‰ (95 %CI)]	*N*	Rate [‰ (95 %CI)]
Total	4,015,431	12,570	3.13 (3.08–3.19)	2,005	0.50 (0.48–0.52)	9,995	2.49 (2.44–2.54)
Age at screening							
45–49	776,513	2,267	2.92 (2.80–3.04)	431	0.56 (0.50–0.61)	1,702	2.19 (2.09–2.30)
50–54	1,246,820	3,681	2.95 (2.86–3.05)	597	0.48 (0.44–0.52)	2,934	2.35 (2.27–2.44)
55–59	1,196,086	3,885	3.25 (3.15–3.35)	589	0.49 (0.45–0.53)	3,107	2.60 (2.51–2.69)
60–64	661,652	2,239	3.38 (3.25–3.53)	303	0.46 (0.41–0.51)	1,854	2.80 (2.68–2.93)
65–70	134,360	498	3.71 (3.39–4.05)	85	0.63 (0.51–0.78)	398	2.96 (2.68–3.27)
Regularity							
Initial	1,264,340	4,826	3.82 (3.71–3.93)	812	0.64 (0.60–0.69)	3,752	2.97 (2.87–3.06)
Successive	2,751,091	7,744	2.81 (2.75–2.88)	1,193	0.43 (0.41–0.46)	6,243	2.27 (2.21–2.33)
Regular	2,373,115	6,448	2.72 (2.65–2.78)	986	0.42 (0.39–0.44)	5,220	2.20 (2.14–2.26)
Irregular	377,976	1,296	3.43 (3.24–3.62)	207	0.55 (0.48–0.63)	1,023	2.71 (2.54–2.88)
Type of mammogram							
SFM	3,910,208	12,134	3.10 (3.05–3.16)	1,910	0.49 (0.47–0.51)	9,674	2.47 (2.43–2.52)
DM	105,223	436	4.14 (3.76–4.55)	95	0.90 (0.73–1.10)	321	3.05 (2.73–3.40)
Relative change[a]			33.53		84.83		23.31
Reading method							
Single	1,506,187	4,327	2.87 (2.79–2.96)	711	0.47 (0.44–0.51)	3,414	2.27 (2.19–2.34)
Double	2,509,244	8,243	3.29 (3.21–3.36)	1,294	0.52 (0.49–0.54)	6,581	2.62 (2.56–2.69)
Relative change[a]			14.35		9.24		15.71
Number of projections							
One	1,240,656	3,625	2.92 (2.83–3.02)	564	0.45 (0.42–0.49)	2,954	2.38 (2.30–2.47)
Two	2,774,775	8,945	3.22 (3.16–3.29)	1,441	0.52 (0.49–0.55)	7,041	2.54 (2.48–2.60)
Relative change[a]			10.33		14.24		6.57

Relative change was calculated for the protocol variables only

SFM screen-film mammogram, *DM* digital mammogram

[a] The reference was the first category

with DM strategy (16.71 ‰) and the lowest with SST (11.11 ‰). Figures 1 and 2 show the cumulative risk of detection for DCIS and invasive cancer, respectively, by protocol strategy. For DCIS (Fig. 1), the highest cumulative risk of detection at 10 years was found with SDT strategy (2.32 ‰), followed by DM strategy (2.27 ‰). For invasive cancer (Fig. 2), the highest cumulative risk was observed with DM strategy (12.07 ‰), followed by SDT strategy (11.21 ‰). The lowest cumulative risk for both DCIS and invasive was observed with SST strategy (1.42 ‰ and 9.24 ‰, respectively).

Table 4 presents the relative change in the cumulative risk of detection at 10 years for the SDT, SDTO, and DDT strategies compared with the SST strategy. The highest relative change was for the DDT strategy compared with SST (50.66 %), the difference being highest for DCIS (59.06 %). The SDT strategy detected 25.48 % more cancers than the SST strategy: 62.69 % more DCIS and 21.22 % more invasive cancers.

Discussion

After 10 years of participation in a population-based breast cancer screening program, the cumulative risk of cancer detection ranged from 11.11 to 16.71 per 1,000 participants, depending on the screening strategy used. The highest cumulative detection rate, both for invasive cancer and DCIS, was found in radiology units using DM or SFM, respectively, double reading and two views. The effect of radiologic protocol variables on cancer detection was higher than that of the distinct protocol strategies. However, although the differences between protocol strategies at each participation were small, the cumulative effect at 10 years showed wide variability.

There are few reports of the risk of developing cancer in the following 10 years after a particular age [18–21] and even fewer deal with DCIS or invasive cancer separately. One of the most commonly used methods was developed by the National Cancer Institute [18, 22, 23]. This

Table 3 The adjusted effect of radiologic protocol-related variables (model 1 and 3) and protocol strategies (model 2 and 4) on risk detection

	Overall		DCIS[a]		Invasive[a]	
	Model 1 Adjusted OR[b]	Model 2 Adjusted OR[b]	Model 3 Adjusted OR[b]	Model 4 Adjusted OR[b]	Model 3 Adjusted OR[b]	Model 4 Adjusted OR[b]
Regularity						
Regular successive	Ref	Ref	Ref	Ref	Ref	Ref
Irregular successive	1.27 (1.19–1.36)	1.29 (1.21–1.38)	1.46 (1.24–1.72)	1.50 (1.27–1.77)	1.25 (1.16–1.35)	1.26 (1.17–1.36)
Age in screening						
44–49	0.85 (0.80–0.90)	0.86 (0.82–0.92)	1.02 (0.89–1.18)	1.05 (0.92–1.21)	0.81 (0.76–0.86)	0.82 (0.76–0.87)
50–54	Ref	Ref	Ref	Ref	Ref	Ref
55–59	1.19 (1.14–1.25)	1.20 (1.14–1.25)	1.15 (1.02–1.29)	1.16 (1.03–1.30)	1.19 (1.13–1.26)	1.20 (1.14–1.26)
60–64	1.45 (1.36–1.54)	1.46 (1.37–1.55)	1.22 (1.04–1.43)	1.23 (1.05–1.44)	1.50 (1.41–1.61)	1.51 (1.41–1.61)
65–70	1.57 (1.41–1.75)	1.57 (1.41–1.75)	1.59 (1.22–2.07)	1.59 (1.22–2.07)	1.58 (1.41–1.79)	1.59 (1.41–1.79)
Type of mammogram						
SFM	Ref		Ref		Ref	
DM	1.30 (1.14–1.48)		1.40 (1.06–1.86)		1.21 (1.04–1.40)	
Reading method						
Single	Ref		Ref		Ref	
Double	1.31 (1.18–1.45)		1.44 (1.17–1.79)		1.27 (1.14–1.42)	
Number of projections						
One	Ref		Ref		Ref	
Two	1.11 (1.04–1.18)		1.12 (0.96–1.32)		1.07 (0.99–1.15)	
Strategy						
SDT		Ref		Ref		Ref
SDTO		1.01 (0.88–1.15)		1.00 (0.76 – 1.32)		0.99 (0.86 – 1.14)
SST		0.80 (0.71–0.90)		0.61 (0.48–0.79)		0.82 (0.72–0.94)
DDT		1.20 (0.98–1.47)		0.98 (0.62–1.54)		1.08 (0.85–1.37)

We estimated each model by protocol variables or protocol strategy

Strategy: *SDT* screen-film mammography, double reading and two views, *SDTO* screen-film mammography, double reading and two views in initial participation and one view in successive; *SST* screen-film mammography, single reading and two views, *DDT* digital mammography, double reading and two views

[a] Woman with unknown cancer histology was excluded of this analysis. Multinomial model with three possible outcomes: No cancer, DCIS cancer or Invasive cancer

[b] Adjusted also for radiologic unit (as random effect), number of screening participation and period (year). In parenthesis, the 95 % confidence interval

methodology has been applied in breast cancer in distinct countries [19–21]. In the USA, the risk of developing cancer in the next 10 years in 50-year-old screened women from 1975 to 2009 was reported to be 19.1 ‰ [19]. In the UK, this risk in 50-year-old women was estimated as 25.5 ‰ [20]. In Flanders, this risk in women aged 50–54 years was 34.0 ‰ [21]. Our estimations are lower, which may be due to the lower baseline incidence of cancer in Spain than in Belgium, the USA and the UK [24].

The effect of radiologic protocol-related variables has been evaluated for a specific period [25], but the cumulative effect has been little studied. Several studies have evaluated double reading versus single reading in each participation [9, 26, 27]. The increased detection due to double reading may result from the detection of smaller cancers [26]. The effect of double reading on DCIS detection, however, has been less extensively evaluated [7]. Our results suggest that double reading had a greater effect on detection of DCIS than on that of invasive cancers, since double reading is more sensitive to smaller tumors, irrespective of type [26].

Digital mammography may increase cancer detection, the increase being higher for DCIS. In part, this effect has been attributed to its improved contrast resolution [28]. In Dutch screening program between 2004 and 2006, DM increased DCIS detection by 80 % and invasive cancers by 12 % [28]. In Florence between 2004 and 2005, DM increased DCIS detection by 93 % and invasive cancer by 9 % [29]. Part of the increase in DCIS with DM can be explained by the greater sensitivity of DM to microcalcifications compared

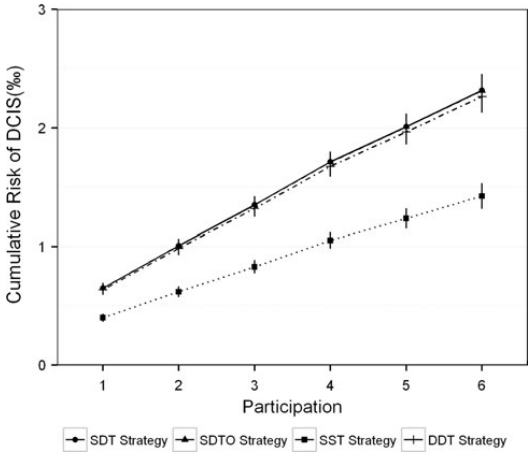


Fig. 1 The cumulative risk of DCIS detection by different protocol strategies

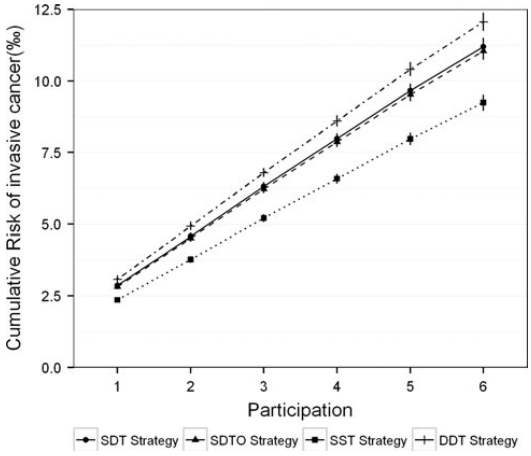


Fig. 2 The cumulative risk of invasive cancer detection by different protocol strategies

with that of SFM [30–32]. Our results are consistent with these results when the radiologic protocol-related variables were analyzed separately. When strategies were compared, however, no differences were observed in DCIS detection while the overall cumulative detection risk was higher with digital strategies. The extent to which DCIS contributes to overdiagnosis is unknown, and there is some evidence that screen-detected DCIS is biologically more aggressive than symptomatic DCIS [33]. The higher cumulative detection rate of strategies using DM could be attributable to greater detection of invasive cancer.

Table 4 Relative change in cancer detection between the SDT, SDTO, and DDT strategies and SST strategy. Cancer detection was lowest with the SST strategy

Cancer detection	Protocol strategy compared to SST	Relative change in cancer detection
Overall	SDT	25.48
	SDTO	26.42
	DDT	50.66
DCIS	SDT	62.69
	SDTO	62.56
	DDT	59.06
Invasive	SDT	21.22
	SDTO	19.53
	DDT	30.59

Strategy: *SDT*: screen-film mammography, double reading and two views, *SDTO* screen-film mammography, double reading and two views in initial participation and one view in successive, *SST* screen-film mammography, single reading and two views, *DDT* digital mammography, double reading and two views

Our study was designed to evaluate the detection rate and screening program characteristics based on routinely collected information in population-based breast cancer screening programs. Therefore, the main cancer-related variables such as age at menarche, age at maternity, the number of children, breast density in the current mammogram, etc. were not available.

This study is one of the first to compare the effect of radiologic protocol-related variables and the cumulative effect of distinct protocol strategies on DCIS and the invasive cancer detection rate in population-based breast cancer screening. Although the differences at each screening may be small, the cumulative effect may have a greater impact on cancer detection.

In conclusion, based on our findings, between 11 and 16 per 1,000 screened women are diagnosed with screen-detected cancer during the first 10 years of follow-up. A better knowledge of the risks and benefits of breast cancer screening, such as understanding the sources of variability produced by the various protocol strategies, could improve cancer prevention policies and allow women to be offered the best possible protocol strategies.

Acknowledgments This study was supported by grants from the Instituto de Salud Carlos III-FEDER (PI09/90251 and PI07/90293). The funding sources had no role in the performance of the study or in the preparation of the manuscript. The authors acknowledge the dedication and support of the entire "Cumulative False- Positive Risk Group". Cumulative False-Positive Risk Group (alphabetical order): Department of Epidemiology and Health Services' Evaluation, Institut Municipal d'Investigació Mèdica-Parc de Salut Mar, Barcelona: Jordi Blanch, Andrea Burón, Xavier Castells, Laia Domingo, Francesc Macià, Marta Román, María Sala. Galician Breast Cancer Screening Program, Directorate for innovation and management of

public health. Health Office, Galicia: Raquel Almazán, Ana Belén Fernández, María Teresa Queiro, Raquel Zubizarreta. Navarre Breast Cancer Screening Program. Public Health Institute, Pamplona: Nieves Ascunce, Ana Barcos, Iosu Delfrade, María Ederra, Nieves Erdozain. General Directorate of Public Health and Center for Public Health Research (CSISP-FISABIO), 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 & Analysis Unit. Health Office, Asturias: Carmen Natal. La Rioja Breast Cancer Screening Program. Fundacion Rioja Salud, Logroño: Araceli Baroja. Head of the Health Promotion and Protection Section. Cancer Screening and Epidemiology Department, UDIAT-CD. Corporació Parc Taulí-Institut Universitari Parc Taulı (UAB), Sabadell: Marisa Baré. Health Promotion and Protection Office, Valladolid: Isabel González. Programas de Salud de la Mujer, Dirección General de Salud Pública, Consejería de Sanidad y Servicios Sociales Gobierno de Cantabria, Santander: Mar Sánchez. Radiology Unit, Hospital Universitario Marqués de Valdecilla, Santander, Spain: Alfonso Vega. Hospital Santa Caterina, Girona, Spain: Joana Ferrer.

Conflict of interest None declared.

References

1. Suhrke P, Maehlen J, Schlichting E, Jorgensen KJ, Gotzsche PC, Zahl PH (2011) Effect of mammography screening on surgical treatment for breast cancer in Norway: comparative analysis of cancer registry data. BMJ 343:d4692
2. Glover JA, Bannon FJ, Hughes CM et al (2012) Increased diagnosis and detection rates of carcinoma in situ of the breast. Breast Cancer Res Treat 133(2):779–784
3. Puig-Vives M, Pollan M, Rue M et al (2012) Rapid increase in incidence of breast ductal carcinoma in situ in Girona, Spain 1983–2007. Breast 21(5):646–651
4. 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(3):170–178
5. Jones JL (2006) Overdiagnosis and overtreatment of breast cancer: progression of ductal carcinoma in situ: the pathological perspective. Breast Cancer Res 8(2):204
6. Taplin S, Abraham L, Barlow WE et al (2008) Mammography facility characteristics associated with interpretive accuracy of screening mammography. J Natl Cancer Inst 100(12):876–887
7. Shaw CM, Flanagan FL, Fenlon HM, McNicholas MM (2009) Consensus review of discordant findings maximizes cancer detection rate in double-reader screening mammography: Irish National Breast Screening Program experience. Radiology 250(2):354–362
8. Blanks RG, Given-Wilson RM, Moss SM (1998) Efficiency of cancer detection during routine repeat (incident) mammographic screening: two versus one view mammography. J Med Screen 5(3):141–145
9. Williams LJ, Hartswood M, Prescott RJ (1998) Methodological issues in mammography double reading studies. J Med Screen 5(4):202–206
10. Sala M, Comas M, Macia F, Martinez J, Casamitjana M, Castells X (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(1): 31–39
11. Nederend J, Duijm LE, Louwman MW, Groenewoud JH, Donkers-van Rossum AB, Voogd AC (2012) Impact of transition from analog screening mammography to digital screening mammography on screening outcome in The Netherlands: a population-based study. Ann Oncol 23(12):3098–3103
12. 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 23(1):104–111
13. Roman R, Sala M, De La Vega M et al (2011) Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening. Breast Cancer Res Treat 130(2):543–552
14. 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, Luxemburg
15. Hosmer DW, Lemeshow S (2002) Applied logistic regression, 2nd edn. Wiley, New York
16. Singer J, Willett JB (2003) Fitting basic discrete-time hazard models. In: Singer J, Willett JB (eds) Applied longitudinal data analysis: modelling change and event occurrence. Oxford University Press, New York, pp 357–467
17. Singer J, Willett JB (2003) Describing discrete-time event occurrence data. In: Singer J, Willett JB (eds) Applied longitudinal data analysis: modelling change and event occurrence. Oxford University Press, New York, pp 325–356
18. Fay MP, Pfeiffer R, Cronin KA, Le C, Feuer EJ (2003) Age-conditional probabilities of developing cancer. Stat Med 22(11): 1837–1848
19. Welch HG, Frankel BA (2011) Likelihood that a woman with screen-detected breast cancer has had her "life saved" by that screening. Arch Intern Med 171(22):2043–2046
20. Pashayan N, Duffy SW, Chowdhury S et al (2011) Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. Br J Cancer 104(10):1656–1663
21. Goossens MC, De Greve J (2010) Individual cancer risk as a function of current age and risk profile. Eur J Cancer Prev 19(6):485–495
22. Fay MP (2004) Estimating age conditional probability of developing disease from surveillance data. Popul Health Metr 2(1):6
23. Wun LM, Merrill RM, Feuer EJ (1998) Estimating lifetime and age-conditional probabilities of developing cancer. Lifetime Data Anal 4(2):169–186
24. Ferlay J, Shin HR, Bray F (2010) Cancer incidence and mortality worldwide: IARC CancerBase No. 10
25. Giordano L, von Karsa L, Tomatis M et al (2012) Mammographic screening programmes in Europe: organization, coverage and participation. J Med Screen 19(Suppl 1):72–82
26. Blanks RG, Wallis MG, Moss SM (1998) A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers, for one and two view mammography: results from the UK National Health Service breast screening programme. J Med Screen 5(4):195–201
27. Anderson ED, Muir BB, Walsh JS, Kirkpatrick AE (1994) The efficacy of double reading mammograms in breast screening. Clin Radiol 49(4):248–251
28. de Gelder R, Fracheboud J, Heijnsdijk EA et al (2011) Digital mammography screening: weighing reduced mortality against increased overdiagnosis. Prev Med 53(3):134–140
29. Pisano ED, Gatsonis C, Hendrick E et al (2005) Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 353(17):1773–1783
30. Del Turco MR, 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(4): 860–866
31. Domingo L, Romero A, Belvis F 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(9):2020–2028

32. Irwig L, Houssami N, Armstrong B, Glasziou P (2006) Evaluating new screening tests for breast cancer. BMJ 332(7543): 678–679
33. de Roos MA, Pijnappel RM, Groote AD, de Vries J, Post WJ, Baas PC (2004) Ductal carcinoma in situ presenting as microcalcifications: the effect of stereotactic large-core needle biopsy on surgical therapy. Breast 13(6):461–467

3.4 2n article

OPEN ACCESS Freely available online

PLOS ONE

Impact of Risk Factors on Different Interval Cancer Subtypes in a Population-Based Breast Cancer Screening Programme

Jordi Blanch[1,2], Maria Sala[1,2,3]*, Josefa Ibáñez[4,5], Laia Domingo[1,3], Belén Fernandez[6], Arantza Otegi[7], Teresa Barata[8], Raquel Zubizarreta[6], Joana Ferrer[9], Xavier Castells[1,2,3], Montserrat Rué[3,10], Dolores Salas[4,5], for the INCA Study Group¶

1 Department of Epidemiology and Evaluation, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain, **2** European Higher Education Area (EHEA) Doctoral Programme in Public Health in Department of Pediatrics, Obstetrics and Gynecology, Preventive Medicine and Public Health, Universitat Autònoma de Barcelona (UAB), Bellaterra, Barcelona, Spain, **3** Research network on health services in chronic diseases (REDISSEC), Barcelona, Spain, **4** General Directorate Public Health, Valencia, Spain, **5** Centre for Public Health Research (CSISP), FISABIO, Valencia, Spain, **6** Galician Breast Cancer Screening Programme, Directorate for innovation and management of public health, Santiago de Compostela, Spain, **7** Osakidetza Breast Cancer Screening Programme, Basque Country Health Service, Bilbao, Spain, **8** General Directorate of Health Care Programmes, Canary Islands Health Service, Las Palmas de Gran Canaria, Spain, **9** Department of Radiology, Hospital de Santa Caterina, Salt, Girona, Spain, **10** Department of Basic Medical Sciences, Biomedical Research Institut of Lleida (IRBLLEIDA)-University of Lleida, Lleida, Spain

Abstract

Background: Interval cancers are primary breast cancers diagnosed in women after a negative screening test and before the next screening invitation. Our aim was to evaluate risk factors for interval cancer and their subtypes and to compare the risk factors identified with those associated with incident screen-detected cancers.

Methods: We analyzed data from 645,764 women participating in the Spanish breast cancer screening program from 2000–2006 and followed-up until 2009. A total of 5,309 screen-detected and 1,653 interval cancers were diagnosed. Among the latter, 1,012 could be classified on the basis of findings in screening and diagnostic mammograms, consisting of 489 true interval cancers (48.2%), 235 false-negatives (23.2%), 172 minimal-signs (17.2%) and 114 occult tumors (11.3%). Information on the screening protocol and women's characteristics were obtained from the screening program registry. Cause-specific Cox regression models were used to estimate the hazard ratios (HR) of risks factors for interval cancer and incident screen-detected cancer. A multinomial regression model, using screen-detected tumors as a reference group, was used to assess the effect of breast density and other factors on the occurrence of interval cancer subtypes.

Results: A previous false-positive was the main risk factor for interval cancer (HR = 2.71, 95%CI: 2.28–3.23); this risk was higher for false-negatives (HR = 8.79, 95%CI: 6.24–12.40) than for true interval cancer (HR = 2.26, 95%CI: 1.59–3.21). A family history of breast cancer was associated with true intervals (HR = 2.11, 95%CI: 1.60–2.78), previous benign biopsy with a false-negatives (HR = 1.83, 95%CI: 1.23–2.71). High breast density was mainly associated with occult tumors (RRR = 4.92, 95%CI: 2.58–9.38), followed by true intervals (RRR = 1.67, 95%CI: 1.18–2.36) and false-negatives (RRR = 1.58, 95%CI: 1.00–2.49).

Conclusion: The role of women's characteristics differs among interval cancer subtypes. This information could be useful to improve effectiveness of breast cancer screening programmes and to better classify subgroups of women with different risks of developing cancer.

Citation: Blanch J, Sala M, Ibáñez J, Domingo L, Fernandez B, et al. (2014) Impact of Risk Factors on Different Interval Cancer Subtypes in a Population-Based Breast Cancer Screening Programme. PLoS ONE 9(10): e110207. doi:10.1371/journal.pone.0110207

Editor: William B. Coleman, University of North Carolina School of Medicine, United States of America

Received May 28, 2014; **Accepted** September 10, 2014; **Published** October 21, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. Data are available upon request because they contain identifying human information and are unsuitable for public deposition. Requests may be made to the corresponding author, Dr. Sala.

Funding: This study has received funding by Instituto de Salud Carlos III-FEDER (PI 09/01153, PI09/02385, PI09/01340). 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.

* Email: msalaserra@parcdesalutmar.cat

¶ Membership of the INCA Study Group is provided in the Acknowledgments.

Introduction

Evaluation of the impact of population-based breast cancer screening programmes is complex and can only be achieved in the long term. Regular screening contributes to reducing mortality from breast cancer, but also has adverse effects, such as false-positives, overdiagnosis and interval cancers [1,2]. The European guidelines for quality assurance in breast cancer screening and diagnosis recommend the use of early performance indicators that provide information of the impact of screening [3]. These

indicators should measure positive and negative effects to achieve a long-term balance between benefits and harms. One of the most important surrogate indicators of the effectiveness of breast cancer screening programmes is the interval cancer rate.

Interval cancers are primary breast cancers diagnosed in women after a negative screening test and before the next screening invitation [3]. Because they are diagnosed by symptoms, affected women lose the benefit of early detection and, in case of false-negative results, suffer delayed diagnosis and treatment.

Numerous studies have examined interval cancer rates in distinct screening programmes [4–13], allowing comparison of the sensitivity of these programmes. However, most studies do not distinguish among interval cancer subtypes, which can only be achieved after a radiological review of both screening and diagnostic mammograms. Half of interval cancers are true interval cancers in our context [14], which are tumours that are undetectable in the last screening participation but become symptomatic before the next participation. Failure to detect these tumours is caused by the limitations of the screening test and is inherent to organized screening process. Retrospectively, false-negative cancers (i.e., missed) can be seen in the screening mammogram and their occurrence is associated with the organization of screening programmes [14–16]. The remaining subtypes are minimal-sign cancers (which show detectable but non-specific sign at latest screen) and occult tumours at mammography (no signs of mammographic abnormalities either at screening or at diagnostic), which have been less studied since they account for less than 25% of interval cancers [14].

Study of interval cancers has focused on detecting differences with screen-detected cancers according to personal and tumor-related characteristics [14,16–18]. Compared with screen-detected cancers, interval cancers show a higher prevalence of features associated with poor prognosis [14,16]. Breast density has also been related to interval cancer [17,19], showing a positive association between higher breast density and interval cancer risk. However, the specific role of breast density on interval cancer subtypes taking into consideration personal and organizational characteristics has not been evaluated.

The aim of our study was to evaluate the factors associated with a higher probability of interval cancer (overall, true interval, false negative, minimal signs and occult tumours) and to compare them with those associated with incident screen-detected cancers. We also studied the influence of breast density on detection mode and the occurrence of subtypes.

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. Further information about our data and the methods used can be requested from the authors.

Settings and study population

Population-based breast cancer screening in Spain is offered individually to 100% of the target population by the National Health Service. This programme adheres to the European Guidelines [3] and its results meet the required standards [20]. Each of the 17 administrative regions in Spain is responsible for the local application and has several radiology units that carry out screening. Despite the high degree of consensus, regional application can vary in the target population definition and in the mammographic screening protocol used (starting age, single or double reading, etc) [20].

All women aged 50 years (or 45 years depending on the region) to 69 years are actively invited to participate by letter every 2 years [20]. The screening programmes stop inviting women with a breast cancer diagnosis.

We built a retrospective cohort of 645,764 screened women from 32 radiology units in five regions of Spain who underwent mammography between January 1, 2000 and December 31, 2006 and who were followed up until June 30, 2009 for interval cancer identification. These women underwent a total of 1,508,584 screening mammograms (Figure 1). During the study period, 5,309 cancers were detected in routine screening mammograms, of which 3,547 were detected in successive participations, and 1,653 emerged as interval cancers. Interval cancers with unknown diagnostic date (n = 16) were excluded. We included both invasive and in situ carcinomas.

Screening outcomes

Screening mammography has three possible outcomes: a negative result, a positive result (abnormal findings requiring further assessments) or early recall (an intermediate mammogram is performed out of sequence with the screening interval at 6 or 12 months). Cancers detected at intermediate mammogram were considered screen-detected cancers [3].

A positive result is considered to be screen-detected tumour if, after further assessments, there is histopathological confirmation of cancer. Otherwise, the result is considered a false-positive and woman is invited at 2 years. Further assessments can include non-invasive procedures (magnetic resonance imaging, ultrasonography, additional mammography) and/or invasive (fine-needle aspiration cytology, core-needle biopsy and open biopsy).

Interval cancer was defined as "a primary breast cancer arising after a negative screening episode and before the next invitation to screening or within 24 months for women who reached the upper age limit" [3]. We extended the definition until the 30th month, because we allowed a 6 month margin for women to attend each round. Interval cancers were identified by merging data from the registers of screening programmes with population-based cancer registries [21], the regional Minimum Data Set (based on hospital discharges with information on the main diagnosis) and hospital-based cancer registries.

Interval cancer classification and breast density assessment

Interval cancers were classified by three panels, formed by three experienced radiologists. The panels performed a semi-informed independent review of both screening and diagnostic mammograms with independent double reading and arbitration. Of 1,653 interval cancers, 1,012 had available screening and diagnostic mammograms, and therefore could be classified into the four interval cancer categories. Briefly, screening mammograms were first reviewed alone and provisionally classified into positive, negative, and minimal-signs. Afterwards, the screening and diagnostic mammograms were reviewed together and interval cancers were definitively classified into: true interval cancers (the screening mammogram showed normal or benign results), false-negatives (an abnormality suspicious for malignancy was retrospectively seen on the screening mammogram), minimal-signs (detectable but non-specific signs were identified on the screening mammogram) and occult tumours (showing no mammographic abnormalities at diagnosis despite clinical signs) [3]. More details of the classification process have been reported in a previous study [15].

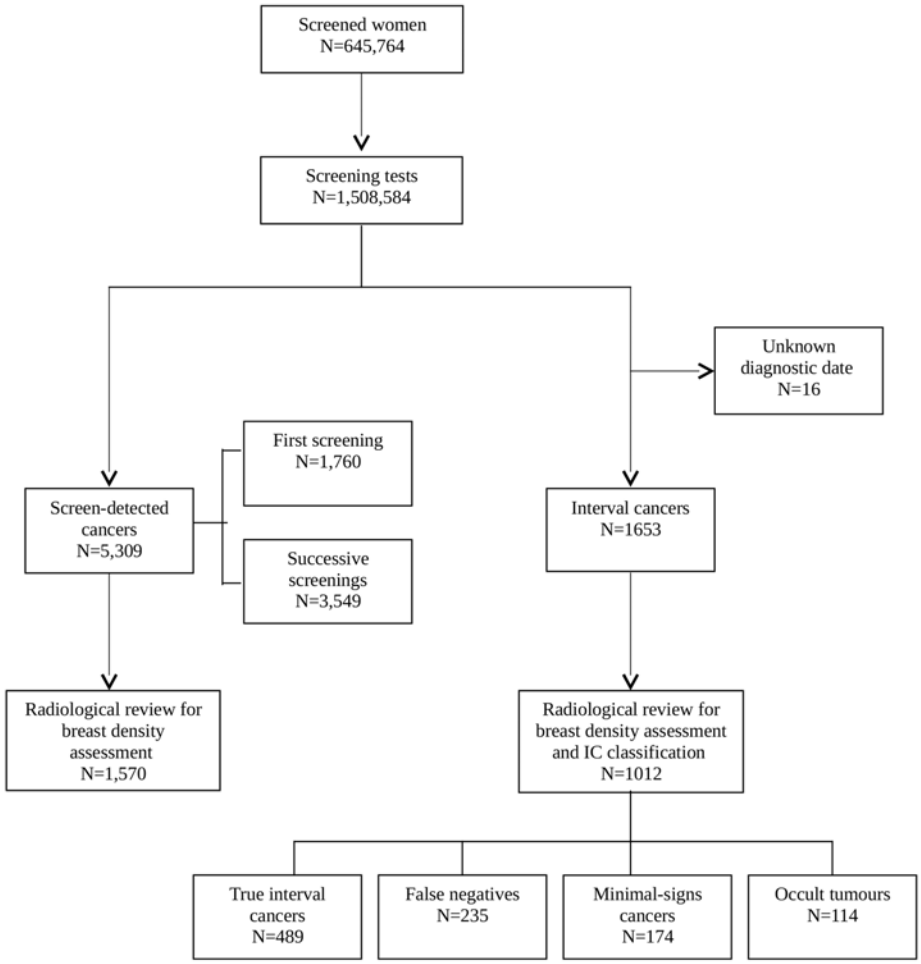


Figure 1. Flow chart of the study populations.
doi:10.1371/journal.pone.0110207.g001

The breast density of the cancer-free breast was determined by one radiologist from each panel, using Boyd's scale, a semi-quantitative score of six categories using percentages of density: A: 0%; B: 1–10%; C:10–25%; D:25–50%; E:50–75%; F:75–100% [22]. For purposes of this study, the first three categories were combined into the <25% group [23]. Breast density was assessed for classified interval cancers (N = 1,012) and from a sample of screen-detected cancers matched by region and year of the last screening (N = 1,570).

Study variables

All information was collected from each woman at each attendance. The variables related to the screening protocol included the reading method (single or double reading) and mammography type (screen-film mammography [SFM] or digital mammography [DM]). The variables related to women's personal characteristics were use of hormonal replacement therapy (HRT) at screening or in the previous 6 months (Yes/No), menopausal status (pre- or postmenopausal), previous benign biopsy outside screening (Yes/No), the presence or absence of a first-degree familial history of breast cancer (Yes/No), mammography date, number of participations, the existence of a previous false-positive (Yes/No), and the existence of an early recall (Yes/No).

Statistical analysis

Breast cancer rates per 1,000 women and per 10,000 mammograms were estimated, for interval cancers and for screen-detected cancers.

The cumulative probability of suffering an interval cancer was calculated using the Kaplan-Meier estimator with time-dependent variables according the distinct study variables. We estimated the hazard ratios (HR) of interval cancer and incident screen-detected cancer using a multivariate cause-specific Cox model with time-dependent variables. Radiology units were included as a random

effect [24]. We built explanatory models including all relevant covariates according to the literature to evaluate their effect and statistical significance on the risk of cancer detection and of interval cancer subtypes.

Given that incidence of breast cancer depends on age, we used age as time scale in the Kaplan-Meier and Cox regression analyses [25]. Age at study entry was defined as the age at the first mammogram in the study period. Age at study exit was the lowest of: age at breast cancer diagnosis, age at the study closure, or age at 30 months after the last mammogram. Observations were censored when the event breast cancer diagnosis did not occur during the study period.

We conducted a case-control analysis to determine whether the study variables differed between interval cancers (cases) and screen-detected cancers (controls), using a multivariate logistic regression model. For the analysis of variables associated to the subtypes of interval cancer, we fitted a multinomial regression model, using the screen-detected cancers as a reference group. The multinomial model estimates k-1 models, comparing each group to the referent group. The exponentiated multinomial logit coefficient provides an estimate of the relative risk and, commonly, is expressed as relative risk ratio (RRR).

All P-values were based on two-sided tests and were considered statistically significant if less than 0.05. Statistical analyses were performed using the R statistical software (version 3.0.1) [26].

Results

Interval cancer subtypes and risk factors

The interval cancer rate was 2.57 per 1,000 screened women and 10.99 per 10,000 screening mammograms (Table 1). The largest proportion of interval cancers was diagnosed in the second year (60.0% of interval cancers). The radiologist teams classified 1,012 interval cancers as follows: 489 true intervals, 235 false-negatives, 174 minimal-signs and 114 occult tumours.

Among classified interval cancers, 42.1% of occult tumours, 32.3% of false-negatives and 31.0% of minimal-signs were diagnosed in the first 12 months after the participation and only 19.6% of true interval cancers emerged in the same period.

Figure 2 presents the cumulative probability of developing an interval cancer according to family history of breast cancer, previous benign biopsy outside screening and the presence of a previous false-positive result. The other covariates did not show statistically significant differences in the bivariate analysis (data not shown). Women with a false-positive result in the previous participation had a higher cumulative probability of developing interval cancer than women without (0.033 vs 0.013). Women with family history of breast cancer or previous benign biopsy result outside screening had a higher cumulative probability than women without (0.021 vs 0.012 and 0.022 vs 0.012, respectively).

The hazard risks from cause-specific survival analyses for incident screen-detected cancers and interval cancers (overall and subtypes) are shown in Table 2 (see also tables S1 and S2). The existence of a false-positive in the previous mammogram showed the highest hazard ratio for developing interval cancer (HR = 2.71; 95% CI: 2.28–3.23), while having an early recall showed the highest hazard ratio for screen-detected cancer (HR = 3.59; 95% CI:3.12–4.14). A previous false-positive result, premenopausal status, a family history of breast cancer and previous benign breast biopsy were risk factors for both interval and screen-detected cancers. HRT use was significantly associated with increased risk of developing interval cancer (HR = 1.27; 95% CI:1.07–1.5). DM was associated with screen-detected cancer (HR = 1.34; 95%

Table 1. Number of screen-detected and interval cancers (by time of diagnosis) with their rates per 1,000 women and 10,000 mammograms and proportions of interval cancer subtypes among radiologically classified interval cancers.

		Rates per 1,000 women	Rates per 10,000 Mammograms	Radiologic information	TI	FN	MS	OT
	N			*N*	*N*	*N*	*N*	*N*
Screened women	645,764							
Mammograms	1,508,584							
Screen-detected cancer	5,307	8.22	35.18	1,570				
In successive participations	3,547	5.49	23.54	1,100				
Interval cancer	1,653	2.57	10.99	1,012	489	235	174	114
Time of diagnosis*	N (%)			N (%)	N (%)	N (%)	N (%)	N (%)
<12 months	484 (29.3)	0.75	3.21	273 (27.0)	96 (19.6)	76 (32.3)	53 (30.5)	48 (42.1)
12–23 months	992 (60.0)	1.54	6.58	626 (61.9)	322 (65.8)	139 (59.1)	105 (60.3)	60 (52.6)
> = 24 months	177 (10.7)	0.27	1.17	113 (11.2)	71 (14.5)	20 (8.5)	16 (9.2)	6 (5.3)

*Time of diagnosis (in months) for interval cancer is the time between the last mammogram and diagnosis of cancer.

Abbreviations: TI: True interval, FN: False-negative, MS: Minimal-signs, OT: Occult tumours.

doi:10.1371/journal.pone.0110207.t001

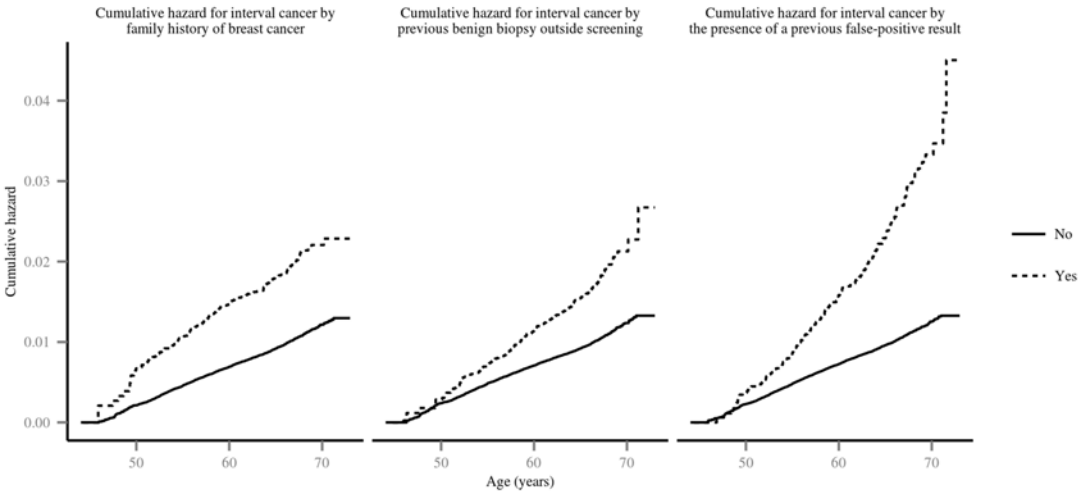


Figure 2. Cumulative hazard for interval cancer by the presence of a previous false-positive result, family history of breast cancer and previous benign biopsy outside screening.
doi:10.1371/journal.pone.0110207.g002

CI:1.12–1.61). Early recall was a protective factor for interval cancer (HR = 0.56; 95% CI:0.40–0.78).

The effect of the risk factors on developing interval cancer differed according to subtypes. DM was associated with true interval cancer (HR = 2.17; 95% CI:1.37–3.45), but not with false-negatives (HR = 1.57; 95% CI:0.78–3.13). The effect of a previous false-positive result was higher for false-negatives (HR = 8.79; 95% CI:6.24–12.40) than for true interval cancer (HR = 2.26; 95% CI:1.59–3.21). HRT use was a statistically significant risk factor for minimal-signs (HR = 1.63; 95% CI:1.03–2.56) and a family history of breast cancer was a risk factor for true interval cancer (HR = 2.11; 95% CI:1.60–2.78) and minimal-signs as well (HR = 1.67; 95% CI:1.02–2.74). Previous benign biopsy outside screening was a risk factor for a false-negative cancers (HR = 1.83; 95% CI:1.23–2.71).

Comparison of breast density and other risk factors between interval cancers and screen detected cancers

The proportion of women with extremely dense breasts (Table 3) was higher for interval cancers than for screen-detected cancers (16.4% and 11.7%, respectively, P<0.001); the highest proportion was found in occult tumours (28.1%) and the lowest in minimal-signs cancers (9.8%).

Table 4 presents the risk factors associated with interval cancers compared with screen-detected cancers. Risk factors for overall interval cancer were the presence of a previous false-positive (OR = 2.11; 95% CI:1.56–2.86), HRT use (OR = 1.57; 95% CI:1.18–2.10) and extremely dense breasts (OR = 1.63; 95% CI:1.24–2.14). By interval cancer subtypes, risk factors for true interval cancer were the same as those for interval cancer: the presence of a previous false-positive (RRR = 1.79; 95% CI:1.21–2.63), HRT use (RRR = 1.57; 95% CI:1.10–2.25) and extremely dense breasts (RRR = 1.67; 95% CI:1.18–2.36). However, for false negatives, only the presence of a previous false-positive result was a risk factor (RRR = 4.55; 95% CI:3.07–6.75), which was the strongest observed association in this analysis. Extremely breast density was at the limit of significance (RRR = 1.58; 95% CI:1.00–2.49). The only statistically significant risk factor for minimal signs was HRT use (RRR = 1.84; 95% CI:1.12–3.02), and that for occult tumours was extremely breast density (RRR = 4.92; 95% CI:2.58–9.38).

Discussion

This study provides information on the determinants of interval cancer and its subtypes and a comparison with screen-detected cancer in a retrospective cohort of screened women. Women's characteristics (premenopausal status, a family history of breast cancer, and previous benign breast biopsy) were risk factors for both interval and screen-detected cancer, showing a similar strength of association. A family history of breast cancer had the strongest association with true interval cancer while previous benign biopsy was a significant risk factor only for false-negatives. The presence of a previous false-positive result was a risk factor for both screen-detected and interval cancer but the association with interval cancer was stronger, especially for false-negative cancers. Double reading, digital mammography and early recall were associated with screen-detected cancer but not with interval cancer.

Few studies have analysed the determinants of interval cancer and even fewer have taken subtypes into account. Previous studies have described interval cancer rates ranging from 1.8 to 2.9 per 1,000 women and from 10.6 to 29.5 per 10,000 mammograms [4–13], which are consistent with our results. The main risk factor for interval cancer was a previous false-positive result, with the strongest association for false-negatives, suggesting that some results interpreted as false-positives may, in fact, be false-negatives [27–29]. In fact, false positive have been found also a risk factor for screen cancer detection especially when cytology or biopsy had been performed as further assessment [30,31].

DM was not associated with increased risk of interval cancer, in line with the study by Hoff et al. [32] and a recent study from Norway [33], showing that DM use did not modify false-negatives rates. The increased risk of true interval cancers with DM could be explained by the higher precision in detecting less advanced and

Table 2. Adjusted hazard ratios from cause-specific survival analyses for incident screen-detected cancers and for interval cancer (overall and subtypes).

	Incident SDC	IC	TI	FN	MS	OT
	HR* (95%CI)	HR* (95%CI)	HR* (95%CI)	HR* (95%CI)	HR* (95%CI)	HR* (95%CI)
Reading Method						
Double	Ref	Ref	Ref	Ref	Ref	Ref
Single	1.31 (0.85–2.02)	0.85 (0.60–1.21)	1.55 (0.66–3.62)	1.09 (0.36–3.25)	1.38 (0.26–7.28)	0.52 (0.15–1.81)
Type of mammogram						
SFM	Ref	Ref	Ref	Ref	Ref	Ref
DM	1.34 (1.12–1.61)	1.00 (0.78–1.28)	2.17 (1.37–3.45)	1.57 (0.78–3.13)	0.97 (0.35–2.67)	0.36 (0.11–1.24)
Early Recall						
No	Ref	Ref	Ref	Ref	Ref	+
Yes	3.59 (3.12–4.14)	0.56 (0.40–0.78)	0.46 (0.22–0.96)	0.36 (0.15-0.83)	0.85 (0.29–2.46)	
Previous false-positive						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	1.34 (1.20–1.60)	2.71 (2.28–3.23)	2.26 (1.59–3.21)	8.79 (6.24–12.40)	1.80 (0.93–3.47)	0.34 (0.08–1.39)
HRT use						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	0.98 (0.86–1.11)	1.27 (1.07–1.50)	1.33 (0.99–1.79)	1.30 (0.82–2.05)	1.63 (1.03–2.56)	0.92 (0.46–1.84)
Menopausal status						
Postmenopausal	Ref	Ref	Ref	Ref	Ref	Ref
Premenopausal	1.26 (1.09–1.45)	1.41 (1.20–1.67)	1.29 (0.94–1.76)	1.28 (0.79–2.07)	0.91 (0.52–1.61)	1.33 (0.78–2.27)
Family history of breast cancer						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	1.75 (1.56–1.94)	1.65 (1.41–1.93)	2.11 (1.60–2.78)	1.44 (0.91–2.27)	1.67 (1.02–2.74)	1.44 (0.77–2.69)
Previous benign biopsy outside screening						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	1.34 (1.18–1.53)	1.73 (1.46–2.04)	1.26 (0.90–1.74)	1.83 (1.23–2.71)	1.44 (0.87–2.37)	1.12 (0.54–2.33)

*Multivariate Cox time-dependent models. The hazard ratios for each variable were adjusted for the other variables in the table, the year of the last mammogram, and for radiology unit (as a random effect).

+We excluded early recall, because there were no cases of early recall in occult tumors.

Abbreviations: SDC: Screen-detected cancer, IC: Interval Cancer, TI: True interval, FN: False-negative, MS: Minimal-signs, OT: Occult tumors, SFM: Screen-film mammography, DM: Digital mammography and HRT: Hormonal Replacement Therapy.

doi:10.1371/journal.pone.0110207.t002

smaller tumours with DM, which has been demonstrated in several studies carried out in mammography screening contexts [32–35].

Premenopausal status, a family history of breast cancer and previous benign breast biopsy result are well-known risk factors for breast cancer [36], which is coherent with our findings and those of previous studies, revealing an association with both incident screen-detected and interval cancer [9,18,37]. However, the role of these factors seemed to differ when we specifically analysed subtypes. HRT use was a risk factor for minimal-signs, but we have no information on HRT type or on treatment length, which could have affected these findings [37]. The effect of family history was greater for true interval cancer, in agreement with the hypothesis that tumours in women with a family history of breast cancer grow faster and are more aggressive [16,38].

Because information on breast density was not available for the whole cohort, we analyzed the effect of breast density as a risk factor for interval cancer subtypes compared with those of screen-detected cancers in a case-control. In previous study in the same cohort we explored the role of breast density within interval cancer subtypes, adjusting by tumour characteristics [15]. In the current work, we considered women-related and organizational characteristic, obtaining consistent results. High breast density was a risk factor for interval cancer, mainly for occult tumours, but also for true interval and false-negatives. Pollan et al. observed that breast density played a greater role in interval cancer than in screen-detected cancer [19]. The strong association of breast density and occult tumours pointed to a masking effect, but breast density appears to play a lesser role in false-negatives. The association with true interval cancer also reinforce the hypothesis that tumours stimulated by growth factors found in dense breasts [39] are more likely to be true interval cancers. Understanding the role of breast density is important in breast cancer screening since it is one of the variables proposed to tailor screening.

This study has some limitations. First, we could not get all diagnostic mammograms from interval cancers mainly because logistic reasons, avoiding a complete interval cancer classification. Moreover, misclassification among interval cancers cannot be excluded because is inherent to radiologist subjectivity. However, this misclassification would attenuate differences among study groups. Breast density was not available for the entire population because it was not systematically collected and registered.

Table 3. Distribution of variables related to the screening protocol and women's characteristics by screen-detected and interval cancer (overall and subtypes).

	SDC	IC	TI	FN	MS	OT
	N (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)
Total	1,570	1,012	489	235	174	114
Reading method						
Double	1,450 (92.4)	889 (87.8)*	424 (86.7)	202 (86.0)	162 (93.1)	101 (88.6)+
Single	120 (7.6)	123 (12.2)	65 (13.3)	33 (14.0)	12 (6.9)	13 (11.4)
Type of Mammogram						
SFM	1,470 (93.6)	941 (93.0)	446 (91.2)	216 (91.9)	168 (96.6)	111 (97.4)+
DM	100 (6.4)	71 (7.0)	43 (8.8)	19 (8.1)	6 (3.4)	3 (2.6)
Participation						
Successive	1,100 (70.1)	742 (73.3)	361 (73.8)	176 (74.9)	128 (73.6)	77 (67.5)
Initial	470 (29.9)	270 (26.7)	128 (26.2)	59 (25.1)	46 (26.4)	37 (32.5)
Early Recall						
No	1,423 (90.6)	985 (97.3)*	474 (96.9)	227 (96.6)	170 (97.7)	114 (100.0)+
Yes	147 (9.4)	27 (2.7)	15 (3.1)	8(3.4)	4 (2.3)	0 (0.0)
Previous false-positive						
No	1,475 (93.9)	904 (89.3)*	444 (90.8)	185 (78.7)	163 (93.7)	112 (98.2)+
Yes	95 (6.1)	108 (10.7)	45 (9.2)	50 (21.3)	11 (6.3)	2 (1.8)
Age group (yrs)						
44–49	81 (5.2)	61 (6.1)*	32 (6.5)	12 (5.1)	8 (4.6)	10 (8.8)+
50–54	421 (26.8)	351 (34.7)	170 (34.8)	73 (31.1)	54 (31.0)	54 (47.4)
55–59	438 (27.9)	271 (26.8)	136 (27.8)	57 (24.3)	55 (31.6)	23 (20.2)
60–64	460 (29.3)	224 (22.1)	103 (21.1)	62 (26.4)	39 (22.4)	20 (17.5)
65–70	170 (10.8)	104 (10.3)	48 (9.8)	31 (13.2)	18 (10.3)	7 (6.1)
HRT use						
No	1,306 (91.5)	771 (87.9)*	366 (87.6)	175 (88.8)	139 (85.8)	91 (91.0)+
Yes	122 (8.5)	106 (12.1)	52 (12.4)	22 (11.2)	23 (14.2)	9 (9.0)
Menopausal status						
Postmenopausal	1,222 (85.5)	725 (81.7)*	346 (81.6)	166 (83.0)	138 (86.8)	75 (72.1)+
Premenopausal	207 (14.5)	162 (18.3)	78 (18.4)	34 (17.0)	21 (13.2)	29 (27.9)
Family history of breast cancer						
No	1,224 (87.5)	760 (87.5)	352 (85.6)	175 (89.3)	139 (88.5)	94 (89.5)
Yes	175 (12.5)	109 (12.5)	59 (14.4)	21 (10.7)	18 (11.5)	11 (10.5)
Previous benign biopsy outside screening						
No	1,168 (88.1)	717 (88.0)	347 (89.4)	158 (83.6)	134 (88.1)	78 (90.7)
Yes	158 (11.9)	98 (12,0)	41 (10.6)	31 (16.4)	18 (11.8)	8 (9.3)
Density						
<25%	607 (38.7)	325 (32.1)*	153 (31.3)	86 (36.6)	68 (39.1)	18 (15.8)+
25–50%	428 (27.3)	169 (26.6)	134 (27.4)	64 (27.2)	49 (28.2)	22 (19.3)
50–75%	352 (22.4)	252 (24.9)	123 (25.2)	47 (20.0)	40 (23.0)	42 (36.8)
>75%	183 (11.7)	166 (16.4)	79 (16.2)	38 (16.2)	17 (9.8)	32 (28.1)

*P-value<0.05, Chi-squared test between screen-detected and interval cancer.
+P-value<0.05, Chi-squared test between screen-detected and interval cancer subtype.
Abbreviations: SDC: Screen-detected Cancer, IC: Interval Cancer, TI: True interval, FN: False-negative, MS: Minimal-signs, OT: Occult tumours, SFM: Screen-film mammography, DM: Digital mammography and HRT: Hormone Replacement Therapy.
doi:10.1371/journal.pone.0110207.t003

Therefore, we could not include this factor in the analysis of determinants. In addition, breast density was visually assessed, implying also some subjectivity and misclassification. Equally, information was unavailable on other important cancer-related variables such as age at menarche, age at maternity, the number of children and body mass index, type of mammographic abnormality and clinico-histopathological features of the breast cancers. The small numbers of occult tumours did not provide sufficient statistical power for this subgroup in some analyses.

Table 4. Adjusted odds ratio (OR) or relative risk ratios (RRR) for interval cancer and subtypes compared with screen-detected cancers.

	IC	TI	FN	MS	OT
	OR* (95%CI)	RRR** (95%CI)	RRR** (95%CI)	RRR** (95%CI)	RRR** (95%CI)
Reading method					
Single vs Double	1.55 (0.84–2.91)	1.60 (0.75–3.42)	1.67 (0.62–4.55)	2.09 (0.45–9.75)	1.03 (0.24–4.47)
Type of Mammogram					
DM vs SFM	0.61 (0.40–0.90)	0.84 (0.52–1.37)	0.56 (0.29–1.11)	0.37 (0.14–0.99)	0.21 (0.06–0.79)
Participation					
Initial vs Successive	0.70 (0.57–0.87)	0.66 (0.50–0.87)	0.78 (0.53–1.13)	0.77 (0.51–1.16)	0.65 (0.40–1.06)
Early Recall					
Yes vs No	0.23 (0.14–0.35)	0.27 (0.15–0.47)	0.25 (0.11–0.55)	0.23 (0.08–0.64)	+
Previous false-positive					
Yes vs No	2.11 (1.56–2.86)	1.79 (1.21–2.63)	4.55 (3.07–6.75)	1.15 (0.59–2.21)	0.32 (0.08–1.33)
Age group (yrs)					
44–49 vs 50–54	1.07 (0.71–1.59)	1.17 (0.71–1.91)	0.87 (0.42–1.79)	1.05 (0.45–2.43)	1.17 (0.53–2.59)
55–59 vs 50–54	0.65 (0.51–0.82)	0.65 (0.48–0.88)	0.69 (0.45–1.06)	0.80 (0.51–1.26)	0.37 (0.21–0.66)
60–64 vs 50–54	0.52 (0.40–0.67)	0.49 (0.35–0.67)	0.71 (0.46–1.10)	0.53 (0.32–0.87)	0.36 (0.19–0.66)
65–70 vs 50–54	0.65 (0.47–0.89)	0.61 (0.41–0.92)	0.92 (0.54–1.55)	0.67 (0.36–1.24)	0.31 (0.13–0.75)
HRT use					
Yes vs No	1.57 (1.18–2.10)	1.57 (1.10–2.25)	1.48 (0.89–2.44)	1.84 (1.12–3.02)	1.26 (0.60–2.64)
Menopausal status					
Pre vs Postmenopausal	0.91 (0.67–1.22)	0.91 (0.63–1.32)	1.00 (0.60–1.68)	0.77 (0.42–1.40)	0.84 (0.46–1.53)
Family history of breast cancer					
Yes vs No	0.98 (0.75–1.28)	1.16 (0.84–1.61)	0.80 (0.49–1.31)	0.91 (0.54–1.53)	0.75 (0.38–1.46)
Previous benign biopsy outside screening					
Yes vs No	0.89 (0.67–1.17)	0.77 (0.53–1.12)	1.19 (0.77–1.85)	0.94 (0.55–1.60)	0.70 (0.32–1.49)
Density					
25–50% vs <25%	1.18 (0.95–1.46)	1.26 (0.96–1.66)	1.12 (0.78–1.61)	0.99 (0.66–1.47)	1.53 (0.80–2.94)
50–75% vs <25%	1.28 (1.03–1.60)	1.33 (1.00–1.77)	0.97 (0.65–1.44)	0.93 (0.61–1.43)	3.54 (1.97–6.36)
>75% vs <25%	1.63 (1.24–2.14)	1.67 (1.18–2.36)	1.58 (1.00–2.49)	0.74 (0.41–1.34)	4.92 (2.58–9.38)

The last category in each row is the reference category.
*Logistic regression between screen-detected (ref) and interval cancer. The ORs are adjusted for the other variables in the table and the year of last mammogram.
**Multinomial regression between screen-detected (ref) and subtypes of interval cancer. The RRRs are adjusted for the other variables in the table and the year of last mammogram.
+We excluded early recall, because there were no cases of early recall in occult tumors. All the estimates are also adjusted by mammogram year.
Abbreviations: SDC: Screen-detected cancer, IC: Interval Cancer, TI: True interval, FN: False-negative, MS: Minimal-signs, OT: Occult tumours, SFM: Screen-film mammography, DM: Digital mammography and HRT: Hormone Replacement Therapy.
doi:10.1371/journal.pone.0110207.t004

A strength of study was the cohort size and the consolidation of the screening programmes in Spain. The participating radiology units had completed at least five rounds ensuring the quality indicators were stable. There are other studies with radiologically classified interval cancers, but the present is, to our knowledge, the largest one with information on breast density for the different subtypes.

In conclusion, the current work provides comprehensive data on the relationship between personal and organizational characteristics and the risk of interval cancer. This information could be useful to better classify subgroups of women at different risks of developing cancer in a moment when personalization of breast cancer screening is being proposed [40]. The strong relationship observed between false-positives results and false negatives, together with the previous knowledge on the relationship between false-positive and cancer, emphasizes the need for return for further screening in women with false-positive results, and calls for more research on this topic.

Supporting Information

Table S1 Incidence of variables related to the screening protocol and women's characteristics by screen-detected and interval cancer (overall and subtypes).
(DOC)

Table S2 Crude hazard ratios from cause-specific survival analyses for incident screen-detected cancers and for interval cancer (overall and subtypes).
(DOC)

Acknowledgments

The authors acknowledge the dedication and support of the entire Interval Cancer (INCA) Study Group. The members of the Interval Cancer (INCA) Study Group are (alphabetical order): IMIM (Hospital del Mar Medical Research Institute), Barcelona: Jordi Blanch, Xavier Castells, Mercè Comas, Laia Domingo, Francesc Macià, Juan Martínez, Ana Rodríguez-Arana, Marta Román, Anabel Romero, Maria Sala. General Directorate Public Health and Centre for Public Health Research (CSISP), FISABIO, Valencia: Carmen Alberich, María Casals, Josefa Ibáñez, Amparo Lluch, Inmaculada Martínez, Josefa Miranda, Javier Morales, Dolores Salas, Ana Torrella. Galician Breast Cancer Screening Programme, Xunta de Galicia: Raquel Almazán, Miguel Conde, Montserrat Corujo, Ana Belén Fernández, Joaquín Mosquera, Alicia Sarandeses, Manuel Vázquez, Raquel Zubizarreta. General Directorate of Health Care Programmes. Canary Islands Health Service: Teresa Barata, Isabel Diez de la Lastra, Juana María Reyes. BasqueCountry Breast Cancer Screening Programme. Osakidetza: Arantza Otegi, Garbiñe Sarriugarte. Corporació Sanitària Parc Taulí, Sabadell: Marisa Baré, Núria Torà. Hospital Santa Caterina, Girona: Joana Ferrer, Francesc Castanyer, Gemma Renart. Epidemiology Unit and Girona Cancer Registry; and University of Girona: Rafael Marcos-Gragera, Montserrat Puig-Vives. Biomedical Research Institut of Lleida (IRBLLEIDA): Carles Forné, Montserrat Martínez-Alonso, Albert Roso, Montse Rué, Ester Vilaprinyó. Universitat Rovira i Virgili, Tarragona: Misericordia Carles, Aleix Gregori, María José Pérez, Roger Pla.

Author Contributions

Conceived and designed the experiments: JB MS JI LD BF AO TB RZ JF XC MR DS. Performed the experiments: JB MS JI LD BF AO TB RZ JF XC MR DS. Analyzed the data: JB LD MS MR. Contributed reagents/materials/analysis tools: JB LD MR MS. Wrote the paper: JB LD MS MR DS. Reviewed the manuscript: JB MS JI LD BF AO TB RZ JF XC MR DS.

References

1. Independent UK Panel on Breast Cancer Screening (2012) The benefits and harms of breast cancer screening: an independent review. Lancet 380: 1778–1786.
2. Njor S, Nyström L, Moss S, Paci E, Broeders M, et al. (2012) Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. J Med Screen 19 Suppl 1: 33–41.
3. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, et al. (2006) European Commission. European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis. 4th ed. Luxembourg: Office for Official Publications of the European Communities.
4. Bulliard J-L, Sasieni P, Klabunde C, De Landtsheer J-P, Yankaskas BC, et al. (2006) Methodological issues in international comparison of interval breast cancers. Int J Cancer 119: 1158–1163.
5. Bucchi L, Ravaioli A, Foca F, Colamartini A, Falcini F, et al. (2008) Incidence of interval breast cancers after 650,000 negative mammographies in 13 Italian health districts. J Med Screen 15: 30–35.
6. Törnberg S, Kemetli L, Ascunce N, Hofvind S, Anttila A, et al. (2010) A pooled analysis of interval cancer rates in six European countries. Eur J Cancer Prev 19: 87–93.
7. Heidinger O, Batzler WU, Krieg V, Weigel S, Biesheuvel C, et al. (2012) The incidence of interval cancers in the German mammography screening program: results from the population-based cancer registry in North Rhine-Westphalia. Dtsch Arztebl Int 109: 781–787.
8. Payne JI, Caines JS, Gallant J, Foley TJ (2013) A review of interval breast cancers diagnosed among participants of the Nova Scotia Breast Screening Program. Radiology 266: 96–103.
9. Lowery JT, Byers T, Hokanson JE, Kittelson J, Lewin J, et al. (2011) Complementary approaches to assessing risk factors for interval breast cancer. Cancer Causes Control 22: 23–31.
10. Bennett RL, Sellars SJ, Moss SM (2011) Interval cancers in the NHS breast cancer screening programme in England, Wales and Northern Ireland. Br J Cancer 104: 571–577.
11. Hofvind S, Yankaskas BC, Bulliard J-L, Klabunde CN, Fracheboud J (2009) Comparing interval breast cancer rates in Norway and North Carolina: results and challenges. J Med Screen 16: 131–139.
12. Seigneurin A, Exbrayat C, Labarère J, Colonna M (2009) Comparison of interval breast cancer rates for two-versus single-view screening mammography: a population-based study. Breast 18: 284–288.
13. 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.
14. Domingo L, Sala M, Servitja S, Corominas JM, Ferrer F, et al. (2010) Phenotypic characterization and risk factors for interval breast cancers in a population-based breast cancer screening program in Barcelona, Spain. Cancer Causes Control 21: 1155–1164.
15. Domingo L, Salas D, Zubizarreta R, Baré M, Sarriugarte G, et al. (2014) Tumor phenotype and breast density in distinct categories of interval cancer: results of population-based mammography screening in Spain. Breast Cancer Res 16: R3.
16. Kirsh VA, Chiarelli AM, Edwards SA, O'Malley FP, Shumak RS, et al. (2011) Tumor characteristics associated with mammographic detection of breast cancer in the Ontario breast screening program. J Natl Cancer Inst 103: 942–950.
17. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, et al. (2000) Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst 92: 1081–1087.
18. Halapy E, Chiarelli AM, Klar N, Knight JA (2005) Accuracy of breast screening among women with and without a family history of breast and/or ovarian cancer. Breast Cancer Res Treat 90: 299–305.
19. Pollán M, Ascunce N, Ederra M, Murillo A, Erdozáin N, et al. (2013) Mammographic density and risk of breast cancer according to tumor characteristics and mode of detection: a Spanish population-based case-control study. Breast Cancer Res 15: R9.
20. Ascunce N, Salas D, Zubizarreta R, Almazán R, Ibáñez J, et al. (2010) Cancer screening in Spain. Ann Oncol 21 Suppl 3: iii43–51.
21. Navarro C, Martos C, Ardanaz E, Galceran J, Izarzugaza I, et al. (2010) Population-based cancer registries in Spain and their role in cancer control. Ann Oncol 21 Suppl 3: iii3–13.
22. Boyd NF, Byng JW, Jong RA, Fishell EK, Little LE, et al. (1995) Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. J Natl Cancer Inst 87: 670–675.
23. Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ (1998) Mammographic densities and breast cancer risk. Cancer Epidemiol Biomarkers Prev 7: 1133–1144.
24. Therneau TM, Grambsch PM (2000) Modeling Survival Data: Extending the Cox Model. Springer, New York.
25. Pencina MJ, Larson MG, D'Agostino RB (2007) Choice of time scale and its effect on significance of predictors in longitudinal studies. Stat Med 26: 1343–1359.
26. R Core Team (2014) R: A Language and Environment for Statistical Computing. Available: http://www.r-project.org/.
27. Ashbeck EL, Rosenberg RD, Stauber PM, Key CR (2007) Benign breast biopsy diagnosis and subsequent risk of breast cancer. Cancer Epidemiol Biomarkers Prev 16: 467–472.
28. Kabat GC, Jones JG, Olson N, Negassa A, Duggan C, et al. (2010) A multi-center prospective cohort study of benign breast disease and risk of subsequent breast cancer. Cancer Causes Control 21: 821–828.
29. Blanch J, Sala M, Román M, Ederra M, Salas D, et al. (2013) Cumulative risk of cancer detection in breast cancer screening by protocol strategy. Breast Cancer Res Treat 138: 869–877.
30. Otten JDM, Fracheboud J, den Heeten GJ, Otto SJ, Holland R, et al. (2013) Likelihood of early detection of breast cancer in relation to false-positive risk in life-time mammographic screening: population-based cohort study. Ann Oncol 24: 2501–2506.
31. Castells X, Román M, Romero A, Blanch J, Zubizarreta R, et al. (2013) Breast cancer detection risk in screening mammography after a false-positive result. Cancer Epidemiol 37: 85–90.
32. Hoff SR, Abrahamsen A-L, Samset JH, Vigeland E, Klepp O, et al. (2012) Breast cancer: missed interval and screening-detected cancer at full-field digital mammography and screen-film mammography— results from a retrospective review. Radiology 264: 378–386.
33. Hofvind S, Skaane P, Elmore JG, Sebuødegård S, Hoff SR, et al. (2014) Mammographic Performance in a Population-based Screening Program: Before, during, and after the Transition from Screen-Film to Full-Field Digital Mammography. Radiology: 131502.
34. De Gelder R, Fracheboud J, Heijnsdijk EAM, den Heeten G, Verbeek ALM, et al. (2011) Digital mammography screening: weighing reduced mortality against increased overdiagnosis. Prev Med (Baltim) 53: 134–140.
35. Domingo L, Romero A, Belvis F, Sánchez 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.
36. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, et al. (1989) Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 81: 1879–1886.

37. Hofvind S, Møller B, Thoresen S, Ursin G (2006) Use of hormone therapy and risk of breast cancer detected at screening and between mammographic screens. Int J Cancer 118: 3112–3117.
38. Domingo L, Blanch J, Servitja S, Corominas JM, Murta-Nascimento C, et al. (2013) Aggressiveness features and outcomes of true interval cancers: comparison between screen-detected and symptom-detected cancers. Eur J Cancer Prev 22: 21–28.
39. Guo YP, Martin LJ, Hanna W, Banerjee D, Miller N, et al. (2001) Growth factors and stromal matrix proteins associated with mammographic densities. Cancer Epidemiol Biomarkers Prev 10: 243–248.
40. Vilaprinyo E, Forné C, Carles M, Sala M, Pla R, et al. (2014) Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. PLoS One 9: e86858.

3.5 3r article

RESEARCH

Multi-state versus semi-parametric Cox and discrete-time models for assessing the effect of false-positive results of mammogram on breast cancer risk

Jordi Blanch[1,2], Ronald B Geskus[3,4,5], Maria Sala[1,6,7], Xavier Castells[1,6,7] and Montserrat Rué[8*]

*Correspondence: montse.rue@cmb.udl.cat
[8]Lleida University-Lleida Biomedical Research Institute (IRBLleida), Avda. Rovira Roure 50, Lleida, Spain
Full list of author information is available at the end of the article

Abstract

Introduction: Screening for breast cancer may lead to interventions in asymptomatic women, resulting in partially and selective data on the natural disease course. This study aims to estimate the biases induced by screening when assessing the effect of a false positive result (FP) and to compare the effect estimates obtained with three different modelling approaches.

Methods: The study data consists of a retrospective cohort of women who underwent mammography exams (INCA-CAT study). We assumed a three states model $S_0 \rightarrow S_p \rightarrow S_c$ (absence of disease, pre-clinical, clinical) where the transition $S_0 \rightarrow S_p$ is interval-censored and $S_p \rightarrow S_c$ is either observed exactly or right-censored. We used multi-state, Cox, and discrete-time models. In addition, we conducted a simulation study aimed to explore the models performance. We were interested in assessing two sources of bias: the one that arises from the models assumptions and the one that originates when the models are applied to data partially observed. Summary measures of performance were absolute and relative bias, mean square error, and coverage of confidence intervals.

Results: Simulations show that in the hypothetical scenario of complete data the three-state model performs very well for estimating the HR of FP for both transitions. The Cox model has also good properties for the $S_0 \rightarrow S_p$ transition but slightly underestimates the true HR for the $S_0 \rightarrow S_c$ transition. Discrete time models do not perform as well as the other, with coverage of the HR intervals for the $S_0 \rightarrow S_p$ transition lower than 80%. When the models were applied to the INCA study, the multi-state and Cox models provided similar estimates for the HR on the $S_0 \rightarrow S_p$ transition ($HR \simeq 1.77$). When considering only two states, with no distinction between S_p and S_c all the three models provided similar results, with HRs varying from 1.68 to 1.73.

Conclusion: Our work suggests that multi-state models are the most appropriate for estimating breast cancer progression within the screening context. The simulation study showed that the Cox or discrete-time models have provided similar results to the multi-state models.

Keywords: Multi-state models; Cox models; Discrete-time models; Simulation; Bresat cancer

Introduction

The assessment of breast cancer screening programs traditionally has been based on indicators of performance, e.g. frequency of detected cancers, false-positive results, and interval cancers, obtained cross-sectionally at specific time periods (usually a screening round). Cross-sectional measures are limited, however, by the fact that they do not capture changes over time that may be associated to the risk of developing breast cancer. In contrast, longitudinal studies of women attending breast cancer screening may assess how past and current information on risk factors affect the risk of developing breast cancer.

In research on breast cancer screening, longitudinal data methods were initially applied to estimate the cumulative rates of false-positive results [1–5]. Later on, longitudinal data analyses have been used to study 1) the frequency of screen-detected cancers over several rounds [6–8], 2) the effect of previous false-positive results on detection rate [9,10], 3) the determinants of interval cancer [6,11], and 4) the association between longitudinal breast density measurements and breast cancer risk [12]. Interval cancers are symptomatic cancers that appear between two screening exams or during a period after the last screening exam.

In a previous work, we studied the determinants of screen-detected cancer and interval cancer -symptomatic breast cancer that appears after negative tests and before the next invitation [11]. We analyzed screen-detected and interval cancers as distinct causes of failure, but, the fact breast cancer is detected in a screening exam precludes an interval cancer, and vice versa. Progression of breast cancer fits better in the multi-state models theory were individuals transition among different health states at variable rates that can be related to their characteristics.

Screening for breast cancer may lead to interventions in asymptomatic women, resulting in data on the natural disease course that are observed partially and selective. As a consequence, screening may cause several biases in the analysis of time to event data. In analysis of survival from diagnosis, the most known and studied biases are *lead-time* and *length* sampling. The lead-time is defined as the time gained by diagnosing the disease before the patient experiences symptoms. Even if early diagnosis and early treatment had no benefit, the survival of early detected cancer cases would be longer than the survival of clinical cases. Length sampling bias arises because screen-detected cancers are more likely to have slower growth than non-screen detected cancers [13]. In addition, when screen detected and symptomatic cancers are combined, screening may interfere in the assessment of risk factors of breast cancer due to 1) earlier time of detection for screen detected tumours; 2) the fact that symptomatically detected tumours have characteristics that are different from screen detected cancers [11]; and 3) there is overdiagnosis of low growth tumours that never would become symptomatic during a woman lifetime [14]. In summary, early detection of breast cancer introduces changes in the natural history of the disease that need to be considered when assessing the effect of interventions or the association of risk factors.

Simulation models can be used to examine how screening interferes with the assessment of risk factors. Recently, Taghipour *et al.* used a partially observable Markov model to estimate relevant parameters of breast cancer progression and early detection in the Canadian National Breast Screening Study [15]. Previously, other authors had modeled the natural history of breast cancer, using

analytic or simulation models that incorporated input data from the literature and made different assumptions [13, 16–25]. The work of the Cancer Intervention and Surveillance Modeling network (CISNET) breast group [19–24] reflects how modeling studies can provide evidence to complement the results of randomized controlled trials and observational studies.

The objectives of this study are 1) To estimate the biases induced by the screening process when assessing the effect of a FP result in a mammogram test on the progression to pre-clinical cancer and clinical cancer; 2) To evaluate the effect estimates obtained via parametric multi-state models that take interval-censoring into account and semi-parametric Cox models or discrete-time models that don't.The paper is organized as follows. Section 2 presents the motivating study, the INCA-CAT, which contains data on sequential screening mammograms in Catalan women. Section 3 is a methods section that defines concepts relevant to screening, describes the methods used to estimate the effect of false-positive results in previous mammographic exams on the hazard of being diagnosed of breast cancer, and also details how a simulation study was performed. Section 4 presents the results of the simulation study and their application to the INCA-CAT study. Section 5 contains the discussion and conclusions.

Motivating data: the INterval CAncer (INCA) study

The aim of the INCA study was to assess the determinants of interval breast cancer in women attending early detection programs in Spain [11,26]. The INCA study also compared the characteristics associated with screen-detected and interval breast cancers. We selected a subset of the INCA dataset, corresponding to four radiology units of the Catalonia region, the INCA-CAT dataset.

Population-based breast cancer screening in Catalonia is offered biennially to all women aged 50-69 years. Screening mammography has three possible outcomes: 1) negative result (normal), 2) positive result (abnormal findings requiring further assessments), and 3) early recall (an intermediate mammogram is performed out of sequence with the screening interval, at 6 or 12 months). Cancers detected at an intermediate mammogram were considered screen-detected cancers [27].

A positive result is considered to be a screen-detected tumour if, after further assessments, there is histopathological confirmation of cancer. Otherwise, the result is considered false-positive and the woman is invited again after two years. Further assessments can include non-invasive procedures (magnetic resonance imaging, ultrasonography, additional mammography) and/or invasive (fine-needle aspiration cytology, core-needle biopsy and open biopsy).

Interval cancer was defined as "a primary breast cancer arising after a negative screening episode and before the next invitation to screening or within 24 months for women who reached the upper age limit" [27]. This definition was extended until the 30th month, to allow a 6 months margin for women to attend each round. Interval cancers were identified by merging the screening programmes databases with hospital discharge databases and population-based or hospital cancer registries.

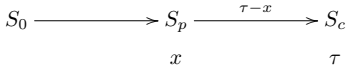
The INCA-CAT dataset consists of a retrospective cohort of 96,636 women who underwent mammography exams between January 1, 2000 and December 31, 2006, and were followed-up until June 30, 2009 for interval cancer assessment. These women underwent a total of 230,742 screening

mammograms. During the study period, 963 cancers were diagnosed, of which 671 were detected in screening exams, and 313 emerged as interval cancers. Both invasive and ductal *in situ* (non-invasive) breast cancer carcinomas are included.

Among several risk factors studied, the existence of a FP result in the previous mammogram showed the highest hazard ratio for developing interval cancer, HR=2.71 [11]. The corresponding hazard ratio for screen-detected cancer was 1.34. The authors also reported that the effect of a previous FP result was higher for interval cancers classified as false negative tumours, HR=8.79, than for true interval cancers, HR=2.26.

Methods

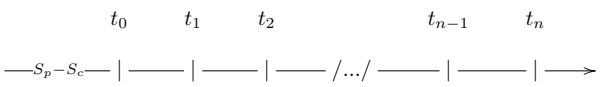
We assume a three-state model



where S_0 indicates absence of breast cancer, S_p indicates the pre-clinical state, where breast cancer is asymptomatic but can be detected with a diagnostic test i.e. mammography, and S_c indicates the clinical state or presence of symptoms [18, 28]. The ages at entering S_p and S_c are x and τ, respectively; $\tau - x$ is the sojourn time in S_p.

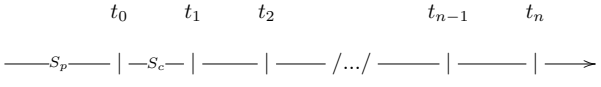
The following examples illustrate some possibilities that result from the interaction of the natural history of the disease and the screening process. The times $t_i, i = 0, ..., n$ indicate when the mammography exams are scheduled. We assume that women attend the exams and that mammogram sensitivity is not 100%.

1 Before starting the screening exams the woman enters S_p and then S_c.



This woman is not included in the study because she is diagnosed of breast cancer before the first screening exam at time t_0, with the exception of being misclassified because of sensitivity of mammography lower than 100%.

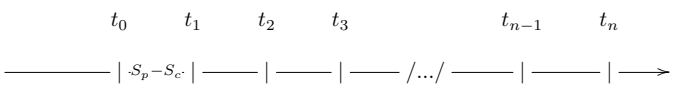
2 The woman enters S_p before the first screening exam and enters S_c between t_0 and t_1.



There are two possibilities: a) an early diagnosis at time t_0 where age at entering S_c will not be observed; or b) the mammography exam misses the tumour and it is symptomatically

diagnosed at time t between t_0 and t_1. If a) the tumour is screen-detected, if b) it is the type of interval cancer called *false negative*.

3 The next example corresponds to the type of interval cancer called *true interval*.



Time scale, censoring, and late entry

For the time to S_p, two time scales can be chosen: time since entry into the screening program, or age. Age is a more relevant time scale, since it directly reflects the biological process. The incidence of pre-clinical cancer since entry additionally also depends on the study design, i.e. at what age women enter the study. For time to S_c, the same time scale can be chosen, but it may be of more interest to study the sojourn time, i.e. time from S_p to S_c. In the multi-state theory, this is called the *clock-reset* approach.

In the data, the transition $S_0 \to S_p$ is interval-censored. We can assume that the transition $S_p \to S_c$ is either observed exactly or right-censored. When the sojourn time is the estimand, such data have been called doubly (interval) censored. The situation is similar to the estimation of time from HIV infection to AIDS. Typically, HIV infection is not observed exactly but only known to lie in the interval between a last HIV negative test and a first HIV positive test. AIDS is usually either observed exactly or right-censored.

In the INCA study not all women entered the screening program at the age of 50. Women that were above 50 when the program started entered only when they had not yet developed symptoms: they had to be free of clinical cancer. Phrased otherwise, women that already developed pre-clinical or clinical cancer before they would enter the screening program are missed. This may give rise to left truncated data.

Women that have a pre-clinical cancer detected are treated. As a consequence, the clinical cancer will not occur. This implies that transitions $S_p \to S_c$ are never observed in these women. This is a form of informative censoring for the estimation of the sojourn time distribution. Women with a short sojourn time are more likely to have an interval cancer, i.e. to have the clinical cancer observed instead of the screen detected cancer. In fact, if mammography was 100% sensitive, women that have a sojourn time of more than two years will always be censored at the first screening visit at which the pre-clinical cancer is observed.

Estimation of the distribution of age at pre-clinical cancer may still be possible. However, if we use the time of interval cancer as right hand side of the detection interval, the observation times are no longer independent of the event times. An alternative to explore is to set the right hand side at the time at which the first subsequent screening exam would be performed, even though this was not actually done.

A Markov multi-state model for breast cancer screening

A Markov multi-state model describes how individuals move between a series of states in continuous time [29, 30]. In a Markov model, the next state and the time at which the transition occurs only depends on the present state [31]. A change of state is called a transition, or an event. Transition intensities that may depend on time t, or also on a set of covariates $z(t)$, represent the instantaneous risk or hazard of moving from state r to state $s \neq r$

$$q_{rs}(t; z(t)) = \lim_{\delta t \to 0} \frac{P(S(t + \delta t) = s | S(t) = r, z(t))}{\delta t}.$$

where $S(t)$ indicates the state of an individual at time t.

Statistical analysis

We used three approaches to analyse the INCA-CAT data, 1) the multi-state model, 2) the Cox model, which is the more frequently used in the literature to assess risk factors of breast cancer, and 3) the discrete-time model which also has been used in similar studies [7, 8].

When using the multi-state model, we started with a 3-state model and assumed that a) the transition intensity $S_0 \to S_p$ increases with age, as it does incidence of breast cancer; b) the time at entering S_p is interval-censored between the time of a negative exam and the time of breast cancer diagnosis; c) the time at entering S_c is exact for women with IC and is right-censored for women with SD cancer; d) there are censored states due to the sensitivity of the mammogram being lower than 100%; e) false-positive result in the screening mammogram is a time-dependent covariate that is measured as 0 in the absence of a FP result and 1 from the time of the first FP result. With this model we estimated the effect of a FP result on the time to entering S_p. Then, we considered a 2-state model where the event was BC diagnosis (SC or IC) and compared the effect of FP on BC diagnosis if the IC was detected between screens or in the next screen. The `msm` package for `R` [29] was used for the multi-state analyses.

When using the Cox or the discrete-time models, we also performed different analyses. First, we applied a cause-specific strategy for SD and IC, separately, as in competing risks analysis. For each cause-specific model, the competing event was censored. Second, we considered that the event was breast cancer diagnosis, either SD or IC. We estimated the effect of a FP result in two scenarios, considering the exact age at the IC diagnosis and assuming that the IC cancer was diagnosed in the next screening exam.

In addition, we conducted a simulation study to explore the performance of the different models for measuring the effect of a FP result on risk of breast cancer. We were interested in assessing two sources of bias, first the one that arises from the assumptions of the statistical models and then the bias that originates when the statistical models are applied to data only partially observed due to the screening process. We worked with two types of simulated datasets, 1) the *complete data* type which contains all the information on the natural history of breast cancer, and 2) the *observed data* type which contains the partially observed data as a consequence of the screening.

Simulation study

We simulated the natural history of breast cancer as an progressive illness multi-state model with the three states and transitions $S_0 \rightarrow S_p \rightarrow S_c$. The following assumptions were made:

1 **Simulation procedures:**
For each specific combination of assumptions, the same database of simulated datasets was used to compare the statistical methods of interest. Non-convergence failures were monitored and used to assess if inadequate assumptions were made. The `R` random sampling functions for specific distributions were used to generate the random variables. To enable replication of the datasets, a seed was specified. We simulated 500 datasets with 62,000 women in each dataset.

2 **Methods for generating the datasets:**
   (a) **Time scale**: Age is the time scale with age 50 years as the origin.
   (b) **Time to the pre-clinical state,** T_p: We assume that it follows a Weibull distribution, We(λ, ν) with λ and ν the scale and shape parameters, respectively, where $h_0(t) = \lambda\nu t^{\nu-1}$ is the baseline risk function of a conventional relative risk model. Times to S_p were generated using the method proposed by Austin [32,33] for Cox models.
   (c) **Time to study entry,** T_e: We will make two assumptions, without and with late entry. For the late entry scenario we assume that the time at entering follows a uniform distribution, $U(0, 15)$, and all the data previous to T_e will not be used.
   (d) **Sojourn time in the pre-clinical state,** T_s: We assume that it follows an exponential distribution Exp(0.25), which corresponds to a mean time in S_p equal to 4 years [28].
   (e) **Time to the clinical state,** T_c: $T_c = T_p + T_s$.
   (f) **Mammogram sensitivity**: We assume two different values, 100% and 85%. We assume that each screening exam is independent within a woman over time.
   (g) **Incidence of FP results**: We assumed binomial distributions with varying probabilities conditional to the exam sequence number, according to the Román study [34].
   (h) **Effect of FP results on incidence of breast cancer**:
   The presence of a FP result will modify the hazard of entering the pre-clinical state according to a relative risks model. Times to S_p were generated using the closed-form expression proposed by Austin [32] for Cox models with time varying covariates.
   (i) **Dropout**: dropout were associated with FP results. That assumption was based on evidence that a FP result produces a decrease in the adherence to screening, and this decrease is more pronounced if the FP result occurs in the early exams. The RAFP study [4], which included women of similar characteristics as in the INCA study and used a multilevel discrete hazard model, was used to simulate a dropout associated to FP results.
   (j) **Number of screening exams**: We assumed ten screening exams as in the majority of the European mass screening programs. We followed-up each woman until the first occurrence of breast cancer diagnosis, administrative censoring at 20 years of follow-up, and dropout during the study.

3 **Scenarios to be investigated**:
   (a) With / without late entry.
   (b) We fixed the parameters for the Weibull distribution using the INCA-Cat dataset, $scale = 0.00025$ and $shape = 1.5712$.
   (c) Hazard ratio of FP result for time to entering S_p equal to 2.

4 **Statistical methods to be evaluated**:
   (a) We used the `msm` package in `R` developed by Jackson [29] to fit the multi-state models. The `msm` package a) assumes time-homogeneous or piecewise constant hazards. This is a limitation given that we have assumed a Weibull distribution for the time to S_p, and an exponential distribution for the sojourn time in S_p; b) allows for interval-censored transitions and this can be considered a strength of the `msm` package; c) allows for misclassification of states, which is also an asset, given that mammography has sensitivity and specificity lower than 100%.
   (b) We estimated the effect of a FP result with the Cox model. We applied a cause-specific strategy for SD and IC, separately. The counting process structure allows to account for left-truncation and time-dependent covariates, as well as non proportional-hazards. We used the `coxph` function of the `survival` package `R`.
   (c) We estimated the discrete-time model using a logit link for the hazard of the event. The model contains a time indicator variable given by the prior number of screening rounds attended which acts as multiple intercepts, one per screening round [4]. The time indicator variable represents the baseline logit hazard function [35]. Age was included as a continuous covariate. As in the Cox model, we applied a cause-specific strategy for SD and IC, separately. We considered that the event time corresponds to the last observed screening round. We used the `glm` function of the `stats` package in `R`.

5 **Summary measures of performance**: absolute bias, relative bias (as percentage of the true value), mean square error, and coverage of confidence intervals.

In each simulation, we focus on quantifying the potential bias due to late-entry and the censoring after detection of pre-clinical cancer. To quantify the bias due to late-entry, we performed two simulations for each scenario, with and without late-entry. The second potential bias is due to the fact that the sojourn time is informatively censored. A woman with a screen detected cancer receives treatment and therefore, it is not possible to observe the time to S_c. We assess the effect of not observing the time to S_c, when evaluating the effect of a FP result on entering the pre-clinical state S_p.

As a secondary analysis we reduced the three-state model to a two-state model with 1: S_0 (absence of breast cancer) and 2: S_p or S_c (breast cancer). Here the time T of interest is age at breast cancer diagnosis (either screen-detected or symptomatically detected). T can be right-censored since each women is followed until breast cancer, or end of study. Lost to follow-up is simulated as non-related or related to the FP status. For clinically diagnosed tumours, we also assessed the effect of extending the follow-up to the next screening exam for estimating the effect of a FP result on the incidence of breast cancer, indistinctly screen-detected or symptomatic.

Results

Simulation study

In this section we present the results of the simulation study and try to relate the models performance to their assumptions and to the specificities of the data. For the complete data analyses we have assumed that there are three states. We assume that women receive 10 biennial mammographic exams at the age interval 50-69 years, age at entering screening may or may not have late entry, and the HR of a false-positive result is 2.

Transition intensities for the three states multi-state model. Complete and observed data

Figures 1 and 2 present the simulated (theoretical) and estimated transition intensities in the age interval 50-69 years for the **complete** and the **observed** datasets, respectively. For complete data, the estimated $S_0 \rightarrow S_p$ piecewise constant transition rates overlap with the theoretical Weibull function and the estimated $S_p \rightarrow S_c$ transition rate is unbiased (Figure 1). Late entry does not change these results. With observed data, the estimated $S_0 \rightarrow S_p$ piecewise constant rates follow the Weibull pattern, similarly to the complete data scenario, but the estimated transition rate overestimates considerably the theoretical rate (Figure 2).

Hazard ratio of a FP result, for the three state models. Complete data

Before interpreting the results it is important to mention that the HR of FP for the transition $S_p \rightarrow S_c$ only can be estimated when using multi-state models. We have simulated the data assuming that a FP result is associated with the transition $S_0 \rightarrow S_p$ with a $HR = 2$ and it is not associated with the transition $S_p \rightarrow S_c$ (HR=1). For the multi-state model these values are our theoretical values. For the Cox and discrete time models only the transitions $S_0 \rightarrow S_p$ and $S_0 \rightarrow S_c$ can be estimated. We have assessed the performance of these models assuming that the theoretical HR value for both transitions is 2.

Table 1 shows the performance of the studied models, when considering **three states** and **complete data**, with respect to the estimation of the theoretical HR of having a FP result in the screening mammogram for the transitions $S_0 \rightarrow S_p$ and $S_p \rightarrow S_c$. The MS3 model performs very well for both transitions, either *with or without late entry*, with very low bias and coverage higher than 95%.

The Cox CS model has also good properties when estimating the HR of a FP result for the $S_0 \rightarrow S_p$ transition. For the $S_0 \rightarrow S_c$ the $\widehat{\overline{\beta}}$ value slightly underestimates the true β value with a bias around 2.5%, independently of the presence/absence of *late entry*. Coverage is slightly lower than 95% for the $S_0 \rightarrow S_c$ transition with *late entry*.

The discrete time model with three states (DT3) does not perform as well as the MS3 and the Cox CS models. In this case the percentage bias of the $\widehat{\overline{\beta}}$ associated to the $S_0 \rightarrow S_c$ transition approaches 5% and the coverage of the HR intervals for the $S_0 \rightarrow S_p$ transition is lower than 80%, for both scenarios *with/without late entry*.

Hazard ratio of a FP result, for the three state models. Observed data
Table 2 shows the performance of the studied models for the HR of a FP, when considering **three states** and **observed data**. *With or without late entry*, both the MS3 and the Cox CS models perform well in terms of bias, MSE, and coverage. Instead, the DT3 model shows considerable bias (around 10% overestimation) and high MSE when estimating the HR of a FP result for the $S_0 \to S_c$ transition. The coverage of the intervals for this HR is lower than 90% *without late entry* and near 95% *with late entry* which probably is due to wide confidence intervals of the estimated HRs as the high MSE indicates.

Hazard ratio of a FP result, for the two state models. Observed data
Table 3 presents the performance of the studied models for the HR of a FP result, when considering **two states** in the **observed data**. Here it is assumed that the time to the event of interest is the time when the tumour is detected, either by screening or clinically. We also have assessed the scenario that, for the clinically detected tumours, assumes that the time to event is the time at the next screening exam.

We observe a good performance of the three studied methods, with a better performance of the multi-state model with *no late entry* followed by the discrete time (DT2 next screen) and the Cox model (either exact time or next screen). The DT2 exact time model slightly overestimates the effect and *with late entry* has lower coverage that the other methods.

Aplication to the INCA-CAT study

Tables 4 and 5 present the estimates of the HR of FP result when applying the models to the INCA-CAT data.

When the studied models were applied to the INCA study and the three states S_0, S_p and S_c were considered, the multi-state and the Cox models provided similar estimates for the HR of FP on the $S_0 \to S_p$ transition ($HR \simeq 1.77$). However, the discrete-time model provided a higher estimate of the HR, 1.93. For the $S_p \to S_c$ transition, we obtained a HR near to 1 as in the simulation study. The $S_0 \to S_c$ transitions estimated with the Cox and the discrete time models provide different values, HR around 1.92 and 1.22, respectively.

When considering only two states, with no distinction between pre-clinical or clinically detected cancer, all the three models provide similar results, with the HR of a FP result varying from 1.68 to 1.73, when the time of clinically detected cancer is considered an exact time. When the time of clinically detected cancer is extended up to the next screening exams, the three estimated values vary from 1.64 for the multi-state model to 1.82 for the discrete-time model.

Discussion

This study assessed the biases induced by early detection on the estimates of the effect of a mammographic FP result on the progression to pre-clinical and clinical BC. In addition, we evaluated the effect estimates obtained via parametric multi-state models, which are seldom used in the screening

literature [36, 37], and semi-parametric Cox models or discrete-time models, which are common in this area [6–11]. A simulation study, has provided information on how the assumptions of the statistical models, or the fact that these models are applied to data partially observed, affect the parameter estimates.

The simulation study showed that, in general, all the models considered had an acceptable performance for estimating the effect of a FP result, in the different scenarios assessed. When we worked with three health states, S_0, S_p and S_c, the multi-state model showed the best performance in all the scenarios, either for complete or observed data. The three-state model was followed by the Cox cause-specific and the discrete-time models with good and fair performances, respectively. The simulation analysis of the two-state models showed good performance of the three modelling approaches considered. The multi-state model with no late entry showed the best performance followed by the discrete-time model with time at the next screen for the clinically diagnosed tumours, and then the Cox model in all the assessed scenarios. Therefore, although with our simulated data all the studied models perform satisfactorily, the multi-state model worked better than the other two.

When the studied methods were applied to the INCA study and three states were considered, the multi-state and the Cox cause-specific models provided similar results for the effect of a FP on the $S_0 \rightarrow S_p$ transition, whereas the discrete time model differed from them. When both pre-clinical and clinical cancers were grouped together, again the multi-state and the Cox model provided similar results. In this scenario of only two states (no cancer/cancer) the discrete time model did not differed much from the multi-state and the Cox models. It is important to mention that the low incidence of breast cancer in women invited to screening may have made it difficult to observe more marked differences between the methods.

There are three main differences between the three models compared, 1) the time metric, 2) the type of censoring for the time to S_p, and 3) the assumption of progressive disease versus competing risks. First, according to Singer and Willet no single time metric is universally appropriate, and even different scales might be used to study the same event [35]. Whereas in the multi-state and Cox models the time scale was age as a continuous variable [6, 11, 36, 37]; in the discrete-time model we used the prior number of screening rounds attended, as is generally done in the literature [7,8,10]. Then, whereas the multi-state and Cox models take into account the elapsed time between participations, the discrete-time model assumes that times between screening exams are equal [35]. This assumption was satisfied in the simulation study; but not in INCA study, were women could have dropped a screening exam and return to screening afterwards. Probably the overestimated effect in the INCA study for the discrete-time model may be related to a violation of this assumption in a subgroup of participants.

The second difference is the type of censoring for the times of interest. All the studied models assume that time to S_c is exact for clinically detected cancers and right-censored for screen detected cancers, S_p. Multi-state models assume that time to S_p is interval-censored between two screening exams and can be estimated for both, screen or clinically, detected cancers. The Cox model also

can handle interval-censoring, so time to S_p could be estimated, but this is not generally done in the early detection literature. Instead, the time to screen detection is used. Given that this time depends on the periodicity of the screening exams, it seems more reasonable to estimate the time to S_p. Other authors also consider that the interval-censoring approach is better, because it can take account for the probability of entering the pre-clinical stage, S_p, between visits or between the last visit and the time to S_c [38]. And, our results show that the multi-state model provides a less biased estimate that the other models, whether there is left-truncation or not.

Third, we have compared a progressive illness multi-state model that models the transition of women among the different states of breast cancer with competing risks models that, in general, assume unrelated competing events [31]. For instance, in the competing risks model, the event *screen detected* cancer is considered a competing event to *symptomatically diagnosed* [31,39]. Even though for some of the simulated scenarios the estimated effect of a FP result is similar in both types of models, the progressive illness model reflects better the natural history of the disease and therefore, conceptually is more adequate.

As in most simulation studies, our assumptions could not cover all the possible scenarios. In particular, 1) we focused on studying the effect of one time-varying variable, a FP result, to the specific transitions among health states. 2) We assumed that the transition intensities were proportional with respect to the covariate values and that lost-to follow-up was missing at random. 3) We assumed that when a woman does not attend a screening test, she never returns to participate, although, in real life women may have intermittent participation in screening. 4) Based on the organization of the public screening program, we assumed that the exam times were independent of the cancer progression process. 5) We assumed that the presence of a FP result does not affect the sojourn time in S_p in women that enter this state. 6) In the INCA study, it is possible that some breast cancers diagnosed in women that participate in population screening are detectable but missed at screening (false negative) and therefore are misclassified as interval cancers. The multi-state models can use a hidden Markov process to account for misclassification. In our case, the assumption of hidden Markov process did not modify the model estimates.

Conclusions

Our work suggests that the multi-state models are the most appropriate for estimating breast cancer progression within the screening context. The simulation study has shown that, in the context of breast cancer screening, other types of models widely used in the literature, like the Cox or discrete-time models, have provided similar results to the multi-state models.

Competing interests
The authors declare that they have no competing interests.

Author's contributions
MR, RG and JB conceived the study. JB carried out the implementation with the supervision of MR. JB, RG and MR drafted the first version of the manuscript. All authors contributed to the writing and approved the final version.

Acknowledgements

This paper was supported by the research grant PRX16/00028 from the Spanish Ministry of Education, Culture and Sports, and partially supported by the research grant PI14/00113 from the Spanish Ministry of Economy and Competitiveness and by GRAES-2014-SGR978 from the Generalitat de Catalunya. We thank the researchers of the RAFP and INCA projects for having provided their data.

Author details

[1]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. [2]Vascular Health Research Group (ISV-Girona). Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain. [3]Amsterdam Public Health Research Institute, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands. [4]Nuffield Department of Medicine, University of Oxford, Oxford, UK. [5]Oxford University Clinical Research Unit, Wellcome Trust Major Overseas Programme, Ho Chi Minh City, Vietnam. [6]Health Services Research on Chronic Patients Network (REDISSEC), Madrid, Spain. [7]Epidemiology and Evaluation Department, Hospital del Mar, Barcelona, Spain. [8]Lleida University-Lleida Biomedical Research Institute (IRBLleida), Avda. Rovira Roure 50, Lleida, Spain.

References

1. Elmore, J.G., Barton, M.B., Moceri, V.M., Polk, S., Arena, P.J., Fletcher, S.W.: Ten-year risk of false positive screening mammograms and clinical breast examinations. The New England Journal of Medicine **338**(16), 1089–1096 (1998)
2. 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. Journal of Epidemiology and Community Health **60**(4), 316–321 (2006). doi:10.1136/jech.2005.042119
3. Hubbard, R.A., Kerlikowske, K., Flowers, C.I., Yankaskas, B.C., Zhu, W., Miglioretti, D.L.: Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography; a cohort study. Annals of Internal Medicine **155**(8), 481–492 (2011). doi:10.7326/0003-4819-155-8-201110180-00004
4. Román, R., Sala, M., De La Vega, M., Natal, C., Galceran, J., González-Román, I., Baroja, a., Zubizarreta, R., Ascunce, N., Salas, D., Castells, X.: Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening. Breast Cancer Research and Treatment **130**, 543–552 (2011). doi:10.1007/s10549-011-1581-4
5. Hofvind, S., Ponti, A., Patnick, J., Ascunce, N., Njor, S., Broeders, M., Giordano, L., Frigerio, A., Törnberg, S., Van Hal, G., Martens, P., Májek, O., Danes, J., von Euler-Chelpin, M., Aasmaa, A., Anttila, A., Becker, N., Péntek, Z., Budai, A., Mádai, S., Fitzpatrick, P., Mooney, T., Zappa, M., Ventura, L., Scharpantgen, A., Hofvind, S., Seroczynski, P., Morais, A., Rodrigues, V., Bento, M.J., Gomes de Carvalho, J., Natal, C., Prieto, M., Sánchez-Contador Escudero, C., Zubizarreta Alberti, R., Fernández Llanes, S.B., Ascunce, N., Ederra Sanza, M., Sarriugarte Irigoien, G., Salas Trejo, D., Ibáñez Cabanell, J., Wiege, M., Ohlsson, G., Törnberg, S., Korzeniewska, M., de Wolf, C., Fracheboud, J., Patnick, J., Lancucki, L., Ducarroz, S., Suonio, E.: False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. Journal of Medical Screening **19 Suppl 1**(SUPPL. 1), 57–66 (2012). doi:10.1258/jms.2012.012083
6. Hofvind, S., Bjurstam, N., Sørum, R., Bjørndal, H., Thoresen, S., Skaane, P.: Number and characteristics of breast cancer cases diagnosed in four periods in the screening interval of a biennial population-based screening programme. Journal of Medical Screening **13**(4), 192–6 (2006)
7. Blanch, J., Sala, M., Roman, M., Ederra, M., Salas, D., Zubizarreta, R., Sanchez, M., Rue, M., Castells, X., Group, C.: Cumulative risk of cancer detection in breast cancer screening by protocol strategy. Breast Cancer Research and Treatment **138**(3), 869–877 (2013). doi:10.1007/s10549-013-2458-5; 10.1007/s10549-013-2458-5
8. Ripping, T.M., Hubbard, R.A., Otten, J.D.M., Den Heeten, G.J., Verbeek, A.L.M., Broeders, M.J.M.: Towards personalized screening: Cumulative risk of breast cancer screening outcomes in women with and without a first-degree relative with a history of breast cancer. International Journal of Cancer **138**(7), 1619–1625 (2016). doi:10.1002/ijc.29912
9. von Euler-Chelpin, M., Risør, L.M., Thorsted, B.L., Vejborg, I.: Risk of breast cancer after false-positive test results in screening mammography. Journal of the National Cancer Institute **104**(9), 682–9 (2012). doi:10.1093/jnci/djs176
10. Castells, X., Roman, M., Romero, A., Blanch, J., Zubizarreta, R., Ascunce, N., Salas, D., Buron, A., Sala, M.: Breast cancer detection risk in screening mammography after a false-positive result. Cancer Epidemiology **37**(1), 85–90 (2013). doi:10.1016/j.canep.2012.10.004
11. Blanch, J., Sala, M., Ibáñez, J., Domingo, L., Fernandez, B., Otegi, A., Barata, T., Zubizarreta, R., Ferrer, J., Castells, X., Rué, M., Salas, D.: Impact of risk factors on different interval cancer subtypes in a population-based breast cancer screening programme. PloS One **9**(10), 110207 (2014). doi:10.1371/journal.pone.0110207
12. Armero, C., Forné, C., Rué, M., Forte, A., Perpiñán, H., Gómez, G., Baré, M.: Bayesian joint ordinal and survival modeling for breast cancer risk assessment. Statistics in Medicine **35**(28), 5267–5282 (2016). doi:10.1002/sim.7065
13. Zelen, M., Feinleib, M.: On the theory of screening for chronic diseases. Biometrika **56**(3), 601–614 (1969)
14. Esserman, L.J., Thompson, I.M., Reid, B.: Overdiagnosis and overtreatment in cancer: an opportunity for improvement. JAMA : the journal of the American Medical Association **310**(8), 797–8 (2013). doi:10.1001/jama.2013.108415
15. Taghipour, S., Banjevic, D., Miller, A.B., Montgomery, N., Jardine, A.K.S., Harvey, B.J.: Parameter estimates for invasive breast cancer progression in the Canadian National Breast Screening Study. British journal of cancer **108**(3), 542–8 (2013). doi:10.1038/bjc.2012.596
16. Chen, H.H., Duffy, S.W., Tabar, L.: A Markov Chain Method to Estimate the Tumour Progression Rate from Preclinical to Clinical Phase, Sensitivity and Positive Predictive Value for Mammography in Breast Cancer Screening. Journal of the Royal Statistical

Society **45**(3), 307–317 (1996)
17. Shen, Y., Zelen, M.: Screening sensitivity and sojourn time from breast cancer early detection clinical trials: mammograms and physical examinations. Journal of clinical oncology **19**(15), 3490–3499 (2001)
18. Lee, S.J., Zelen, M.: Modelling the early detection of breast cancer. Annals of Oncology **14**(8), 1199–1202 (2003). doi:10.1093/annonc/mdg323
19. Berry, D.a., Inoue, L., Shen, Y., Venier, J., Cohen, D., Bondy, M., Theriault, R., Munsell, M.F.: Modeling the impact of treatment and screening on U.S. breast cancer mortality: a Bayesian approach. Journal of the National Cancer Institute. Monographs **4009**(36), 30–6 (2006). doi:10.1093/jncimonographs/lgj006
20. Fryback, D.G., Stout, N.K., Rosenberg, M.a., Trentham-Dietz, A., Kuruchittham, V., Remington, P.L.: The Wisconsin Breast Cancer Epidemiology Simulation Model. Journal of the National Cancer Institute. Monographs **53726**(36), 37–47 (2006). doi:10.1093/jncimonographs/lgj007
21. Hanin, L.G., Miller, A., Zorin, A.V., Yakovlev, A.Y.: Chapter 10: The University of Rochester Model of Breast Cancer Detection and Survival. JNCI Monographs **2006**(36), 66–78 (2006). doi:10.1093/jncimonographs/lgj010
22. Lee, S., Zelen, M.: Chapter 11: A Stochastic Model for Predicting the Mortality of Breast Cancer. JNCI Monographs **2006**(36), 79–86 (2006). doi:10.1093/jncimonographs/lgj011
23. Mandelblatt, J., Schechter, C.B., Lawrence, W., Yi, B., Cullen, J.: The SPECTRUM population model of the impact of screening and treatment on U.S. breast cancer trends from 1975 to 2000: principles and practice of the model methods. Journal of the National Cancer Institute. Monographs **20007**(36), 47–55 (2006). doi:10.1093/jncimonographs/lgj008
24. Plevritis, S.K., Sigal, B.M., Salzman, P., Rosenberg, J., Glynn, P.: Chapter 12: A Stochastic Simulation Model of U.S. Breast Cancer Mortality Trends From 1975 to 2000. JNCI Monographs **2006**(36), 86–95 (2006). doi:10.1093/jncimonographs/lgj012
25. Weedon-Fekjaer, H., Lindqvist, B.H., Vatten, L.J., Aalen, O.O., Tretli, S.: Breast cancer tumor growth estimated through mammography screening data. Breast cancer research : BCR **10**(3), 41 (2008). doi:10.1186/bcr2092
26. Domingo, L., Salas, D., Zubizarreta, R., Baré, M., Sarriugarte, G., Barata, T., Ibáñez, J., Blanch, J., Puig-Vives, M., Fernández, A., Castells, X., Sala, M.: Tumor phenotype and breast density in distinct categories of interval cancer: results of population-based mammography screening in Spain. Breast Cancer Research **16**(1), 3 (2014). doi:10.1186/bcr3595
27. 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. Fourth edition–summary document. Annals of Oncology **19**(4), 614–22 (2008). doi:10.1093/annonc/mdm481
28. Lee, S.J., Zelen, M.: Scheduling periodic examinations for the early detection of disease: Applications to breast cancer. Journal of the American Statistical Association **93**(444), 1271–1281 (1998)
29. Jackson, C.H.: Multi-State Models for Panel Data: The msm Package for R. Journal of Statistical Software **38**(8), 1–28 (2011). doi:10.18637/jss.v038.i08
30. Geskus, R.B.: Data Analysis with Competing Risks and Intermediate States. Chapman and Hall/CRC Biostatistics Series, Boca Raton, FL (2016)
31. Putter, H., Fiocco, M., Geskus, R.B.: Tutorial in biostatistics: competing risks and multi-state models. Statistics in Medicine **26**(11), 2389–2430 (2007). doi:10.1002/sim.2712
32. Austin, P.C.: Generating survival times to simulate Cox proportional hazards models with time-varying covariates. Statistics in Medicine **31**(29), 3946–3958 (2012). doi:10.1002/sim.5452
33. Bender, R., Augustin, T., Blettner, M.: Generating survival times to simulate Cox proportional hazards modelse. Statistics in Medicine **24**(11), 1713–1723 (2005). doi:10.1002/sim.2059
34. Román, R., Sala, M., Salas, D., Ascunce, N., Zubizarreta, R., Castells, X., 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, a.B., Galceran, J., González-Román, I., Ibáñez, J., Macià, F., Natal, C., Queiro, M.T., Román, R., Sala, M., Salas, D., Velarde, J.M., Zubizarreta, R.: Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. Annals of Oncology **23**, 104–111 (2012). doi:10.1093/annonc/mdr032
35. Singer, J.D., Willett, J.B.: Applied Longitudinal Data Analysis. Modeling Change and Event Occurrence. Oxford University Press, Inc, New York, NY (2003)
36. Uhry, Z., Hédelin, G., Colonna, M., Asselain, B., Arveux, P., Rogel, A., Exbrayat, C., Guldenfels, C., Courtial, I., Soler-Michel, P., Molinié, F., Eilstein, D., Duffy, S.: Multi-state Markov models in cancer screening evaluation: a brief review and case study. Statistical Methods in Medical Research **19**(5), 463–486 (2010). doi:10.1177/0962280209359848
37. Putter, H., van der Hage, J., de Bock, G.H., Elgalta, R., van de Velde, C.J.H.: Estimation and Prediction in a Multi-State Model for Breast Cancer. Biometrical Journal **48**(3), 366–380 (2006). doi:10.1002/bimj.200510218
38. Leffondré, K., Touraine, C., Helmer, C., Joly, P.: Interval-censored time-to-event and competing risk with death: Is the illness-death model more accurate than the cox model? International Journal of Epidemiology **42**(4), 1177–1186 (2013). doi:10.1093/ije/dyt126
39. Allignol, A., Schumacher, M., Wanner, C., Drechsler, C., Beyersmann, J.: Understanding competing risks: a simulation point of view. BMC Medical Research Methodology **11**(1), 86 (2011). doi:10.1186/1471-2288-11-86

Figures
Tables

Figure 1: Transition intensities for the multi-state with three states (MS3) model. Complete data.

Figure 2: Transition intensities for the multi-state with three states (MS3) model. Observed data.

Table 1: Simulation results for the complete data. Three state models.

Method	**HR**	**Parameter**	**Performance**[a,b]			
			Bias	**Percentage Bias**	**Accuracy MSE**	**Coverage**
MS3, no LE	2	$HR_{S_0 \to S_p}$	0.0043	0.22	0.0104	97.20
	1	$HR_{S_p \to S_c}$	0.0010	0.10	0.0042	95.80
MS3, LE	2	$HR_{S_0 \to S_p}$	0.0050	0.25	0.0118	94.80
	1	$HR_{S_p \to S_c}$	0.0049	0.49	0.0046	95.60
Cox CS, no LE	2	$HR_{S_0 \to S_p}$	-0.0203	-1.01	0.0136	97.20
		$HR_{S_0 \to S_c}$	-0.0539	-2.69	0.0173	95.00
Cox CS, LE	2	$HR_{S_0 \to S_p}$	-0.0212	-1.06	0.0154	95.40
		$HR_{S_0 \to S_c}$	-0.0493	-2.47	0.0179	93.40
DT3, no LE	2	$HR_{S_0 \to S_p}$	-0.0431	-2.16	0.0176	77.60
		$HR_{S_0 \to S_c}$	0.0824	4.12	0.0239	91.60
DT3, LE	2	$HR_{S_0 \to S_p}$	-0.0495	-2.48	0.0198	74.80
		$HR_{S_0 \to S_c}$	0.1063	5.18	0.0282	88.80

HR: hazard ratio of a false positive result, MS3: multi-state models with three states, LE: late entry, CS: cause-specific, DT: discrete time event.

[a] Bias=$\delta = \overline{\hat{\beta}} - \beta$; Percentage bias=$\left(\frac{\overline{\hat{\beta}} - \beta}{\beta} * 100 \right)$; Accuracy or mean square error (MSE): $(\overline{\hat{\beta}} - \beta)^2 + (\text{SE}(\hat{\beta}))^2$; Coverage: percentage of confidence intervals that contain the true HR.

[b] Assumptions:.....

Table 2: Simulation results for the observed data. Three state models.

Method	HR	Parameter	Performance[a,b]			
			Bias	Percentage Bias	Accuracy MSE	Coverage
MS3, no LE	2	$HR_{S_0 \to S_p}$	0.0072	0.36	0.0106	97.40
	1	$HR_{S_p \to S_c}$	0.0186	1.86	0.0150	94.80
MS3, LE	2	$HR_{S_0 \to S_p}$	0.0075	0.37	0.0123	94.40
	1	$HR_{S_p \to S_c}$	0.0242	2.42	0.0153	94.00
Cox CS, no LE	2	$HR_{S_0 \to S_p}$	-0.0197	-0.98	0.0149	96.00
		$HR_{S_0 \to S_c}$	-0.0233	-1.51	0.0479	95.80
Cox CS, LE	2	$HR_{S_0 \to S_p}$	-0.0211	-1.06	0.0174	95.00
		$HR_{S_0 \to S_c}$	-0.0152	-0.76	0.0477	95.60
DT3, no LE	2	$HR_{S_0 \to S_p}$	0.0073	0.36	0.0136	96.80
		$HR_{S_0 \to S_c}$	0.1899	9.49	0.0974	87.20
DT3, LE	2	$HR_{S_0 \to S_p}$	0.0035	0.17	0.0142	93.60
		$HR_{S_0 \to S_c}$	0.2282	11.41	0.1119	84.80

HR: hazard ratio of a false positive result, MS3: multi-state models with 3 states, LE: late entry, CS: cause-specific, DT: discrete time event.

[a] Bias=$\delta = \overline{\hat{\beta}} - \beta$; Percentage bias=$\left(\frac{\overline{\hat{\beta}} - \beta}{\beta} * 100 \right)$; Accuracy or mean square error (MSE): $(\overline{\hat{\beta}} - \beta)^2 + (\text{SE}(\hat{\beta}))^2$; Coverage: percentage of confidence intervals that contain the true HR.

[b] Assumptions: .....

Table 3: Simulation results for the observed data. Two state models.

Method	Parameter $HR_{S_0 \to S_c}$	Performance[a,b]			
		Bias	Percentage Bias	Accuracy MSE	Coverage
MS2 exact time, no LE		0.0070	0.35	0.0106	97.40
MS2 exact time, LE		0.073	0.36	0.0132	94.40
MS2 next screen, no LE		0.0064	0.32	0.0106	97.40
MS2 next screen, LE		0.0066	0.33	0.0122	94.40
Cox exact time, no LE		-0.0225	-1.13	0.0106	96.80
Cox exact time, LE		-0.0216	-1.08	0.0122	95.60
Cox next screen, no LE		-0.0225	-1.13	0.0106	96.60
Cox next screen, LE		-0.0216	-1.08	0.0122	95.60
DT2 exact time, no LE		0.0497	2.49	0.0125	94.80
DT2 exact time, LE		0.0565	2.83	0.0147	92.80
DT2 next screen, no LE		-0.0035	-0.17	0.0095	97.60
DT2 next screen, LE		0.0028	0.14	0.0109	95.60

HR: hazard ratio of a false positive result, MS3: multi-state models with 3 states, LE: late entry, CS: cause-specific, DT: discrete time event.

[a] Bias=$\delta = \overline{\hat{\beta}} - \beta$; Percentage bias=$\left(\frac{\overline{\hat{\beta}} - \beta}{\beta} * 100 \right)$; Accuracy or mean square error (MSE): $(\overline{\hat{\beta}} - \beta)^2 + (\text{SE}(\hat{\beta}))^2$; Coverage:.....

[b] Assumptions: HR=2 for the $S_0 \to S_c$ transition.

Table 4: INCA study: Three state models

Model	**Parameter**	HR (95% CI)
MS3	$HR_{S_0 \to S_p}$	1.7685 (1.2115 - 2.5818)
	$HR_{S_p \to S_c}$	0.9997 (0.5192 - 1.9247)
Cox, CS	$HR_{S_0 \to S_p}$	1.7722 (1.3389 - 2.3457)
	$HR_{S_0 \to S_c}$	1.6152 (1.1207 - 2.3279)
DT3, CS	$HR_{S_0 \to S_p}$	1.9500 (1.4283 - 2.6625)
	$HR_{S_0 \to S_c}$	1.2185 (0.6646 - 2.2341)

Table 5: INCA study: Two state models

Model	$HR_{S_0 \to S_c}$ (95% CI)
MS2, exact time	1.6811 (1.1526 - 2.4520)
MS2, next screen	1.6472 (1.1296 - 2.4020)
Cox, exact time	1.7051 (1.3643 - 2.1311)
Cox, next screen	1.7099 (1.3696 - 2.1346)
DT2, exact time	1.7450 (1.3233 - 2.3011)
DT2, next screen	1.8243 (1.4305 - 2.3265)

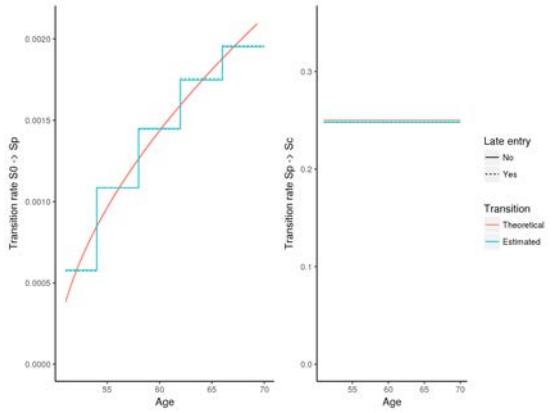


Figura 3.1: Transition intensities for the multi-state with three states (MS3) model. Complete data.

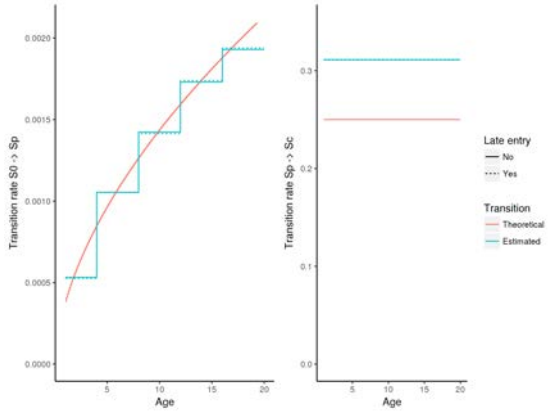


Figura 3.2: Transition intensities for the multi-state with three states (MS3) model. Observed data.

3.6 4t article

Breast Cancer RESEARCH

RESEARCH ARTICLE Open Access

Tumor phenotype and breast density in distinct categories of interval cancer: results of population-based mammography screening in Spain

Laia Domingo[1,2*], Dolores Salas[3,4], Raquel Zubizarreta[5], Marisa Baré[2,6,7], Garbiñe Sarriugarte[8], Teresa Barata[9], Josefa Ibáñez[3,4], Jordi Blanch[1], Montserrat Puig-Vives[10], Ana Belén Fernández[5], Xavier Castells[1,2,7], Maria Sala[1,2,7] and on behalf of the INCA Study Group

Abstract

Introduction: Interval cancers are tumors arising after a negative screening episode and before the next screening invitation. They can be classified into true interval cancers, false-negatives, minimal-sign cancers, and occult tumors based on mammographic findings in screening and diagnostic mammograms. This study aimed to describe tumor-related characteristics and the association of breast density and tumor phenotype within four interval cancer categories.

Methods: We included 2,245 invasive tumors (1,297 screening-detected and 948 interval cancers) diagnosed from 2000 to 2009 among 645,764 women aged 45 to 69 who underwent biennial screening in Spain. Interval cancers were classified by a semi-informed retrospective review into true interval cancers (n = 455), false-negatives (n = 224), minimal-sign (n = 166), and occult tumors (n = 103). Breast density was evaluated using Boyd's scale and was conflated into: <25%; 25 to 50%; 50 to 75%; >75%. Tumor-related information was obtained from cancer registries and clinical records. Tumor phenotype was defined as follows: luminal A: ER+/HER2- or PR+/HER2-; luminal B: ER +/HER2+ or PR+/HER2+; HER2: ER-/PR-/HER2+; triple-negative: ER-/PR-/HER2-. The association of tumor phenotype and breast density was assessed using a multinomial logistic regression model. Adjusted odds ratios (OR) and 95% confidence intervals (95% CI) were calculated. All statistical tests were two-sided.

Results: Forty-eight percent of interval cancers were true interval cancers and 23.6% false-negatives. True interval cancers were associated with HER2 and triple-negative phenotypes (OR = 1.91 (95% CI:1.22-2.96), OR = 2.07 (95% CI:1.42-3.01), respectively) and extremely dense breasts (>75%) (OR = 1.67 (95% CI:1.08-2.56)). However, among true interval cancers a higher proportion of triple-negative tumors was observed in predominantly fatty breasts (<25%) than in denser breasts (28.7%, 21.4%, 11.3% and 14.3%, respectively; <0.001).
False-negatives and occult tumors had similar phenotypic characteristics to screening-detected cancers, extreme breast density being strongly associated with occult tumors (OR = 6.23 (95% CI:2.65-14.66)). Minimal-sign cancers were biologically close to true interval cancers but showed no association with breast density.
(Continued on next page)

* Correspondence: LDomingo@parcdesalutmar.cat
[1]Department of Epidemiology and Evaluation, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
[2]Research network on health services in chronic diseases (REDISSEC), Barcelona, Spain
Full list of author information is available at the end of the article

(Continued from previous page)

Conclusions: Our findings revealed that both the distribution of tumor phenotype and breast density play specific and independent roles in each category of interval cancer. Further research is needed to understand the biological basis of the overrepresentation of triple-negative phenotype among predominantly fatty breasts in true interval cancers.

Introduction

The main goal of mammographic screening is to reduce mortality and morbidity from breast cancer through early detection. However, women with interval cancer do not benefit from early detection, as their tumors are detected clinically after a negative screening episode and before the following screening invitation [1].

Interval cancers can be distinguished into four categories by the retrospective review of both screening and diagnostic mammograms: a) true interval cancers are those that showed normal or benign features in the previous screening mammogram; b) false-negative cancers are detected when signs suspicious for malignancy are retrospectively seen on a mammogram; c) minimal-signs are cancers showing detectable but non-specific signs at the latest screening; and d) occult tumors are those that present clinical signs of the disease despite a lack of mammographic abnormalities either at screening or at diagnosis. The European guidelines recommend first reviewing the screening films without histopathological information, and then using the screening and diagnostic films for the definitive classification. This practice involves substantial effort and is not normally routinely performed [1,2]. This explains why there are few large series with specific information on interval cancer categories, especially series providing biological information [3]. Studies evaluating interval cancers and following the recommendations of the European guidelines have found that about half are true interval cancers, over 20% are false negatives [3-5], and fewer than 20% are occult tumors and minimal-sign cancers [5,6].

There is evidence that interval cancers are more likely to have less favorable molecular features than screening-detected cancers, such as a high proportion of tumors not expressing estrogen receptor (ER negative, ER-) or progesterone receptor (PR negative, PR-) [4,7-9]. Some studies have reported a higher proportion of triple-negative cancers (ER-, PR-, human epidermal growth factor receptor 2 (HER2)-) among interval cancers [7,10] and this increase is even higher if only the subset of true interval cancers is considered in comparison to screening-detected cancers [4]. So far, this tumor phenotype lacks the benefit of specific adjuvant therapy and is associated with an aggressive behavior pattern and poor prognosis [11].

Breast density has also been related to interval cancer. There is increasing evidence that women with dense breasts are more likely to be diagnosed with interval cancer [12-14], but the role of breast density has not yet been elucidated [13,15]. A masking effect, which would contribute to hide the tumors [15], as well as a biological effect related to tumor growth [16], has been purposed. Because breast density influences both the risk and detection of breast cancer, as well as the likelihood of developing certain pathological subtypes [17,18], studying this factor in interval cancers would be of great interest.

We hypothesized that the roles of tumor phenotype and that of breast density differ in distinct categories of interval cancers. The aim of this study was to describe the tumor-related characteristics of true interval cancers, false negatives, minimal-sign cancers and occult tumors, and to assess the association of breast density and tumor phenotype in the four interval cancer categories. This study provides a comprehensive approach to the four categories of interval breast cancer identified from one of the largest cohorts of women participating in population-based breast cancer screening.

Methods

Setting

We performed a case–control study nested in a cohort of 645,764 women aged 45 to 69 years, screened in Spain between 1 January 2000 and 31 December 2006, and followed up until June 2009. These women underwent a total of 1,508,584 screening mammograms. During the study period, 5,309 cancers were detected in routine screening mammograms and 1,669 emerged as interval cancers, including both invasive and *in situ* carcinomas.

All women resident in Spain aged 50 to 69 years are actively invited to participate in the population-based screening program by a written letter every 2 years, following the European guidelines for Quality Assurance in Mammographic Screening Recommendations [1]. This nationwide program achieves the required standards [19].

We gathered data from five Spanish regions (Basque Country, Canary Islands, Catalonia, Galicia, and Valencia), covering a population of 752,487 women in 2005. Two mammographic projections (mediolateral-oblique and craniocaudal views) were made both in the initial and in successive rounds, except in one program. All mammograms were read by two radiologists, except in two programs, and the classification used for mammogram

reading was BI-RADS [20]. Two regions switched to digital mammography during 2003 to 2005.

All screening programs keep mammography registers with data from participants and the final outcome of screening. Once a tumor is histologically confirmed, the woman is referred to a hospital for treatment and follow up. They are not further invited to screening, as they are controlled in the health care system.

Study data were collected using a protocol approved by the ethics committee of Parc de Salut Mar (CEIC-Parc de Salut MAR), Barcelona. Specific patient consent was not required because we used retrospective data from screening participants who had previously signed information release documents.

Study population: case and control definitions

Case subjects with interval cancer and control subjects with screening-detected cancer were drawn from women enrolled in any of the screening programs. We used the definition of interval cancers purposed in the European guidelines: "primary breast cancer arising after a negative screening episode, with or without further assessment, and before the next invitation to screening, or within 24 months for women who reached the upper age limit" [1]. The overall 1,669 interval cancers were matched by screening program and the year of the last screening mammogram to one screening-detected cancer, that is, a pathologically-confirmed malignant lesion identified during the screening process. We excluded those cases and controls with no available information on screening and diagnostic (only for interval cancers) mammograms. Finally, we analyzed 948 interval cancers, and 1,297 screening-detected cancers. Ductal *in situ* carcinomas were excluded from the analysis.

Assessment of interval cancers and breast density classification

Interval cancers were identified by merging data from the registers of screening programs with population-based cancer registries, the regional Minimum Basic Data Set (MBDS) and hospital-based cancer registries. The use of different data sources ensured the quality and homogeneity of the process across the study period and regions. Population-based cancer registries covered four out of five regions. The MBDS (based on hospital discharges with information on the principal diagnosis) is updated yearly and is available in all regions. All data sources kept information on the time of diagnosis, which allowed us to ensure that all interval cancers fitted the case definition.

For interval cancer classification, three panels with three experienced radiologists performed a semi-informed retrospective review of both screening and diagnostic mammograms through independent double reading with arbitration. Screening mammograms were first reviewed alone, without the radiologists seeing the diagnostic mammogram and without histological information (blind review). Interval cancers were provisionally classified into positive (abnormality clearly visible and warrants assessment), negative (normal mammogram), and minimal-sign (subtle abnormality, not necessarily regarded as warranting assessment). Later, the diagnostic and screening mammograms were reviewed together and interval cancers were definitively classified into true interval cancers, false negatives, minimal-sign cancers, and occult tumors [1]. In the definitive classification, we ensured that the site where the minimal signs were identified correlated with the site of the interval cancer. When there was no correlation, the case was considered a true interval cancer.

One radiologist from each panel determined the breast density of the cancer-free breast, for both interval and screening-detected cancers. Breast density was evaluated using Boyd's scale, a semiquantitative score of six categories using percentages of density: A: 0%; B: 1 to 10%; C: 10 to 25%; D: 25 to 50%; E: 50 to 75%; F: 75 to 100% [21]. For purposes of assessing the impact of predominately fatty versus increasingly dense breasts, the first three categories were combined into the <25% group [22].

Study variables

The woman's age at diagnosis was obtained from the date of birth and date of the screening mammogram. Tumor-related information (the tumor histology, grade, size, lymph node involvement, and ER, PR, HER2, p53 and Ki67 status) was obtained from the cancer registries, hospital-based registers, and from the clinical records. Biomarker assessment was performed as part of the diagnostic process in the hospitals. The positivity criteria used by each hospital followed international recommendations and their updates throughout the study period [23,24]. Tumors were considered positive when more than 20% and 10% of cells stained positive for Ki67 and p53, respectively. For the histological classification, we used ICD-O, 3rd edition. Histological grade was defined according to the Scarff-Bloom-Richardson criteria, modified by Elson [25].

Based on the expression of ER, PR and HER2, tumors were classified into four phenotypes: 1) luminal A: ER+/HER2- or PR+/HER2-; 2) luminal B: ER+/HER2+ or PR+/HER2+; 3) HER2: ER-/PR-/HER2+; and 4) triple-negative: ER-, PR-, HER2- [26].

Statistical analysis

Comparisons were established between screening-detected cancers, true interval cancers, false negatives, minimal-sign cancers, and occult tumors. Statistical significance was assessed using the Chi-square or Fisher exact test for categorical variables, and one-way analysis of variance (ANOVA) for continuous variables. If a significant

difference was found, we calculated standardized Pearson residuals as a measure of deviation between the observed and expected values to determine which cells contributed most to the Chi-square estimator [27]. Clinical features, age at diagnosis, breast density, biomarker expression, and the phenotypic classification were compared between study groups. Then, we carried out a stratified analysis of tumor phenotype and breast density by study groups.

A multinomial regression analysis was computed to determine the effect of tumor phenotype and breast density on the odds of developing a true interval cancer, a false negative, a minimal-sign cancer, or an occult tumor versus screening-detected cancers. Our final multinomial regression model was adjusted for screening program (categorical), age (continuous), and tumor size (categorical, <11 mm; 11 to 20 mm; 21 to −50 mm; >50 mm). The outputs were plotted, showing the adjusted odds ratio (OR) and the 95% CI for each category of interval cancer, which served as the endpoints of the multinomial model.

We conducted sensitivity analyses by including or excluding screening-detected cancers diagnosed in prevalent screening. We tested different reference categories for breast density (≤10%, ≤50%), and we checked the inclusion of covariates into the multivariate models (year of screening mammogram, histological grade, Ki67 and p53 status, the use of digital or analog mammography, and menopausal status). The sensitivity analyses showed no significant differences with respect to the definitive multinomial model. We examined the interaction between breast density and phenotype and found a non-significant effect within the multiple endpoints of the multinomial model.

All *P*-values were based on two-sided tests and were considered statistically significant if <0.05. Statistical analyses were performed using the SPSS (version 12.0) and R statistical software programs.

Results

A total of 1,297 screening-detected cancers and 948 interval cancers were included in the analyses. Most interval cancers were true interval cancers (n = 455, 48.0%), followed by false negatives (n = 224, 23.6%), minimal-sign cancers (n = 166, 17.5%) and occult tumors (n = 103, 10.9%).

Table 1 summarizes information on age at diagnosis and tumor-related characteristics of screening-detected cancers and interval cancer categories. Women with true interval cancers and occult tumors were younger (mean age 56.4 years and 55.1 years, respectively) than women in the remaining subsets (*P* <0.001). Over 80% of true interval cancers were detected 12 months after the last screening or later, whereas 42.7% of occult tumors developed within the first 12 months. As expected, the highest percentage of tumors ≤10 mm in size was found among screening-detected cancers (34.8%; *P* <0.001). Among interval cancers, the percentage ranged from 7.9 to 13.3% in true interval cancers and occult tumors, respectively. Extremely dense breasts (>75%) were most frequently associated with occult tumors followed by false-negative cancers and true interval cancers (28.2, 17.0 and 16.5%, respectively, versus 11.6% in screening-detected cancers; *P* <0.001).

The expression of biomarkers among study groups is detailed in Table 2. True interval cancers were less likely to express ER and PR than screening-detected cancers but were more likely to overexpress HER2, p53, and Ki67. In contrast, the molecular profile observed among occult tumors revealed a higher percentage of ER + cancers (88.4 versus 82.5%) and a lower percentage of HER2+ cancers (14.1 versus 21.9%) compared with screening-detected cancers. Molecularly, false-negative tumors were similar to screening-detected cancers, although they showed a higher proportion of tumors overexpressing Ki67 (50.3 versus 40.2%). Almost 35% of minimal-sign cancers overexpressed p53.

The distribution of tumor phenotypes among study groups is shown in Table 3. True interval cancers and minimal-sign tumors showed a higher proportion of triple-negative cancers (19.9 and 17.3%, respectively), whereas false-negative and occult tumors showed a similar tumor phenotype profile to screening-detected cancers.

In Table 4 is shown the distribution of tumor phenotypes among study groups, stratified by breast density. According to breast density, differences in phenotype distribution were statistically significant among true interval cancers. The highest proportion of triple-negative cancers among true interval cancers was observed in breasts with 25% lower density than in denser breasts (28.7, 21.4, 11.3 and 14.3%, respectively; *P* <0.001).

Adjusted OR and 95% CI estimated by multinomial regression analysis are plotted in Figure 1. True interval cancers were associated with HER2 and triple-negative phenotypes (OR 1.91, 95% CI 1.22, -2.96; OR 2.07, 95% CI 1.42, 3.01, respectively) and extremely dense breasts (OR 1.67, 95% CI 1.08, 2.56). Occult tumors were over six times more likely to develop in extremely dense breasts (OR 6.23, 95% CI 2.65, 14.66). False-negative cancers showed a non-significant tendency to occur in extremely dense breasts, whereas in the adjusted model, minimal-sign cancers showed no association with either breast density or tumor phenotype.

Discussion

This comprehensive study suggests that true interval and minimal-sign cancers showed similar tumor phenotype distribution, with almost 20% of these tumors being

Table 1 Comparison of age at diagnosis and tumor characteristics at diagnosis between screening-detected cancers (n = 1,297) and interval cancers (n = 948)

	Screening-detected cancers	True interval cancers	False negatives	Minimal-sign cancers	Occult tumors	
	n = 1,297	**n = 455**	**n = 224**	**n = 166**	**n = 103**	***P*-value[†]**
Interval cancer entities, n (%)[‡]		455 (48.0)	224 (23.6)	166 (17.5)	103 (10.9)	
Time since last screening, n (%)						
<=12 months		89 (19.6)	73 (32.7)	53 (32.1)	44 (42.7)	
>12 months		364 (80.4)	150 (67.3)	112 (67.9)	59 (57.3)	<0.001
Age, y, mean (95% CI)	57.6 (57.3, 57.9)	56.4 (55.9, 57.0)	57.4 (56.6, 58.1)	56.8 (56.0, 57.6)	55.1 (54.0, 56.2)	<0.001
Tumor size, mm, mean (95% CI)	15.7 (15.1, 16.3)	25.3 (23.6, 26.9)	23.9 (22.1, 25.8)	22.7 (20.5, 24.8)	19.3 (17.0, 21.6)	<0.001
Focality, n (%)						
Unifocal	1030 (82.8)	341 (79.1)	171 (78.4)	118 (74.7)	83 (85.6)	
Multifocal and/or multicentric	214 (17.2)	90 (20.9)	47 (21.6)	40 (25.3)	14 (14.4)	0.041
Unknown	53	24	6	8	6	
Tumor size, n (%)						
<= 10 mm	452 (34.8)*	36 (7.9)*	18 (8.0)*	22 (13.3)*	13 (12.6)*	
11 to 20 mm	521 (40.2)	147 (32.3)	79 (35.3)	53 (31.9)	45 (43.7)	
21 to 50 mm	233 (18.0)*	171 (37.6)*	78 (34.8)*	62 (37.3)*	23 (22.3)	
>50 mm	91 (7.0)*	101 (22.2)*	49 (21.9)*	29 (17.5)	22 (21.4)*	<0.001
Unknown	0	0	0	0	0	
Lymph node involvement, n (%)						
Negative	872 (70.2)*	195 (50.4)*	102 (54.5)	76 (49.7)*	54 (62.1)	
Positive	371 (29.8)*	192 (49.6)*	85 (45.5)	77 (50.3)*	33 (37.9)	<0.001
Unknown	54	68	37	13	16	
Histological type, n (%)						
Ductal	1039 (80.5)	349 (77.6)	165 (74.0)	129 (77.7)	70 (68.6)	
Lobular	109 (8.4)*	54 (12.0)	36 (16.1)*	16 (9.6)	21 (20.6)*	
Other	143 (11.1)	47 (10.4)	22 (9.9)	21 (12.7)	11 (10.8)	<0.001
Unknown	6	5	1	0	1	
Histological grade, n (%)						
I	390 (34.9)*	57 (14.9)*	41 (21.4)	33 (22.9)	19 (22.6)	
II	474 (42.4)	149 (39.0)	88 (45.8)	62 (43.1)	42 (50.0)	
III	241 (21.6)*	171 (44.8)*	61 (31.8)	47 (32.6)	20 (23.8)	
NA	13 (1.2)	5 (1.3)	2 (1.0)	2 (1.4)	3 (3.6)	<0.001

Table 1 Comparison of age at diagnosis and tumor characteristics at diagnosis between screening-detected cancers (n = 1,297) and interval cancers (n = 948) *(Continued)*

Breast density, n (%)						
<25%	510 (39.3)*	139 (30.5)*	81 (36.2)	64 (38.6)	15 (14.6)*	
25 to 50%	359 (27.7)	127 (27.9)	60 (26.8)	47 (28.3)	20 (19.4)	
51 to 75%	277 (21.4)*	114 (25.1)	45 (20.1)	39 (23.5)	39 (37.9)*	
>75%	151 (11.6)	75 (16.5)	38 (17.0)	16 (9.6)	29 (28.2)*	<0.001
Unknown	0	0	0	0	0	

Missing values were excluded from the calculations of percentages. *Standardized Pearson residuals with statistically significant deviation between observed and expected values.
†*P*-values for comparison of characteristics among the five study groups were obtained by one-way analysis of variance for continuous variables and the Chi-square test for categorical variables. All tests were two-sided. ‡Row percentages.

Table 2 Biomarker expression among screening-detected cancers (n = 1,297) and distinct categories of interval cancers (n = 948)

	Screening-detected cancers	True interval cancers	False negatives	Minimal-sign cancers	Occult tumors	
	n = 1,297	n = 455	n = 224	n = 166	n = 103	*P*-value[†]
Estrogen receptor	1022 (82.5)*	283 (63.2)*	178 (81.7)	114 (71.3)	84 (88.4)	<0.001
Missing values	58	7	6	6	8	
Progesterone receptor	775 (63.7)*	214 (48.2)*	128 (59.0)	86 (54.4)	60 (64.5)	<0.001
Missing values	81	11	7	8	10	
HER2	203 (21.9)	113 (29.1)*	44 (24.0)	31 (23.1)	11 (14.1)	0.018
Missing values	371	67	41	32	25	
p53	149 (22.7)	86 (36.6)*	23 (21.5)	27 (34.6)	17 (28.8)	<0.001
Missing values	641	228	117	88	44	
Ki67	381 (40.2)	169 (52.5)*	83 (50.3)	50 (41.7)	26 (39.4)	0.001
Missing values	349	133	59	46	37	

Number of cases and percentage of tumors with positive biomarker expression. *Standardized Pearson residuals with statistically significant deviation between observed and expected values. [†]*P*-values for comparison of characteristics among the five study groups were obtained by two-sided Chi-square test. HER2, human epidermal growth factor receptor 2.

triple negative. In contrast, false-negative and occult tumors were phenotypically closer to screening-detected cancers. High breast density was mainly associated with occult tumors, and to a lesser extent, to true interval cancers and false negatives. However, among true interval cancers, those with the triple-negative phenotype were more likely to occur in predominately fatty breasts than in extremely dense breasts.

The proportion of tumors classified as true interval cancers, false negatives, minimal-sign cancers, or occult tumors is in line with previous series [3-6], that followed the European guidelines for the definition and classification of interval cancers. The percentage of false negatives slightly exceeded the limit of 20% recommended by the European guidelines, but is lower than that in other contexts [3,28]. Nevertheless, the lack of a standardized method for the radiological classification of interval cancers, together with the subjective nature of mammography interpretation, hamper valid comparisons between screening programs [2,28].

As expected by the lead time, all interval cancers were larger at diagnosis and were more likely to show lymph node involvement than screening-detected cancers. In agreement with previous work [4,8], true interval cancers were those with the longest waiting time to breast cancer diagnosis and were also the largest. However, some studies that analyzed occult tumors and true interval cancers together have reported that the clinical features of this subset differed less than those of screening-detected tumors [8]. As occult tumors were those detected earliest after screening, resulting in a higher proportion of small carcinomas and showing a molecular pattern similar to screening-detected cancers, grouping true interval and occult tumors together may lead to underestimation of the less prognostically favorable features of true interval cancers.

Table 3 Distribution of tumor phenotypes among screening-detected cancers (n = 1,297) and categories of interval cancers (n = 948)

	Screening-detected cancers	True interval cancers	False negatives	Minimal-sign cancers	Occult tumors	
	n = 1,297	n = 455	n = 224	n = 166	n = 103 (%)	P value[†]
Tumor phenotype						
Luminal A	629 (68.3)	197 (50.9)*	124 (68.1)	79 (59.4)	62 (79.5)	
Luminal B	139 (15.1)	60 (15.5)	29 (15.9)	18 (13.5)	8 (10.3)	
HER2	62 (6.7)	53 (13.7)*	14 (7.7)	13 (9.8)	3 (3.8)	
Triple-negative	91 (9.9)*	77 (19.9)*	15 (8.2)	23 (17.3)	5 (6.4)	<0.001
Unknown	376	68	42	33	25	

Results are expressed as number (%). Tumor phenotype = Luminal A (ER+/HER2- or PR+/HER2-); Luminal B (ER+/HER2+ or PR+/HER2+); HER2 (ER-/PR-/HER2+); Triple-negative (ER-/PR-/HER2-). *Standardized Pearson residuals with statistically significant deviation between observed and expected values. [†]The distribution of tumor phenotype was compared among the study groups using the two-sided Chi-square test. ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.

Table 4 Distribution of tumor phenotypes among screening-detected cancers (n = 1,297) and categories of interval cancers (n = 948) stratified by breast density

	Breast density				
	<25%	25 to 50%	50 to 75%	>75%	*P*-value†
Tumor phenotype					
Screening-detected cancers					
Luminal A	247 (68.6)	165 (65.5)	142 (70.3)	75 (71.4)	
Luminal B	51 (36.7)	49 (19.3)	24 (11.9)	15 (14.3)	
HER2	20 (5.6)	16 (6.3)	17 (8.4)	9 (8.6)	
Triple negative	42 (11.7)	24 (9.4)	19 (9.4)	6 (5.7)	
Unknown	150	105	25	46	0.306
True interval cancers					
Luminal A	60 (52.2)	51 (45.5)	49 (50.5)	37 (58.7)	
Luminal B	13 (11.3)	17 (15.2)	15 (15.5)	15 (23.8)	
HER2	9 (7.8)	20 (17.9)	22 (22.7)*	2 (3.2)*	
Triple negative	33 (28.7)*	24 (21.4)	11 (11.3)	9 (14.3)	
Unknown	24	15	17	12	<0.001
False negatives					
Luminal A	40 (62.5)	39 (75.0)	22 (64.7)	23 (71.9)	
Luminal B	10 (15.6)	4 (7.7)	9 (26.5)	6 (18.8)	
HER2	6 (9.4)	4 (7.7)	2 (5.9)	2 (6.3)	
Triple negative	8 (12.5)	5 (9.6)	1 (2.9)	1 (3.1)	
Unknown	17	8	11	6	0.375
Minimal-sign cancers					
Luminal A	30 (60.0)	26 (66.7)	16 (55.2)	7 (46.7)	
Luminal B	6 (12.0)	3 (7.7)	5 (17.2)	4 (26.7)	
HER2	2 (4.0)	5 (12.8)	2 (6.9)	4 (26.7)	
Triple negative	12 (24.0)	5 (12.8)	6 (20.7)	0 (0)	
Unknown	14	8	10	1	0.081
Occult tumors					
Luminal A	8 (80.0)	12 (80.0)	26 (78.8)	16 (80.0)	
Luminal B	0 (0)	2 (13.3)	4 (12.1)	2 (10.0)	
HER2	0 (0)	1 (6.7)	0 (0)	2 (10.0)	
Triple negative	2 (20.0)	0 (0)	3 (9.1)	0 (0)	
Unknown	5	5	6	9	0.296

Results are expressed as number (%). Tumor phenotypes = Luminal A (ER+/HER2- or PR+/HER2-); Luminal B (ER+/HER2+ or PR+/HER2+); HER2 (ER-/PR-/HER2+); Triple-negative (ER-/PR-/HER2-). *Standardized Pearson residuals with statistically significant deviation between observed and expected values. †*P*-value assesses the distribution of tumor phenotype distribution among breast density categories within study groups, using the two-sided Chi-square test, or Fisher exact test when appropriate. ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.

True interval and minimal-sign cancers showed similarities in their patterns of biomarker expression and tumor phenotype. Our results confirm that true interval cancers were less likely to express hormonal receptors [4,8,9,29] and support previous series reporting overexpression of HER2, p53, and Ki67 [4,9,30]. To our knowledge, this is the first study that provides complete molecular characterization of minimal-sign cancers. In line with previous evidence for the overrepresentation of triple-negative tumors among interval cancers [4,7,10], we found that most triple-negative tumors were concentrated among true interval and minimal-sign cancers. The biological similarities shared by both entities suggest that some minimal-sign tumors could be a more advanced form of true interval cancer, whereas false negatives seem to be a clearly distinct entity from minimal-sign cancers. These two entities should not be classified together, as has been done in some previous studies [31,32].

Breast density is a well-known risk factor for breast cancer and particularly interval cancer [13,14], but its

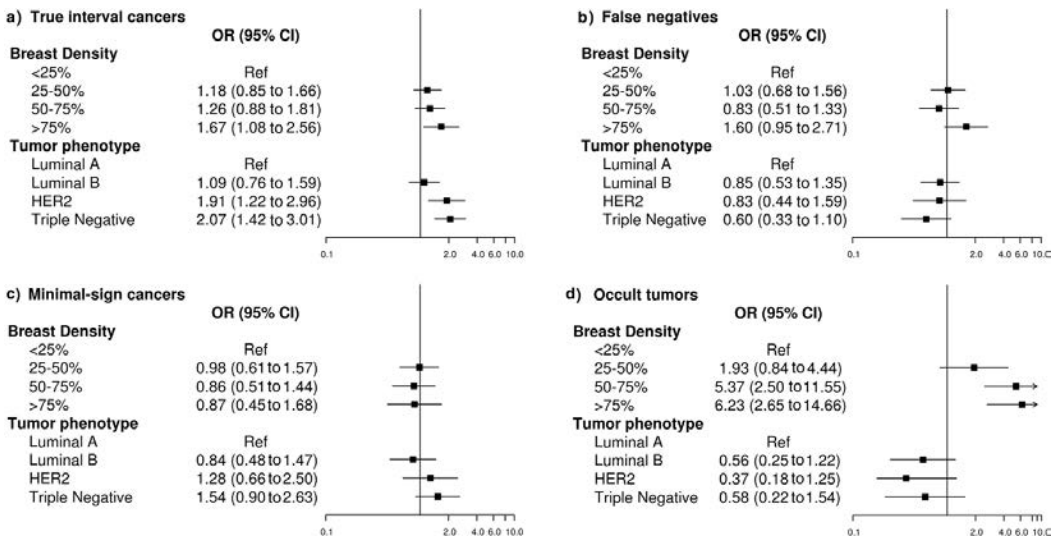


Figure 1 Multinomial logistic regression model of the association of breast density and tumor phenotypes with categories of interval cancer, adjusted for age at screening, screening program, and tumor size. The association of breast density with tumor phenotype, adjusted by screening program (categorical), age (continuous), and tumor size (categorical, <11 mm; 11 to 20 mm; 21 to 50 mm; >50 mm), is shown for the multiple endpoints of the multinomial logistic regression models, which are **(a)** true interval cancers; **(b)** false negatives; **(c)** minimal-sign cancers and **(d)** occult tumors. The reference category (Ref) is screening-detected cancers. The black squares and the horizontal lines represent the odds ratios (OR) and corresponding 95% CI, respectively. ORs are presented on the log scale. Tumor phenotype = Luminal A: ER/HER2- or PR+/HER2-; Luminal B: ER+/HER2+ or PR+/HER2+; HER2: ER-/PR-/HER2+; Triple-negative: ER-/PR-/HER2-.

association with tumor phenotypes remains controversial. Our findings revealed that luminal cancers were more likely to be detected in extremely dense breasts than in predominately fatty breasts, in agreement in with some previous studies [10,33,34], but contrasting with others [18]. Yanhjyan *et al*. [18] reported a higher proportion of triple-negative cancers among women with dense breasts. However, their study design was not comparable with ours, as these authors did not take into account whether the cancers were detected by screening. Unless the detection mode is considered, the association of triple-negative cancers and breast density may be overestimated, because tumors detected between two screenings are more likely to be detected in women with dense breasts and to be triple negative [15,17].

Our findings support the association of breast density and interval cancer independently of phenotype. The association of breast density and true interval cancers reinforces the hypothesis that some tumors are stimulated by growth factors found in dense breasts [35]. However, the overrepresentation of triple-negative tumors among predominantly fatty breasts in true interval cancers may reflect the aggressive behavior, rapid carcinogenesis and nonlinear progression of this tumor phenotype, regardless of breast density [11,36]. Further research is still needed to understand the biological basis of the association of breast density and tumor phenotypes, taking into account the mode of detection. The knowledge of epidemiological factors and radiological features predictive of an aggressive tumor subtype, such as the triple-negative phenotype, could add information for future personalized screening programs in women at risk of interval cancer.

The strong association of breast density and occult tumors pointed to a masking effect, confirming the assumptions noted years ago by Houssami [2]. Our findings also reinforce the idea that a masking effect mainly affects cancers that developed up to 12 months after screening [15]. Nevertheless, breast density appears to play a lesser role in false negatives, in line with previous series [13,37]. Breast density remains a major issue in breast cancer screening because it is one of the variables proposed to tailor screening [38]. Information on its role among interval cancer categories along with data on its relationship with tumor phenotypes may be useful to estimate the potential benefit of personalizing screening strategies on the basis of this factor.

The strengths of the current study are the large sample size and the completeness of the information. These factors have allowed us to study the role of breast density and tumor phenotype for each interval cancer category and to describe some features that may help to better understand their etiology.

There are, however, some limitations that should be considered. First, misclassification among interval cancers cannot be excluded. Some interval cancers could be classified as screening-detected if symptomatic women waited for the screening visit instead of making an immediate appointment with a physician. However, such misclassification would attenuate differences in tumor characteristics among study groups. Second, not all cases would have been phenotypically classified. Since this lack of information affects both screening-detected cancers and interval cancers, and was similar in all screening programs we do not believe that it affects the results. However, data on p53 and Ki67 were not always available because they were not routinely checked in all centers. Given that their lack of availability was not random, these data were not entered into the multinomial model. Third, grouping breast density into four categories reduced the sample size in the stratified analyses, but allowed the role of extremely dense breasts to be assessed. Collapsing breast density into two categories (≤50 and >50%) diminished the magnitude of the association of breast density and distinct categories of interval cancer (data not shown). Fourth, some important variables associated with breast density, such as body mass index, age at menarche or childbirth, are not routinely collected by screening programs, and therefore we could not adjust for these potential confounders.

Conclusions

Our findings revealed that both the distribution of tumor phenotype and breast density play specific and independent roles in each category of interval cancer. Almost half of the interval cancers were true interval cancers, which encompassed a high percentage of tumors with a molecular profile associated with poor prognosis on the one hand and were more likely to be detected among women with extremely dense breasts on the other. False-negative and occult tumors had similar phenotypic characteristics to screening-detected cancers, high breast density being strongly associated with occult tumors. Minimal-sign cancers were biologically close to true interval cancers but showed no association with breast density. In view of the heterogeneity within interval cancers, further studies aiming to characterize interval cancers should avoid grouping true interval cancers and occult tumors, or false-negative and minimal-sign cancers. Knowledge of the clinical and biological particularities of interval cancers and of the role of breast density may be useful for the design of new risk-based screening strategies.

Abbreviations

ANOVA: analysis of variance; CI: confidence intervals; ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; MBDS: Minimum Basic Data Set OR: odds ratio; PR: progesterone receptor.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LD drafted the first version of the manuscript, performed the statistical analysis, and participated in the design of the study. MS, DS, RZ, and XC, conceived the study and participated in its design and coordination, and critically revised the manuscript for important intellectual content. MB, GS, TB, JI, JB, LD, MPV and ABF participated in the acquisition of data from screening programs and from clinical records, helped in the interpretation of the results, and helped to draft the manuscript. JB supported the statistical analysis, gathered data, and validated the whole database. ABF, JI, LD, DS and MS coordinated the review process for interval cancer classification and for the assessment of breast density. All authors critically reviewed the manuscript. All of them read and approved the final manuscript.

Acknowledgments

This work was supported by Instituto de Salud Carlos III-FEDER (PI 09/01153, PI09/02385, PI09/01340). The authors acknowledge the dedication and support of the entire Interval Cancer (INCA) Study Group (alphabetical order): IMIM (Hospital del Mar Medical Research Institute), Barcelona: Jordi Blanch, Xavier Castells, Mercè Comas, Laia Domingo, Francesc Macià, Juan Martínez, Ana Rodríguez-Arana, Marta Román, Anabel Romero, Maria Sala. General Directorate Public Health and Centre for Public Health Research (CSISP), FISABIO, Valencia: Carmen Alberich, María Casals, Josefa Ibáñez, Amparo Lluch, Inmaculada Martínez, Josefa Miranda, Javier Morales, Dolores Salas, Ana Torrella. Galician Breast Cancer Screening Program, Xunta de Galicia: Raquel Almazán, Miguel Conde, Montserrat Corujo, Ana Belén Fernández, Joaquín Mosquera, Alicia Sarandeses, Manuel Vázquez, Raquel Zubizarreta. General Directorate of Health Care Programmes. Canary Islands Health Service: Teresa Barata, Isabel Díez de la Lastra, Juana María Reyes. Basque Country Breast Cancer Screening Program. Osakidetza: Arantza Otegi, Garbiñe Sarriugarte. Corporació Sanitària Parc Taulí, Sabadell: Marisa Baré, Núria Torà. Hospital Santa Caterina, Girona: Joana Ferrer, Francesc Castanyer, Gemma Renart. Epidemiology Unit and Girona Cancer Registry; and University of Girona: Rafael Marcos-Gragera, Montserrat Puig-Vives. Biomedical Research Institut of Lleida (IRBLLEIDA): Carles Forné, Montserrat Martínez-Alonso, Albert Roso, Montse Rué, Ester Vilaprinyó. Universitat Rovira i Virgili, Tarragona: Misericordia Carles, Aleix Gregori, María José Pérez, Roger Pla.

Author details

[1]Department of Epidemiology and Evaluation, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. [2]Research network on health services in chronic diseases (REDISSEC), Barcelona, Spain. [3]General Directorate Public Health, Valencia, Spain. [4]Centre for Public Health Research (CSISP), FISABIO, Valencia, Spain. [5]Galician Breast Cancer Screening Program, Directorate for innovation and management of public health, Santiago de Compostela, Spain. [6]Epidemiology and Assessment Unit UDIAT-Diagnostic Centre, Corporació Sanitària Parc Taulí, Sabadell, Spain. [7]Department of Pediatrics, Obstetrics and Gynecology, Preventive Medicine and Public Health, Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain. [8]Osakidetza Breast Cancer Screening Programme, Basque Country Health Service, Bilbao, Spain. [9]General Directorate of Health Care Programmes, Canary Islands Health Service, Las Palmas de Gran Canaria, Spain. [10]Epidemiology Unit and Girona Cancer Registry, University of Girona, Girona, Spain.

Received: 23 May 2013 Accepted: 6 January 2014
Published: 10 January 2014

References

1. Perry N, Broeders M, de Wolf C, Törnberg C, Holland R, von Karsa L: *European guidelines for quality assurance in breast cancer screening and diagnosis.* 4th edition. Luxembourg: Office for Official Publications of the European Communities; 2006.
2. Houssami N, Irwig L, Ciatto S: **Radiological surveillance of interval breast cancers in screening programmes.** *Lancet Oncol* 2006, **7:**259–265.
3. Hofvind S, Geller B, Skaane P: **Mammographic features and histopathological findings of interval breast cancers.** *Acta Radiol* 2008, **49:**975–981.

4. Domingo L, Sala M, Servitja S, Corominas JM, Ferrer F, Martinez J, Macia F, Quintana MJ, Albanell J, Castells X: **Phenotypic characterization and risk factors for interval breast cancers in a population-based breast cancer screening program in Barcelona, Spain.** *Cancer Causes Control* 2010, **21:**1155–1164.
5. Vitak B: **Invasive interval cancers in the Ostergotland Mammographic Screening Programme: radiological analysis.** *Eur Radiol* 1998, **8:**639–646.
6. Bare M, Sentis M, Galceran J, Ameijide A, Andreu X, Ganau S, Tortajada L, Planas J: **Interval breast cancers in a community screening programme: frequency, radiological classification and prognostic factors.** *Eur J Cancer Prev* 2008, **17:**414–421.
7. Collett K, Stefansson IM, Eide J, Braaten A, Wang H, Eide GE, Thoresen SO, Foulkes WD, Akslen LA: **A basal epithelial phenotype is more frequent in interval breast cancers compared with screen detected tumors.** *Cancer Epidemiol Biomarkers Prev* 2005, **14:**1108–1112.
8. Kirsh VA, Chiarelli AM, Edwards SA, O'Malley FP, Shumak RS, Yaffe MJ, Boyd NF: **Tumor characteristics associated with mammographic detection of breast cancer in the Ontario breast screening program.** *J Natl Cancer Inst* 2011, **103:**942–950.
9. Musolino A, Michiara M, Conti GM, Boggiani D, Zatelli M, Palleschi D, Bella MA, Sgargi P, Di Blasio B, Ardizzoni A: **Human epidermal growth factor receptor 2 status and interval breast cancer in a population-based cancer registry study.** *J Clin Oncol* 2012, **30:**2362–2368.
10. Caldarella A, Puliti D, Crocetti E, Bianchi S, Vezzosi V, Apicella P, Biancalani M, Giannini A, Urso C, Zolfanelli F, Paci E: **Biological characteristics of interval cancers: a role for biomarkers in the breast cancer screening.** *J Cancer Res Clin Oncol* 2013, **139:**181–185.
11. Chacon RD, Costanzo MV: **Triple-negative breast cancer.** *Breast Cancer Res* 2010, **12:**S3.
12. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V: **Effect of age, breast density, and family history on the sensitivity of first screening mammography.** *JAMA* 1996, **276:**33–38.
13. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, White E: **Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers.** *J Natl Cancer Inst* 2000, **92:**1081–1087.
14. Pollan M, Ascunce N, Ederra M, Murillo A, Erdozain N, Ales-Martinez JE, Pastor-Barriuso R: **Mammographic density and risk of breast cancer according to tumor characteristics and mode of detection: a Spanish population-based case–control study.** *Breast Cancer Res* 2013, **15:**R9.
15. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, Jong RA, Hislop G, Chiarelli A, Minkin S, Yaffe MJ: **Mammographic density and the risk and detection of breast cancer.** *N Engl J Med* 2007, **356:**227–236.
16. Martin LJ, Boyd NF: **Mammographic density. Potential mechanisms of breast cancer risk associated with mammographic density: hypotheses based on epidemiological evidence.** *Breast Cancer Res* 2008, **10:**201.
17. Eriksson L, Czene K, Rosenberg L, Humphreys K, Hall P: **The influence of mammographic density on breast tumor characteristics.** *Breast Cancer Res Treat* 2012, **134:**859–866.
18. Yaghjyan L, Colditz GA, Collins LC, Schnitt SJ, Rosner B, Vachon C, Tamimi RM: **Mammographic breast density and subsequent risk of breast cancer in postmenopausal women according to tumor characteristics.** *J Natl Cancer Inst* 2011, **103:**1179–1189.
19. Ascunce N, Salas D, Zubizarreta R, Almazan R, Ibanez J, Ederra M: **Cancer screening in Spain.** *Ann Oncol* 2010, **21:**iii43–iii51.
20. American College of Radiology (ACR): *Breast Imaging Reporting and Data System Atlas (BI-RADS®Atlas).* Reston, VA: American College of Radiology (ACR); 2003.
21. Boyd NF, Byng JW, Jong RA, Fishell EK, Little LE, Miller AB, Lockwood GA, Tritchler DL, Yaffe MJ: **Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study.** *J Natl Cancer Inst* 1995, **87:**670–675.
22. Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ: **Mammographic densities and breast cancer risk.** *Cancer Epidemiol Biomarkers Prev* 1998, **7:**1133–1144.
23. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, Fitzgibbons PL, Francis G, Goldstein NS, Hayes M, Hicks DG, Lester S, Love R, Mangu PB, McShane L, Miller K, Osborne CK, Paik S, Perlmutter J, Rhodes A, Sasano H, Schwartz JN, Sweep FC, Taube S, Torlakovic EE, Valenstein P, Viale G, Visscher D, Wheeler T, Williams RB, *et al*: **American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version).** *Arch Pathol Lab Med* 2010, **134:**e48–e72.
24. Wheeler TM, Hayes DF, Van DV, Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, McShane LM, Paik S, Pegram MD, Perez EA, Press MF, Rhodes A, Sturgeon C, Taube SE, Tubbs R, Vance GH: **American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer.** *Arch Pathol Lab Med* 2007, **131:**18–43.
25. Elston CW, Ellis IO: **Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up.** *Histopathology* 1991, **19:**403–410.
26. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ: **Strategies for subtypes–dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011.** *Ann Oncol* 2011, **22:**1736–1747.
27. Agresti A: *An Introduction to Categorical Data Analysis.* 2nd edition. New York: Willey; 2007.
28. Ciatto S, Catarzi S, Lamberini MP, Risso G, Saguatti G, Abbattista T, Martinelli F, Houssami N: **Interval breast cancers in screening: the effect of mammography review method on classification.** *Breast* 2007, **16:**646–652.
29. Rayson D, Payne JI, Abdolell M, Barnes PJ, MacIntosh RF, Foley T, Younis T, Burns A, Caines J: **Comparison of clinical-pathologic characteristics and outcomes of true interval and screen-detected invasive breast cancer among participants of a Canadian breast screening program: a nested case–control study.** *Clin Breast Cancer* 2011, **11:**27–32.
30. Crosier M, Scott D, Wilson RG, Griffiths CD, May FE, Westley BR: **Differences in Ki67 and c-erbB2 expression between screen-detected and true interval breast cancers.** *Clin Cancer Res* 1999, **5:**2682–2688.
31. Payne JI, Caines JS, Gallant J, Foley TJ: **A review of interval breast cancers diagnosed among participants of the Nova Scotia Breast Screening Program.** *Radiology* 2013, **266:**96–103.
32. Vitak B, Olsen KE, Manson JC, Arnesson LG, Stal O: **Tumour characteristics and survival in patients with invasive interval breast cancer classified according to mammographic findings at the latest screening: a comparison of true interval and missed interval cancers.** *Eur Radiol* 1999, **9:**460–469.
33. Conroy SM, Pagano I, Kolonel LN, Maskarinec G: **Mammographic density and hormone receptor expression in breast cancer: the Multiethnic Cohort Study.** *Cancer Epidemiol* 2011, **35:**448–452.
34. Ding J, Warren R, Girling A, Thompson D, Easton D: **Mammographic density, estrogen receptor status and other breast cancer tumor characteristics.** *Breast J* 2010, **16:**279–289.
35. Guo YP, Martin LJ, Hanna W, Banerjee D, Miller N, Fishell E, Khokha R, Boyd NF: **Growth factors and stromal matrix proteins associated with mammographic densities.** *Cancer Epidemiol Biomarkers Prev* 2001, **10:**243–248.
36. Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, Gaudet M, Schmidt MK, Broeks A, Cox A, Fasching PA, Hein R, Spurdle AB, Blows F, Driver K, Flesch-Janys D, Heinz J, Sinn P, Vrieling A, Heikkinen T, Aittomaki K, Heikkila P, Blomqvist C, Lissowska J, Peplonska B, Chanock S, Figueroa J, Brinton L, Hall P, Czene K, *et al*: **Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies.** *J Natl Cancer Inst* 2011, **103:**250–263.
37. Ciatto S, Visioli C, Paci E, Zappa M: **Breast density as a determinant of interval cancer at mammographic screening.** *Br J Cancer* 2004, **90:**393–396.
38. 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, **155:**10–20.

doi:10.1186/bcr3595
Cite this article as: Domingo *et al.*: Tumor phenotype and breast density in distinct categories of interval cancer: results of population-based mammography screening in Spain. *Breast Cancer Research* 2014 **16**:R3.

Capítol 4

Discussió

4.1 Resultats principals

Els resultats dels articles presentats donen suport a la hipòtesi principal. Sobretot, remarquen que les característiques dels programes i de la dona modifiquen tant les taxes de detecció de càncer dins el cribratge com la taxa de CI. A més, presenten característiques diferents tant entre sí com entre els diferents tipus de CI.

En resum, els principals resultats d'aquesta tesi són:

1. El risc acumulat de detectar un càncer en el cribratge durant 10 anys de participar-hi és d'entre 11,11 i 16,71 per cada 1.000 participants, depenent del protocol de cribratge utilitzat.

2. El protocol de cribratge que detecta més, tan invasiu com in situ, utilitza doble lectura i dues projeccions. Les diferències entre protocols a cada participació són petites, però l'efecte acumulat al cap de 10 anys té una gran impacte.

3. Les característiques de les dones són factors de risc tant per a la detecció en el cribratge com per als CI amb el mateix ordre de magnitud.

4. La història familiar de CM és el factor de risc amb major efecte en els CI, mentre que les biòpsies prèvies només és factor de risc per als FN.

5. La presència d'un resultat FP és un factor de risc tan per als càncers detectats en el cribratge com per als CI. Aquesta associació és més forta amb els CI, sobretot amb els FN.

6. La doble lectura, la mamografia digital i els controls avançats estan associats amb la detecció en el cribratge i no amb els CI.

7. Pel que fa a les característiques dels tumors, els veritables intervals i els de signes mínims tenen una distribució dels fenotips similars i presenten una major proporció de fenotips triples negatius que els càncers detectats en el cribratge.

8. La distribució dels fenotips per als FN i els tumors ocults és més semblant a la dels tumors detectats en el cribratge.

9. Les mames denses estan associades als tumors ocults. Els veritables intervals i els FN també mostren una associació amb la densitat mamària de menor magnitud.

10. Els models multi-estat i el model de Cox estimen correctament l'efecte del FP sobre l'aparició del CM.

11. Els models multi-estat permeten estimar les transicions de no tenir càncer a estadi preclínic i d'estadi preclínic a clínic.

4.2 Discussió conjunta

La discussió de cada treball es troba en cada un dels articles, juntament amb les seves limitacions i fortaleses. En aquesta discussió conjunta es respondran els objectius generals de la tesi.

Hem observat que el risc acumulat de detectar varia segons les característiques dels programes. La doble lectura és més sensible que la simple. La doble lectura té un major efecte en la detecció de tumors in situ que en la d'invasius, ja que és més sensible a tumors petits[69].

Altres autors han trobat que la mamografia digital detecta més càncers dins el cribratge, sobretot in situ. Aquest efecte s'ha atribuït a la millora del contrast en la digital[70] i a una major sensibilitat per detectar microcalcificacions[71–73]. Quan hem analitzat els diferents protocols, els que utilitzaven mamògrafs digitals van detectar més tumors, però no hem observat diferències en la detecció de tumors in situ dins el cribratge. Per tant, creiem que la millor estratègia seria la utilització de mamògrafs digitals, la doble lectura i dues projeccions en totes les participacions. No obstant, molts programes encara utilitzen la mamografia analògica o la lectura

simple, ja sigui en totes les participacions o només en les successives.

L'estudi dels determinants del CI ha permès estudiar la relació entre les característiques dels programes o de les dones i la seva aparició. Tenir un FP previ augmenta el risc d'aparició d'un CI. En particular, s'ha observat una associació més forta amb els FN, s'ha observat que alguns dels FP podrien ser, en realitat, FN[74–76]. També hem observat que el FP està associat a una major detecció dins el cribratge, especialment si era un FP invasiu. De les característiques dels programes, no hem observat diferències en la doble lectura en cap tipus de càncer ni per la mamografia digital pels CI. En canvi, la mamografia digital està associada amb un augment dels càncers detectats en el cribratge. De les característiques de la pròpia dona, les dones post-menopàusiques, amb història familiar o biòpsies prèvies tenen un risc major de tenir un CI. És un resultat esperable ja que són factors de risc conegut del CM. La història familiar té una associació més forta amb els verdaders intervals, i reforça la hipòtesis que els tumors en dones amb antecedents familiars creixen més ràpid i són més agressius. Pensem que a les dones amb un FP se les hauria d'incentivar a participar més. A part, s'haurien de repensar les estratègies de cribratge per a les dones amb antecedents familiars que a més tenen algun FP.

Els CI són més grans i tenien major afectació ganglionar que els detectats en el cribratge. Dels subtipus de CI, la majoria de verdaders intervals es diagnostiquen passats 12 mesos des de la última mamografia i tenen un mida més gran que la resta. La majoria de tumors ocults es diagnostiquen durant els primers 12 mesos després de la última mamografia i tenen un patró molecular semblant als càncers detectats en el cribratge.

Un dels factors de risc principals d'aparició de CI és tenir la mama densa, especialment en els tumors ocults. La forta associació entre la densitat mamaria i els tumors ocults és deguda a un efecte d'emmascarament[77]. No s'observa una associació entre la densitat i els FN i confirma que l'emmascarament només afecta en els càncers apareguts fins al cap de 12 mesos després del cribratge[18], normalment tumors ocults.

Els veritables intervals i els signes mínims mostren patrons semblants en l'expressió de biomarcadors i el fenotip dels tumors. Els veritables intervals expressen menys receptors hormonals i una sobreexpressió de HER-2, p53 i Ki67. La majoria de triple negatius són veritables intervals o signes mínims. Diversos articles apunten que les característiques dels signes mínims són similars a les dels FN[78, 79], però en el nostre estudi són més semblants als verdaders intervals.

Hem estudiat la relació entre la densitat mamaria i els fenotips dels tumors. Hem trobat que els càncers luminals tenen més probabilitat de ser detectats en les mames extremadament denses, i que existeix una associació entre la densitat mamaria i els CI, independentment del fenotip. Reforça la hipòtesi que alguns tumors estan estimulats per factors de creixement que es troben en les mames denses[80]. És necessari fer més investigació per estudiar la relació entre la densitat mamaria i els fenotips, segons la via diagnòstica. Conèixer el paper de la densitat mamaria en l'aparició dels subtipus d'interval i la seva relació amb els fenotips serà útil a l'hora de dissenyar una estratègia personalitzada de cribratge.

Finalment, l'estudi de simulació ens ha permès comparar diferents mètodes per a estimar l'efecte del FP sobre l'aparició del CM. Podem concloure que els tres models estudiats estimen correctament l'efecte del FP, però el model amb temps discret té un major biaix, una menor precisió i una menor cobertura que el model multi-estat i el model de Cox. A l'analitzar les dades de l'estudi INCA, hem obtingut tres estimacions diferents, però aquestes diferències no són clínicament rellevants. El model multi-estat ens sembla el més adequat per estudiar l'aparició del CM en dones participants en el cribratge, perquè permet combinar la naturalesa discreta del cribratge amb la història natural del CM.

Dels treballs presentats, es conclou que l'avaluació del cribratge de CM és complexa. Aquesta complexitat és deguda, en els treballs presentats, a la limitació produïda per la impossibilitat d'observar alguns dels esdeveniments d'interès -com, per exemple, no observar quan es produeix el diagnòstic clínic si el tumor es detecta en un examen. Juntament, amb els habituals biaxos del cribratge com el biaix de selecció o de participant sa, de durada (length bias en anglès), o d'avanç de diagnòstic. S'ha de tenir en compte, que milions de dones d'arreu del món participen en programes de detecció precoç i que petits canvis poden tenir un impacte potencialment gran en el conjunt de la població, de manera que petites variacions en l'organització dels programes poblacionals poden tenir efectes importants a nivell poblacional.

4.3 Limitacions

La limitació més gran de la tesi prové del seu disseny de cohort retrospectiva. En una cohort retrospectiva estem limitats a la informació disponible en la font de dades.

El primer article de la tesi la informació prové del projecte RAFP. Tot i que, la informació es va homogeneïtzar, les variables sobre les característiques de les dones (THS, menopausa i antecedents familiars o de biòpsies prèvies) presentaven un tant per cent de valors desconeguts que pot arribar al 40%. A més, els diferents programes no recullen variables relacionades amb el CM com l'edat de menarquia, edat de la maternitat, el nombre de fills, la densitat mamària, etc.

En el segon i tercer articles de la tesi la informació prové del projecte INCA. L'estudi INCA tenia les mateixes limitacions que el RAFP. A més, una altra limitació dels estudis sobre els CI és el possible error de classificació, ja que la definició de CI ve determinada per la periodicitat de les proves de cribratge. Pot haver-hi dones amb un CM detectat en el cribratge amb símptomes i dones amb un CI asimptomàtiques. Aquests errors de classificació poden produir una atenuació de les diferències biològiques i clíniques entre els càncers detectats en el cribratge i els CI. Una altra limitació és l'absència de registres de càncer poblacionals que cobreixin tota la població. En el cas de l'INCA, n'existeixen tan sols a Girona i al País Basc.

També pot haver-hi errors de classificació en tipus de CI, ja que no es van aconseguir totes les mamografies del diagnòstic del CI. A més, no es pot excloure l'error de classificació dels tipus de CI, perquè està subjecte a l'opinió del radiòleg. Aquest error atenua les diferències entre els diferents tipus de CI. A part, el reduït nombre de tumors ocults no ens va permetre fer totes les anàlisis que caldria per manca de potència estadística.

La densitat mamària no està disponible per a tota la població, perquè no es recull i registra sistemàticament. No es va poder incloure en l'anàlisi dels determinants. La densitat mamaria va ser estimada visualment i pot estar afectada per la subjectivitat del radiòleg. Tot i així, es va utilitzar l'escala BI-RADS la més utilitzada per als radiòlegs on només es requereix la interpretació del radiòleg per a classificar bé les dones amb mames heterogèniament denses.

4.4 Fortaleses

Els projectes RAFP i INCA han permès disposar per primera vegada a Espanya de dues bases de dades recollides en diferents programes de cribratge on es té informació de les dones cribrades en les diferents participacions. Els dos disposen de llargs períodes temporals, des de la posada en marxa dels programes fins a finals de la dècada del 2000. El període temporal estudiat fou de 25 anys per al projecte RAFP

i de 9 anys per al projecte INCA.

El primer article és dels pocs que avaluen l'efecte de les variables de protocol radiològic sobre la incidència acumulada de càncer detectat dins el cribratge i el primer en ajustar per diferents variables. El segon article és dels pocs que estudien els determinants dels diferents tipus de CI i el primer en considerar a la vegada els diferents subtipus de CI i densitat.

4.5 Línies futures

Aquesta tesi s'ha realitzat dins de la línia de recerca sobre l'avaluació del cribratge del Servei d'Epidemiologia i Avaluació de l'Hospital del Mar-IMIM i dins la línia d'avaluació de les intervencions sanitàries de la UdL-IRBLleida. S'han demanat i concedit diferents projectes per a la continuïtat de les qüestions presentades, fet que ha permès l'aprofundiment en diferents aspectes del cribratge.

Una de les conclusions dels projectes RAFP i INCA és l'augment de risc de tenir un CM en dones amb un resultat FP previ. L'any 2012 s'inicià el projecte BELE, finançat pel FIS (PI11/01296) on hi participen 8 programes de detecció precoç del CM d'arreu de l'estat espanyol. Té per objectiu aprofundir en el coneixement de les lesions benignes que sovint hi ha darrera els FP i les sospites radiològiques identificades en el cribratge i la seva posterior evolució a càncer. S'espera que els seus resultats permetin adequar el cribratge segons el nivell de risc de desenvolupar un CM. Els primers resultats mostren que el risc de CM és superior en les dones amb lesió benigna prèvia i que el risc augmenta amb el grau de proliferació de la lesió benigna prèvia[81, 82].

Una de les qüestions que han sorgit del projecte INCA és sobre l'evolució clínica dels tumors segons la via diagnòstica. L'any 2013 s'inicià el projecte CAMISS (FIS PI12/00387) per avaluar de forma integrada el conjunt de serveis sanitaris per atendre el CM. Aquest projecte integra des del diagnòstic precoç fins al tractament i seguiment del CM. En la part del cribratge, es recollirà la informació clínica de les dones que formen part de la cohort INCA. Permetrà estudiar el període lliure de malaltia i la supervivència global segons via diagnòstica, els diferents tipus de CI o les característiques clíniques i biològiques dels tumors.

El coneixement dels efectes adversos del cribratge i els factors associats permeten informar millor a les dones de la població diana. S'està desenvolupant el projecte

InforMa (FIS PI14/00113) amb l'objectiu de dissenyar material informatiu i avaluar l'efecte d'informar les dones sobre beneficis i efectes adversos del cribratge.

Actualment, un dels aspectes més discutits del cribratge és la seva estratègia universal, on a totes les dones de 50 a 69 anys se'ls aplica el mateix protocol. S'està plantejant passar a estratègies personalitzades segons les característiques de les dones. La millor manera d'estudiar l'efectivitat de la detecció precoç personalitzada és realitzant assajos clínics. Tot i la complexitat i dificultat en realitzar-los, actualment als EEUU s'està duent a terme l'estudi WISDOM, que té un disseny pragmàtic que combina assaig clínic amb estudi observacional[83]. Una alternativa són els estudis que utilitzen models matemàtics, analítics o la simulació. La informació extreta dels projectes RAFP i INCA s'ha utilitzat per estudiar el cost-efectivitat o la relació dany-benefici que tindria la detecció precoç basada en el risc de CM[60].

Capítol 5

Conclusions

5.1 Conclusions

1. Entre 11 i 16 dones participants de cada 1.000 seran diagnosticades amb un càncer detectat pel programa durant els primers 10 anys de seguiment.

2. Un millor coneixement dels beneficis i efectes adversos del cribratge de CM, com la variabilitat a causa de les diferents estratègies de protocol i de les característiques de la dona, ha de permetre millorar el cribratge i oferir a les dones estratègies millors i més ben informades.

3. Els factors de risc del CI són els mateixos que els dels CM detectats en el cribratge.

4. La relació entre les característiques personals i organitzatives amb el risc de CI permet identificar subgrups de dones amb diferent risc de desenvolupar CM.

5. Gairebé la meitat dels CI van ser veritables intervals, entre els quals hi ha un alt percentatge de tumors amb fenotip associat a un mal pronòstic.

6. La distribució dels fenotips per als FN i els tumors ocults és semblant a la distribució en els CM detectats en el cribratge. Els tumors ocults estan fortament associats a les mames denses.

7. Els models multi-estat permeten modelar la història natural del CM i incloure les participacions de les dones en el cribratge dins el model per avaluar millor l'efecte de les variables en el risc de desenvolupar CM.

5.2 Implicacions per a la salut pública

Els resultats presentats aporten nova informació per a l'avaluació del balanç entre risc i beneficis del cribratge. Aquests resultats estan contribuint a adequar la informació proporcionada a les dones convidades al programa de detecció precoç del CM, poden contribuir a millorar l'efectivitat del cribratge i a entendre millor els determinants del risc de tenir un CM per a poder proposar estratègies personalitzades.

Un dels aspectes controvertit del cribratge és la sensació de tranquil·litat davant d'un resultat negatiu o d'un FP. En el cas del resultat negatiu, aquesta sensació de seguretat pot produir un retard diagnòstic en les dones amb símptomes després d'una prova negativa. S'ha d'informar aquestes dones que existeix la possibilitat de l'aparició d'un CM entre dues proves de cribratge i que han d'estar atentes davant possibles símptomes de malaltia. En el cas de les dones amb un FP, se les ha d'informar que tenen més risc de tenir un CM, animar-les a què continuïn participant en el cribratge i que han d'estar alerta quan aparegui algun símptoma de CM.

La classificació radiològica dels CI aporta informació rellevant per estudiar si els factors de risc són diferents entre tipus de CI. Si bé s'hauria de reduir el nombre de FN, també cal identificar les dones amb més risc de veritable interval. A més, cal promoure la creació de registres de càncer de base poblacional que cobreixin la totalitat de la població catalana i espanyola, ja que són l'eina principal per a l'avaluació del cribratge i per poder conèixer l'efectivitat.

Apèndix A

Altres articles relacionats amb la tesi en el que ha participat el doctorand

A.1 Relació entre el fals-positiu i els càncers detectats en el cribratge

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
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).

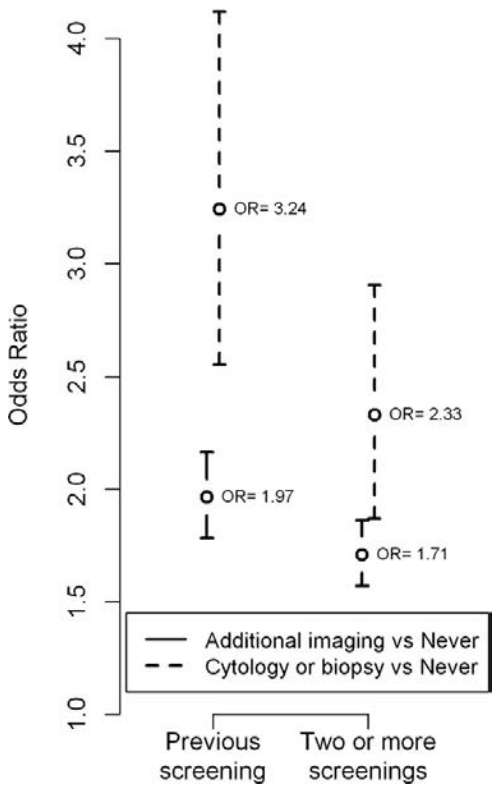


Fig. 1. Adjusted odds ratios (OR) for the cancer detection risk depending on whether the false-positive test occurred in the previous screening round or two or more screenings in advance.

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

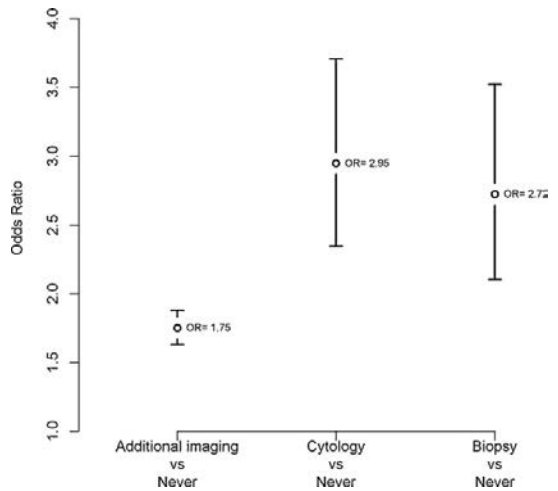


Fig. 2. Adjusted odds ratios (OR) for the cancer detection risk depending on the type of additional procedure leading to the false-positive test.

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.

A.2 Característiques clíniques i radiològiques dels càncers detectats en el cribratge segons el fals-positiu

Cancer Epidemiology 37 (2013) 660–665

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Clinical and radiological features of breast tumors according to history of false-positive results in mammography screening

Laia Domingo [a,b,*], Anabel Romero [a], Jordi Blanch [a,b], Dolores Salas [c], Mar Sánchez [d], Ana Rodríguez-Arana [e], Joana Ferrer [f], Josefa Ibáñez [c], Alfonso Vega [g], M. Soledad Laso [h], Xavier Castells [a,b], Maria Sala [a,b]

[a] Department of Epidemiology and Evaluation, IMIM (Hospital del Mar Research Institute), C/ Dr. Aiguader, 88, 08003 Barcelona, Spain
[b] Red de investigación en servicios de salud en enfermedades crónicas (REDISSEC), Spain
[c] General Directorate Public Health and Centre for Public Health Research (CSISP), Avda. Catalunya, 21, 46002 Valencia, Spain
[d] General Directorate of Public Health, Department of Health, Government of Cantabria, C/ Federico Vial, 13, 39009 Santander, Spain
[e] Radiology and Nuclear Medicine Service, Hospital del Mar, Passeig Marítim 25-29, 08003 Barcelona, Spain
[f] Radiology Unit, Hospital Santa Caterina, C/ Doctor Castany s/n, 17190 Girona, Spain
[g] Radiology Unit, Hospital Universitario Marqués de Valdecilla, Avda. Valdecilla, 25, 39008 Santander, Spain
[h] Breast Cancer Screening Unit Burjassot and Center for Public Health Research (CSISP), C/ Sixto Cámara, 12, 46100 Burjasot, Valencia, Spain

ARTICLE INFO

Article history:
Received 22 March 2013
Received in revised form 22 July 2013
Accepted 23 July 2013
Available online 17 August 2013

Keywords:
Breast neoplasms
Mass screening
False positives
Biopsy
Fine-needle
Carcinoma in situ

ABSTRACT

Background: Women with a false-positive result after a screening mammogram have an increased risk of cancer detection in subsequent participations, especially after assessments involving cytology or biopsy. We aimed to compare women's personal characteristics, tumoral features and the radiological appearance of cancers with and without a previous false-positive result generated by additional imaging or invasive procedures.

Methods: From 1996 to 2007, 111,098 women aged 45–69 years participated in four population-based breast cancer screening programs in Spain, and 1281 cancers were detected. We included all cancers detected in subsequent screenings ($n = 703$) and explored the occurrence of previous false-positive results. We identified false-positives requiring additional imaging or invasive procedures. Differences on tumoral features (invasiveness, tumor size, and lymph node status) and radiological appearance were assessed by Chi-square test, and agreement between the location of cancer and prior suspicious by Cohen's kappa coefficient. A multivariate analysis was preformed to evaluate the effect of previous screening results and age on the odds of presenting an in situ carcinoma.

Results: Among the 703 cancers detected in subsequent screenings, 148 women (21.1%) had a previous false-positive result. Of these, 105 were by additional imaging and 43 by invasive procedures. Women with prior false-positive result requiring invasive assessment, compared to women with negative tests, and women with prior false-positive requiring additional imaging, had a higher proportion of in situ carcinomas (31.7%, 15.3%, 12.9%, respectively; $p = 0.014$) and microcalcifications (37.2%, 20.2%, 9.5%, respectively; $p = 0.003$). The proportion of in situ carcinomas was even higher in women over 60 years (39.2%, 12.5%, 13.0%, respectively; $p = 0.001$). Ipsilateral cancer was observed in 65.7% of cases with prior cytology or biopsy ($k = 0.479$; 95%CI: 0.330–0.794).

Conclusion: A large number of in situ malignancies and calcification patterns were found among women with prior false-positive result in mammography screening requiring cytology or biopsies, suggesting progression from a previously benign lesion.

* Corresponding author at: Department of Epidemiology and Evaluation, IMIM (Hospital del Mar Research Institute), Dr. Aiguader, 88, 08003 Barcelona, Spain. Tel.: +34 933160794; fax: +34 932483496.
E-mail addresses: LDomingo@parcdesalutmar.cat, laia.dt@gmail.com (L. Domingo), arm.anabel@gmail.com (A. Romero), jblanch@parcdesalutmar.cat (J. Blanch), salas_dol@gva.es (D. Salas), sanchez_mm@gobcantabria.es (M. Sánchez), CRC0713@parcdesalutmar.cat (A. Rodríguez-Arana), joana_ferrer@ias.scs.es (J. Ferrer), ibanyez_jos@gva.es (J. Ibáñez), avegab@telefonica.net (A. Vega), laso_sol@gva.es (M.S. Laso), Xcastells@parcdesalutmar.cat (X. Castells), Msalaserra@parcdesalutmar.cat (M. Sala).

1877-7821/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.canep.2013.07.006

1. Introduction

In breast cancer screening programs, most women with a screen-detected abnormality will not be diagnosed with cancer after being recalled for further assessment [1]. This further assessment may imply additional imaging tests or invasive procedures, which will usually rule out a malignancy and will therefore generate a false-positive result. However, women with false-positive results have been reported to be at higher risk for cancer detection in subsequent screening rounds [2–5]. Although this association seems consistent from a statistical point of view, a clear explanation is lacking and few studies have specifically evaluated this relationship.

The high risk of breast cancer after a false-positive result may be partly explained by the false-negative hypothesis (that is, cancers missed after further assessments in the screening that are diagnosed at the next screening) [6]. Another explanation is that women with benign breast disease have a greater risk of developing invasive breast carcinoma [7,8]. Specific information on imaging and pathological features might help to improve understanding of this event. Nonetheless, to date, few studies have evaluated information on tumoral characteristics, radiological appearance, and the location of the two lesions [4,9,10], and none have distinguished between a false-positive involving invasive procedures or just additional imaging.

This study aimed to compare women's personal characteristics, tumoral features and their radiological appearance among patients with and without a previous false-positive result generated by additional imaging or invasive procedures, and to assess the agreement between the location of the false-positive lesion and that of the subsequent malignant tumor.

2. Materials and methods

2.1. Settings

Data were obtained from five radiology units from four different population-based breast cancer screening programs in Spain (Cantabria, Barcelona, Girona and Valencia) between 1996 and 2007. More detailed information on screening in the programs involved in the study and on database construction has been previously reported [11]. Briefly, women in the target population, aged between 45 and 69 years, receive information on screening and are invited to undergo mammography at 2-yearly intervals. All programs are based on the European guidelines for quality assurance in screening mammography, and their results meet the required standard [12,13]. Since 2007, the programs have obtained two views (mediolateral oblique and craniocaudal). Previously, a single view was obtained for subsequent screenings in one program. Reading methods were single reading in one program, double reading with consensus in two programs, and double reading with arbitration in one program. All radiology units began their screening activities between 1996 and 1998 using screen-film radiographic technology and switched to full-field digital mammography between September 2004 and January 2005.

This study was approved by the Ethics Committee of our institution. Informed consent was not required.

2.2. Study population

Women participating in at least two screening rounds (thus having the chance of a false-positive result) and diagnosed with cancer in the screening process were included in this study. Of 291,218 mammograms performed during the study period, 1281 cancers were detected. Of these 1281 cancers, 578 were detected at the first screening and were therefore excluded from the analysis. A total of 703 women with cancer were included in the analysis. These women had a median of three screening participations (interquartile range (IQR): 2–4). Interval cancers – i.e. cancers manifesting clinically between two screening mammograms – were not included in the study.

2.3. Screening results: definition of a false positive results and cancer diagnosis

Within the screening program, women with a negative mammographic reading (equivalent to BI-RADS score of 1 and 2) are recalled for a new screening mammogram at 24 months. Women with a positive mammographic reading (equivalent to a BI-RADS score of 3, 4, 5 and 0) are immediately recalled for further assessments to exclude malignancy within the screening program. The diagnostic work-up for additional evaluation takes place within a maximum of 2 months after screening.

A positive result in the screening test was considered a false-positive result if, after immediate further assessments – which can include invasive or non-invasive procedures – cancer was not diagnosed. Non-invasive procedures may include additional mammography, magnetic resonance imaging, or ultrasonography; invasive procedures encompass fine-needle aspiration cytology, core-biopsy, or open biopsy. A definitive diagnosis of breast cancer was always histopathologically confirmed. If cancer is ruled out after the additional evaluation, women are routinely invited to participate again 24 months after the initial screening test.

2.4. Study variables

Women's personal characteristics (age, familial history of malignant breast disease, or menopausal status) were collected at each woman's attendance through a questionnaire. Tumor-related characteristics (invasiveness, size, lymph node status) were identified from screening databases.

Information on location and radiological patterns was collected for both the false-positive episode and the cancer diagnosis from the radiological reports. There were three possible locations: left breast, right breast or both breasts. Radiological patterns were classified into masses, distortions, calcifications, asymmetries, and multiple patterns.

2.5. Statistical analysis

Women were classified according to whether they had a false-positive result or not. Among women with a previous false-positive result, we differentiated between those that involved a non-invasive or an invasive procedure. We calculated Pearson's χ^2 test to compare women's personal characteristics and tumoral features between study groups. An analysis stratified by age was performed to assess the percentage of in situ carcinomas among study groups. For the purpose of this subanalysis, age was collapsed into two categories: <60 years old, and ≥ 60 years old. A logistic regression analysis was performed to examine the association between previous screening results on the odds of presenting an in situ carcinoma, adjusting by age. The potential interaction between age and screening results was tested in the logistic regression model. The interaction term was not statistically significant.

We conducted a sensitivity analysis to evaluate the inclusion of false-positive results in the previous screening round or in two or more screening rounds before the cancer diagnosis. The analysis confirmed equivalent results regarding women's personal characteristics and tumoral features. Consequently, in the above-mentioned analyses we included any false-positive result reported before the cancer diagnosis to enhance the study power.

A concordance analysis was performed to evaluate the degree of agreement between the breast where the suspicious lesion leading to false positive was observed and the breast where cancer was subsequently detected. For this analysis, only 118 women with information about the location of the lesions identified in the false-positive result and in the screening-detected cancer were included. This analysis was carried out using contingency tables and Cohen's kappa coefficients (κ), and their 95% confidence intervals were calculated. All the statistical analyses were performed with the statistical software package SPSS, version 12.0. Graphics were generated using R statistical software (version 2.12.2).

3. Results

Among the 703 cancers detected in subsequent screening rounds, 148 women (21.1%) had a previous false-positive result. Of these 148 women, 84 (56.8%) experienced a false-positive result in the previous round, while the remaining 43.2% did so in previous episodes (Fig. 1).

We compared women's personal characteristics and tumoral features among women with previous negative screening tests, women who had a false-positive result for an additional imaging test, and women with a previous false-positive result involving invasive procedures. The distribution of women's characteristics (Table 1) showed no statistically significant differences among the study groups. Cancer diagnoses among women with a false-positive after additional imaging tended to occur at earlier ages compared with women with negative screening tests and women with a prior invasive procedure ($p = 0.084$).

Tumor-related features (Table 2) showed some differences among study groups. We detected a higher percentage of ductal carcinoma in situ among women with previous false-positive results after an invasive procedure compared with the remaining study groups (31.7% vs. 15.3% and 12.87%; $p = 0.014$), as well as a higher percentage of invasive tumors of more than 20 mm. Women with prior negative tests had the highest percentage of invasive tumors smaller than 10 mm (32.9% vs. 25.4% and 20.8%; $p = 0.030$). The percentage of invasive cancers with lymph node involvement was similar among the study groups ($p = 0.755$).

Radiological patterns reported at diagnosis showed statistically significant differences among the study groups ($p = 0.003$). Masses were the most frequent radiological pattern in women with a previous negative screening mammogram and women with previous additional imaging procedures (32.7% and 38.1%, respectively, $p = 0.003$). In women with previous invasive procedures, the most frequent pattern (37.2%) was calcifications. The percentage of distortions was higher among women with previous negative tests than in the remaining participants (16.8%, 11.4%, and 7.0%, respectively).

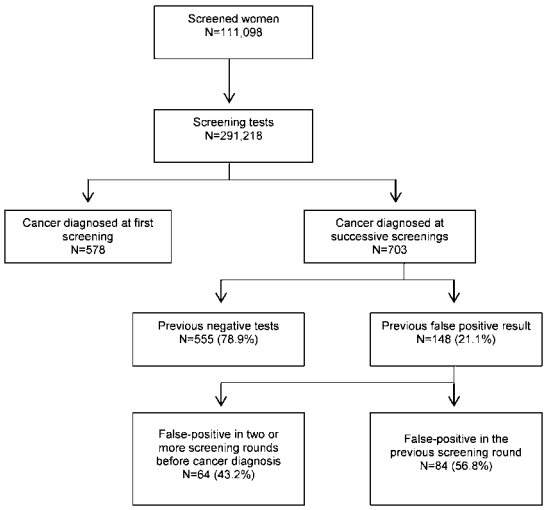


Fig. 1. Flowchart of the study.

The component bar graph in Fig. 2 shows the percentage of in situ carcinomas stratified by age, according to the results of previous screening mammograms. The data show a percentage of in situ carcinomas of 39.2% among elderly women $\geq$60 years with prior invasive procedures in previous screenings compared with 12.5% and 13.0% in the same age group with previous negative screening or previous additional imaging procedures ($p = 0.001$). For women under 60 years, the percentage of in situ cancers remained below 20% ($p = 0.493$).

The results from logistic regression analysis are shown in Table 3. Women with previous invasive procedure were more likely to present an in situ carcinoma, even after the adjustment for age [OR = 2.65 (1.32–5.35)].

A total of 57.8% of tumors in women with a previous false-positive result after an additional imaging procedure were located

Table 1
Personal characteristics in women with a prior false-positive result, by additional imaging or invasive procedures.

	Previous negative screening tests		Previous additional imaging procedures		Previous invasive procedures		*p*-Value
	N	%	*n*	%	*n*	%	
Total	555		105		43		
Age groups							
<55	94	16.9	26	24.8	7	16.3	
[55–60)	131	23.6	33	31.4	7	16.3	
[60–65)	170	30.6	25	23.8	15	34.9	
[65–70]	160	28.8	21	20.0	14	32.6	0.084
Family history of breast cancer							
Yes	57	10.3	8	7.6	4	9.3	
No	496	89.7	97	92.4	39	90.7	0.693
Unknown	2		0		0		
Menopausal status							
Post menopausal	498	91.7	90	87.4	37	92.5	0.348
Pre menopausal	45	8.3	13	12.6	3	7.5	
Unknown	12		2		3		

Table 2
Tumoral features in women with a previous false-positive result (by additional imaging or invasive procedures).

	Previous negative screening tests		Previous additional imaging procedures		Previous invasive procedures		*p*-Value
	n	% (IQR)[a]	*n*	% (IQR)[a]	*n*	% (IQR)[a]	
Total	555		105		43		
Median time between lesions[b]			2.52	(2.00–4.09)	3.98	(1.99–5.99)	
Women <60 years			2.80	(2.01–4.08)	2.11	(1.93–3.88)	
Women ≥60 years			2.45	(1.99–4.28)	4.03	(2.00–6.04)	
Invasiveness							
Ductal carcinoma in situ	83	15.3	13	12.9	13	31.7	0.014
Invasive cancer	459	84.7	88	87.1	28	68.3	
Unknown	13		4		2		
Tumor size[c]							
<10 mm	125	32.9	18	25.4	5	20.8	0.030
10–20 mm	177	46.6	33	46.5	10	41.7	
20–50 mm	64	16.8	18	25.4	5	20.8	
>50 mm	14	3.7	2	2.8	4	16.7	
Unknown	92		21		6		
Lymph nodes[c]							
Negative	287	75.5	57	79.2	19	79.2	0.755
Positive	93	24.5	15	20.8	5	20.8	
Unknown	92		21		6		
Radiological patterns							
Masses	177	32.7	40	38.1	13	30.2	0.003
Distortions	91	16.8	12	11.4	3	7.0	
Calcifications	109	20.2	10	9.5	16	37.2	
Asymmetry	25	4.6	6	5.7	0	0.0	
Multiple patterns	139	25.7	37	35.2	11	25.6	
Unknown	14		0		0		

[a] IQR: interquartile range.
[b] Median time is expressed in years.
[c] Only invasive cancers.

Table 3
Association of previous screening results and age with breast cancer histology.

	Ductal carcinoma in situ	
	Unadjusted OR (95%CI)	Adjusted OR (95%CI)
Previous screening results		
Previous negative screening tests	Ref.	Ref.
Previous additional imaging procedures	0.82 (0.44–1.53)	0.78 (0.42–1.46)
Previous invasive procedures	2.57 (1.28–5.16)	2.65 (1.32–5.35)
Age at cancer detection		
<60 years	Ref.	Ref.
>=60 years	0.77 (0.51–1.16)	0.73 (0.48–1.11)

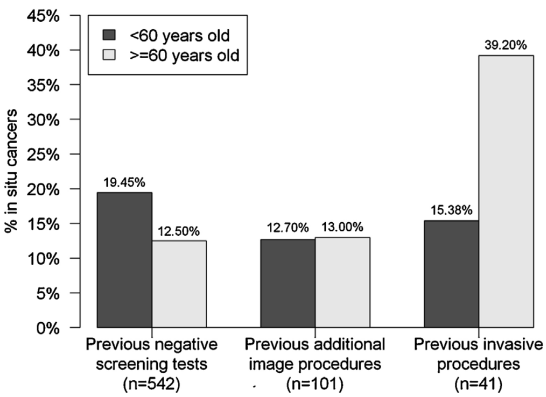


Fig. 2. Bar graph showing the percentage of in situ carcinomas by age group, and by the results of previous screening tests.

in the ipsilateral breast where the first suspicious lesion was detected, showing weak agreement (kappa index of 0.357; 95% confidence interval: 0.195–0.519). However, the percentage of agreement was higher (65.7%) when the previous false-positive occurred after an invasive procedure (kappa index of 0.479; 95% confidence interval: 0.330–0.794).

4. Discussion

Our data reveal that the percentage of in situ cancers is higher among women with a previous invasive assessment that excluded a malignancy, especially among women older than 60 years. Cancers detected after an invasive assessment were also associated with a higher percentage of microcalcifications and two out of three cancers were detected in the same breast where the first suspicious lesion was assessed. Tumors in women with a previous assessment with an additional imaging procedure were similar to those detected among women with prior negative tests, and the agreement of the laterality of both lesions was weaker.

Several studies have reported that women with false-positive results are at higher risk for cancer detection in subsequent screening rounds [2–5]. In our context, the risk dramatically increases after assessments involving cytology or biopsy [3]. The exclusion of cancer after an invasive procedure does not indicate the absence of a benign abnormality. Given a high index of suspicion, most biopsies and cytological analyses have a final diagnosis of a benign breast lesion [14]. The high percentages of in situ carcinomas found among women with prior invasive procedures, reinforced by the association observed in the multivariate model considering age, could reflect progression of a benign breast lesion that subsequently became cancerous. This model was originally based on evidence of gradual histological continuity over long periods, from hyperplasia, atypical hyperplasia, ductal carcinoma in situ and invasive breast cancer, through the random accumulation of genetic mutations [15]. Although this model of linear progression may be conceptually flawed [16], benign breast disease is known to be an important risk factor for breast cancer, for at least 20 years after the first benign breast biopsy [7,17]. This hypothesis is supported by the median time between lesions of almost 4 years. Moreover, it also seems consistent that elderly women were more likely to have a higher percentage of in situ cancers, since tumor progression is slower at older ages [18].

There is evidence that the elevated risks associated with benign breast lesions and subsequent malignant lesions are bilateral, suggesting that they may be both risk factors and precursors [19]. Hartmann et al. [7] described an excess of ipsilateral cancer during the first 10 years, which may explain the moderate agreement between the location of cancer among women with prior invasive procedures, and the location of the prior suspicious lesion. Unfortunately, more specific data on the location of both lesions, along with histological information from the cytology and biopsy, was not available to help us to better understand the biological mechanisms underlying the relationship between false-positives and cancer.

The comparison of women-related and tumor-related characteristics of patients with prior additional imaging pointed to slight differences with patients with prior negative tests. Women with prior additional imaging tended to be younger at diagnosis, and the proportion of invasive and larger tumors was greater. Previous studies showing an association between younger ages and false-negative assessments have related such events to the high proportion of dense breasts among young women [20]. These findings are consistent with studies that have hypothesized that some false-positive results might, in fact, be false-negative results [5,6,9,10,20,21] resulting in a delay in diagnosis, which could explain the higher proportion of larger invasive tumors. Notably, the above-mentioned studies included cancers that appeared before 24 months and, given that tumoral growth is time-dependent, these results are not fully comparable with our own, which are restricted to tumors detected at least 24 months after the last screening. In addition, the high recall rates reported in Spain, especially those involving additional imaging procedures [1], could lead to a less selected population, with a greater number of true negative tests, which could attenuate differences when compared with those in with women with previous negative results.

If the false-negative hypothesis could explain most cases with a previous false-positive result, high agreement for laterality would be expected. Our results showed a weak-to-moderate agreement for laterality, the highest agreement being observed in patients with previous invasive procedures. The agreement would probably be lower if we took specific breast quadrants into account.

Regarding radiological appearance, cancers with prior additional imaging were more likely to present as masses than cancers with prior negative assessments. Masses are one of the most common patterns found in cancers with previous false-negative assessment [6,10]. Calcifications and distortions, which are the two radiological patterns with the highest positive predictive value in our setting [22], accounted for a small proportion of tumors with prior additional imaging. Nevertheless, a high proportion of calcifications was found among patients with previous invasive procedures, who also showed higher agreement for laterality. This finding is in agreement with studies showing a relationship between calcifications and in situ carcinomas [23,24]. Baker et al. [24] indicate that benign tissue calcifications are likely to lead to an in situ carcinoma, which is an immediate precursor of invasive breast cancer.

Our study has some limitations. Firstly, even though the data were drawn from a project including more than 240,000 screening tests, the sample size was relatively small, since the occurrence of cancer after a false-positive result is fairly infrequent. Secondly, some important variables were not routinely collected by breast cancer screening program databases. Information on breast density was not available and consequently its potential as a confounder could not be tested. However, Barlow et al. [2] suggested that the association of a previous false-positive and cancer is independent of this variable. Other important variables not included in the databases were histopathological data from the biopsies, and the location of the lesion within each breast. The lack of more specific information, such as quadrants, might have overestimated the concordance between the false-positive result and cancer detection. Breast Imaging-Reporting and Data System information of the mammograms was not included in the analysis since some of the programs participating in the study report this data in a different way. Third, interval cancers were not included in this study and some of the lesions identified as false-positives could also progress to an interval cancer [5,25]. Future studies should concentrate on analysis of these cancers, which could aid comprehension of cancers that appear shortly after a false-positive assessment.

To date, this is the first work providing data on tumoral features that also analyzes the history of false-positive screening results and the type of false-positive results. Our results provide new insights into the association between a false-positive result and subsequent cancer detection within a screening program, which can probably be explained by a combination of factors including false-negative results and progression of benign lesions.

In conclusion, the high proportion of in situ malignancies and calcification patterns found among women with a previous false-positive screening result involving invasive procedures, seems to indicate that certain cancers could develop from previously benign lesions. This information reinforces the importance of encouraging women with false-positives results to attend regular screening. The study of cohorts of screened women with exhaustive data on screening episodes would be very useful to improve the knowledge of the natural history of breast cancer.

Conflict of interest statement

The authors declare that they have no conflict of interest.

Acknowledgments

We are grateful to Sonia Sánchez, Paula Merino, Mercè Comas, Marta Román, and Cristina Hernández for their contribution to data collection from the screening programmes. This study was supported by grants from the Instituto de Salud Carlos III FEDER (PI07/90293).

References

[1] 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.
[2] 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;98(17):1204–14.
[3] Castells X, Roman M, Romero A, Blanch J, Zubizarreta R, Ascunce N, et al. Breast cancer detection risk in screening mammography after a false-positive result. Cancer Epidemiol 2013;37(1):85–90.
[4] 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;104(9):682–9.
[5] 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.
[6] Meeson S, Young KC, Wallis MG, Cooke J, Cummin A, Ramsdale ML. Image features of true positive and false negative cancers in screening mammograms. Br J Radiol 2003;76(901):13–21.
[7] 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;353(3):229–37.
[8] Zhou WB, Xue DQ, Liu XA, Ding Q, Wang S. The influence of family history and histological stratification on breast cancer risk in women with benign breast disease: a meta-analysis. J Cancer Res Clin Oncol 2011;137(7):1053–60.
[9] Allgood PC, Duffy SW, Warren R, Hunnam G. Audit of negative assessments in a breast-screening programme in women who later develop breast cancer-implications for survival. Breast 2006;15(4):503–9.
[10] 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;105(1):37–43.
[11] Sala M, Salas D, Belvis F, Sanchez M, Ferrer J, Ibanez J, et al. Reduction in false-positive results after introduction of digital mammography: analysis from four population-based breast cancer screening programs in Spain. Radiology 2011;258(2):388–95.
[12] Ascunce N, Salas D, Zubizarreta R, Almazan R, Ibanez J, Ederra M. Cancer screening in Spain. Ann Oncol 2010;21(Suppl. 3):iii43–51.
[13] Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L, eds. European guidelines for quality assurance in breast cancer screening and diagnosis. Luxenburg: Office for Official Publications of the European Communities, 2006.
[14] 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;21(6):821–8.
[15] Wellings SR, Jensen HM. On the origin and progression of ductal carcinoma in the human breast. J Natl Cancer Inst 1973;50(5):1111–8.
[16] Lopez-Garcia MA, Geyer FC, Lacroix-Triki M, Marchio C, Reis-Filho JS. Breast cancer precursors revisited: molecular features and progression pathways. Histopathology 2010;57(2):171–92.
[17] Ashbeck EL, Rosenberg RD, Stauber PM, Key CR. Benign breast biopsy diagnosis and subsequent risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2007;16(3):467–72.
[18] Tabar L, Fagerberg G, Chen HH, Duffy SW, Smart CR, Gad A, et al. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. Cancer 1995;75(10):2507–17.
[19] Allred DC, Mohsin SK, Fuqua SA. Histological and biological evolution of human premalignant breast disease. Endocr Relat Cancer 2001;8(1):47–61.
[20] Murphy IG, Dillon MF, Doherty AO, McDermott EW, Kelly G, O'Higgins N, et al. Analysis of patients with false negative mammography and symptomatic breast carcinoma. J Surg Oncol 2007;96(6):457–63.
[21] Duijm LE, Groenewoud JH, de Koning HJ, Coebergh JW, van Beek M, Hooijen MJ, et al. Delayed diagnosis of breast cancer in women recalled for suspicious screening mammography. Eur J Cancer 2009;45(5):774–81.
[22] Domingo L, Romero A, Belvis F, Sanchez M, Ferrer J, Salas D, et al. 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 2011;21(9):2020–8.
[23] Allred DC. Ductal carcinoma in situ: terminology, classification, and natural history. J Natl Cancer Inst Monogr 2010;2010(41):134–8.
[24] Baker R, Rogers KD, Shepherd N, Stone N. New relationships between breast microcalcifications and cancer. Br J Cancer 2010;103(7):1034–9.
[25] Bennett ML, Welman CJ, Celliers LM. How reassuring is a normal breast ultrasound in assessment of a screen-detected mammographic abnormality? A review of interval cancers after assessment that included ultrasound evaluation. Clin Radiol 2011;66(10):928–39.

A.3 Supervivència del càncer de mama segons via diagnòstica

Aggressiveness features and outcomes of true interval cancers: comparison between screen-detected and symptom-detected cancers

Laia Domingo[a,d,e], Jordi Blanch[a,d], Sònia Servitja[b], Josep Maria Corominas[c], Cristiane Murta-Nascimento[a], Antonio Rueda[f], Maximino Redondo[d,g], Xavier Castells[a,d] and Maria Sala[a,d,e]

The question of whether screen detection confers an additional survival benefit in breast cancer is unclear and subject to several biases. Our aim was to examine the role of the diagnostic method (screen-detected, symptom-detected, and true interval cancers) and the clinical–pathological features in relapse-free survival and overall survival in breast cancer patients. We included 228 invasive breast cancers diagnosed in Barcelona from 1996 to 2008 among women aged 50–69 years. Ninety-seven patients were screen detected within the screening, 34 truly arose between 2-year screening mammograms (true interval cancers), and 97 were symptom detected outside the screening. The clinical–pathological features at diagnosis were compared. The overall and disease-free survival probabilities were computed using the Kaplan–Meier method. Cox proportional hazard models were applied, with adjustment by clinical–pathological variables. At diagnosis, symptom-detected and true interval cancers were in more advanced stages and were less differentiated. The highest proportion of triple-negative cancers was detected among true interval cancers ($P=0.002$). At 5 years of follow-up, the disease-free survival rates for screen-detected, true interval, and symptom-detected cancers were 87.5% (95% confidence interval, 80.5–95.2%), 64.1% (46.4–88.5%), and 79.4% (71.0–88.8%), respectively, and the overall survival rates were 94.5% (89.3–99.9%), 65.5% (47.1–91.2%), and 85.6% (78.3–93.6%), respectively. True interval cancers had the highest hazard ratio for relapse prediction (1.89; 0.67–5.31) and a hazard ratio of death of 5.55 (1.61–19.15) after adjustment for tumor–node–metastasis stage and phenotype. Clinically detected tumors, especially true interval cancers, more frequently showed biological features related to worse prognosis and were associated with poorer survival even after adjustment for clinical–pathological characteristics. *European Journal of Cancer Prevention* 22:21–28 © 2012 Wolters Kluwer Health | Lippincott Williams & Wilkins.

European Journal of Cancer Prevention 2013, **22**:21–28

Keywords: breast cancer screening, interval cancer, phenotype, prognosis, survival

[a]Epidemiology and Evaluation Department, [b]Medical Oncology Department, [c]Pathology Department, Hospital del Mar-IMIM, [d]CIBER Epidemiology and Public Health (CIBERESP), [e]Department of Pediatrics, Obstetrics and Gynecology, EHEA Doctoral Program in Public Health, Preventive Medicine and Public Health, Autonomous University of Barcelona (UAB), Barcelona, [f]Medical Oncology Department and [g]Department of Biochemistry, Hospital Costa del Sol, University of Malaga, Málaga, Spain

Correspondence to Laia Domingo, MSc, Epidemiology and Evaluation Department, Hospital del Mar-IMIM, MAR Health Park, Pg. Maritim 25-29, 08003 Barcelona, Spain
Tel: +34 933 160 794; fax: +34 932 483 496;
e-mail: ldomingo@parcdesalutmar.cat

Received 22 February 2012 **Accepted** 6 April 2012

Introduction

In the last two decades, breast cancer survival has improved. Reduced mortality from this disease has been attributed to a combination of widespread mammographic screening, the introduction of effective systemic treatment modalities, and progress in radiotherapy and surgery (Berry *et al*., 2005; Benson *et al*., 2009; Autier *et al*., 2010). However, the contribution of screening to this reduction is unclear, as outcomes may be affected by several biases such as lead-time bias (screen-detected cancers are detected earlier in their natural course than those found outside of screening), length bias (screening tends to detect slow-growing tumors, which spend longer time in the asymptomatic phase), and selection bias (the screened population is not representative of the general population).

Previous works that have focused on breast cancer survival and detection mode have reported that screen detection confers an additional survival benefit beyond the stage shift (Shen *et al*., 2005; Wishart *et al*., 2008) and reduces the risk of systemic recurrence compared with symptomatic cancers at a similar stage (Joensuu *et al*., 2004). However, part of this benefit remains unexplained as it is difficult to obtain molecular information from a large series of tumors for inclusion in statistical models (Vitak *et al*., 1997; Immonen-Räihä *et al*., 2005; Shen *et al*., 2005; Bordas *et al*., 2007; Zackrisson *et al*., 2007; Dawson *et al*., 2009; Lawrence *et al*., 2009).

It is already known that tumors detected during screening are related to clinical–pathological features with a better prognosis, such as low grade or hormone-receptor

0959-8278 © 2012 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/CEJ.0b013e328354d324

expression, compared with those detected by other means (i.e. outside of screening or as interval cancers; Bordas *et al*., 2007). Wider differences have been described between screen-detected and interval cancers (tumors that manifest clinically between a normal screening result and the following invitation for screening), which are associated with a delay in diagnosis that could potentially worsen prognosis and consequently impair the program's efficiency. Nevertheless, the greatest differences have been found between screen-detected cancers and true interval cancers (tumors with a short preclinical phase and no suspicious finding after radiological review of the previous mammogram). Among this subgroup, increased tumor cell proliferation (Vitak *et al*., 1997) and a higher occurrence of triple-negative tumors [tumors that lack expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2)] have been reported (Domingo *et al*., 2010; Rayson *et al*., 2011). To date, this biomarker pattern lacks the benefit of specific adjuvant therapy and is associated with a poor prognosis (Chacón and Costanzo, 2010).

Although differences at diagnosis have been reported between screen-detected and true interval cancers, only two studies (Vitak *et al*., 1997; Rayson *et al*., 2011) have taken this latter group into consideration when evaluating clinical outcomes, and both have reported significantly poorer survival. However, neither of these studies included information on phenotype and a group of women who did not participate in the screening.

The aim of the present study was to evaluate the role of the diagnostic method and biological characteristics in relapse-free survival and overall survival in breast cancer patients.

Methods

Setting

The study was carried out among women diagnosed with breast cancer in a reference hospital in the city of Barcelona (Spain) between 1995 and 2008. This center is publicly funded, serves an area with around 300 000 inhabitants, and has run a population-based breast cancer screening program since 1995.

The screening program's target population ($\sim$90 000 women) is women aged 50–69 years, and involves women who are invited to undergo mammography every 2 years. Mediolateral, oblique, and craniocaudal views are available for each breast. All mammograms are read by two radiologists, and when double reading leads to different assessments, a third radiologist serves as a tiebreaker. The program is based on the European Guidelines for Quality Assurance in Mammographic Screening (Perry *et al*., 2006) and its results meet the Europe Against Cancer standards, although the participation rate barely reached the standard level of 70%.

This study was approved by the Ethics Committee and informed consent was obtained to supply tumor biopsy material for pathologic evaluation.

Study population

A total of 1432 patients with breast cancer aged 50–69 years were identified from the hospital-based cancer registry during the study period. Of these cancer cases studied, 740 were detected in women attending the breast cancer screening program (screen-detected cancers), 98 emerged as interval cancers between screening mammograms, and 594 cancers were diagnosed among women who did not participate in the screening program (symptom-detected cancers) and who were referred with breast abnormalities, typically palpable lesions, by their primary care physicians or were self-referred to the hospital.

Interval cancers were retrieved by merging the mammography register with the hospital-based cancer registry; in this process, we identified 98 interval cancers diagnosed and treated in our setting. For these, we aimed to perform a radiological review of both screening and diagnostic mammograms to classify them into the five subtypes, following European Guideline recommendations (Perry *et al*., 2006). Both mammograms were available in 80 cases, and two experienced radiologists performed the retrospective review. Screening mammograms were first reviewed independently, and in a second review, the radiologists assessed screening and diagnostic mammograms with the histological information to determine whether any abnormalities detected on the screening films corresponded to the site of the subsequent interval cancer. Interval breast cancers were definitively classified into five subtypes: true interval cancer ($n = 34$), false negative ($n = 13$), occult tumors ($n = 14$), minimal signs ($n = 4$), and unclassifiable ($n = 15$). A complete description of the interval cancer identification process has been published previously (Domingo *et al*., 2010).

Because of difficulties in obtaining retrospective information on such a large number of cases, a random sample of screen-detected ($n = 97$) and symptom-detected cancers ($n = 97$) was selected. These two samples were compared with all true interval cancers ($n = 34$). In-situ cancers and patients with stage IV breast cancer at diagnosis were excluded from the analyses.

Data collection

Tumor-related data [pathological tumor–node–metastasis (TNM) status, histological type, histological grade], patient-related data (age), and vital status at the end of follow-up were obtained from the hospital-based cancer registry. Data on treatment, recurrences, and immunohistochemical information (ER, PR, HER, and p53 status) were obtained from the registry of pathology, if available, and from review of clinical records. Further immunohistochemical analyses were carried out in tumor

samples in which biomarker expression had not been determined previously. Information on adjuvant treatment (radiotherapy, chemotherapy, hormonotherapy, and targeted therapies) and follow-up was obtained from the clinical records.

Laboratory methods

ER, PR, p53, and HER2 are routinely determined during the diagnostic process in our hospital. Samples lacking information on biomarker expression were analyzed following the same procedures as those used in clinical practice. Immunohistochemical staining was performed on paraffin block sections with the tissue specimens fixed in 10% neutral-buffered formalin for 24 h [ER clone ID5, 1 : 50 (Dako, Glostrup, Denmark); PR clone PgR636, 1 : 200 (Dako); p53 clone DO7, 1 : 50 (Novocastra Laboratories, Newcastle, UK); for Her2/neu protein overexpression (HercepTest; Dako)]. In accordance with standard guidelines, positivity for ER, PR, and p53 was based on more than 10% of the cells testing positive. For HER2, scores of 0–1 in the HerceptTest kit were considered negative, whereas a HerceptTest score of 3 was considered positive. Equivocal scores of 2 were confirmed by fluorescent in-situ hybridization as positive when *HER2/neu* oncogene amplification was detected.

On the basis of the expression of ER, PR, and HER2, the tumors were classified into four phenotypes: (a) luminal A (ER +, PR +, HER–); (b) luminal B (ER or PR +, HER– or ER +, PR +/–, HER +); (c) HER2 (ER–, PR–, HER2 +); and (d) triple negative (ER–, PR–, HER–) (Perou *et al.*, 2000; Cheang *et al.*, 2009).

Follow-up of cancer cases

Locoregional recurrence was defined as disease recurrence within the ipsilateral breast or chest wall, in the ipsilateral axillary nodes, internal mammary nodes, or supraclavicular nodes. Distant recurrence was defined as disease recurrence in sites other than the breast or regional lymph nodes (bone, skin, or visceral metastasis).

Disease-free survival was defined as the time from diagnosis to the first occurrence of one or more of the following: a local or regional recurrence, cancer in the contralateral breast, distant metastasis, and second primary carcinoma, whichever occurred first. Overall survival was defined from the date of diagnosis to death from any cause.

When disease-free survival was computed, women lost to follow-up or those who died were censored either at last visit or at death. For overall survival, patients were censored at the date of their last hospital visit. The median follow-up period was 5.13 years.

Statistical analysis

Contingency tables were calculated to compare possible differences in patient and histopathological characteristics among the study groups. Statistical significance was assessed using χ^2-tests. Survival curves were generated using the Kaplan–Meier method and were compared by the log-rank test. Kaplan–Meier estimates of 5-year disease-free and overall survival rates after diagnosis were computed with 95% confidence intervals (CI).

Cox proportional hazard regression analyses were carried out to evaluate survival differences between screen-detected, true interval, and symptom-detected cancers, controlling for known prognostic and predictive factors such as age, TNM stage, and phenotype in an attempt to control lead time and length biases related to screening. Unadjusted and adjusted hazard ratios and 95% CI were computed and are shown for our main variable of interest (detection method). We computed a baseline regression model that included only the detection method, and gradually, all other study variables were added to control for their potential effect on survival times. The proportional hazards assumption was ascertained by assessment of log–log survival plots. All calculations were carried out using the statistical software SPSS, version 12.0 (SPSS Inc., Chicago, Illinois, USA) and R, version 2.12.2 (R Development Core Team, 2011). All *P*-values were two-sided and values less than 0.05 were considered statistically significant.

Results

The analyses included 228 patients with breast cancer, of which 97 were screen-detected, 34 were true interval cancers, and 97 were cancers detected symptomatically outside of the screening program. The clinical and pathological characteristics for the three detection groups are shown in Table 1. At diagnosis, clinically detected tumors (symptom-detected and true interval cancers) were at more advanced stages, larger, more frequently classified as lymph node positive, and poorly differentiated compared with screen-detected cancers. The most frequent phenotype in all three groups was luminal A, but the highest percentage was observed among screen-detected cancers (66.3%). However, the most frequent occurrence of the triple-negative phenotype was found among true interval cancers (28.1%), in contrast to the proportions observed among screen-detected and symptom-detected cancers (3.5 and 10.7%, respectively; $P = 0.002$).

The median follow-up was 6.7 years (range, 0.1–14.0) for women with screen-detected cancer, 3.9 years (range, 0.5–11.7) for women with true interval cancer, and 6.2 years (range, 0.5–15.2) for symptomatic women. Figure 1 shows the disease-free survival curves by detection mode and molecular phenotypes. Figure 2 shows the overall survival curves by these same factors. Disease-free survival and overall survival were worse in true interval cancers than in screen-detected and symptomatic cancers. Disease-free survival was longer in luminal cancers (both A and B) than in HER2 and triple-negative tumors

Table 1 Clinical–pathological characteristics of screen-detected cancers, true interval cancers, and symptom-detected cancers

	Screen-detected cancers $n=97$ (%)	True interval cancers $n=34$ (%)	Symptom-detected cancers $n=97$ (%)	P-value[a]
Age group (years)				
50–59	45 (46.4)	18 (52.9)	44 (45.4)	
60–69	52 (53.6)	16 (47.1)	53 (54.6)	0.741
pTNM stage				
I	56 (58.9)	7 (20.6)	34 (37.0)	
II	34 (35.8)	18 (52.9)	36 (39.1)	
III	5 (5.3)	9 (26.5)	22 (23.9)	<0.001
Unknown[b]	2	0	5	
Tumor size (mm)				
<20	69 (72.6)	13 (40.6)	41 (50.6)	
20–50	22 (23.2)	13 (40.6)	31 (38.3)	
>50	4 (4.2)	6 (18.8)	9 (11.1)	0.003
Unknown[b]	2	2	16	
Lymph node involvement				
Negative	71 (74.0)	16 (47.1)	47 (50.5)	
Positive	25 (26.0)	18 (52.9)	46 (49.5)	0.001
Unknown[b]	1	0	4	
Histological grade				
I	45 (47.4)	9 (26.5)	27 (29.7)	
II	36 (37.9)	14 (41.2)	36 (39.6)	
III	14 (14.7)	11 (32.3)	28 (30.8)	0.023
Unknown[b]	2	0	6	
ER status				
Negative	12 (13.0)	13 (39.4)	20 (24.4)	
Positive	80 (87.0)	20 (60.6)	62 (75.6)	0.005
Unknown[b]	5	1	15	
PR status				
Negative	29 (31.5)	21 (63.6)	34 (42.0)	
Positive	63 (68.5)	12 (36.4)	47 (58.0)	0.005
Unknown[b]	5	1	16	
HER2 status				
Negative	76 (87.4)	28 (87.5)	56 (71.8)	
Positive	11 (12.6)	4 (12.5)	22 (28.2)	0.023
Unknown[b]	10	2	19	
p53 status				
Negative	75 (84.3)	21 (65.6)	59 (76.6)	
Positive	14 (15.7)	11 (34.4)	18 (23.4)	0.081
Unknown[b]	8	2	20	
Phenotype				
Luminal type A	57 (66.3)	12 (37.5)	34 (45.3)	
Luminal type B	19 (22.1)	7 (21.9)	23 (30.7)	
HER2	7 (8.1)	4 (12.5)	10 (13.3)	
Triple negative	3 (3.5)	9 (28.1)	8 (10.7)	0.002
Unknown[b]	11	2	22	
Systemic treatment				
Chemotherapy alone	13 (13.4)	14 (41.2)	20 (20.6)	
Hormonal therapy alone	37 (38.1)	4 (11.8)	29 (29.9)	
Chemotherapy and hormonal therapy and/or trastuzumab[c]	36 (37.1)	15 (44.1)	41 (42.3)	
Local treatment only[d]	11 (11.3)	1 (2.9)	7 (7.2)	0.006
Relapses				
No relapses	85 (87.6)	26 (76.5)	75 (77.3)	
Local relapses	2 (2.1)	0 (0)	1 (1.0)	
Distant relapses	6 (6.2)	8 (23.5)	19 (19.6)	
Second malignances	4 (4.1)	0 (0)	2 (2.1)	0.067
Death				
No	89 (91.8)	26 (76.5)	74 (75.5)	
Yes	8 (8.2)	8 (23.5)	23 (23.7)	0.009

ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; PR, progesterone receptor; TNM, tumor–node–metastasis.
[a]Screen-detected, true interval, and symptom-detected cancers are compared. χ^2 two-sided test.
[b]Tumors with missing information were excluded for the calculation of percentage.
[c]Trastuzumab for HER2-positive breast cancer since 2006.
[d]Includes only local therapies (such as surgery and/or radiotherapy).

(log-rank test = 12.1; $P = 0.007$). Kaplan–Meier estimates of the 5-year disease-free survival rates after diagnosis for screen-detected, true interval, and symptom-detected cancers were 87.5% (95% CI, 80.5–95.2%), 64.1% (95% CI, 46.4–88.5%), and 79.4% (95% CI, 71.0–88.8%), respectively. Kaplan–Meier estimates of the 5-year overall survival rates after diagnosis for screen-detected, true interval, and symptom-detected cancers were 94.5% (95% CI, 89.3–99.9%), 65.5% (95% CI, 47.1–91.2%), and 85.6% (95% CI, 78.3–93.6%), respectively.

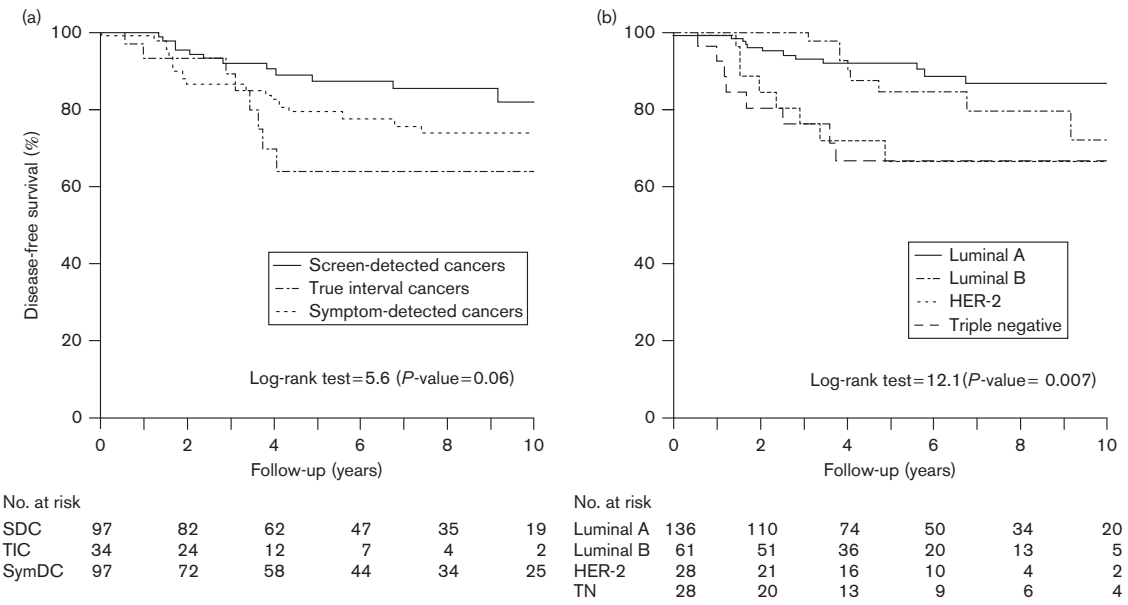
Estimates of the Cox regression models to evaluate the determinants associated with relapse-free survival and overall survival are presented in Tables 2 and 3, respectively. The unadjusted model showed that true interval cancers and symptom-detected cancers had a higher risk of relapse than screen-detected cancers, the estimated hazard ratios being statistically significant for true interval cancers. However, in the adjusted model, the prognostic effect of the detection method was attenuated, especially when biological factors were included in the multivariate model. The hazard risk of death was 5.02 (95% CI, 1.87–13.48) times higher among true interval cancers, and was 2.78 (95% CI, 1.24–6.24) times higher among symptom-detected cancers, compared with screen-detected cancers. Adjustment for TNM stage attenuated both values, which nevertheless remained statistically significant. Finally, in the model including phenotype, only true interval cancers were associated with a worse prognosis.

Discussion

Our results suggest that tumors detected clinically, especially true interval cancers, encompass a subgroup of tumors with different biological characteristics leading to poorer prognosis and worse outcomes than in screen-detected cancers. When adjustment for age, TNM stage, and phenotype was carried out, the detection mode remained as an independent factor for overall survival.

As expected, the pathological characteristics of screen-detected tumors were related to a better prognosis. At diagnosis, screen-detected cancers were smaller, more frequently lymph node-negative, and of lower grade than clinically detected tumors. These findings are in agreement with several publications (Vitak *et al*., 1997; Joensuu *et al*., 2004; Collett *et al*., 2005; Shen *et al*., 2005; Bordas *et al*., 2007; Pálka *et al*., 2008; Chiarelli *et al*., 2012). Among clinically detected tumors, 18.8% of true interval cancers were larger than 50 mm, whereas this percentage was 11.1% for symptom-detected cancers. This finding is especially important, given that true interval cancers have a short preclinical phase (sojourn time) and truly arise in less than 2 years. ER and PR positivity was more frequent in screen-detected than in clinically detected cancers, as observed in other studies (Sihto *et al*., 2008; Dawson *et al*., 2009; Mook *et al*., 2011; Nagtegaal *et al*., 2011), but not in

Fig. 1



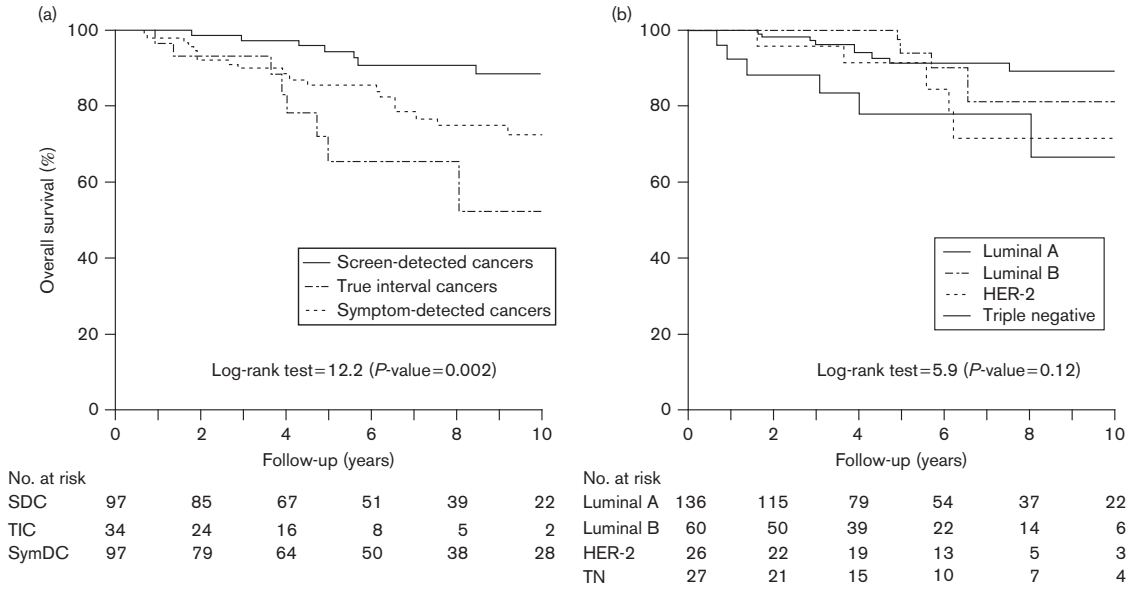
Kaplan–Meier curves comparing disease-free rates by detection mode and phenotypes. Kaplan–Meier curves for disease-free survival by detection method (a) and phenotypic classification (b) of breast tumors. The number of women at risk every 2 years is presented on the time axis. HER2, human epidermal growth factor receptor 2; SDC, screen-detected cancers; SymDC, symptom-detected cancers; TIC, true interval cancers; TN, triple negative.

all (Joensuu *et al.*, 2004). In agreement with previous series (Domingo *et al.*, 2010; Van der Vegt *et al.*, 2010; Rayson *et al.*, 2011), we found the highest percentage of triple-negative tumors among true interval cancers, with substantial differences when compared with both screen-detected and symptom-detected cancers.

These results, together with the findings of other researchers that relate true interval cancers with tumors with high mitotic capacity and a high cell-proliferation phase (Ki-67 and S-phase fraction; Vitak *et al.*, 1997; Crosier *et al.*, 1999; Kirsh *et al.*, 2011), support the hypothesis that true interval cancers constitute a subgroup of breast cancers with rapid growth and high aggressiveness that are less likely to be detected on screening mammograms. These features may determine the probability of detection by screening programs; tumors with a shorter sojourn time (which is the time taken by tumors to grow from a mammographically detectable size to a clinically detectable size) are more likely to be detected between screening intervals (Weedon-Fekjaer *et al.*, 2005). This is in part a reflection of the length bias associated with screening practices, in agreement with the idea that tumors detected on routine screening are not simply tumors diagnosed earlier, before becoming symptomatic, but also show biological differences. However, a percentage of screen-detected cancers also show features of worse prognosis (41.1% in stages II–III, 26.3% lymph node positive, and 3.5% presenting a triple-negative phenotype) that could benefit from screening.

Several studies have shown that screen detection remains an independent prognostic factor after adjustment for disease stage (Joensuu *et al.*, 2004; Shen *et al.*, 2005; Wishart *et al.*, 2008) and some biological characteristics (Gill *et al.*, 2004; Sihto *et al.*, 2008; Dawson *et al.*, 2009). Most of these studies, which do not differentiate interval cancers, recommend taking the detection method into account when estimating individual prognosis (Joensuu *et al.*, 2004; Mook *et al.*, 2011). In agreement with previous series, our results showed better survival outcomes among tumors detected by mammography than among those detected by other means (Vitak *et al.*, 1997; Joensuu *et al.*, 2004; Bordas *et al.*, 2007; Zackrisson *et al.*, 2007). Nonetheless, most of the studies considered all interval cancers together, which may have attenuated the worse outcome in the true interval cancer subgroup. Studies revealed that true interval cancers (Vitak *et al.*, 1997; Van der Vegt *et al.*, 2010; Rayson *et al.*, 2011) showed a trend toward decreased relapse-free and overall survival when compared with cancers detected by other means. However, only Vitak *et al.* (1997) compared true interval with both screen-detected

Fig. 2



Kaplan–Meier curves comparing overall survival rates by detection mode and phenotypes. Kaplan–Meier curves for overall survival by detection method (a) and phenotypic classification (b) of breast tumors. The number of women at risk every 2 years is presented on the time axis. HER2, human epidermal growth factor receptor 2; SDC, screen-detected cancers; SymDC, symptom-detected cancers; TIC, true interval cancers.

Table 2 Comparison of disease-free survival regression models, including the detection method, adjusted by different factors

Model variables	Categories	Number of Patients	Number of Events	HR adj (95% CI)
Detection method	SDC	97	12	Ref.
	TIC	34	8	2.75 (1.12–6.77)
	SymDC	97	22	1.85 (0.91–3.76)
Detection method + age + TNM stage	SDC	95	12	Ref.
	TIC	34	8	2.27 (0.90–5.72)
	SymDC	92	22	1.84 (0.89–3.82)
Detection method + age + TNM stage + phenotype	SDC	84	11	Ref
	TIC	32	8	1.89 (0.67–5.31)
	SymDC	72	12	0.91 (0.37–2.27)
Detection method + age + TNM stage + phenotype + treatment	SDC	84	11	Ref.
	TIC	32	8	1.95 (0.66–5.79)
	SymDC	72	12	0.93 (0.37–2.35)

CI, confidence interval; HR adj, adjusted hazard ratio; ref., reference; SDC, screen-detected cancers; SymDC, symptom-detected cancers; TIC, true interval cancers; TNM, tumor–node–metastasis.

Table 3 Comparison of overall survival regression models, including the detection method, adjusted by different factors

Model variables	Categories	Number of Patients	Number of Events	HR adj (95% CI)
Detection method	SDC	97	8	Ref.
	TIC	34	8	5.02 (1.87–13.48)
	SymDC	97	23	2.78 (1.24–6.24)
Detection method + age + TNM stage	SDC	95	8	Ref
	TIC	34	8	3.87 (1.40–10.75)
	SymDC	92	23	2.73 (1.20–6.21)
Detection method + age + TNM stage + phenotype	SDC	84	6	Ref.
	TIC	32	8	5.55 (1.61–19.15)
	SymDC	72	12	2.14 (0.72–6.30)
Detection method + age + TNM stage + phenotype + treatment	SDC	84	6	Ref.
	TIC	32	8	7.67 (2.07–28.50)
	SymDC	72	12	2.44 (0.79–7.54)

CI, confidence interval; HR adj, adjusted hazard ratio; ref., reference; SDC, screen-detected cancers; SymDC, symptom-detected cancers; TIC, true interval cancers; TNM, tumor–node–metastasis.

and symptom-detected cancers, whereas the remaining studies did not include this last group.

Unadjusted Cox models for disease-free survival showed the method of detection as an independent factor for relapse prediction, consistent with previous series including true interval cancers (Vitak *et al*., 1997; Rayson *et al*., 2011). However, when tumor size and biological characteristics were included in the multivariate analyses to adjust for lead and length bias, as expected, the prognostic value of the detection mode lost its effect for recurrences. Nevertheless, for overall survival, our trends suggest that true interval cancers trigger a subgroup of breast cancers with an independent unfavorable prognostic significance beyond that explained by the conventional factors included in the present study. If confirmed, this finding may merit further investigation into its underlying biological mechanisms. Recently, some

genetic and epigenetic mechanisms have been related to interval malignancies, such as the methylation process of specific genes (Suijkerbuijk *et al*., 2011) or the action of growth factors produced in the breast stroma in response to tumor aggressiveness (Li *et al*., 2005). Further efforts should be made in the future to improve understanding of the genetic mechanisms related to cancer aggressiveness, tumor cell proliferation, and the process of carcinogenesis to improve the early detection of fast-growing tumors.

This study has some limitations, the main one being the small sample size. Given the difficulties in identifying and classifying interval cancers, to date, few studies have focused on true interval cancers. However, the current study adds to previous analyses limited by incomplete pathological data or a failure to restrict the analyses to true interval cancers. Second, misclassification of the detection mode cannot be excluded. Some interval cancers could be classified as screen-detected if symptomatic women waited for the screening visit instead of making an appointment with a physician. This misclassification would attenuate the effect studied. Thus, the survival difference in favor of screened women might be greater than that observed because of the inclusion of some women with symptomatic cancers in the screened cohort. In addition, the sociodemographic characteristics of women participating in the screening program, whose interval cancer was diagnosed and treated in other hospitals, did not differ from those of women treated in our setting. However, as participation in screening practices may be affected by several selection factors, we considered a group of women not affected by screening in the analyses.

This study has some strengths. To our knowledge, this is the first work focused on true interval cancer characterization that analyzes prognosis by considering cancers detected both inside and outside screening, in addition to the molecular profile. This design allows us to control the main biases that affect the outcomes of the screening practices. Moreover, all analyses were carried out with information on phenotype and other biological markers from recently diagnosed patients (1996–2008) who received homogeneous adjuvant treatment on the basis of the oncology protocol of our institution: chemotherapy based on anthracyclines or anthracyclines and taxanes, hormone therapies (tamoxifen or aromatase inhibitors), and targeted therapies (trastuzumab has been prescribed to patients with HER2 overexpression since 2006) in the modern era.

Conclusion

These results suggest that screening programs detect a major proportion of tumors with more favorable biological characteristics, which could partly explain the better survival of patients with screening-detected tumors. Moreover, our results show that true interval cancers share a subset of features of worse prognosis, related both to their growth rate and short sojourn time, and to their clinical course. Further understanding of their biologic features and individual determinants could increase the benefits of screening: in the short term, by increasing the sensitivity of the programs and aiding the choice of optimal screening interval for specific subsets of women at high risk, and in the long term, by increasing the early detection of tumors with a less favorable natural history.

Acknowledgements

The authors acknowledge the contribution of the Hospital del Mar Tumor Registry (Barcelona) in providing tumor-related data. They also thank Cristina Hernández and Marta Román for their assistance in data management and Teresa Baró for her technical assistance in performing the immunohistochemical analysis.

This study was partially supported by CIBER de Epidemiología y Salud Pública (AE08_004).

Conflicts of interest

There are no conflicts of interest.

References

Autier P, Boniol M, La Vecchia C, Vatten L, Gavin A, Héry C, *et al.* (2010). Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. *BMJ* **341**:1–7.

Benson JR, Jatoi I, Keisch M, Esteva FJ, Makris A, Jordan VC (2009). Early breast cancer. *Lancet* **373**:1463–1479.

Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, *et al.* (2005). Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* **353**:1784–1792.

Bordas P, Jonsson H, Nystrom L, Lenner P (2007). Survival from invasive breast cancer among interval cases in the mammography screening programmes of northern Sweden. *Breast* **16**:47–54.

Chacón RD, Costanzo MV (2010). Triple-negative breast cancer. *Breast Cancer Res* **12 (Suppl 2)**:1–9.

Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, *et al.* (2009). Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst* **101**:736–750.

Chiarelli AM, Edwards SA, Sheppard AJ, Mirea L, Chong N, Paszat L, *et al.* (2012). Favourable prognostic factors of subsequent screen-detected breast cancers among women aged 50–69. *Eur J Cancer Prev* **23**:1–8.

Collett K, Stefansson IM, Eide J, Braaten A, Wang H, Eide GE, *et al.* (2005). A basal epithelial phenotype is more frequent in interval breast cancers compared with screen detected tumors. *Cancer Epidemiol Biomarkers Prev* **14**:1108–1112.

Crosier M, Scott D, Wilson RG, Griffiths CD, May FE, Westley BR (1999). Differences in Ki67 and c-erbB2 expression between screen-detected and true interval breast cancers. *Clin Cancer Res* **5**:2682–2688.

Dawson SJ, Duffy SW, Blows FM, Driver KE, Provenzano E, LeQuesne J, *et al.* (2009). Molecular characteristics of screen-detected vs. symptomatic breast cancers and their impact on survival. *Br J Cancer* **101**:1338–1344.

Domingo L, Sala M, Servitja S, Corominas JM, Ferrer F, Martínez J, *et al.* (2010). Phenotypic characterization and risk factors for interval breast cancers in a population-based breast cancer screening program in Barcelona, Spain. *Cancer Causes Control* **21**:1155–1164.

Gill PG, Farshid G, Luke CG, Roder DM (2004). Detection by screening mammography is a powerful independent predictor of survival in women diagnosed with breast cancer. *Breast* **13**:15–22.

Immonen-Räihä P, Kauhava L, Parvinen I, Holli K, Kronqvist P, Pylkkänen L, *et al.* (2005). Mammographic screening reduces risk of breast carcinoma recurrence. *Cancer* **103**:474–482.

Joensuu H, Lehtimaki T, Holli K, Elomaa L, Turpeenniemi-Hujanen T, Kataja V, *et al.* (2004). Risk for distant recurrence of breast cancer detected by mammography screening or other methods. *JAMA* **292**:1064–1073.

Kirsh VA, Chiarelli AM, Edwards SA, O'Malley FP, Shumak RS, Yaffe MJ, *et al.* (2011). Tumor characteristics associated with mammographic detection of breast cancer in the Ontario Breast Screening Program. *J Natl Cancer Inst* **103**:942–950.

Lawrence G, Wallis M, Allgood P, Nagtegaal ID, Warwick J, Cafferty FH, *et al.* (2009). Population estimates of survival in women with screen-detected and symptomatic breast cancer taking account of lead time and length bias. *Breast Cancer Res Treat* **116**:179–185.

Li T, Sun L, Miller N, Nicklee T, Woo J, Hulse-Smith L, *et al.* (2005). The association of measured breast tissue characteristics with mammographic density and other risk factors for breast cancer. *Cancer Epidemiol Biomarkers Prev* **14**:343–349.

Mook S, van't Veer LJ, Rutgers EJ, Ravdin PM, van de Velde AO, van Leeuwen FE, *et al.* (2011). Independent prognostic value of screen detection in invasive breast cancer. *J Natl Cancer Inst* **103**:585–597.

Nagtegaal ID, Allgood PC, Duffy SW, Kearins O, Sullivan EO, Tappenden N, *et al.* (2011). Prognosis and pathology of screen-detected carcinomas: how different are they? *Cancer* **117**:1360–1368.

Pálka I, Kelemen G, Ormandi K, Lazar G, Nyari T, Thurzo L, *et al.* (2008). Tumor characteristics in screen-detected and symptomatic breast cancers. *Pathol Oncol Res* **14**:161–167.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, *et al.* (2000). Molecular portraits of human breast tumours. *Nature* **406**:747–752.

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.* 4th ed. Luxembourg: Office for Official Publications of the European Communities.

Rayson D, Payne JI, Abdolell M, Barnes PJ, MacIntosh RF, Foley T, *et al.* (2011). Comparison of clinical-pathologic characteristics and outcomes of true interval and screen-detected invasive breast cancer among participants of a Canadian breast screening program: a nested case-control study. *Clin Breast Cancer* **11**:27–32.

R Development Core Team (2011). *R: A language and environment for statistical computing. ISBN 3-900051-07-0.* Vienna, Austria: R Foundation for Statistical Computing. Available at: *http://www.R-project.org/*.

Shen Y, Yang Y, Inoue LY, Munsell MF, Miller AB, Berry DA (2005). Role of detection method in predicting breast cancer survival: analysis of randomized screening trials. *J Natl Cancer Inst* **97**:1195–1203.

Sihto H, Lundin J, Lehtimäki T, Sarlomo-Rikala M, Bützow R, Holli K, *et al.* (2008). Molecular subtypes of breast cancers detected in mammography screening and outside of screening. *Clin Cancer Res* **14**:4103–4110.

Suijkerbuijk KP, van Diest PJ, van der WE (2011). Improving early breast cancer detection: focus on methylation. *Ann Oncol* **22**:24–29.

Van der Vegt B, Wesseling J, Pijnappel RM, Dorrius MD, den Heeten GJ, de Roos MA, *et al.* (2010). Aggressiveness of 'true' interval invasive ductal carcinomas of the breast in postmenopausal women. *Mod Pathol* **23**: 629–636.

Vitak B, Stal O, Manson JC, Thomas BA, Arnesson LG, Ekelund L, *et al.* (1997). Interval cancers and cancers in non-attenders in the Ostergotland Mammographic Screening Programme. Duration between screening and diagnosis, S-phase fraction and distant recurrence. *Eur J Cancer* **33**: 1453–1460.

Weedon-Fekjaer H, Vatten LJ, Aalen OO, Lindqvist B, Tretli S (2005). Estimating mean sojourn time and screening test sensitivity in breast cancer mammography screening: new results. *J Med Screen* **12**:172–178.

Wishart GC, Greenberg DC, Britton PD, Chou P, Brown CH, Purushotham AD, *et al.* (2008). Screen-detected vs. symptomatic breast cancer: Is improved survival due to stage migration alone? *Br J Cancer* **98**:1741–1744.

Zackrisson S, Janzon L, Manjer J, Andersson I (2007). Improved survival rate for women with interval breast cancer – results from the breast cancer screening programme in Malmo, Sweden 1976–1999. *J Med Screen* **14**:138–143.

Bibliografia

[1] J Ferlay, I Soerjomataram, M Ervik, R Dikshit, S Eser, C Mathers, M Rebelo, DM Parkin, D Forman, and F Bray. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11, 2013.

[2] Ahmedin Jemal, Melissa M Center, Carol DeSantis, and Elizabeth M Ward. Global patterns of cancer incidence and mortality rates and trends. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 19(8):1893–907, aug 2010.

[3] Barbara S Hulka and Patricia G Moorman. Breast cancer: hormones and other risk factors. *Maturitas*, 61(1-2):203–13; discussion 213, jan.

[4] Michelle D Althuis, Jaclyn M Dozier, William F Anderson, Susan S Devesa, and Louise A Brinton. Global trends in breast cancer incidence and mortality 1973-1997. *International journal of epidemiology*, 34(2):405–12, apr 2005.

[5] Peter M Ravdin, Kathleen A Cronin, Nadia Howlader, Christine D Berg, Rowan T Chlebowski, Eric J Feuer, Brenda K Edwards, and Donald A Berry. The decrease in breast-cancer incidence in 2003 in the United States., apr 2007.

[6] Kathleen A Cronin, Peter M Ravdin, and Brenda K Edwards. Sustained lower rates of breast cancer in the United States. *Breast cancer research and treatment*, 117(1):223–4, sep 2009.

[7] Donald Maxwell Parkin. 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)*, 45(9):1649–53, jun 2009.

[8] B Séradour, H Allemand, A Weill, and P Ricordeau. Changes by age in breast cancer incidence, mammography screening and hormone therapy use in France from 2000 to 2006. *Bulletin du cancer*, 96(4):E1–6, apr 2009.

[9] Philippe Autier, Mathieu Boniol, Carlo La Vecchia, Carlo LaVecchia, Lars Vatten, Anna Gavin, Clarisse Héry, and Mary Heanue. Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. *BMJ (Clinical research ed.)*, 341:c3620, jan 2010.

[10] M J Sánchez, T Payer, R De Angelis, N Larrañaga, R Capocaccia, and C Martinez. Cancer incidence and mortality in Spain: estimates and projections for the period 1981-2012. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 21 Suppl 3(Supplement 3):iii30–36, may 2010.

[11] Marina Pollán, Roberto Pastor-Barriuso, Eva Ardanaz, Marcial Argüelles, Carmen Martos, Jaume Galcerán, María-José Sánchez-Pérez, María-Dolores Chirlaque, Nerea Larrañaga, Ruth Martínez-Cobo, María-Cres Tobalina, Enrique Vidal, Rafael Marcos-Gragera, Antonio Mateos, Isabel Garau, María-Dolores Rojas-Martín, Rosario Jiménez, Ana Torrella-Ramos, Josefina Perucha, Maria-Eugenia Pérez-de Rada, Susana González, María-José Rabanaque, Joan Borràs, Carmen Navarro, Esther Hernández, Angel Izquierdo, Gonzalo López-Abente, and Carmen Martínez. Recent changes in breast cancer incidence in Spain, 1980-2004. *Journal of the National Cancer Institute*, 101(22):1584–1591, nov 2009.

[12] R Clèries, L Esteban, J Borràs, R Marcos-Gragera, A Freitas, M Carulla, M Buxó, A Puigdefàbregas, A Izquierdo, R Gispert, J Galceran, and J Ribes. Time trends of cancer incidence and mortality in Catalonia during 1993-2007. *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico*, 16(1):18–28, jan 2014.

[13] Izquierdo Font A, Marcos-Graguera R, Vilardell Gil M, Buixó M, and Fuentes J. El càncer a Girona 2007-2009: projeccions de la incidència 2013-2014. *CanGir*, (4):1–32, 2013.

[14] Jaume Galceran, Josep Gumà, Marià Carulla, Alberto Ameijide, Francina Saladié, and Joan Borràs Col. *EL CÀNCER A TARRAGONA 2013 DADES i XIFRES*. Fundació Lliga per a la Investigació i Prevenció del Càncer. Lliga

Contra el Càncer de les Comarques de Tarragona i Terres de l'Ebre., Reus, 2013.

[15] Milena Sant, Maria Dolores Chirlaque Lopez, Roberto Agresti, Maria José Sánchez Pérez, Bernd Holleczek, Magdalena Bielska-Lasota, Nadya Dimitrova, Kaire Innos, Alexander Katalinic, Hilde Langseth, Nerea Larrañaga, Silvia Rossi, Sabine Siesling, and Pamela Minicozzi. Survival of women with cancers of breast and genital organs in Europe 1999–2007: Results of the EUROCARE-5 study. *European Journal of Cancer*, 51(15):2191–2205, 2015.

[16] VT DeVita, TS Lawrence, and SA Rosenberg. *Cancer: Principles & Practice of Oncology.* Wolters Kluwer/Lippincott Williams & Wilkins, Philadelphia, 10th edition, 2014.

[17] SR Lakhani, IO Ellis, SJ Schnitt, Ph Tan, and MJ van de Vijver. *WHO Classification of Tumours of the Breast.* World Health Organization. International Agency for Research on Cancer, Lyon, 4th edition, 2012.

[18] N Hamajima, K Hirose, K Tajima, T Rohan, E E Calle, C W Heath, R J Coates, J M Liff, R Talamini, N Chantarakul, S Koetsawang, D Rachawat, A Morabia, L Schuman, W Stewart, M Szklo, C Bain, F Schofield, V Siskind, P Band, A J Coldman, R P Gallagher, T G Hislop, P Yang, L M Kolonel, A M Y Nomura, J Hu, K C Johnson, Y Mao, S De Sanjosé, N Lee, P Marchbanks, H W Ory, H B Peterson, H G Wilson, P A Wingo, K Ebeling, D Kunde, P Nishan, J L Hopper, G Colditz, V Gajalanski, N Martin, T Pardthaisong, S Silpisornkosol, C Theetranont, B Boosiri, S Chutivongse, P Jimakorn, P Virutamasen, C Wongsrichanalai, M Ewertz, H O Adami, L Bergkvist, C Magnusson, I Persson, J Chang-Claude, C Paul, D C G Skegg, G F S Spears, P Boyle, T Evstifeeva, J R Daling, W B Hutchinson, K Malone, E A Noonan, J L Stanford, D B Thomas, N S Weiss, E White, N Andrieu, A Brêmond, F Clavel, B Gairard, J Lansac, L Piana, R Renaud, A Izquierdo, P Viladiu, H R Cuevas, P Ontiveros, A Palet, S B Salazar, N Aristizabel, A Cuadros, L Tryggvadottir, H Tulinius, A Bachelot, M G Lê, J Peto, S Franceschi, F Lubin, B Modan, E Ron, Y Wax, G D Friedman, R A Hiatt, F Levi, T Bishop, K Kosmelj, M Primic-Zakelj, B Ravnihar, J Stare, W L Beeson, G Fraser, R D Bullbrook, J Cuzick, S W Duffy, I S Fentiman, J L Hayward, D Y Wang, A J McMichael, K McPherson, R L Hanson, M C Leske, M C Mahoney, P C Nasca, A O Varma, A L Weinstein, T R Moller, H Olsson, J Ranstam, R A Goldbohm, P A van den Brandt, R A Apelo, J Baens, J R de la Cruz, B Javier, L B Lacaya, C A Ngelangel, C La Vecchia, E Negri, E Marubini, M Ferraroni, M Gerber,

S Richardson, C Segala, D Gatei, P Kenya, A Kungu, J G Mati, L A Brinton, R Hoover, C Schairer, R Spirtas, H P Lee, M A Rookus, F E van Leeuwen, J A Schoenberg, M McCredie, M D Gammon, E A Clarke, L Jones, A Neil, M Vessey, D Yeates, P Appleby, E Banks, V Beral, D Bull, B Crossley, A Goodill, J Green, C Hermon, T Key, N Langston, C Lewis, G Reeves, R Collins, R Doll, R Peto, K Mabuchi, D Preston, P Hannaford, C Kay, L Rosero-Bixby, Y T Gao, F Jin, J-M Yuan, H Y Wei, T Yun, C Zhiheng, G Berry, J Cooper Booth, T Jelihovsky, R MacLennan, R Shearman, Q-S Wang, C-J Baines, A B Miller, C Wall, E Lund, H Stalsberg, X O Shu, W Zheng, K Katsouyanni, A Trichopoulou, D Trichopoulos, A Dabancens, L Martinez, R Molina, O Salas, F E Alexander, K Anderson, A R Folsom, B S Hulka, L Bernstein, S Enger, R W Haile, A Paganini-Hill, M C Pike, R K Ross, G Ursin, M C Yu, M P Longnecker, P Newcomb, A Kalache, T M M Farley, S Holck, and O Meirik. Alcohol, tobacco and breast cancer–collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *British journal of cancer*, 87(11):1234–45, nov 2002.

[19] P Boyle, P Autier, H Bartelink, J Baselga, P Boffetta, J Burn, H J G Burns, L Christensen, L Denis, M Dicato, V Diehl, R Doll, S Franceschi, C R Gillis, N Gray, L Griciute, A Hackshaw, M Kasler, M Kogevinas, S Kvinnsland, C La Vecchia, F Levi, J G McVie, P Maisonneuve, J M Martin-Moreno, J Newton Bishop, F Oleari, P Perrin, M Quinn, M Richards, U Ringborg, C Scully, E Siracka, H Storm, M Tubiana, T Tursz, U Veronesi, N Wald, W Weber, D G Zaridze, W Zatonski, and H zur Hausen. European Code Against Cancer and scientific justification: third version (2003). *Annals of Oncology*, 14(7):973–1005, jul 2003.

[20] Ellen Warner. Breast-Cancer Screening. *The New England journal of medicine*, 365(11):1025–32, sep 2011.

[21] I C Bennett, M Gattas, and B T Teh. The genetic basis of breast cancer and its clinical implications. *The Australian and New Zealand journal of surgery*, 69(2):95–105, feb 1999.

[22] John T Schousboe, Karla Kerlikowske, Andrew Loh, and Steven R Cummings. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Annals of internal medicine*, 155(1):10–20, jul 2011.

[23] Norman F Boyd, Lisa J Martin, Martin Yaffe, and Salomon Minkin. Mammographic density. *Breast cancer research : BCR*, 11 Suppl 3:S4, jan 2009.

[24] N F Boyd, G A Lockwood, J W Byng, D L Tritchler, and M J Yaffe. Mammographic densities and breast cancer risk. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 7(12):1133–44, dec 1998.

[25] Norman F Boyd, Helen Guo, Lisa J Martin, Limei Sun, Jennifer Stone, Eve Fishell, Roberta A Jong, Greg Hislop, Anna Chiarelli, Salomon Minkin, and Martin J Yaffe. Mammographic Density and the Risk and Detection of Breast Cancer. *The New England journal of medicine*, 356(3):227–36, 2007.

[26] S R Wellings and H M Jensen. On the origin and progression of ductal carcinoma in the human breast. *Journal of the National Cancer Institute*, 50(5):1111–8, may 1973.

[27] S R Wellings, H M Jensen, and R G Marcum. An atlas of subgross pathology of the human breast with special reference to possible precancerous lesions. *Journal of the National Cancer Institute*, 55(2):231–73, aug 1975.

[28] Fulvia Farabegoli, Marie-Helene Champeme, Ivan Bieche, Donatella Santini, Claudio Ceccarelli, Massimo Derenzini, and Rosette Lidereau. Genetic pathways in the evolution of breast ductal carcinoma in situ. *The Journal of pathology*, 196(3):280–6, mar 2002.

[29] D Craig Allred, Yun Wu, Sufeng Mao, Iris D Nagtegaal, Sangjun Lee, Charles M Perou, Syed K Mohsin, Peter O'Connell, Anna Tsimelzon, and Dan Medina. Ductal carcinoma in situ and the emergence of diversity during breast cancer evolution. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 14(2):370–8, jan 2008.

[30] Harold J Burstein, Kornelia Polyak, Julia S Wong, Susan C Lester, and Carolyn M Kaelin. Ductal carcinoma in situ of the breast. *The New England journal of medicine*, 350(14):1430–41, apr 2004.

[31] Laura J. Esserman, Ian M. Thompson, Brian Reid, Peter Nelson, David F. Ransohoff, H. Gilbert Welch, Shelley Hwang, Donald A. Berry, Kenneth W. Kinzler, William C. Black, Mina Bissell, Howard Parnes, and Sudhir Srivastava. Addressing overdiagnosis and overtreatment in cancer: A prescription for change. *The Lancet Oncology*, 15(6):e234–e242, 2014.

[32] C W Elston and I O Ellis. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, 19(5):403–10, nov 1991.

[33] Emad A Rakha, Jorge S Reis-Filho, Frederick Baehner, David J Dabbs, Thomas Decker, Vincenzo Eusebi, Stephen B Fox, Shu Ichihara, Jocelyne Jacquemier, Sunil R Lakhani, José Palacios, Andrea L Richardson, Stuart J Schnitt, Fernando C Schmitt, Puay-Hoon Tan, Gary M Tse, Sunil Badve, and Ian O Ellis. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Research*, 12(4):207, aug 2010.

[34] G Canavese, A Catturich, C Vecchio, D Tomei, M Gipponi, P Bruzzi, and F Badellino. Prognostic role of lymph-node level involvement in patients undergoing axillary dissection for breast cancer. *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*, 24(2):104–9, apr 1998.

[35] G Canavese, M Gipponi, A Catturich, C Di Somma, C Vecchio, F Rosato, D Tomei, F Cafiero, L Moresco, G Nicolò, F Carli, G Villa, F Buffoni, and F Badellino. Sentinel lymph node mapping opens a new perspective in the surgical management of early-stage breast cancer: a combined approach with vital blue dye lymphatic mapping and radioguided surgery. *Seminars in surgical oncology*, 15(4):272–7, dec 1998.

[36] Milena Sant, Claudia Allemani, Riccardo Capocaccia, Timo Hakulinen, Tiiu Aareleid, Jan Willem Coebergh, Michel P. Coleman, Pascale Grosclaude, Carmen Martinez, Janine Bell, Judith Youngson, and Franco Berrino. Stage at diagnosis is a key explanation of differences in breast cancer survival across Europe. *International Journal of Cancer*, 106(3):416–422, sep 2003.

[37] A Goldhirsch, W C Wood, A S Coates, R D Gelber, B Thürlimann, and H-J Senn. Strategies for subtypes–dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 22(8):1736–47, aug 2011.

[38] Mohammad Jahanzeb. Adjuvant Trastuzumab Therapy for HER2-Positive Breast Cancer. *Clinical Breast Cancer*, 8(4):324–333, aug 2008.

[39] P S Frame and S J Carlson. A critical review of periodic health screening using specific screening criteria. 3. Selected diseases of the genitourinary system. *The Journal of family practice*, 2(3):189–94, jun 1975.

[40] P S Frame, S J Carlson, and A Critical. A critical review of periodic health screening using specific screening criteria. Part 1: Selected diseases of respiratory, cardiovascular, and central nervous systems. *The Journal of family practice*, 2(1):29–36, feb 1975.

[41] P S Frame and S J Carlson. A critical review of periodic health screening using specific screening criteria. Part 2: Selected endocrine, metabolic and gastrointestinal diseases. *The Journal of family practice*, 2(2):123–9, apr 1975.

[42] Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet*, 380(9855):1778–86, nov 2012.

[43] X Castells, M Sala, N Ascunce, D Salas, R Zubizarreta, and Montserrat Casamitjana, editors. *Descripción del cribado del cáncer en España. Proyecto DESCRIC*. Plan de Calidad para el Sistema Nacional de Salud. Ministerio de Sanidad y Consumo. Agència d'Avaluació de Tecnologia i Recerca Mèdiques de Cataluña; 2007. Informes de Evaluación de Tecnologías Sanitarias, AATRM núm. 2006/01, Madrid, 2006.

[44] Gonzalo López-Abente, N Aragonés, B Pérez-Gómez, M Pollán, J García-Pérez, R Ramis, and P Fernández-Navarro. Time trends in municipal distribution patterns of cancer mortality in Spain. *BMC cancer*, 14(535), apr 2014.

[45] Katrina Armstrong, Elizabeth Moye, Sankey Williams, Jesse A Berlin, and Eileen E Reynolds. Screening Mammography in Women 40 to 49 Years of Age : A Systematic Review for the American College of Physicians. *Annals of Internal Medicine*, 146(7):516–526, 2007.

[46] Eugenio Paci. 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*, 19 Suppl 1:5–13, jan 2012.

[47] N Perry, M Broeders, C de Wolf, S Törnberg, R Holland, and L von Karsa. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition–summary document. *Annals of Oncology*, 19(4):614–622, 2008.

[48] S Hofvind, B M Geller, J Skelly, and P M Vacek. Sensitivity and specificity of mammographic screening as practised in Vermont and Norway. *The British journal of radiology*, 85(1020):e1226–32, dec 2012.

[49] R Román, M Sala, D Salas, N Ascunce, R Zubizarreta, and X Castells. Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 23(1):104–11, jan 2012.

[50] R Román, M Sala, M De La Vega, C Natal, J Galceran, I González-Román, A Baroja, R Zubizarreta, N Ascunce, D Salas, and X Castells. Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening. *Breast cancer research and treatment*, 130(2):543–52, dec 2011.

[51] Sune Bangsbøll Andersen, Ilse Vejborg, and My von Euler-Chelpin. Participation behaviour following a false positive test in the Copenhagen mammography screening programme. *Acta oncologica (Stockholm, Sweden)*, 47(4):550–5, jan 2008.

[52] M L Burman, S H Taplin, D F Herta, and J G Elmore. Effect of false-positive mammograms on interval breast cancer screening in a health maintenance organization. *Annals of internal medicine*, 131(1):1–6, jul 1999.

[53] Donella Puliti, Stephen W Duffy, Guido Miccinesi, Harry de Koning, Elsebeth Lynge, Marco Zappa, and Eugenio Paci. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *Journal of medical screening*, 19 Suppl 1:42–56, jan 2012.

[54] Sven Törnberg, Levent Kemetli, Nieves Ascunce, Solveig Hofvind, Ahti Anttila, Brigitte Sèradour, Eugenio Paci, Cathrine Guldenfels, Edward Azavedo, Alfonso Frigerio, Vitor Rodrigues, and Antonio Ponti. A pooled analysis of interval cancer rates in six European countries. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP)*, 19(2):87–93, mar 2010.

[55] Laia Domingo, Jordi Blanch, Sònia Servitja, Josep Maria Corominas, Cristiane Murta-Nascimento, Antonio Rueda, Maximino Redondo, Xavier Castells, and Maria Sala. 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)*, 22(1):21–8, jan 2013.

[56] Paul D. Allison. Discrete-time Methods for the anaylsis of event histories. *American Sociological MEthodology*, 13:61–98°, 1982.

[57] J D Singer and J B Willett. SingerWillet.pdf. *Journal of educational Statistics*, 18(2):155–195, 1993.

[58] DG Kleinbaum and M Klein. *Survival Analysis- A Self-Learning Text, Third Edition*. Springer Publishers New York, Inc, New York, 3rd edition, 2011.

[59] R B Geskus. *Data Analysis with Competing Risks and Intermediate States*. Chapman and Hall/CRC, 2014.

[60] Ester Vilaprinyo, Carles Forné, Misericordia Carles, Maria Sala, Roger Pla, Xavier Castells, Laia Domingo, and Montserrat Rue. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. *PloS one*, 9(2):e86858, jan 2014.

[61] Mercè Comas, Arantzazu Arrospide, Javier Mar, Maria Sala, Ester Vilaprinyó, Cristina Hernández, Francesc Cots, Juan Martínez, and Xavier Castells. Budget Impact Analysis of Switching to Digital Mammography in a Population-Based Breast Cancer Screening Program: A Discrete Event Simulation Model. *PLoS ONE*, 9(5):e97459, may 2014.

[62] Rianne de Gelder, Eveline A M Heijnsdijk, Nicolien T van Ravesteyn, Jacques Fracheboud, Gerrit Draisma, and Harry J de Koning. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiologic reviews*, 33(1):111–21, 2011.

[63] S J Lee and M Zelen. Scheduling periodic examinations for the early detection of disease: Applications to breast cancer. *Journal of the American Statistical Association*, 93(444):1271–1281, 1998.

[64] S J Lee and M Zelen. Modelling the early detection of breast cancer. *Annals of Oncology*, 14(8):1199–1202, 2003.

[65] Christopher H Jackson. Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software*, 38(8):1–28, 2011.

[66] R B Geskus. *Data analysis with competing risks and intermediate states*. Chapman and Hall/CRC Biostatistics Series, Boca Raton, FL, 2016.

[67] H Putter, M Fiocco, and R B Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11):2389–2430, may 2007.

[68] Andrea Burton, Douglas G Altman, Patrick Royston, and Roger L Holder. The design of simulation studies in medical statistics. *Statistics in Medicine*, 25(24):4279–4292, dec 2006.

[69] R G Blanks, M G Wallis, and S M Moss. A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers, for one and two view mammography: results from the UK National Health Service breast screening programme. *Journal of medical screening*, 5(4):195–201, jan 1998.

[70] Rianne de Gelder, Jacques Fracheboud, Eveline A M Heijnsdijk, Gerard den Heeten, André L M Verbeek, Mireille J M Broeders, Gerrit Draisma, and Harry J de Koning. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. *Preventive medicine*, 53(3):134–40, sep 2011.

[71] Marco Rosselli Del Turco, Paola Mantellini, Stefano Ciatto, Rita Bonardi, Francesca Martinelli, Barbara Lazzari, and Nehmat Houssami. Full-field digital versus screen-film mammography: comparative accuracy in concurrent screening cohorts. *AJR. American journal of roentgenology*, 189(4):860–6, oct 2007.

[72] Laia Domingo, Anabel Romero, Francesc Belvis, Mar Sánchez, Joana Ferrer, Dolores Salas, Josefa Ibáñez, Alfonso Vega, Francesc Ferrer, M Soledad Laso, Francesc Macià, Xavier Castells, and Maria Sala. Differences in radiological patterns, tumour characteristics and diagnostic precision between digital mammography and screen-film mammography in four breast cancer screening programmes in Spain. *European radiology*, 21(9):2020–8, oct 2011.

[73] Les Irwig, Nehmat Houssami, Bruce Armstrong, and Paul Glasziou. Evaluating new screening tests for breast cancer. *BMJ (Clinical research ed.)*, 332(7543):678–9, mar 2006.

[74] Erin L Ashbeck, Robert D Rosenberg, Patricia M Stauber, and Charles R Key. Benign breast biopsy diagnosis and subsequent risk of breast cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 16(3):467–72, mar 2007.

[75] Geoffrey C Kabat, Joan G Jones, Neal Olson, Abdissa Negassa, Catherine Duggan, Mindy Ginsberg, Rita A Kandel, Andrew G Glass, and Thomas E Rohan. A multi-center prospective cohort study of benign breast disease and

risk of subsequent breast cancer. *Cancer causes & control : CCC*, 21(6):821–8, jun 2010.

[76] J Blanch, M Sala, M Román, M Ederra, D Salas, R Zubizarreta, M Sanchez, M Rué, and X Castells. Cumulative risk of cancer detection in breast cancer screening by protocol strategy. *Breast cancer research and treatment*, 138(3):869–77, apr 2013.

[77] Nehmat Houssami, Les Irwig, and Stefano Ciatto. Radiological surveillance of interval breast cancers in screening programmes. *The lancet oncology*, 7(3):259–65, mar 2006.

[78] Jennifer I Payne, Judy S Caines, Julie Gallant, and Theresa J Foley. A review of interval breast cancers diagnosed among participants of the Nova Scotia Breast Screening Program. *Radiology*, 266(1):96–103, jan 2013.

[79] B. Vitak, K. E. Olsen, J. C. Månson, L. G. Arnesson, and O. Stål. Tumour characteristics and survival in patients with invasive interval breast cancer classified according to mammographic findings at the latest screening: a comparison of true interval and missed interval cancers. *European Radiology*, 9(3):460–469, mar 1999.

[80] Y P Guo, L J Martin, W Hanna, D Banerjee, N Miller, E Fishell, R Khokha, and N F Boyd. Growth factors and stromal matrix proteins associated with mammographic densities. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 10(3):243–8, mar 2001.

[81] Xavier Castells, Laia Domingo, Josep María Corominas, Isabel Torá-Rocamora, María Jesús Quintana, Marisa Baré, Carmen Vidal, Carmen Natal, Mar Sánchez, Francina Saladié, Joana Ferrer, Mar Vernet, Sonia Servitja, Ana Rodríguez-Arana, Marta Roman, Josep Alfons Espinàs, and María Sala. Breast cancer risk after diagnosis by screening mammography of nonproliferative or proliferative benign breast disease: a study from a population-based screening program. *Breast Cancer Research and Treatment*, 149(1):237–244, jan 2015.

[82] Xavier Castells, Isabel Torá-Rocamora, Margarita Posso, Marta Román, Maria Vernet-Tomas, Ana Rodríguez-Arana, Laia Domingo, Carmen Vidal, Marisa Baré, Joana Ferrer, María Jesús Quintana, Mar Sánchez, Carmen Natal, Josep A. Espinàs, Francina Saladié, and María Sala. Risk of Breast Cancer in

Women with False-Positive Results according to Mammographic Features. *Radiology*, 280(2):379–386, aug 2016.

[83] Yiwey Shieh, Martin Eklund, Lisa Madlensky, Sarah D. Sawyer, Carlie K. Thompson, Allison Stover Fiscalini, Elad Ziv, Laura J. van't Veer, Laura J. Esserman, and Jeffrey A. Tice. Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. *Journal of the National Cancer Institute*, 109(5):djw290, jan 2017.