



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

Universitat Autònoma de Barcelona

Facultat de Medicina

Departamento de Pediatría, Obstetricia y Ginecología y de Medicina Preventiva

Doctorado en Metodología de la Investigación Biomédica y Salud Pública

TESIS DOCTORAL

**CRIBADO POBLACIONAL DEL CÁNCER DE MAMA: ASPECTOS
RELACIONADOS CON LA EFECTIVIDAD Y COSTE-EFECTIVIDAD**

Doctoranda

Margarita Posso Rivera

Directores

Xavier Bonfill Cosp

Teresa Puig Reixach

Barcelona, Diciembre 2017

AGRADECIMIENTOS

Al Dr. Xavier Bonfill quien ha estado presente durante una gran parte de mi vida académica, primero en mi residencia y después durante el desarrollo de mi tesis doctoral. Ha sido un privilegio contar con su dirección y buen criterio.

A la Dra. Teresa Puig por su decisivo apoyo en este proceso y por ser un modelo a seguir en mi vida profesional. Gracias por su paciencia, dedicación y motivación durante todos estos años.

A todos mis compañeros del Servicio de Epidemiología Clínica y Salud Pública del Hospital de la Santa Creu i Sant Pau y del Centro Cochrane Iberoamericano. Siempre estuvieron ahí para alentarme en muchas etapas tanto de mi vida profesional como personal.

A todos los investigadores que han participado en este trabajo, junto a ellos he aprendido el valor del trabajo en equipo y sus aportes me han ayudado a llevar a cabo cada uno de los artículos con excelencia.

Al Dr. Xavier Castells, por sus valiosos consejos y por brindarme la oportunidad de continuar afianzando mi carrera profesional. A todos mis compañeros del Hospital del Mar a los que veo día a día, porque con su compañía me enseñan seguir trabajando en equipo con entusiasmo y buen humor.

A mis compañeros durante mis años de tesis, Dimelza, Mónica y Héctor por todo el tiempo que hemos compartido y las experiencias que hemos vivido que me han dado fuerza para recorrer este camino con una actitud positiva.

A mi pareja por su infinita paciencia y ánimos que han sido indispensables para alcanzar esta meta en mi carrera.

Finalmente, a mis padres y mis hermanos, que a pesar de estar a miles de kilómetros, me han dado su apoyo incondicional y su confianza animándome siempre a seguir adelante con todos mis proyectos.

“El científico no es aquella persona que da las respuestas correctas, sino aquél quien hace las preguntas correctas”

Claude Lévi-Strauss

ÍNDICE

Índice.....	VII
Índice de tablas	IX
Índice de figuras.....	IX
Abstract	3
Resumen	5
Resum	7
Introducción	11
1.1. El cáncer de mama.....	11
1.1.1. Problema prioritario en salud pública	11
1.1.2. Factores de riesgo.....	12
1.2. El cribado del cáncer de mama.....	16
1.2.1. Programas de cribado poblacional con mamografía	16
1.2.2. Balance entre riesgos y beneficios del cribado del cáncer de mama.....	19
1.2.3. La mamografía digital en el cribado	21
1.2.4. La doble lectura mamográfica en los programas de cribado	22
1.2.5. La detección de lesiones benignas de la mama y evaluación de sus biomarcadores	24
1.3. Las evaluaciones económicas en salud.....	26
1.3.1. Análisis de coste-consecuencia y coste-efectividad	27
1.3.2. Revisión sistemática de evaluaciones económicas.....	29
Justificación.....	37
Objetivos.....	41
3.1. Objetivos generales.....	41
3.2. Objetivos específicos	41
Métodos.....	45
4.1. Estudio 1. Análisis de coste-consecuencia de la doble lectura mamográfica	45
4.2. Estudio 2. Análisis de coste-efectividad de la doble lectura mamográfica	47
4.3. Estudio3. Revisión sistemática de la efectividad y coste-efectividad de la doble lectura mamográfica	49
4.4. Estudio 4. Expresión de biomarcadores en las lesiones benignas de la mama	51
Resultados.....	55
Artículos que conforman el trabajo de tesis	55
5.1. Artículo 1.....	56
5.1.1. Resumen de resultados.....	56
Publicación 1	59
Posso M, Puig T, Quintana MJ, Solà-Roca J, Bonfill X. Double versus single reading of mammograms in a breast cancer screeningn programme: a cost-consequence analysis. Eur Radiol. 2016; 26 (9): 3262-71.	61
5.2. Artículo 2.....	71

5.2.1. Resumen de resultados.....	71
Publicación 2.....	75
Posso M, Carles M, Rué M, Puig T, Bonfill X. Cost-effectiveness of double reading versus single reading of mammograms in a breast cancer screening programme. PlosOne. 2016; 11(7): e0159806	77
5.3. Artículo 3.....	91
5.3.1. Resumen de resultados.....	91
Publicación 3.....	95
Posso M, Puig T, Carles M, Rué M, Canelo-Aybar C, Bonfill X. Effectiveness and cost-effectiveness of double reading in digital mammography screening: a systematic review and meta-analysis. Eur J Rad. 2017; 96: 40-9.....	97
5.4. Artículo 4.....	107
5.4.1. Resumen de resultados.....	107
Publicación 4.....	111
Posso M, Corominas JM, Serrano L, Román M, Torá-Rocarmora I, et al. Biomarkers expression in benign breast disease and risk of subsequent breast cancer: a case-control study. Cancer Med. 2017; 6(6): 1482-89.....	117
Discusión.....	123
6.1. Efectividad de la doble lectura mamográfica	123
6.2. Coste-efectividad de la doble lectura mamográfica.....	125
6.3. Nuevas tecnologías y otras variables relacionadas con la efectividad y coste-efectividad del cribado	127
6.4. Lesiones benignas y personalización en el cribado	129
6.5. Limitaciones y fortalezas.....	131
6.6. Implicaciones para los programas de cribado.....	133
6.7. Implicaciones para la investigación.....	134
Conclusiones.....	139
Conclusions.....	143
Referencias	147
Anexos.....	159
Abreviaturas	159
Otras publicaciones relacionadas con el tema de tesis	161

ÍNDICE DE TABLAS

Tabla 1. Factores de riesgo no modificables asociados a un mayor riesgo de desarrollo de cáncer de mama.	14
Tabla 2. Factores de riesgo modificables asociados a un mayor riesgo de desarrollo de cáncer de mama.	15
Tabla 3. Coste estimado (Euros €, 2010) correspondiente a 57.157 participaciones según estrategias de la lectura.	60
Tabla 4. Efectos en salud estimados para 57.157 participaciones según estrategias de lectura.	61
Tabla 5. Cálculo de la relación de coste-efectividad incremental (RCEI) de las estrategias incluidas en el análisis.	77

ÍNDICE DE FIGURAS

Figura 1. Tasas estandarizadas por edad de incidencia y mortalidad por cáncer de mama en Europa. 1975-2010.	12
Figura 2. Efecto sobre la mortalidad del cribado con mamografía en mujeres de 50 a 69 años.	20
Figura 3. Representación gráfica de la hipótesis de evolución lineal del cáncer de mama.	25
Figura 4. Plano de coste-efectividad y umbral de decisión utilizados en las evaluaciones económicas.	28
Figura 5. Algoritmo seguido para realizar el análisis de coste-consecuencia.	47
Figura 6. Análisis combinado de la tasa de detección de cáncer con la doble lectura y con la lectura simple.	97
Figura 7. Expresión superior a 90% de receptores de estrógenos en una lesión no proliferativa de la mama.	112
Figura 8. Expresión superior a 80% de receptores de progesterona en una lesión no proliferativa de la mama.	112
Figura 9. Expresión inferior a 1% de Ki67 en una lesión proliferativa de la mama.	113



SUMMARY

ABSTRACT

Background

In breast cancer screening programmes, double reading of digital mammograms is used since it can detect more lesions. Some of these lesions are not aggressive and may be considered false positives or overdiagnosis. Therefore, it is relevant to evaluate the effectiveness and cost-effectiveness of double reading and to characterize benign lesions.

Objectives

To evaluate the effectiveness and cost-effectiveness of double reading of digital mammograms and to characterize the benign breast lesions detected in a breast cancer screening programme.

Methods

Four studies with different methodological designs were performed. To determine the effectiveness and cost-effectiveness of double reading of digital mammograms, we performed: i) a cost-consequence study; ii) a cost-effectiveness study; and iii) a systematic review of observational studies and economic evaluations. To determine whether the expression of estrogen (ER), progesterone (PR) and Ki-67 receptors in benign lesions increases the risk of cancer, iv) a case-control study was conducted.

Results

The cost-consequence study included 57,157 digital mammograms. The average cost of the screening process with double reading with consensus and arbitration was 15% higher (44.9 €) than with single reading (39.1 €). Double reading had 0.3% (N = 181) more false-positive results than single reading (P < 0.01). Both reading strategies had similar cancer detection rates (4.6/1000 and 4.2/1000 mammograms, respectively; P = 0.28) and similar positive predictive values (9.3% and 9.1%, respectively, P = 0.81).

In the cost-effectiveness study, 28,636 digital mammograms were analysed. The number of cancers detected was 148 (5.2/1000) with double reading, 138 (4.8/1000) with double reading in prevalent screening and single reading in incident screening, and 137 (4.8/1000) with single reading (P = 0.77). The average cost of detecting one cancer was 8,912 €, 8,614 € and 8,287 €, respectively. The incremental cost-effectiveness ratio (ICER) of double reading in prevalent screening and single reading in incident screening was 53,312 € per detected cancer. This strategy was dominated by extension so it was eliminated from the analysis. The ICER of double versus single reading was 16,684 €.

In the systematic review four studies with low risk of bias were included. Three of them were made in Europe and one in Japan. The pooled rate of cancer detection with double reading was 6.0/1000 mammograms and it was 5.7/1000 with single reading ($P = 0.76$). The proportion of false positives with double reading was 4.7% and 4.1% with single reading ($P = 0.12$). Single reading plus a diagnosis-aid system was cost-effective in Japan. No studies evaluating mortality, quality of life, or overdiagnosis were identified.

The case-control study included 258 benign lesions, 86 were cases and 172 were controls. Compared with benign lesions with low ER ($\leq 70\%$) expression, moderate (71% to 90%) and high ($> 90\%$) expression showed an increased risk of breast cancer (OR = 1.98 and OR = 2.63, respectively). Lesions with high PR ($> 80\%$) showed an increased risk of breast cancer (OR = 2.22) compared to those with low expression ($\leq 40\%$). Proliferative lesions with $\geq 1\%$ Ki67 expression did not show an increased risk of cancer (OR = 1.16).

Conclusion

In the era of digital mammography, double reading significantly increases the frequency of false positives without an important increase in the cancer detection rate. Double reading is not cost-effective when applied only in the prevalent screening and it also appears not to be cost-effective in the incident screening. Benign lesions with high ER or RP expression have a higher risk of developing breast cancer.

RESUMEN

Antecedentes

En los programas de cribado con mamografía digital, la doble lectura mamográfica es la estrategia de elección al ser más sensible en la detección de lesiones. Algunas de estas lesiones no son agresivas y pueden considerarse falsos positivos o sobrediagnóstico. Por tanto, es relevante evaluar la efectividad y coste-efectividad de la doble lectura y caracterizar las lesiones benignas.

Objetivos

Evaluar la efectividad y coste-efectividad de la doble lectura mamográfica en el contexto de la mamografía digital y valorar el comportamiento de lesiones benignas detectadas en un programa de cribado.

Métodos

Se realizaron cuatro estudios con diferentes diseños metodológicos. Para determinar la efectividad y el coste-efectividad de la doble lectura, se realizó un estudio de coste-consecuencia, uno de coste-efectividad y una revisión sistemática. Para determinar si la expresión de receptores de estrógeno (RE), progesterona (RP) y Ki-67 en las lesiones benignas aumenta el riesgo de cáncer, se realizó un estudio de casos y controles.

Resultados

El estudio de coste-consecuencia incluyó 57.157 mamografías digitales. El coste medio del proceso de cribado con la doble lectura con consenso y arbitraje fue un 15% superior (44,9 €) que con la lectura simple (39,1 €). La doble lectura tuvo un 0,3% (N = 181) más de resultados falsos positivos (P < 0,01). Ambas estrategias de lectura tuvieron tasas de detección de cáncer similares (4,6/1000 y 4,2/1000 mamografías, respectivamente; P = 0,28).

El estudio de coste-efectividad incluyó 28.636 mamografías digitales. Se detectaron 148 (5,2/1000) cánceres con la doble lectura, 138 (4,8/1000) con la doble lectura en el cribado prevalente y lectura simple en el cribado incidente y 137 (4,8/1000) con la lectura simple (P = 0,77). El coste medio de la detección de un cáncer fue 8.912 €, 8.614 € y 8.287 €, respectivamente. La relación de coste-efectividad incremental (RCEI) de la doble lectura en el cribado prevalente y lectura simple en el cribado incidente fue de 53.312 € por cáncer detectado. Esta estrategia fue dominada por extensión siendo excluida del análisis. La RCEI de la doble lectura frente a la lectura simple fue 16.684 €.

En la revisión sistemática se incluyeron estudios con bajo riesgo de sesgo, tres europeos y uno japonés. La tasa combinada de detección de cáncer con la doble lectura fue 6,0/1000 mamografías y 5,7/1000 con la lectura simple (P = 0,76). La proporción de falsos positivos con la doble lectura

fue 4,7% y 4,1% con la lectura simple ($P = 0,12$). La lectura simple más un sistema de ayuda al diagnóstico fue coste-efectiva en Japón. No se identificaron estudios sobre mortalidad, calidad de vida o sobrediagnóstico.

En el estudio de casos y controles se incluyeron 258 lesiones benignas; 86 casos y 172 controles. Comparado con las lesiones que presentaron una expresión baja de RE ($\leq 70\%$), la expresión moderada (71% a 90%) y alta ($> 90\%$) mostró un riesgo mayor de cáncer de mama (OR = 1,98 y OR = 2,63; respectivamente). Las lesiones con alta expresión de RP ($> 80\%$) mostraron un mayor riesgo de cáncer de mama (OR = 2,22) comparado con las de baja expresión ($\leq 40\%$). Las lesiones proliferativas con expresión $\geq 1\%$ de Ki67 no mostraron un riesgo mayor de cáncer (OR = 1,16).

Conclusiones

En el contexto de la mamografía digital, la doble lectura incrementa significativamente la frecuencia de falsos positivos y de forma más modesta la tasa de detección de cáncer. La doble lectura no es coste-efectiva cuando se aplica solo en el cribado prevalente y tampoco parece serlo en el cribado incidente. Las lesiones benignas con alta expresión de RE o RP tienen un mayor riesgo de desarrollar un cáncer de mama.

RESUM

Antecedents

Als programes de cribratge amb mamografia digital, la doble lectura mamogràfica és l'estratègia d'elecció en ser més sensible en la detecció de lesions. Algunes d'aquestes lesions no són agressives i poden considerar-se falsos positius o sobrediagnòstic. Per tant, és rellevant avaluar l'efectivitat i cost-efectivitat de la doble lectura i caracteritzar les lesions benignes.

Objectius

Avaluar l'efectivitat i cost-efectivitat de la doble lectura en el context de la mamografia digital i valorar el comportament de lesions benignes detectades en un programa de cribratge.

Mètodes

Es varen realitzar quatre estudis amb diferents dissenys metodològics. Per determinar l'efectivitat i el cost-efectivitat de la doble lectura, es va realitzar un estudi de cost-conseqüència, un de cost-efectivitat i una revisió sistemàtica. Per determinar si l'expressió de receptors d'estrogen (RE), progesterona (RP) i Ki-67 en les lesions benignes augmenta el risc de càncer, es va realitzar un estudi de casos i controls.

Resultats

L'estudi de cost-conseqüència va incloure 57.157 mamografies digitals. El cost mitjà del procés de cribratge amb la doble lectura amb consens i arbitratge va ser un 15% superior (44,9 €) que amb la lectura simple (39,1 €). La doble lectura va tenir un 0,3% (N = 181) més falsos positius (P < 0,01). Ambdues estratègies de lectura van tenir taxes de detecció de càncer similars (4,6/1000 i 4,2/1000 mamografies, respectivament; P = 0,28).

L'estudi de cost-efectivitat va incloure 28.636 mamografies digitals. Es van detectar 148 (5,2/1000) càncers amb la doble lectura, 138 (4,8/1000) amb la doble lectura en el cribratge prevalent i lectura simple en el cribratge incident i 137 (4,8/1000) amb la lectura simple (P = 0,77). El cost mitjà de la detecció d'un càncer va ser 8.912 €, 8.614 € i 8.287 €, respectivament. La relació de cost-efectivitat incremental (RCEI) de la doble lectura en el cribratge prevalent i lectura simple en el incident va ser de 53.312 € per càncer detectat. Aquesta estratègia va ser dominada per extensió sent exclosa de l'anàlisi. La RCEI de la doble lectura enfront de la lectura simple va ser 16.684 €.

En la revisió sistemàtica es van incloure estudis amb baix risc de biaix, tres europeus i un de japonès. La taxa combinada de detecció de càncer amb la doble lectura va ser 6,0/1000 mamografies i 5,7/1000 amb la lectura simple (P = 0,76). La proporció de falsos positius amb la doble lectura va ser 4,7% i 4,1% amb la lectura simple (P = 0,12). La lectura simple més un

sistema d'ajuda al diagnòstic va ser cost-efectiva a Japó. No es varen identificar estudis sobre mortalitat, qualitat de vida o sobrediagnòstic.

En l'estudi de casos i controls es van incloure 258 lesions benignes; 86 casos i 172 controls. Comparat amb les lesions que van presentar una expressió baixa de RE ($\leq 70\%$), l'expressió moderada (71% a 90%) i alta ($> 90\%$) va mostrar un risc major de càncer (OR = 1,98 i OR = 2,63; respectivament). Les lesions amb alta expressió de RP ($> 80\%$) van mostrar un major risc de càncer (OR = 2,22) comparat amb les de baixa expressió ($\leq 40\%$). Les lesions proliferatives amb expressió $\geq 1\%$ de Ki67 no van mostrar un risc major de càncer (OR = 1,16).

Conclusions

En el context de la mamografia digital, la doble lectura incrementa significativament la freqüència de falsos positius i de forma més modesta la taxa de detecció de càncer. La doble lectura no és cost-efectiva quan s'aplica solament en el cribratge prevalent i tampoc sembla ser-ho en el cribratge incident. Les lesions benignes amb alta expressió de RE o RP tenen un major risc de desenvolupar un càncer de mama.



INTRODUCCIÓN

INTRODUCCIÓN

1.1. EL CÁNCER DE MAMA

1.1.1. Problema prioritario en salud pública

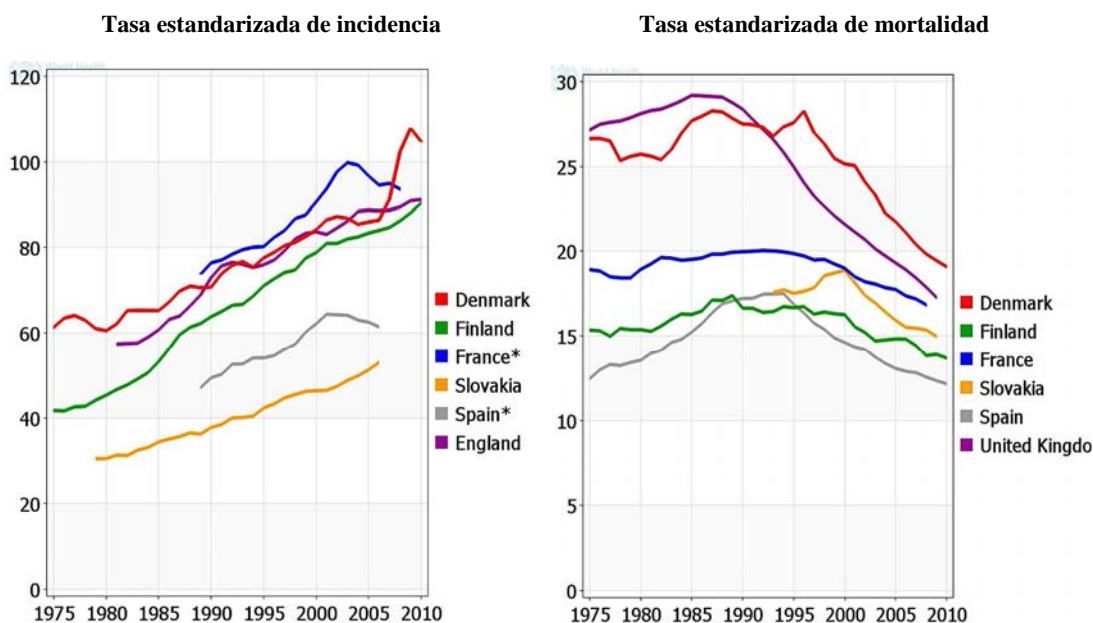
A nivel mundial el cáncer de mama representa un problema prioritario para la salud pública. Es el segundo cáncer más frecuentemente diagnosticado en el mundo y el primero entre las mujeres ¹. Las cifras más actuales publicadas en 2012 estiman una tasa de incidencia ajustada por edad de 43,1 por 100.000 mujeres lo que representó el diagnóstico de 1,7 millones de casos nuevos (25% de todos los cánceres) ¹.

Durante las décadas de los 80, 90 y 2000, particularmente en los países del norte de Europa y en Estados Unidos, se ha observado un incremento en la incidencia del cáncer de mama ². Diversos factores se han propuesto para explicar esta tendencia creciente y actual estabilización. Destacan los cambios en los estilos de vida, reproductivos y hormonales, el envejecimiento de la población y las mejoras en los sistemas de registro y técnicas de detección precoz. También se ha sugerido que una parte de este incremento estaría explicada por el sobrediagnóstico de casos que de no ser por el cribado no se hubieran manifestado clínicamente ³.

Además de tener una incidencia alta en Europa (367.090 mujeres diagnosticadas en 2012), este cáncer también destaca por haber presentado una clara reducción en la mortalidad durante las últimas décadas (Figura 1) y una alta supervivencia. Según datos recogidos por EURO-CARE-5, la supervivencia a 5 años de mujeres europeas diagnosticadas durante el período 2000-2007 fue de 81,3% ⁴. Así, en su conjunto la carga que representa esta enfermedad ocupa el cuarto lugar entre los cánceres con mayor carga de morbilidad en Europa ⁵.

España, con aproximadamente 27.700 casos nuevos durante el año 2015, se sitúa entre los países con menor incidencia a nivel europeo ⁶. Ya desde principios de los años 80 presentaba una de las incidencias más bajas con una tasa ajustada por edad de 54,7 por 100.000 mujeres/año ⁷. En el curso de los años 90 y 2000, coincidiendo con la implementación de los programas de cribado, esta cifra se incrementó en un 20% a 30% ⁸ hasta llegar a 85,5 por 100.000 mujeres/año durante el período 1998-2002 ⁶. Una vez que se alcanzó la cobertura completa del cribado la incidencia se ha estabilizado alrededor de 87,3 y 88,3 por 100.000 mujeres/año durante el período 2003-2015 ⁶.

Figura 1. Tasas estandarizadas por edad de incidencia y mortalidad por cáncer de mama en Europa. 1975-2010.



*Información regional
NORDCAN (www.ancr.nu)
ECO (eco.iarc.fr)
England: www.ons.gov.uk

WHO (www.who.int/healthinfo/en/)

Fuente: CANCER FACT SHEETS: BREAST CANCER ⁶. <http://eco.iarc.fr/today/data/pdf/fact-sheets/cancers/cancer-fact-sheets-15.pdf>

Con la mortalidad y la supervivencia sucede algo similar. La tasa de mortalidad ajustada de 11,8 casos por 100.000 mujeres/año estimada en el año 2012 es una de las más bajas en Europa¹ y la supervivencia a 5 años de 82,8%⁴ es más alta que la estimada en mujeres europeas diagnosticadas en el periodo 2000-2007. Estos buenos resultados se han atribuido a múltiples factores, destacando la detección más temprana de los tumores y la mejora de los tratamientos disponibles para este cáncer.

1.1.2. Factores de riesgo

Gracias al gran esfuerzo que se dedica a la investigación en cáncer de mama se conoce en gran parte su historia natural y numerosos factores de riesgo que influyen en su aparición (Tabla 1 y 2).

Sin embargo, aproximadamente en el 75% de las mujeres no se encuentra ningún factor de riesgo determinante.

A los antecedentes genéticos se les atribuye entre 5% y 10% de todos los cánceres de mama ^{9,10}. Así, las mutaciones en los genes BRCA1 y BRCA2 continúan siendo la principal causa de cáncer de mama y de ovario hereditarios ¹¹. Otro factor importante y ampliamente estudiado es la presencia de lesiones benignas de la mama. Estas lesiones se han evaluado por su valor como potenciales precursores de tumores ¹² pero también por su potencial significado como marcadores de mayor riesgo de cáncer ¹³. De hecho, actualmente se conoce que la aparición de alguna lesión benigna de la mama representa un incremento de entre 1,5 y 4 veces en el riesgo de cáncer respecto a la población general ¹⁴.

La edad, el origen étnico y el nivel socioeconómico también se han evaluado, encontrándose que el 75% de todos los casos ocurre en mayores de 50 años, con predominio en las clases económicas menos favorecidas y poblaciones de origen caucásico ¹⁵.

Algunos factores hormonales y reproductivos como la menarquia temprana, la menopausia tardía, no tener hijos o tenerlos después de los 30 años de edad, la utilización de terapia de reemplazo hormonal durante la menopausia especialmente de forma prolongada han sido propuestos como factores de riesgo de cáncer de mama ¹⁶⁻¹⁹.

Tabla 1. Factores de riesgo no modificables asociados a un mayor riesgo de desarrollo de cáncer de mama.

Factor	Estimación de riesgo	Potencial vía de carcinogénesis	Referencia
Mutaciones genéticas (BRCA1 y BRCA2)	15-200 más riesgo	Mutación autosómica dominante	Robson2007 ²⁰ , Foulkes2008 ²¹
Historia personal de cáncer de ovario o endometrio	> 5 más riesgo. Varía de acuerdo con las líneas germinales	Mutaciones germinales	Collaborative Group2001 ²²
Edad ≥ 65	5,8 más riesgo. Incremento lineal a partir de los 50 años	Multifactorial relacionado con el daño del ADN	NCI-SEER2016 ²³
Tratamiento con radioterapia	5,2 más riesgo. Incremento lineal de acuerdo a la dosis	Daño del ADN	Henderson2010 ²⁴ , Clemons2000 ²⁵
Densidad mamaria	4,7 más riesgo si la densidad es > 75% vs < 10%	Aumento del tejido glandular	Boyd2007 ²⁶
Lesiones benignas de la mama	1,9 a 5,3 más riesgo en lesiones hiperplásicas	Presencia de células con potencial proliferativo	Nelson2012 ²⁷ , Dupont1987 ²⁸ , Dyrstad2015 ¹⁴
Historia familiar de cáncer de mama	1,8 a 3,6 más riesgo si familiar de primer grado o múltiples	Incierto	Colditz1993 ²⁹ , Pharoah1997 ³⁰
Primer embarazo > 30 años	1,7 a 1,9 más riesgo si > 30 años vs. < 20 años	Diferenciación de células protectoras en el embarazo	Hulka1997 ³¹
Densidad ósea	1,6 a 1,8 más riesgo en mujeres con mayor densidad	Marcador de estrógeno circulante	Zhang1997 ³² , Qu2013 ³³
Menarquia < 12 años	1,3 más riesgo si < 12 años vs. ≥ 15 años	Exposición prolongada a hormonas	Brinton1988 ³⁴ , Hulka1997 ³¹
Menopausia > 55 años	1,2 a 1,5 más riesgo si ≥ 45 años vs. < 45 años	Exposición prolongada a hormonas	Brinton1988 ³⁴
Origen étnico	Mayor incidencia en mujeres blancas y negras no hispanas	Multifactorial (genética y ambiental)	NCI-SEER2016 ²³
Nivel socioeconómico	Mayor incidencia en niveles bajos	Multifactorial (falta de acceso)	NCI-SEER2016 ²³ , Harper2009 ³⁵

Adaptado a partir de Peairs et al. Screening for breast cancer. Seminars in Oncology, 2017; 44(1): 60-72³⁶.

Otros factores como la exposición a la radiación, particularmente en el tórax y durante la infancia ^{24,25}, la obesidad ³⁷, una dieta con exceso de ácidos grasos omega 6 y el consumo prolongado de alcohol han sido relacionados con la aparición del cáncer ³⁸. Debido a la inexistencia de una prevención primaria eficaz, las estrategias de detección precoz cumplen un papel fundamental en el control del cáncer. Hoy en día es ampliamente aceptado que los programas de detección precoz por un lado y los avances terapéuticos por otro, han contribuido favorablemente al descenso paulatino en la mortalidad.

Tabla 2. Factores de riesgo modificables asociados a un mayor riesgo de desarrollo de cáncer de mama.

Factor	Estimación de riesgo	Potencial vía de carcinogénesis	Referencia
Tratamiento de reemplazo hormonal	1,3 si usuario habitual > 5 años vs. nunca	Exposición a hormonas	Nelson2002 ³⁹
Lactancia materna	1,3 más riesgo si no lactancia vs. > 16 meses	La diferenciación celular se produce en embarazo y lactancia	Enger1998 ⁴⁰
Consumo de alcohol	1,2 más riesgo si 2 bebidas al día vs. No bebedor	Estimula el metabolismo carcinógeno	Ellison2001 ⁴¹
Obesidad antes de la menopausia	1,2 más riesgo si percentil de obesidad superior a 80 th vs. percentil 20	El tejido adiposo es fuente de estrógeno, insulina y factor de crecimiento	Tretli1989 ⁴² , Nelson2012 ²⁷
Tabaquismo	1,1 más riesgo	Incierto	Gaudet2013 ⁴³ , Gram2015 ⁴⁴
Sedentarismo	Incremento en la actividad física puede ser protector	La actividad física reduce el factor de crecimiento, insulina y estrógeno	Pizot2016 ⁴⁵ , Wu2013 ⁴⁶

Adaptado a partir de Peairs et al. Screening for breast cancer. *Seminars in Oncology*, 2017; 44(1): 60-72 ³⁶.

1.2. EL CRIBADO DEL CÁNCER DE MAMA

1.2.1. Programas de cribado poblacional con mamografía

En 1968, la Organización Mundial de la Salud publicó la monografía de Wilson y Jungner sobre los Principios y Procedimientos para evaluar la implementación de los programas de cribado ⁴⁷. Desde entonces, los 10 principios (Box 1) propuestos en esta monografía han ido ganando reconocimiento, se han revisado y actualizado al contexto actual ^{48,49} y su uso se ha extendido a nivel mundial.

Box A. Resumen de los criterios propuestos por Wilson y Jungner

1. *La enfermedad debe ser un problema de salud importante.*
2. *Debe existir un tratamiento aceptado para los pacientes.*
3. *Deben existir instalaciones para diagnóstico y tratamiento.*
4. *La enfermedad debe presentar una fase latente reconocible.*
5. *Debe existir una prueba de cribado adecuada.*
6. *La prueba debe ser aceptable para la población.*
7. *La historia natural de la enfermedad debe ser conocida.*
8. *Debe existir una política sobre a quienes tratar como pacientes.*
9. *El coste de cada caso detectado debe ser equilibrado en relación con los costes totales de la atención de la enfermedad.*
10. *El cribado debe ser un proceso continuo.*

En el caso concreto del cáncer de mama, se puede aceptar que esta enfermedad cumple con todos los criterios clásicamente propuestos por Wilson y Jungner lo que favorece el despliegue de programas de detección precoz mediante el cribado mamográfico. Particularmente, su historia natural presenta una fase preclínica de aproximadamente 2 años ⁵⁰, por lo que el cribado bienal o trienal sería recomendable.

Actualmente la mamografía es la prueba recomendada como test de cribado debido a que permite identificar lesiones sospechosas de manera precoz ⁵¹. Se ha reportado una sensibilidad de la mamografía que varía con la edad entre un 86% en el grupo de 40 a 49 años y un 94% en el grupo de 50 a 69 años ^{52,53} mientras que la especificidad supera el 80% ^{52,53}. El valor predictivo positivo (VPP) de la mamografía depende de muchos factores como son: la calidad de la radiografía, la

experiencia del radiólogo, la edad de la mujer y los signos mamográficos presentes, entre otros. Un estudio español ha reportado que el VPP de la prueba puede variar entre el 15% y el 40% ⁵⁴.

El cribado poblacional del cáncer de mama mediante mamografía es una estrategia de detección precoz que tiene como principal objetivo la reducción de la mortalidad por esta causa. Los programas de detección precoz de cáncer se basan en que el tratamiento de la enfermedad es más efectivo cuando ésta es detectada en sus primeras fases de desarrollo.

En Europa los programas de cribado mamográfico de base poblacional se instauraron a finales de los 80 e inicios de los 90, inicialmente a nivel regional, extendiéndose luego a nivel nacional con coberturas que actualmente superan el 90% ⁵⁵. Durante los años 2000 a 2010, aproximadamente la mitad de los programas alcanzaron el 70% de participación recomendado en las Guías europeas para la garantía de la calidad en la detección y el diagnóstico del cáncer de mama ⁵⁵.

En España, aunque ya existían algunas iniciativas como el programa del Ayuntamiento de Barcelona de 1987, el primer programa de cribado del cáncer de mama implementado a nivel poblacional se fue el de Navarra de 1990 ⁵⁶. Este primer programa consideraba como población diana a mujeres con edades entre 50 y 65 años, posteriormente este rango de edad se amplió hasta los 69 años. En el resto de España nuevos programas fueron puestos en marcha de forma progresiva. En el año 2006 se completó la extensión a todas las comunidades autónomas ofreciendo el cribado al 100% de las mujeres del grupo de edad de 50 a 64 años, al 48% del grupo 65 a 69 de años y al 23% del grupo 45 a 49 años ⁵⁷.

Actualmente todas las comunidades autónomas reconocen y utilizan las Guías europeas como documento de referencia y organizan sus esfuerzos para alcanzar una participación del 70%, límite establecido como mínimo para obtener un efecto a nivel poblacional ⁵⁸. En Cataluña, dentro de las líneas estratégicas del Departamento de Salud, se cuenta con el Plan Director de Oncología que elabora periódicamente el Plan Contra el Cáncer. Uno de los principales objetivos establecidos para el periodo 2015-2020, una vez alcanzada la cobertura geográfica del 100% y una participación próxima al 70%, es mejorar la calidad del cribado de cáncer de mama y reducir su variabilidad ⁵⁹.

El cribado se ofrece individualmente a la totalidad de la población diana y es financiado por el Sistema Nacional de Salud. Se estima que aproximadamente cuatro millones de mujeres son invitadas anualmente a participar en los programas de cribado poblacional. Con una frecuencia bienal se invita a mujeres entre los 50 y los 69 años, aunque existe cierta variabilidad en la edad de inicio, la cual puede ser 45 años en algunos programas ^{57,58}.

También de acuerdo con las Guías europeas ^{51,60}, se realizan dos proyecciones mamográficas, una cráneo caudal y otra medio lateral. Las imágenes son valoradas habitualmente por dos radiólogos entrenados siguiendo un proceso de doble lectura que termina en un único resultado de la mamografía ⁵⁷. Cuando el resultado de la mamografía es negativo, la mujer es invitada a realizar una nueva mamografía a los dos años. En el que caso que el resultado de la mamografía sea

positivo o exista sospecha de una lesión, se realizan pruebas complementarias para confirmar o descartar la existencia de cáncer.

Entre las pruebas complementarias que pueden realizarse se cuenta con procedimientos de tipo no invasivos como son las mamografías adicionales y las ecografías. Los procedimientos invasivos habitualmente son las biopsias mínimamente invasivas y las biopsias quirúrgicas. En España, durante el periodo 2007-2008, el 4,9% de las participantes en los programas de cribado se sometieron a pruebas adicionales de cualquier tipo y el 0,7% a pruebas de tipo invasivo⁵⁸.

Más de cuarenta años han pasado desde la publicación de los criterios de Wilson y Jungner y la actualización de los mismos obliga a revisar la efectividad de los programas en base a la evidencia científica disponible^{48,49}. En el año 2011, Harris et al.⁴⁹ formularon las siguientes preguntas como método para cuestionar el valor actual de los programas de cribado.

- i) Si se aplicara en las condiciones actuales, ¿el programa de cribado ofrecería un beneficio neto suficiente que justifique su inicio (o continuación) en relación con los recursos requeridos?
- ii) Si la respuesta es afirmativa, ¿qué se necesita para optimizar la implementación del programa?

Box B. Resumen de las consideraciones para estimar los beneficios y los riesgos de un programa de cribado

Magnitud de los beneficios potenciales

- 1. Probabilidad de reducir los daños que ocurrirían sin el cribado*
- 2. Probabilidad de identificar correctamente a aquellas personas que sufrirían daños sin el cribado*
- 3. Magnitud de los beneficios para la salud de los tratamientos en estadios precoces frente a estadios avanzados*

Magnitud de los riesgos potenciales

- 1. Frecuencia de los exámenes de detección falsos positivos*
- 2. Experiencia de personas con resultados falsos positivos*
- 3. Frecuencia de sobrediagnóstico*
- 4. Experiencia de personas con sobrediagnóstico*
- 5. Frecuencia y gravedad de los daños del tratamiento y tratamiento*

Adaptado a partir de Harris et al. Reconsidering the Criteria for Evaluating Proposed Screening Programs: Reflections From 4 Current and Former Members of the U.S. Preventive Services Task Force. *Epidemiologic Reviews*. 2011; 33 (1):20-35⁴⁹.

En el caso de los programas de cribado del cáncer de mama la respuesta a estas preguntas pasa por evaluar el balance entre sus riesgos y beneficios (Box 2) y por sopesar el valor neto de los beneficios obtenidos en relación con los recursos utilizados.

1.2.2. Balance entre riesgos y beneficios del cribado del cáncer de mama

Los beneficios del cribado del cáncer de mama mediante mamografía han sido ampliamente estudiados en diversos ensayos clínicos aleatorizados ⁶¹⁻⁷⁷ y posteriormente en los metanálisis realizados a partir de sus resultados ^{51,78-82}. Los estudios fueron desarrollados en varios países e incluyeron más de 500.000 mujeres. Se diferenciaron con respecto al periodo en que se realizaron, el tipo de mamografía utilizada, el intervalo entre exámenes mamográficos, el método de asignación de mujeres a los grupos examinados, el número de visitas de cribado, la edad de las mujeres y los métodos de análisis.

A pesar de las diferencias entre los ensayos clínicos, los metanálisis coinciden en mostrar una reducción de la mortalidad por cáncer de mama de alrededor del 20% con los programas de cribado. Esta reducción muestra variaciones de acuerdo a la edad, los años de seguimiento, el número de mujeres incluidas y la periodicidad de la mamografía. Particularmente, en el grupo de mujeres de 50 a 69 años de edad, la reducción es estadísticamente significativa independientemente del tipo de metanálisis utilizado -incluyendo o no a los estudios de mayor riesgo de sesgo y/o utilizando una aproximación de efecto aleatorio o efecto fijo-. ^{51,78-82}.

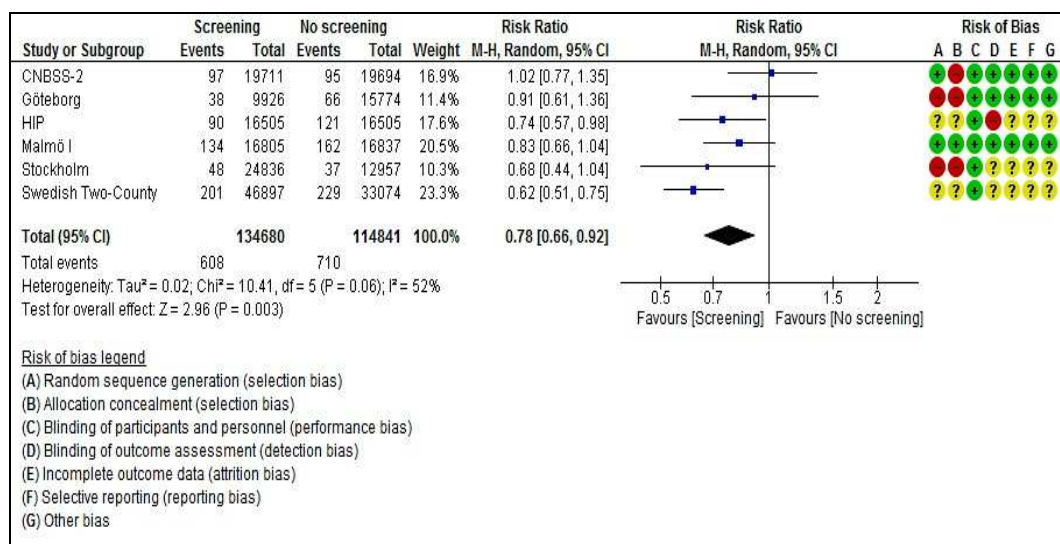
En este grupo de 50 a 69 años, se cuenta con seis ensayos clínicos ⁶¹⁻⁷¹ que tuvieron una duración media de 6,3 años, incluyeron a 249.160 mujeres y proporcionaron información sobre las muertes observadas en las mujeres diagnosticadas durante el periodo de cribado (short case accrual method). El metanálisis de efecto aleatorio mostró una reducción de la mortalidad del 22% (Figura 2). Para traducir esto en términos absolutos debemos considerar el riesgo basal de la población. Así, una reducción del 22% se traduce en 132 a 462 muertes evitadas por cada 100.000 mujeres invitadas al cribado y seguidas durante 17,6 años, si asumimos que el riesgo basal es de 0,6% (riesgo reportado en los ensayos clínicos) a 2,1% (riesgo estimado para la población actual), respectivamente ⁵¹.

Además de su efecto sobre la mortalidad por cáncer de mama, la detección precoz ofrece otros beneficios, por ejemplo, la opción de realizar un tratamiento más conservador ⁸³, la mejora en el pronóstico y el efecto positivo de saber que una prueba negativa se asocia con un riesgo reducido de cáncer de mama.

A pesar de los múltiples beneficios de los programas de cribado del cáncer de mama, desde el momento en que se establecieron ha existido un intenso debate en relación a sus posibles daños. Clásicamente se han descrito consecuencias negativas en las mujeres que son clasificadas como falsos positivos⁸⁴ y más recientemente el debate gira en torno al sobrediagnóstico.

El sobrediagnóstico se ha definido como la detección de lesiones, incluidos algunos carcinomas, que debido a su historia natural nunca se habrían manifestado clínicamente durante la vida de la mujer^{79,85}. Los efectos negativos derivados del sobrediagnóstico son frecuentes y pueden afectar de forma significativa la calidad de vida de las mujeres sometidas a pruebas diagnósticas invasivas o tratamientos quirúrgicos innecesarios.

Figura 2. Efecto sobre la mortalidad del cribado con mamografía en mujeres de 50 a 69 años.



Metanálisis desarrollado por el grupo responsable de la revisión sistemática para la recomendación sobre el cribado en mujeres de 50 a 69 años. JRC. Recommendations on Breast Cancer Screening. 2017. <http://ecibc.jrc.ec.europa.eu/recommendations>⁵¹.

Un factor importante que se debe tener en cuenta es que el cribado no beneficia por igual a todas las mujeres, por ejemplo, las que presentan tumores más agresivos o de rápido crecimiento tienen más riesgo de no ser detectadas en los cribados bienales, siendo diagnosticadas muchas veces como cánceres de intervalo^{86,87}. En contraste, aquellas mujeres con tumores muy poco agresivos, de muy lento crecimiento y que quizás nunca se manifiesten clínicamente, corren el riesgo de ser sobrediagnosticadas y el consiguiente sobretratamiento.

También se deben considerar otros daños potenciales como son los falsos positivos, los falsos negativos, la exposición a la radiación, los efectos psicológicos negativos en términos de ansiedad durante el tiempo de espera de el informe de la mamografía y la falsa tranquilidad en las mujeres diagnosticadas posteriormente de un cáncer de intervalo^{51,88}.

Respecto a los falsos positivos, un estudio realizado en base a 4.739.498 mamografías del programa de cribado en España, ha estimado que la tasa acumulada de falsos positivos en mujeres que iniciaron el cribado entre 50 y 51 años fue de 20,4% ⁸⁹. Una cifra similar se ha reportado en un estudio realizado en base al programa poblacional noruego, donde se estimó una tasa acumulada de falsos positivos de mujeres entre 50 y 69 años de 20,8% ⁹⁰. Elmore et al, reportaron un número mayor de falsos positivos con una tasa de 47,3% después de analizar un total de 10 cribados regulares en mujeres de 50 a 79 años en Estados Unidos ⁹¹.

El impacto de los falsos positivos ha sido estudiado y se ha valorado como uno de los principales problemas del cribado. Sus consecuencias se derivan en una alta frecuencia de evaluaciones adicionales que incluyen biopsias con resultados negativos. La incomodidad en la mujer sometida innecesariamente a estas pruebas complementarias puede conducir a una disminución en la probabilidad de participar rutinariamente del cribado. Además, es importante tener en cuenta el incremento en los costes que todo esto supone ^{52,92}.

Por tanto, debido al delicado balance entre beneficios y daños del cribado, existe la necesidad constante de seguir investigando y evaluando estrategias que mejoren este balance en beneficio de las mujeres participantes en los programas de cribado.

1.2.3. La mamografía digital en el cribado

En el contexto del cribado del cáncer de mama, permanentemente se investiga sobre nuevas técnicas radiológicas que demuestren una mayor sensibilidad y especificidad. Gracias a este esfuerzo, hoy en día la mamografía de tipo digital ha remplazado a la mamografía analógica. La mamografía digital ha demostrado ventajas en la obtención, despliegue y almacenamiento de las imágenes ⁹³. La tecnología digital permite ajustar el contraste y mejorar la calidad de las imágenes.

Comparado con la mamografía analógica, la mamografía digital ha demostrado una mayor precisión diagnóstica, particularmente en mujeres jóvenes, con mamas densas ⁹⁴ y presencia de microcalcificaciones en las imágenes ⁹⁵. La tasa de detección de neoplasias en la población general del cribado es similar entre ambas técnicas ⁹⁶⁻⁹⁸ con la ventaja que la tasa de recitación y la proporción de falsos positivos es menor con la mamografía digital ^{99,100} y sin ocasionar cambios en la tasa de cánceres de intervalo ^{98,99}.

Respecto a los costes que representa el cribado con mamografía digital, teniendo en cuenta las ventajas señaladas anteriormente, es razonable proponer que su uso represente un ahorro a largo plazo. Así lo confirma un estudio europeo basado en modelos de simulación de eventos discretos que reportó unos costes significativamente más bajos con la mamografía digital que con la analógica ¹⁰¹. En este estudio, la reducción de costes se produjo a expensas de reducir los costes de

pruebas adicionales y de tratamiento. El ahorro total fue de 2.866.124 euros al año 20 del horizonte temporal, lo que representó el 8,1% del coste total con mamografía analógica ¹⁰¹.

Aunque la mamografía digital se está utilizando en la mayoría de países europeos, la tomosíntesis también está siendo continuamente evaluada, sobretodo en mujeres con mamas densas. En este grupo de mujeres la tomosíntesis parece ser superior a la mamografía aunque actualmente la evidencia no es concluyente y las Guías europeas recientemente publicadas no se posicionan ni a favor ni en contra de su uso en el cribado por lo que la mamografía digital continúa siendo la técnica de elección ⁵¹.

Las estrategias para realizar la lectura de las mamografías también han sido evaluadas en varios países europeos. Debido a que ha demostrado incrementar la sensibilidad para la detección de neoplasias, en la mayoría de los países europeos se realiza la doble lectura mamográfica ⁵⁵. Sin embargo, vale la pena señalar que estas estrategias fueron principalmente estudiadas en el contexto de la mamografía analógica siendo, hasta la fecha, muy escasa su evaluación en el contexto de la mamografía digital.

1.2.4. La doble lectura mamográfica en los programas de cribado

La sensibilidad y especificidad de la mamografía condicionan la efectividad de los programas de cribado. Respecto a la sensibilidad, un valor por debajo del 70% podría limitar la eficacia del cribado. Mientras que un valor bajo en la especificidad podría conducir a un incremento de efectos indeseables como los falsos positivos, las evaluaciones invasivas innecesarias, etc.

Una de las estrategias que se han adoptado para mejorar la sensibilidad del cribado con mamografía es la doble lectura. Así, la doble lectura mamográfica es la estrategia recomendada por las Guías europeas para la garantía en el cribado ⁶⁰.

La doble lectura consiste en que la valoración de cada imagen mamográfica la realizan dos lectores, frecuentemente radiólogos. Esta valoración puede ser independiente cuando el segundo radiólogo desconoce el informe del primero o no independiente cuando el segundo radiólogo tiene acceso al informe del primero.

La doble lectura no independiente otorga la ventaja de contar con un resultado de la mamografía inmediatamente posterior a la valoración del segundo lector. No obstante, su principal desventaja es la menor sensibilidad comparado con la doble lectura independiente ¹⁰². Es por esto que en la mayoría de programas europeos se ha implementado la doble lectura independiente.

Por su parte, la doble lectura independiente tiene como desventaja que en algunos casos no se puede contar con un resultado inmediato tras la segunda valoración del lector. Aunque en la

mayoría de los casos ambos radiólogos coinciden en sus lecturas, existen situaciones en que las valoraciones son discordantes. Por ejemplo, uno de los radiólogos puede identificar una lesión como sospechosa y el otro no, en estos casos deben adoptarse diversas estrategias para resolver esta discordancia.

Una primera estrategia para resolver las discordancias entre los lectores consiste en recitar a todas las mujeres en quienes se presente la discordancia con el objetivo de realizarles nuevos exámenes. Aunque esta estrategia ha demostrado mejorar la sensibilidad, ha sido desestimada debido al gran incremento de exámenes complementarios con resultados negativos ¹⁰³.

Una segunda estrategia es realizar una sesión de consenso entre los dos radiólogos que han evaluado la mamografía para que realicen una valoración única. Esta estrategia permite incrementar la sensibilidad controlando el número de recitaciones por lo que se considera como superior a la doble lectura sin consenso ¹⁰⁴.

Finalmente, una tercera estrategia utilizada para resolver los problemas de discordancia es el arbitraje, que consiste en la participación de un tercer radiólogo (o un equipo de radiólogos) quien se encarga de realizar la valoración definitiva de la mamografía. La efectividad del arbitraje también ha sido evaluada reportando una menor tasa de recitación en comparación con la doble lectura sin consenso ¹⁰⁵⁻¹⁰⁸.

En general todas las estrategias de doble lectura que se han detallado previamente han mostrado tener una mayor sensibilidad que la lectura simple. Sin embargo, el incremento en la sensibilidad se acompaña necesariamente de una reducción en la especificidad y el consiguiente incremento de potenciales daños. En este sentido, la doble lectura es cuestionada debido a su potencial para incrementar los efectos indeseables del cribado. La doble lectura incrementa el número de mujeres que son recitadas para realizar pruebas adicionales y por consiguiente también incrementa el riesgo de falsos positivos, sobrediagnóstico y los costes.

Es razonable plantear que la utilidad de la doble lectura podría quedar restringida a situaciones en las que existe una gran discordancia entre ambos lectores como son: la primera participación de la mujer en el cribado donde no se cuenta con una mamografía previa, las lesiones difíciles de identificar en la mamografía y la participación en la lectura de radiólogos poco experimentados ^{103,109}.

Otro punto importante a tener en cuenta es que los estudios que han evaluado la efectividad y coste-efectividad de la doble lectura realizaron el análisis en el contexto de la mamografía analógica ¹⁰³. Actualmente, la tecnología digital ha desplazado a la analógica y la evaluación de la doble lectura en este nuevo contexto es una tarea pendiente.

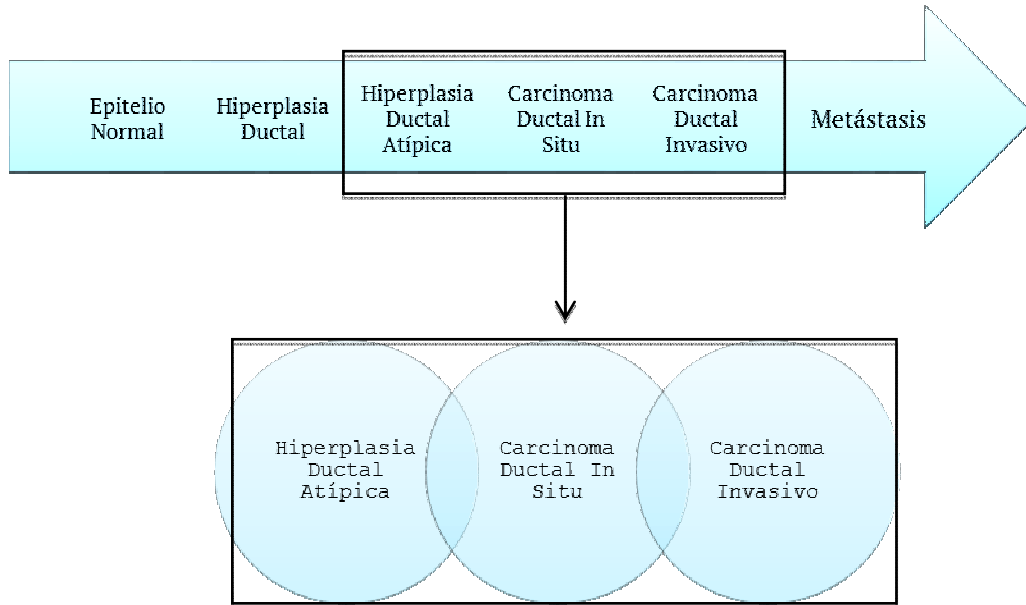
1.2.5. La detección de lesiones benignas de la mama y evaluación de sus biomarcadores

La introducción del cribado con mamografía digital ha generado un incremento en la detección de lesiones benignas de la mama ¹¹⁰⁻¹¹². Existen una serie de hallazgos mamográficos específicos que al ser considerados como sospechosos de malignidad conducen a la realización de pruebas diagnósticas complementarias ¹¹³. La biopsia de mama se considera el patrón de oro debido a que permite establecer la verdadera naturaleza de la lesión. Por lo tanto, es una de las pruebas diagnósticas que más comúnmente se emplea para evaluar estos hallazgos mamográficos. Esto implica que además de la detección de cánceres, en los programas de cribado se pueden identificar mujeres asintomáticas con lesiones benignas.

Las lesiones benignas de la mama constituyen un grupo heterogéneo de lesiones en el que se incluyen fibroadenomas, quistes, fibrosis, papilomas, adenosis e hiperplasias con o sin atipia, entre otros ¹¹⁴. A menudo tanto a nivel científico como clínico estas lesiones se clasifican en tres grupos i) lesiones no proliferativas; ii) proliferativas sin atipia y iii) proliferativas con atipia. Esta clasificación se basa en las características de anatomía patológica de las lesiones y en el riesgo posterior de cáncer estudiados y propuestos por Page y Dupont en la década de los 80 ^{28,115,116}.

Todavía hoy en día la clasificación de Dupont y Page sigue vigente. Una revisión sistemática ¹⁴ recientemente publicada utilizó esta clasificación y estimó el riesgo de cáncer asociado a la presencia de lesiones benignas. Los resultados confirmaron que las lesiones no proliferativas no están asociadas a un incremento importante del riesgo mientras que las lesiones proliferativas presentaron un riesgo significativo y variable dependiendo de la presencia o no de atipia ¹⁴. Las lesiones proliferativas sin atipia se asociaron a un riesgo relativo [RR] de 1,76 (intervalo de confianza [IC] del 95% = 1,58 a 1,95) mientras que las lesiones proliferativas con atipia se asociaron a un RR de 3,93 (IC 95% = 3,24 a 4,76) ¹⁴ (Figura 3).

Figura 3. Representación gráfica de la hipótesis de evolución lineal del cáncer de mama.



El carcinoma ductal *in situ* (DCIS) representa un estado intermedio entre la hiperplasia ductal atípica y el cáncer invasivo. En algunos casos el diagnóstico diferencial entre el DCIS y la hiperplasia ductal atípica pueden coexistir. Modificado de Sakorafas and Tsiotou. Cancer Treatment Reviews. 2000; 26: 103–125 ¹¹⁷.

Con el fin de comprender mejor las características biológicas de las lesiones benignas, algunos autores se han centrado en los receptores de estrógenos (ER) ^{118,119}, el receptor de progesterona (PR) ¹²⁰, y Ki67, ^{118,121,122} como biomarcadores expresados en las lesiones benignas. El uso de estos biomarcadores en la práctica clínica común nació debido a la relevancia que tienen con respecto al pronóstico y tratamiento de las mujeres con cáncer de mama ^{123,124}.

Los trabajos pioneros sobre la clasificación molecular del cáncer de mama publicados por Perou y Sorlie et al ¹²³ a comienzos de los años 2000 utilizando perfiles globales de expresión genética identificaron 5 subtipos moleculares de cáncer de mama invasivo: luminal A, luminal B, de tipo normal breast-like, con expresión de HER2 y de tipo basal-like. Cada uno con características bien definidas y comportamiento diferente respecto a su incidencia, supervivencia y respuesta al tratamiento.

Debido a que la evaluación de la expresión genética en la práctica diaria no es factible, en la actualidad, se utilizan biomarcadores inmunohistoquímicos como sustitutos para determinar la clasificación molecular de los cánceres de mama. Los sustitutos inmunohistoquímicos más comúnmente utilizados son el receptor de estrógenos (RE) y el receptor de progesterona (RP), que

permite diferenciar los cánceres de subtipos luminales y triples negativos. La adición de Ki67, separa el luminal B de los subtipos luminales A ¹²⁵.

Sin embargo, en las mujeres con lesiones benignas el estudio de la expresión de RE, RP, y Ki67 como predictores de cáncer de mama subsecuente ha mostrado resultados inconsistentes ^{118-120,122}. Aunque algunos estudios informaron un mayor riesgo de cáncer posterior en las mujeres con lesiones benignas que tenían una alta expresión de RE, RP, o Ki67 ^{118,120,122}, otros estudios informaron resultados contradictorios ^{118,126,127}.

1.3. LAS EVALUACIONES ECONÓMICAS EN SALUD

Las evaluaciones económicas en salud consisten en la comparación de los costes y los beneficios en salud de dos o más alternativas ¹²⁸. El objetivo final de las evaluaciones económicas es evaluar y definir el coste de los beneficios obtenidos y así promover el uso de las alternativas más eficientes.

Con la escalada de los costes sanitarios y el continuo incremento de tratamientos disponibles, las políticas de salud se sirven cada día más de las evaluaciones económicas. Si bien la eficiencia en la asignación de recursos no es el único aspecto a tener en cuenta para recomendar una intervención, está claro que añade información valiosa para la toma de decisiones.

En la mayoría de países europeos la toma de decisiones sobre las tecnologías o intervenciones sanitarias está sujeta a una evaluación económica. Un ejemplo destacado en el Reino Unido es el Instituto Nacional para la Salud y Excelencia en la Atención (NICE) que ha incentivado el uso exponencial del análisis de coste-efectividad desde su creación en 1999 ¹²⁹. En España, desde el 2010 se cuenta con una guía metodológica para el desarrollo de evaluaciones económicas de tecnologías sanitarias ¹³⁰ aunque su uso no es obligatorio a nivel nacional.

Todas las formas de evaluación económica evalúan los insumos y los resultados de dos o más alternativas que se van a comparar. Los insumos se calculan como los costes relacionados con el uso de una intervención menos los costes que se evitan como resultado de su uso. Los resultados son los beneficios en salud obtenidos gracias al uso de la intervención ¹³¹. Los costes varían en función de la cantidad de recursos utilizados y su precio unitario. Los efectos en salud son más difíciles de estimar principalmente porque las intervenciones a menudo conllevan múltiples efectos en diferentes momentos del tiempo.

Las evaluaciones económicas se clasifican de acuerdo a la forma en que se miden los resultados en salud. Éstos se expresan en unidades físicas en los análisis de coste-efectividad y coste-consecuencia, en términos de valores de utilidad en los análisis de coste-utilidad o en términos monetarios en los análisis de coste-beneficio ¹³². El análisis de minimización de costes es un caso

específico del análisis de coste-efectividad donde se asume una igualdad en los resultados de salud de ambas alternativas y lo que se investiga es con cuál de las dos alternativas se pueden obtener esos resultados al menor coste.

Las dos medidas más comúnmente utilizadas en las evaluaciones económicas son los años de vida ganados (AVG), en los análisis de coste-efectividad, y los años de vida ajustados por calidad (AVAC) en los análisis de coste-utilidad. Un AVAC es calculado a partir del ajuste o ponderación de los años de vida ganados a los cuales se les incorpora una utilidad cuantificada a partir de la calidad de vida relacionada con la salud ¹³¹.

No obstante, existen diversas circunstancias en las que la evaluación de los resultados por medio de AVG o AVAC no es posible o no es la mejor forma de evaluar una intervención. En la evaluación de los programas de cribado del cáncer de mama, por ejemplo, la tasa de detección de cáncer es un resultado en salud que representa correctamente el valor de la intervención por lo cual es frecuentemente utilizado.

1.3.1. Análisis de coste-consecuencia y coste-efectividad

El concepto de coste de oportunidad es la base de los estudios de coste-consecuencia y coste-efectividad. Este concepto presupone que, si se opta por una intervención equivocada -aquella que genera pocos beneficios adicionales- se pierde la oportunidad de utilizar la misma cantidad de dinero o recursos en otras alternativas que generen mayores beneficios para la población.

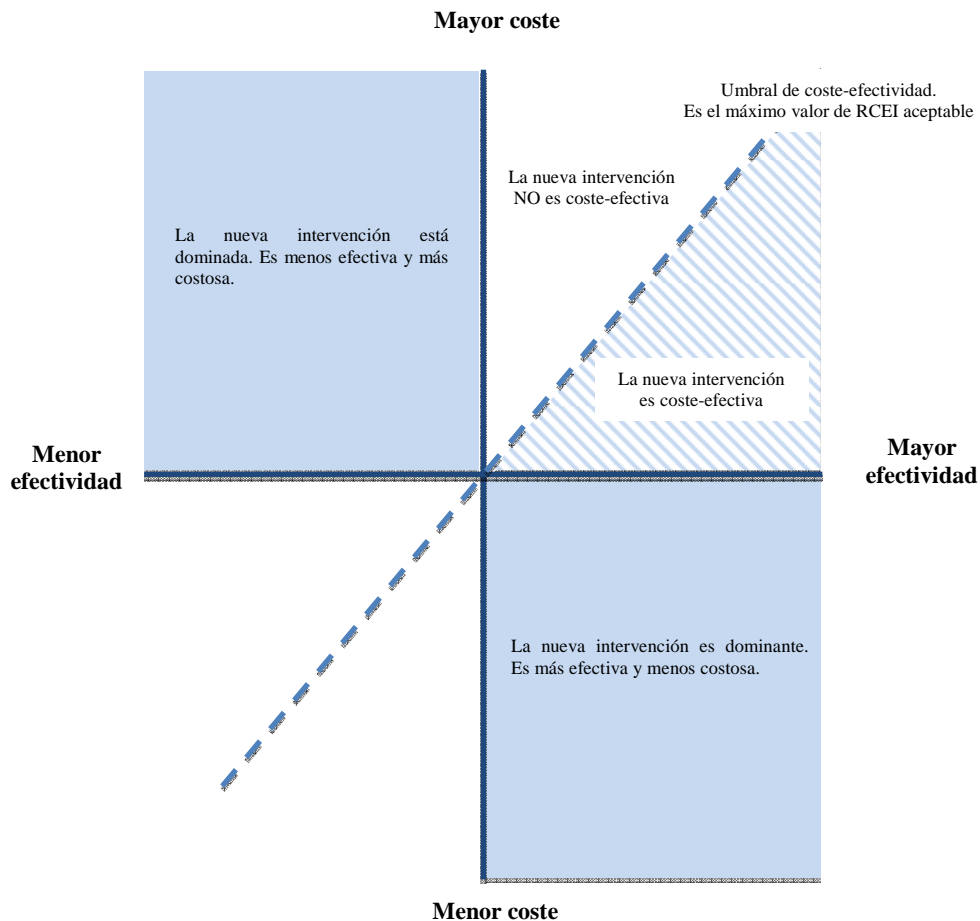
Los análisis de coste-consecuencia presentan de forma desagregada tanto los costes como los efectos en salud que tienen interés para los tomadores de decisiones. Mediante este análisis se presentan los costes por paciente o para toda la población. Bajo esta perspectiva se pueden representar los costes atribuibles a resultados intermedios en salud como serían la gravedad de la enfermedad al momento del diagnóstico o los falsos positivos de una prueba diagnóstica, entre otros ¹³².

La principal ventaja de la presentación desagregada de cada uno de los costes y resultados en salud es que permite al decisor sopesar explícitamente los recursos empleados en una intervención en relación con sus efectos ¹³². De una forma intuitiva los decisores pueden calcular el incremento en costes que supone elegir un componente de una intervención y relacionarlo con los beneficios esperados.

Otra ventaja de este tipo de análisis es que permite una mejor valoración de la incertidumbre de los parámetros utilizados. La desventaja del enfoque es que, en ausencia de una medida agregada como es la relación de coste-efectividad incremental (RCEI), no es posible evaluar la eficiencia de una intervención.

La RCEI es la relación entre la diferencia de costes y la diferencia en resultados entre una alternativa y su comparador. A partir del cálculo de la RCEI se puede determinar el perfil de eficiencia de una intervención. Si una alternativa es más eficaz y menos costosa, se denomina "dominante" y se recomienda priorizarla sobre su comparador. Cuando la elección es entre intervenciones donde una es más efectiva, pero también más costosa, la información relevante para tomar la decisión es el coste adicional de lograr el resultado adicional representado por la RCEI^{128,131} (Figura 4).

Figura 4. Plano de coste-efectividad y umbral de decisión utilizados en las evaluaciones económicas.



El conocimiento de las RCEI también permite identificar alternativas dominadas por extensión. La dominancia por extensión se presenta cuando entre tres alternativas, siendo la primera la menos costosa y menos efectiva, la segunda presenta una RCEI más costosa que la tercera ^{128,131}.

Los análisis de coste-efectividad tienen como objetivo estimar cuál es la RCEI por cada unidad de salud ganada con una alternativa frente a otra. En la mayoría de los casos, la pregunta radica en si se sustituye o no una intervención existente por otra más eficaz, pero también más costosa. En tales casos, se necesita una estimación de los recursos adicionales que se deben gastar para obtener el beneficio adicional. En la evaluación de los programas de cribado, por ejemplo, la RCEI podría indicar el coste adicional que se requiere para detectar un caso adicional de cáncer entre las mujeres participantes en el programa.

1.3.2. Revisión sistemática de evaluaciones económicas

Una revisión sistemática es una investigación científica que tiene como objetivo analizar y sintetizar la evidencia empírica para responder una pregunta específica de investigación. Utiliza métodos explícitos y sistemáticos previamente establecido con miras a reducir al mínimo los sesgos y así proporcionar conclusiones más fiables que permitan tomar decisiones informadas ^{133,134}.

Una revisión sistemática de evaluaciones económicas presenta algunas características que la diferencian de una revisión sistemática de eficacia o efectividad. Particularmente, destacan las diferencias en la evaluación del riesgo de sesgo donde se incorpora también la transferibilidad de los resultados de un contexto a otro ¹²⁹ y la dificultad o imposibilidad de realizar un metanálisis de los resultados ¹³⁵.

Las revisiones sistemáticas de evaluaciones económicas pueden ser de utilidad para diversos propósitos. Entre los más frecuentes se encuentran los siguientes:

- i) Revisión sistemática para sintetizar la evidencia de evaluaciones económicas.
- ii) Revisión sistemática para informar el modelo de una evaluación económica *de novo*, identificar el estudio más relevante para un nodo de decisión específico, entre otros.

El desarrollo y uso de las revisiones sistemáticas de evaluaciones económicas es cada vez más extendido. De hecho, actualmente diversos organismos internacionales han publicado guías metodológicas para el desarrollo de éstas revisiones ^{135,136}. También el Manual Cochrane para Revisiones Sistemáticas de Intervenciones ha incorporado una metodología para identificar y resumir los componentes económicos –uso de recursos y costes- de las intervenciones mientras se realiza una revisión Cochrane de efectividad ¹³⁷.

Aunque el valor de las revisiones sistemáticas de evaluaciones económicas es reconocido, se ha de tener en cuenta algunas limitaciones que podrían limitar su utilización. Primero, la generalización de los resultados de las evaluaciones económicas suele ser limitada en el tiempo y entre los sistemas de salud. Segundo, dado que los modelos de decisión representan en sí mismos una síntesis de diversas posibilidades, en la mayoría de los casos una revisión sistemática exhaustiva de evaluaciones económicas es injustificada ¹³⁸.

A pesar de sus limitaciones, las revisiones sistemáticas de evaluaciones económicas cobran mayor relevancia cuando la valoración de los recursos necesarios para apoyar el uso de una alternativa es crítica y se requiere de información rigurosa y exhaustiva que ayude a la toma de decisiones.



JUSTIFICACIÓN

JUSTIFICACIÓN

El cribado con mamografía digital se ofrece a millones de mujeres a nivel mundial. Por tanto, el impacto que tiene una pequeña mejora en la reducción de riesgos o costes, un incremento en la efectividad o el mejor conocimiento del riesgo en las mujeres con lesiones benignas puede representar un beneficio importante cuando se traduce al conjunto de la población.

En el contexto de la mamografía digital, aunque la mayoría de programas europeos utilizan la doble lectura, diversas razones cuestionan su superioridad frente a la lectura simple. Por un lado, el beneficio de la doble lectura respecto al incremento de la detección de neoplasias podría quedar restringido a situaciones puntuales en las que existe gran discordancia entre los radiólogos. Por otro lado, la doble lectura podría incrementar los potenciales daños del cribado como son la frecuencia de recitaciones y falsos positivos.

Las evaluaciones económicas constituyen una herramienta válida para conocer la magnitud de los recursos necesarios para implementar o continuar con una intervención, proporcionan evidencia sobre el valor de la utilización de esos recursos considerando los beneficios en salud obtenidos y finalmente ayudan a adecuar la asignación de los recursos y elegir las intervenciones más eficientes. Con todo, las evaluaciones económicas sirven para complementar el resto de criterios sobre los que se basan las decisiones en salud y por lo tanto tienen una gran relevancia a nivel social y sanitario.

Aunque se conoce la considerable utilidad de las evaluaciones económicas en salud y en la toma de decisiones sanitarias, lo cierto es que hasta la fecha, se cuenta con escasa evidencia económica en el contexto la mamografía digital. De hecho, respecto a la doble lectura mamográfica, los estudios que han evaluado la efectividad y coste-efectividad de esta intervención, realizaron el análisis en el contexto de la mamografía analógica y no se cuenta con estudios de coste-efectividad de la doble lectura en el contexto de la mamografía digital.

El valor pronóstico y terapéutico de la expresión de receptores de estrógeno (RE), progesterona (RP) y Ki67 en el cáncer de mama es ampliamente conocido. Sin embargo, en las mujeres con lesiones benignas de la mama el estudio de la expresión de RE, RP, y Ki67 como predictores de cáncer de mama ha mostrado resultados heterogéneos. Cabe señalar que estos estudios han sido realizados en contextos diferentes al cribado poblacional.

La evaluación de la doble lectura y la detección de lesiones benignas en esta tesis doctoral se justifican por la ausencia de estudios en el contexto de la mamografía digital y sobre todo por la necesidad de proponer estrategias que mejoren el perfil de efectividad y coste-efectividad de los programas de cribado.

Los resultados de este trabajo de tesis permitirán disponer de un conocimiento riguroso sobre los costes y las consecuencias de la doble lectura, conocimiento necesario para la toma de decisiones

en los programas de cribado poblacional. Así también se aportará información novedosa sobre la expresión de biomarcadores en mujeres con lesiones benignas en el contexto del cribado. La identificación de estos biomarcadores podría mejorar tanto la capacidad de estratificar el riesgo individual de cáncer de mama como ayudar a conducir estrategias más precisas de seguimiento en las mujeres con lesiones benignas.



OBJETIVOS

OBJETIVOS

3.1. OBJETIVOS GENERALES

Los objetivos generales de este trabajo de tesis son:

1. Evaluar la efectividad y el coste-efectividad de la doble lectura mamográfica en el contexto de la mamografía digital.
2. Valorar el comportamiento de lesiones benignas de la mama detectadas en el contexto de un programa de cribado poblacional.

3.2. OBJETIVOS ESPECÍFICOS

Los objetivos específicos de este trabajo de tesis son:

1. Evaluar los efectos de la doble lectura mamográfica sobre la efectividad de un programa de cribado poblacional así como su efecto sobre los recursos utilizados.
2. Evaluar el coste-efectividad de la doble lectura mamográfica sobre la tasa de detección de neoplasias en un programa de cribado poblacional.
3. Realizar una revisión sistemática sobre la efectividad y coste-efectividad de la doble lectura mamográfica en el contexto del cribado con mamografía digital.
4. Evaluar la relación entre los biomarcadores expresados por las lesiones benignas detectadas en un programa de cribado y el desarrollo posterior de cáncer.



MÉTODOS

MÉTODOS

Este trabajo de tesis se presenta por compendio de publicaciones. Los métodos son los correspondientes a cada uno de los estudios realizados. El diseño de cada estudio se eligió de acuerdo con la pregunta que se pretendía responder.

Para determinar la efectividad y el coste-efectividad de las diferentes estrategias de lectura mamográfica se diseñaron: un análisis de coste-consecuencia, un análisis de coste-efectividad y una revisión sistemática de estudios observacionales y evaluaciones económicas. Para determinar si la expresión de receptores de estrógeno, progesterona y Ki-67 en las lesiones benignas conlleva un incremento de riesgo de desarrollo de cáncer, se diseñó un estudio de casos y controles.

4.1. ESTUDIO 1. ANÁLISIS DE COSTE-CONSECUENCIA DE LA DOBLE LECTURA MAMOGRÁFICA

DISEÑO

Análisis de coste-consecuencia.

POBLACIÓN DE ESTUDIO

La población de este estudio está constituida por las mamografías realizadas a las mujeres entre 50 y 69 años que participaron en el programa de cribado del cáncer de mama del área de influencia del Hospital de la Santa Creu i Sant Pau de Barcelona, España.

Se incluyeron 57.157 mamografías digitales realizadas entre junio de 2009 y mayo de 2013, las cuales tuvieron seguimiento hasta mayo de 2015. Del total de mamografías incluidas, 11.353 (19.9%) correspondieron a mujeres que participaron por primera vez en el programa de cribado del cáncer de mama y 45.804 (80.1%) correspondieron a participaciones posteriores.

ESTRATEGIAS DE LECTURA MAMOGRÁFICA EVALUADAS

Se evaluaron cuatro estrategias de lectura. Dos de ellas fueron estrategias de doble lectura -con y sin consenso y arbitraje- y dos fueron estrategias de lectura simple -primer y segundo lector-.

1) Estrategias de doble lectura. Se comparó la doble lectura con consenso y arbitraje frente a la doble lectura sin consenso ni arbitraje.

2) Estrategias de lectura simple. Dado que en el programa de cribado la doble lectura es independiente y el segundo lector desconoce la valoración del primero, se comparó la primera lectura frente a la segunda lectura, tomándolas como lecturas independientes.

INDICADORES DE EFECTIVIDAD EVALUADOS

En este estudio se evaluaron los siguientes indicadores:

Falsos positivos. Un resultado falso positivo se definió como una mamografía sospechosa que conllevó exploraciones adicionales gracias a las cuales se confirmó un resultado no maligno.

Tasa de detección de cáncer. Los cánceres detectados en el cribado fueron confirmados patológicamente como cánceres de mama, tanto carcinoma ductal *in situ* como carcinomas invasivos. La tasa de detección de cáncer se calculó como el número de cánceres detectados en el cribado dividido por el número de mujeres participantes.

Valor predictivo positivo de la recitación. El valor predictivo positivo se calculó como el número de cánceres detectados dividido por el número de mujeres que fueron recitadas para realizar pruebas adicionales.

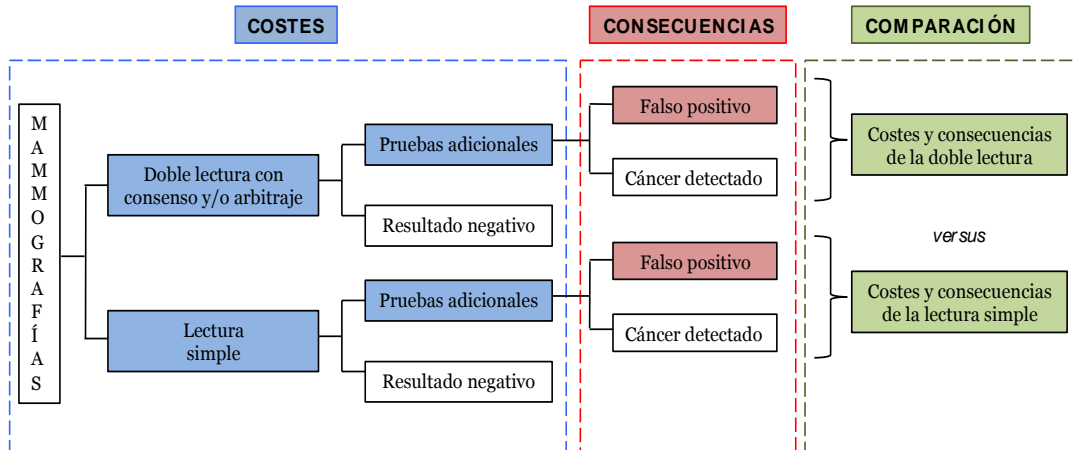
USO DE RECURSOS Y COSTES EVALUADOS

De acuerdo con la perspectiva del financiador^{129,130}, se incluyeron los costes directos derivados de la atención brindada en el programa de cribado. Los costes totales se calcularon utilizando un método de cálculo de costes ascendente. Se calculó los costes del cribado con mamografía a partir de la información de la base de datos del programa y los costes de las pruebas complementarias y el proceso de diagnóstico a partir de la información del departamento de finanzas del hospital. Todos los costes se calcularon en euros (€) según su valor en 2010.

ESTRATEGIA DE ANÁLISIS

Primero se determinaron los costes y las consecuencias de cada una de las estrategias de lectura mamográfica. Posteriormente, se compararon las dos estrategias de doble lectura entre ellas (doble lectura con consenso y arbitraje frente a doble lectura sin consenso) y las dos estrategias de lectura simple entre ellas (primer lector frente a segundo lector).

Figura 2. Algoritmo seguido para realizar el análisis de coste-consecuencia.



A continuación, se eligió la mejor opción de las dos estrategias de doble lectura y la mejor opción de las dos estrategias de lectura simple. Con el fin de encontrar la estrategia, se comparó sólo la mejor opción de las estrategias de doble lectura frente a la mejor de las estrategias de lectura única. La estrategia considerada como más ventajosa fue la estrategia menos costosa, con la menor tasa de falsos positivos y la mayor tasa de detección de cáncer (Figura 6).

Las diferencias entre las estrategias de lectura se probaron mediante la prueba t de Student, la prueba de Chi cuadrado y la prueba exacta de Fisher. Se consideró estadísticamente significativo un valor de p menor a 0,05. Los análisis se realizaron utilizando Microsoft Excel (2011) y el paquete estadístico SPSS de IBM versión 21.0 (SPSS Inc., Chicago, Illinois, Estados Unidos, 2013).

4.2. ESTUDIO 2. ANÁLISIS DE COSTE-EFECTIVIDAD DE LA DOBLE LECTURA MAMOGRÁFICA

DISEÑO

Análisis de coste-efectividad.

POBLACIÓN DE ESTUDIO

La población de este estudio está constituida por las mamografías realizadas a las mujeres entre 50 y 69 años que participaron en el programa de cribado del cáncer de mama del área de influencia del Hospital de la Santa Creu i Sant Pau de Barcelona, España.

Se incluyeron 28.636 mamografías digitales realizadas entre junio de 2009 y mayo de 2011, las cuales tuvieron seguimiento hasta mayo de 2013. Del total de mamografías incluidas, 5.978 (20,9%) correspondieron a mujeres que participaron por primera vez en el programa de cribado del cáncer de mama y 22.658 (79,1%) correspondieron a participaciones posteriores.

Los resultados de cada mamografía se clasificaron de la siguiente manera: (1) recitación, en cuyo caso se solicitaron pruebas adicionales para confirmar o descartar malignidad; (2) control avanzado, en cuyo caso la mujer fue programada para una mamografía de cribado adicional en 12 meses; o (3) control rutinario, en cuyo caso la mujer fue programada para una nueva mamografía de cribado en dos años.

ESTRATEGIAS DE LECTURA MAMOGRÁFICA EVALUADAS

Se evaluaron tres estrategias de lectura:

- 1) Doble lectura. La doble lectura con consenso y arbitraje fue el procedimiento realizado en el programa de detección de cáncer de mama.
- 2) Doble lectura en el cribado prevalente y lectura simple en el cribado incidente. Esta estrategia simuló una doble lectura para las mamografías de mujeres en su primera participación, y una lectura simple para las mamografías de mujeres en sus participaciones sucesivas.
- 3) Lectura simple. Esta estrategia simuló una lectura simple igual a la lectura del primer radiólogo. Con el fin de evaluar el efecto del cambio del primer al segundo lector, se incluyeron los resultados del segundo radiólogo en el análisis de sensibilidad.

INDICADORES DE EFECTIVIDAD EVALUADOS

En este estudio se evaluaron los siguientes indicadores:

Tasa de detección de cáncer. Los cánceres detectados en el cribado fueron confirmados patológicamente como cánceres de mama, tanto carcinoma ductal *in situ* (DCIS) como carcinomas invasivos. La tasa de detección de cáncer se calculó como el número de cánceres detectados en el cribado dividido por el número de mujeres participantes.

USO DE RECURSOS Y COSTES EVALUADOS

De acuerdo con la perspectiva del sistema sanitario ^{129,130}, se incluyeron los costes directos derivados de la atención brindada en el programa de cribado. Los costes totales se calcularon utilizando un método de cálculo de costes ascendente. Se calcularon los costes del cribado con mamografía a partir de la información de la base de datos del programa. Los costes de las pruebas complementarias y el proceso de diagnóstico se obtuvieron a partir de la información del

departamento de finanzas del hospital. Todos los costes se calcularon en euros (€) según su valor en 2010.

ESTRATEGIA DE ANÁLISIS

Se utilizó un modelo de árbol de decisión. En este modelo, el principal resultado de efectividad fue la detección de cáncer. El horizonte temporal fue de cuatro años -de junio de 2009 a mayo de 2013-, que incluyó una ronda de cribado bienal (2009-2011) más dos años de seguimiento para confirmar los resultados negativos. Este modelo supone que las participantes no sufrieron de otras condiciones que les impidan completar todo el horizonte temporal.

No se aplicó tasa de descuento a los costes ni a la tasa de detección de cáncer debido a los siguientes motivos ^{129,139,140}: i) en un programa de cribado se espera que los beneficios en salud ocurran a largo plazo por lo que el descuento no representaría correctamente las preferencias de la población; ii) el horizonte temporal de dos años es muy corto como para mostrar diferencias ocasionadas por el descuento.

Las estrategias de lectura se compararon mediante la relación de coste-efectividad incremental (RCEI). Primero se calculó el coste-efectividad medio dividiendo los costes totales de cada estrategia por el número correspondiente de cánceres detectados. En segundo lugar, se calculó la RCEI comenzando con el coste más bajo por cáncer adicional detectado. Esto permitió identificar una estrategia con menor efectividad y mayor RCEI y descartarla debido a la dominancia extendida. Finalmente, se compararon las RCEI para las estrategias no descartadas. La estrategia más rentable mostró la RCEI más bajo.

Se realizó un análisis de sensibilidad determinístico para evaluar la robustez de las RCEI de acuerdo con la variación en las siguientes variables: tasa de participación, prevalencia de cáncer de mama, sensibilidad de la estrategia de lectura, valor predictivo positivo de la recitación, proporción de controles avanzados y costes de personal.

Se consideró estadísticamente significativo un valor de p menor a 0,05. Los análisis se realizaron utilizando Microsoft Excel (2011) y el paquete estadístico SPSS de IBM versión 21.0 (SPSS Inc., Chicago, Illinois, Estados Unidos, 2013).

4.3. ESTUDIO 3. REVISIÓN SISTEMÁTICA DE LA EFECTIVIDAD Y COSTE-EFECTIVIDAD DE LA DOBLE LECTURA MAMOGRÁFICA

DISEÑO

Revisión sistemática de estudios observacionales y evaluaciones económicas.

ESTRATEGIA DE BÚSQUEDA Y SELECCIÓN DE ESTUDIOS

Se realizó una búsqueda de la literatura publicada entre enero de 1990 y abril de 2017 en MEDLINE, Embase y en The Cochrane Library. Se combinaron una serie de términos relacionados con ‘Breast Neoplasms’, ‘Mass Screening’ y ‘double reading’. La búsqueda no se limitó por idioma.

Se incluyeron revisiones sistemáticas, estudios experimentales, estudios observacionales con grupo control y evaluaciones económicas en los que se comparó la doble lectura con la lectura simple en el contexto de la mamografía digital.

Los estudios incluidos presentaron información sobre alguna de las siguientes variables: sensibilidad, especificidad, tasa de detección de cáncer, tasa de falsos positivos, cánceres de intervalo, mortalidad, sobrediagnóstico, calidad de vida, relación de coste efectividad incremental (RCEI) o valor actual neto.

Se excluyeron los estudios sin grupo control, estudios realizados fuera del contexto de cribado poblacional con mamografía digital o sin confirmación patológica de las neoplasias.

EVALUACIÓN DE LA CALIDAD DE LOS ESTUDIOS Y EXTRACCIÓN DE DATOS

Un autor extrajo la información de los estudios incluidos utilizando un formulario *ad hoc* y un segundo revisor evaluó la calidad de la extracción. La calidad de los estudios diagnóstico se determinó mediante la herramienta QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) ¹⁴¹ y mediante la lista de verificación CHEERS (Consolidated Health Economic Evaluation Reporting Standards) ¹⁴² para las evaluaciones económicas.

ANÁLISIS Y SÍNTESIS DE LOS RESULTADOS

Se describieron y tabularon los hallazgos de cada estudio y se identificaron similitudes y diferencias entre ellos. Se reportó el diseño del estudio, la calidad metodológica, las estrategias de lectura, características de las participantes y los resultados. Los costes se convirtieron y expresaron en valores de paridad de poder de compra de 2015 en unidades de Dólar de los Estados Unidos de América (US \$ PPP) ¹⁴³.

Se realizó metanálisis de proporciones para la sensibilidad, especificidad, tasa de detección de neoplasias y falsos positivos. Se utilizó el comando metaprop en Stata utilizando el método binomial exacto para el cálculo de las proporciones agrupadas. Los valores P de la prueba de heterogeneidad fueron reportados ¹⁴⁴.

4.4. ESTUDIO 4. EXPRESIÓN DE BIOMARCADORES EN LESIONES BENIGNAS DE LA MAMA

DISEÑO

Estudio de casos y controles anidados en una cohorte.

POBLACIÓN DE ESTUDIO

La población de estudio está conformada por un subgrupo de mujeres con lesiones benignas de la mama identificadas en una cohorte de cribado. La cohorte inicial de cribado consta de 549.422 mujeres con al menos una mamografía realizada durante el periodo 1994 y 2011 en España¹⁴⁵.

En esta cohorte se identificaron 10.262 mujeres con lesiones benignas de la mama. Se excluyeron 4.251 mujeres debido a que tenían una clasificación histológica indeterminada (por ejemplo, negativas para células malignas). Por tanto, 6.011 mujeres con lesiones benignas tuvieron una descripción patológica precisa y fueron clasificadas en uno de los siguientes grupos: (1) lesión no proliferativa; (2) lesión proliferativa sin atipia; Y (3) lesión proliferativa con atipia.

Los casos y los controles fueron seleccionados a partir de estas 6011 mujeres. Los casos fueron mujeres con lesiones benignas que desarrollaron un cáncer de mama posterior a la lesión, mientras que los controles fueron mujeres con lesiones benignas que permanecieron libres de cáncer. El periodo de seguimiento se extendió desde el 2011 hasta el 31 de diciembre de 2015.

Durante el periodo de seguimiento, 201 (3,3%) de las 6011 mujeres desarrollaron cáncer de mama (invasivo o *in situ*) y fueron consideradas casos. Se excluyeron 114 casos debido a que la muestra de anatomía patológica fue insuficiente para evaluar la expresión de los biomarcadores y un caso adicional se excluyó debido a la presencia de carcinoma mucinoso en la muestra de lesión benigna. Así, se incluyeron en el análisis 86 casos.

Se seleccionaron al azar dos controles por cada caso. Los casos y controles fueron apareados por la edad en el momento del diagnóstico de la lesión benigna, el tipo de lesión benigna (no proliferativa, proliferativa sin atipia o proliferativa con atipia) y el tiempo de seguimiento entre el diagnóstico de lesión benigna y el final del seguimiento.

En total, el tamaño de la muestra incluida en el análisis fue de 258 mujeres, 86 casos emparejados con 172 controles.

BIOMARCADORES EVALUADOS

Dos patólogos revisaron de forma independiente las muestras patológicas de las lesiones benignas. Los patólogos desconocían si la muestra correspondía a un caso o a un control. De cada muestra se evaluó la expresión de receptores de estrógeno (RE) y receptores de progesterona (RP). En el caso de las lesiones proliferativas adicionalmente se evaluó la expresión de Ki67.

En caso de desacuerdo entre patólogos, la expresión de los biomarcadores se determinó por consenso. En los casos en los que una mujer tuvo más de una biopsia, se utilizó solo la primera. El porcentaje de expresión de RE, RP y Ki67 se cuantificó en base al número de núcleos teñidos positivamente por cada 100 células observadas al microscopio.

Primero se visualizó toda la muestra para identificar las áreas correspondientes a la lesión benigna. En caso de coexistencia de dos lesiones benignas, se evaluó la más grave (lesión proliferativa con atipia seguida de sin atipia y de no proliferativa). En el caso de las lesiones proliferativas, la expresión de los biomarcadores se evaluó en las áreas con mayor actividad proliferativa. En el caso de las lesiones no proliferativas, la expresión de los biomarcadores se evaluó en toda el área de la lesión.

ESTRATEGIA DE ANÁLISIS

El porcentaje de células con expresión de RE, RP y de Ki67 en cada muestra se obtuvo mediante la evaluación microscópica. La distribución de los porcentajes de RE y RP se categorizó en terciles obteniéndose los siguientes grupos:

Expresión de RE: i) baja de 0% a 70%; ii) moderada de 71% a 90%; y iii) alta superior a 90%. Respecto a la expresión de RP: i) baja de 0% a 40%; ii) moderada de 41% a 80%; y iii) alta superior a 80%. Respecto a las lesiones proliferativas, la expresión de Ki67 se clasificó en dos grupos: i) expresión baja de 0% a $< 1\%$; y ii) expresión alta $\geq 1\%$.

Se comparó la distribución de la expresión de RE, RP y Ki67 entre los casos y controles para cada tipo de lesión benigna. Se utilizó una regresión logística condicional para estimar las odds ratios (OR) y el 95% de intervalos de confianza del riesgo de cáncer de mama de acuerdo a la expresión de los biomarcadores.

Se consideró que el valor $P < 0,05$ como estadísticamente significativo. Los análisis estadísticos se realizaron utilizando el paquete estadístico SPSS de IBM versión 21.0 (SPSS Inc., Chicago, Illinois, Estados Unidos, 2013).



RESULTADOS

RESULTADOS

Los resultados de este trabajo de tesis son los correspondientes a cada uno de los estudios desarrollados. Cada uno de ellos ha dado lugar a un artículo y todos han sido publicados en revistas científicas internacionales.

ARTÍCULOS QUE CONFORMAN EL TRABAJO DE TESIS

ARTICULO 1: Posso M, Puig T, Quintana MJ, Solà-Roca J, Bonfill X. Double versus single reading of mammograms in a breast cancer screening programme: A cost-consequence analysis. *Eur Radiol.* 2016; 26(9): 3262-71. Factor de Impacto (FI) 3,967.

ARTICULO 2: Posso M, Carles M, Rué M, Puig T, Bonfill X. Cost-effectiveness of double reading versus single reading of mammograms in a breast cancer screening programme. *PlosOne.* 2016; 11(7): E0159806. FI 2,806.

ARTICULO 3: Posso M, Puig T, Carles M, Rué M, Canelo-Aybar C, Bonfill X. Effectiveness and cost-effectiveness of double reading in digital mammography screening: A systematic review and meta-analysis. *Eur J Rad.* 2017; 96: 40-9. FI 2,462.

ARTICULO 4: Posso M, Corominas JM, Serrano L, Román M, Torá-Rocamora I, et al. Biomarkers expression in benign breast diseases and risk of subsequent breast cancer: a case-control study. *Cancer Med.* 2017; 6(6): 1482-89. FI 3,362.

5.1. ARTÍCULO 1. ANÁLISIS DE COSTE-CONSECUENCIA DE LA DOBLE LECTURA MAMOGRÁFICA

5.1.1. RESUMEN DE RESULTADOS

CARACTERÍSTICAS DE LA POBLACIÓN DE ESTUDIO

Se analizaron 28.636 y 28.521 mamografías digitales realizadas durante las rondas de cribado de 2009-2011 y 2011-2013, respectivamente. Del total de mamografías incluidas (N = 57.157), 11.353 (19,9%) correspondieron a mujeres que participaron por primera vez en el programa de cribado y 45.804 (80,1%) correspondieron a participaciones posteriores.

En un 4,5% (N = 2.556) de las mamografías se observó desacuerdo entre los radiólogos. El 98,1% (N = 2.508) de las discordancias se resolvió por consenso y el 1,9% (N = 48) por arbitraje.

USO DE RECURSOS Y COSTES

El coste total de obtener 57.157 mamografías fue de 322.770 euros (€). El coste del personal (incluida la lectura mamográfica) y de las pruebas adicionales varió de acuerdo a la estrategia de lectura. El coste de las estrategias con doble lectura fue mayor al de las estrategias con lectura simple debido a la mayor frecuencia de pruebas complementarias realizadas.

El coste medio del proceso de cribado (cribado con mamografía y pruebas diagnósticas cuando eran necesarias) en la estrategia de doble lectura con consenso y arbitraje fue de 44,9 €. En la estrategia de doble lectura sin consenso y arbitraje fue de 45,6 €. En la estrategia con lectura simple en base a los del primer lector fue de 39,1 € y en base a los datos del segundo lector fue de 40,1 €.

ANÁLISIS DE COSTE-CONSECUENCIA

El análisis de costes y consecuencias de las dos estrategias de doble lectura demostró que la doble lectura con consenso y arbitraje fue menos costosa que la doble lectura sin consenso ni arbitraje. Por su parte entre las dos lecturas simples la primera lectura mostró ser menos costosa. Para la comparación de costes y consecuencias de la doble lectura frente a la lectura simple, se eligió la doble lectura con consenso y arbitraje y la lectura simple con solo el primer lector.

Respecto a los costes, la doble lectura con consenso y arbitraje fue un 15% (334.341 €) más costosa que la lectura simple con solo el primer lector (Tabla 3).

Tabla 3. Coste estimado (Euros €, 2010) correspondiente a 57.157 participaciones según estrategias de la lectura.

	Doble lectura €	Lectura simple €	Coste incremental €
Cribado con mamografía	2.172.426	1.930.484	241.942 (12,5%)
Pruebas adicionales	399.441	307.043	92.398 (30,1%)
Coste total	2.571.867	2.237.527	334.341 (15,0%)

Costes estimados a partir de la información proporcionada por la base de datos del programa de cribado y por el Departamento de Finanzas del Hospital de la Santa Creu i Sant Pau.

Respecto a los efectos en salud, la doble lectura con consenso y arbitraje tuvo un 0,3% (N = 181) más de resultados falsos positivos que la lectura simple con solo el primer lector (P < 0,01). Ambas estrategias de lectura tuvieron tasas de detección de cáncer similares (4,6/1000 y 4,2/1000 mamografías, respectivamente; P = 0,28) y valores predictivos positivos similares (9,3% y 9,1% respectivamente, P = 0,81) (Tabla 4).

Tabla 4. Efectos en salud estimados para 57.157 participaciones según estrategias de lectura.

	Doble lectura (N)	Lectura simple (N)	Efecto incremental (N)	P
Falsos positivos	4,5% (2.559)	4,2% (2.378)	0,3% (181)	0,001
Valor predictivo positivo	9,3% (263)	9,1% (239)	0,2% (24)	0,812
Mujeres con pruebas adicionales	4,9% (2.822)	4,6% (2.617)	0,3% (205)	0,004
Tasa de detección de cáncer	4,6/1000 (263)	4,2/1000 (239)	0,4/1000 (24)	0,283

Efectos estimados a partir de la información de 57.157 participaciones en el programa de cribado del Hospital de la Santa Creu i Sant Pau durante las rondas 2009-2011 y 2011-2013.

PUBLICACIÓN 1

POSSO M, PUIG T, QUINTANA MJ, SOLÀ-ROCA J, BONFILL X. DOUBLE VERSUS SINGLE READING OF MAMMOGRAMS IN A BREAST CANCER SCREENING PROGRAMME: A COST-CONSEQUENCE ANALYSIS. EUR RADIOL. 2016; 26(9): 3262-71.

European radiology. 2016.

Factor de impacto 3,967. Q1 (15/126). Radiology, nuclear medicine & medical imaging.

Double versus single reading of mammograms in a breast cancer screening programme: a cost-consequence analysis

Margarita C. Posso^{1,4} · Teresa Puig^{1,2} · Ma Jesus Quintana^{1,3} · Judit Solà-Roca¹ · Xavier Bonfill^{1,2,3}

Received: 27 July 2015 / Revised: 30 November 2015 / Accepted: 15 December 2015
© European Society of Radiology 2016

Abstract

Objectives To assess the costs and health-related outcomes of double versus single reading of digital mammograms in a breast cancer screening programme.

Methods Based on data from 57,157 digital screening mammograms from women aged 50–69 years, we compared costs, false-positive results, positive predictive value and cancer detection rate using four reading strategies: double reading with and without consensus and arbitration, and single reading with first reader only and second reader only. Four highly trained radiologists read the mammograms.

Results Double reading with consensus and arbitration was 15 % (Euro 334,341) more expensive than single reading with first reader only. False-positive results were more frequent at double reading with consensus and arbitration than at single reading with first reader only (4.5 % and 4.2 %, respectively; $p < 0.001$). The positive predictive value (9.3 % and 9.1 %;

$p = 0.812$) and cancer detection rate were similar for both reading strategies (4.6 and 4.2 per 1000 screens; $p = 0.283$).

Conclusions Our results suggest that changing to single reading of mammograms could produce savings in breast cancer screening. Single reading could reduce the frequency of false-positive results without changing the cancer detection rate. These results are not conclusive and cannot be generalized to other contexts with less trained radiologists.

Key Points

- Double reading of digital mammograms is more expensive than single reading.
- Compared to single reading, double reading yields a higher proportion of false-positive results.
- The cancer detection rate was similar for double and single readings.
- Single reading may be a cost-effective strategy in breast cancer screening programmes.

Margarita C. Posso is a Ph.D. candidate at the Methodology of Biomedical Research and Public Health program Universitat Autònoma de Barcelona (UAB), Barcelona, Spain.

✉ Margarita C. Posso
mposso@cochrane.es

Keywords Costs and cost analysis · Mammography · Mass screening · Early detection of cancer · Breast neoplasms

Abbreviations

DCIS Ductal carcinoma in situ.

Introduction

Screening with digital mammography has been widely implemented in European programmes as it can detect breast cancer in early stages [1–3]. However, routine mammography screening has potential risks [4–6]. The most frequent risk is false-positive results, [4, 7] defined as ‘abnormal mammograms in women ultimately found to have no cancer’ [8]. Evidence from a systematic review shows that false-positive

¹ Epidemiology Department, Hospital de la Santa Creu i Sant Pau, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain
² Universitat Autònoma de Barcelona (UAB), Barcelona, Spain
³ CIBER of Epidemiology and Public Health (CIBERESP), Barcelona, Spain
⁴ Iberoamerican Cochrane Centre, Biomedical Research Institute Sant Pau (IIB Sant Pau), Hospital de la Santa Creu i Sant Pau, C/ Sant Antoni Maria Claret, 167. Pavelló 18, planta 0, CP: 08025 Barcelona, Spain

results produce psychological distress [9]. Furthermore, unnecessary additional testing to rule out malignancy in women with false-positive results wastes resources.

The usual practice in breast cancer screening programmes in Europe for mammogram interpretation is to perform double reading as it can increase the cancer detection rate [3, 10–12]. However, compared to single reading, double reading may imply a higher proportion of false-positive results as a consequence of a lower specificity [13, 14]. Consensus between the two radiologists or arbitration with a third radiologist can improve the specificity of double reading [15–17]. A recently published study by Klompenhouwer et al. [18] showed that arbitration can improve the recall rate and positive predictive value but that it reduces the sensitivity of blinded double reading of digital mammograms. Nevertheless, in this era of digital mammography, there is no evidence that the proportion of false-positive results at double reading with consensus and arbitration is lower than at single reading.

Policy makers should take studies of resources and costs into account when deciding to implement a single or double reading in breast cancer screening programmes. A cost-consequences analysis [19] is the first approach to determine whether the cost of double reading is justified on the basis of increases in cancer detection rates and decreases in false-positive results. The use of a cost–consequence approach depends on the fact that this analysis measures the health-related consequences without combining them into one summary effectiveness measure. Few studies have assessed the use of additional tests and costs in healthcare services for women after a false positive result. However, these studies were performed in North America and the results may not be transferable to the European context [20, 21]. Therefore, it seems relevant to conduct an analysis of these costs in our context.

There is evidence that cancer detection rate can be higher with digital mammography than with screen-film mammography [22], but double reading has not been compared to single reading in digital screening mammography in Spain. Despite the lack of information, it is reasonable to hypothesize that the cancer detection rate with single reading could be as good as that with double reading, but less expensive. The purpose of this study was to evaluate the costs and health related outcomes of double versus single reading of digital mammograms in a breast cancer screening programme.

Material and methods

Setting and study population

This study was performed in women participating in a population-based breast cancer screening programme of the Hospital Sant Pau in an area of 390,000 inhabitants in the city

of Barcelona, Spain. As part of our protocol, women aged 50–69 years were invited to have a breast cancer screening examination every two years. In this study, we included all digital mammograms performed from June 2009 until May 2013, and followed up until May 2015. We excluded two mammograms with incomplete information and 1,783 intermediate mammograms that were performed between two screening examinations. A total of 57,157 mammograms were analysed, 11,353 (19.9 %) corresponding to women participating for the first time in the breast cancer screening programme (prevalent screening,) and 45,804 (80.1 %) corresponding to subsequent participations (subsequent screening) (Table 1).

Screening programme procedures

As is routine practice in our screening protocol, two projections per breast (mediolateral-oblique and craniocaudal) were taken in all screening examinations. Mammograms were read by four certified screening radiologists who read at least 5,000 mammograms per year in the breast cancer screening programme. Two radiologists read each mammogram separately. The second reader was not aware of the first reader's decision (blinded double reading). The four radiologists were randomly assigned as first or second readers of each mammogram, and then one radiologist who is first reader in one mammogram can be second reader in another mammogram. Each radiologist determined whether the mammogram was normal, in which case the woman was programmed for a further mammography screening in two years, or whether findings were abnormal, in which case additional tests were requested to confirm or rule out malignancy. Additional tests were imaging procedures (mainly mammography and ultrasound) and invasive procedures (fine-needle aspiration, core-needle biopsy, and surgical biopsy). In case of disagreement

Table 1 Characteristics of study population by screening round

	2009-2011		2011-2013		P value
	n	%	n	%	
Population					
Invited women	48,803	100.0	46,819	100.0	
Participants	28,636	58.7	28,521	60.9	<0.001
Prevalent screening	5,978	20.9	5,375	18.8	<0.001
Subsequent screening	22,658	79.1	23,146	81.2	
Age at screening (years)					
Mean (SD)	59.0	(5.9)	59.5	(5.7)	<0.001
50-54	8,181	28.6	7,913	27.7	0.008
55-59	6,947	24.3	7,244	25.4	
60-64	7,047	24.6	7,031	24.7	
65-69	6,461	22.6	6,333	22.2	

SD = standard deviation.

between radiologists, the result of the mammogram was determined by consensus. When consensus was insufficient to reach a final result, we proceeded to arbitration by a third senior radiologist who had extensive experience in mammogram interpretation. All mammograms were available for comparison at the next screening round. The Ethics Committee at our institution approved the study and informed consent to participate was waived.

Evaluation of reading strategies

Based on data from the 57,157 mammograms of the breast cancer screening programme, we evaluated four reading strategies. Two of them were double reading strategies: double reading with and without consensus and arbitration, and two were single reading strategies: single reading with first and second reader only.

- 1) Double reading strategies. Double reading with consensus and arbitration was the procedure performed in the breast cancer screening programme. Double reading with consensus and arbitration involved four radiologists who were randomly assigned as first or second readers of each mammogram and one Double reading without consensus and arbitration was based on the assumption that women were recalled for additional tests if one of the radiologists determined abnormal findings in the mammogram.
- 2) Single reading strategies. Since the second reader was unaware of the first reader's decision, each mammogram had two results; one from the first reader and one from the second reader. For the purpose of the study, single reading with first or second reader only was based on the assumption that the first or second radiologist alone determined the result of the mammogram.

Costs

As this study was conducted from the perspective of the public health system, we included direct healthcare costs. Total costs were calculated using a bottom-up costing method in two stages: mammography screening and additional tests. Unitary costs for the first stage were obtained from the programme database. Costs for the second stage were provided by the financial department at the hospital. All costs were calculated in Euro (€) (2010 value) (Table 2).

Health-related outcomes

A false-positive result was defined as a mammogram with abnormal findings in the absence of confirmed breast cancer.

Screen-detected cancers were pathologically confirmed breast cancers, both ductal carcinoma in situ (DCIS) and invasive carcinomas. We did not include breast malignancies other than primary breast cancers. The cancer detection rate was calculated as the number of screen-detected cancers divided by the number of participants.

Positive predictive value (PPV) was calculated as the number of screen-detected cancers divided by the number of women recalled for additional tests.

Cost-consequence analysis

We compared costs and health-related outcomes of reading strategies described above. Figure 1 describes the algorithm followed to perform the cost-consequence analysis. We determined costs and health-related outcomes of reading strategies and compared double reading with and without consensus and arbitration and single reading with first and second reader only. We then chose the best option of the two double reading strategies and the best option of the two single reading strategies. In order to find the strategy of choice, we compared only the best option of the double reading strategies versus the best one of the single reading strategies. The strategy of choice was the less expensive strategy with the lower rate of false-positive results and the higher cancer detection rate. Finally, we estimated the incremental effect attributable to using the strategy of choice of double reading relative to the strategy of choice of single reading. Differences between reading strategies were tested using Student's t-test, Chi-square test, and Fisher exact test. Statistical tests were two-sided and all *p* values of less than 0.05 were considered statistically significant. Analyses were performed using Microsoft Excel (2011) and IBM SPSS software version 21.0 (SPSS, 2013).

Results

Table 1 shows the characteristics of the study population during the 2009–2011 and the 2011–2013 rounds. From the 57,157 mammograms included in the study, disagreement between radiologists was given in 4.5 % ($N=2,556$); 98.1 % ($N=2,508$) of these were resolved by consensus and 1.9 % ($N=48$) by arbitration (Table 2).

Cost comparison

The mean cost of each screening examination of the four strategies was: € 44.9 for double reading with consensus and

Table 2 Unit and total costs (Euro) stratified by reading strategy

	Unit cost (Euro)	Single reading		Second reader only		Double reading	
		First reader only		With consensus and arbitration		Without consensus and arbitration	
		N° (%)	Cost	N° (%)	Cost	N° (%)	Cost
<i>Mammography screening</i>							
Digital mammogram	€ 5.6	57,157	€ 322,770.0	57,157	€ 322,770.0	57,157	€ 322,770.0
Medical and technician staff	NA	NA	€ 1,408,651.9	NA	€ 1,411,385.6	NA	€ 1,418,260.3
Radiologic reading	€ 3.5	57,157	€ 199,061.8	57,157	€ 199,061.8	114,314	€ 398,123.7
Consensus session	€ 6.9	0	0.0	0	0.0	2,508	€ 17,469.3
Arbitration	€ 10.5	0	0.0	0	0.0	48	€ 501.5
<i>Additional tests</i>							
Women with additional tests	NA	2,617 (100)	NA	2,920 (100)	NA	2,822 (100)	NA
Additional mammogram	€ 32.7	1,295 (49)	€ 42,307.7	1,473 (50)	€ 48,122.9	1,473 (52)	€ 48,122.9
Ultrasonography	€ 50.1	1,928 (74)	€ 96,669.9	2,487 (85)	€ 124,698.2	2,590 (92)	€ 129,862.6
FNAC	€ 141.8	507 (19)	€ 71,902.7	540 (18)	€ 76,582.8	605 (21)	€ 85,801.1
Core biopsy	€ 131.7	322 (12)	€ 42,429.9	343 (12)	€ 45,197.1	449 (16)	€ 59,164.7
Open surgical biopsy	€ 1,536.0	14 (1)	€ 21,504.0	19 (1)	€ 29,184.0	20 (1)	€ 30,720.0
Other diagnostic test	€ 194.9	163 (6)	€ 31,759.7	188 (6)	€ 36,630.9	232 (8)	€ 45,204.0
Imputed cost [†]	NA	2	€ 468.8	1	€ 247.4	4	€ 565.8
Total cost			€ 2,237,526.5		€ 2,293,880.7		€ 2,571,867.1

NA = not applicable. [†] Fine-needle aspiration cytology. [‡] Average cost imputed in cancers cases diagnosed out of the programme. N° = number of units.

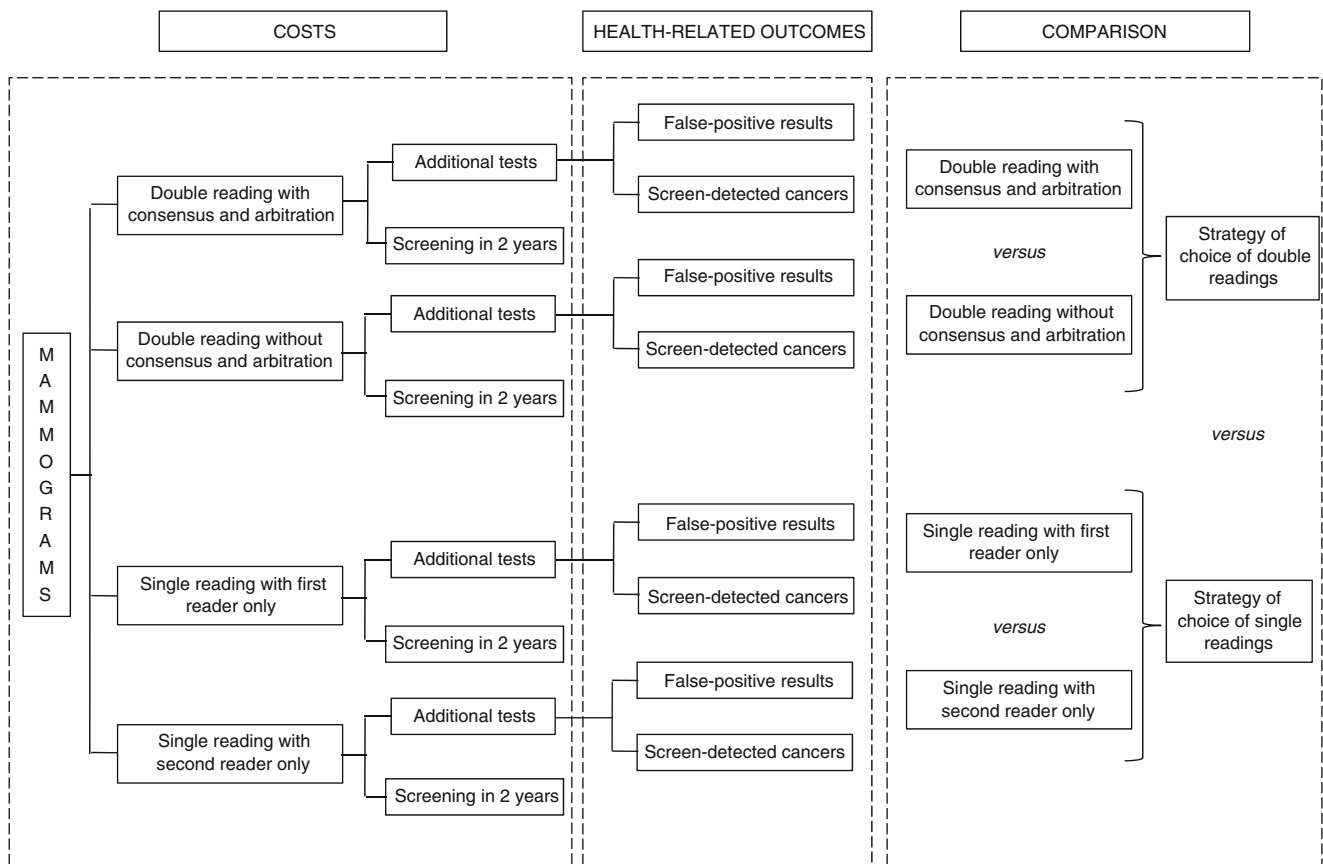


Fig. 1 Algorithm followed to perform the cost-consequence analysis

arbitration, € 45.6 for double reading without consensus and arbitration, € 39.1 for single reading with first reader only, and € 40.1 for single reading with second reader only (Table 2).

Cost-consequence analysis

Table 3 presents the costs and health-related outcomes of each reading strategy. The most relevant data are the following:

Comparison between double reading strategies:

Cost. Double reading without consensus and arbitration was 14 % (€ 36,341) more expensive than double reading with consensus and arbitration.

Health-related outcomes. Double reading without consensus and arbitration had 1.5 % more false positive results than double reading with consensus and arbitration ($p < 0.001$). Both reading strategies had similar cancer detection rates ($p = 0.986$). Double reading with consensus and arbitration was the strategy of choice for both double reading strategies.

Comparison between single reading strategies:

Cost. Single reading with second reader only was 25 % (€ 56,354) more expensive than single reading with first reader only.

Health-related outcomes. Single reading with second reader only had 0.5 % more false positive results than single reading with first reader only ($p < 0.001$). The cancer detection rate was the same with both reading strategies ($p = 1.000$). Single reading with first reader only was the strategy of choice for both single reading strategies.

Comparison of double versus single reading strategies:

We compared the strategy of choice of double readings versus the strategy of choice of single readings.

Cost. Double reading with consensus and arbitration was 15 % (€ 334,341) more expensive than single reading with first reader only.

Health-related outcomes. Double reading with consensus and arbitration had 0.3 % more false positive results than single reading with first reader only ($p = 0.001$). Both reading strategies had similar cancer detection rates (4.6 % and 4.2 %,

Table 3 Comparison between costs and health related outcomes of double and single reading strategies

Costs	Health related outcomes (from 57,157 participants)						
	DR with C/A	DR without C/A	Incremental cost	DR with C/A	DR without C/A	Incremental effect	p value
Mammography screening	€ 2,172,426.0	€ 2,139,154.0	€ 33,272.0 (1.6 %)	4.5 %	6.0 %	-1.5 %	<0.001
Additional tests	€ 399,441.2	€ 469,053.7	€ -69,612.5 (-14.9 %)	2,559	3,419	-860	
Total cost	€ 2,571,867.1	€ 2,608,207.7	€ -36,340.6 (-1.4 %)	9.3 %	7.1 %	2.2 %	0.001
				4.6%	4.7%	-0.1%	0.986
				263	266	-3	
	1st reader	2nd reader	Incremental cost	1st reader	2nd reader	Incremental effect	p value
Mammography screening	€ 1,930,483.8	€ 1,933,217.4	€ -2,733.6 (0.1 %)	4.2 %	4.7 %	-0.5 %	<0.001
Additional tests	€ 307,042.7	€ 360,663.3	€ -53,620.6 (17.5 %)	2,378	2,681	-303	
Total cost	€ 2,237,526.5	€ 2,293,880.7	€ -56,354.2 (2.5 %)	9.1 %	8.2 %	0.9 %	0.210
				4.2%	4.2%	0.0%	1.000
				240	239	1	
	1st reader	DR with C/A	Incremental cost	DR with C/A	1st reader	Incremental effect	p value
Mammography screening	€ 1,930,483.8	€ 2,172,426.0	€ 241,942.2 (12.5 %)	4.5 %	4.2 %	0.3 %	0.001
Additional tests	€ 307,042.7	€ 399,441.2	€ 92,398.4 (30.1 %)	2,559	2,378	181	
Total cost	€ 2,237,526.5	€ 2,571,867.1	€ 334,340.6 (15.0 %)	9.3 %	9.1 %	0.2 %	0.812
				2,822	2,617	205	0.004
				4.6%	4.2%	0.4%	0.283
				263	239	24	

DR = double reading. C/A = consensus and arbitration.

respectively; $p=0.283$) and positive predictive values (9.3 % and 9.1 %, respectively, $p=0.812$).

Characteristics of findings of double reading with consensus and arbitration and single reading with first reader only

A total of 2,822 (4.9 %) women underwent additional tests after double reading with consensus and arbitration, and 2,617 (4.6 %) underwent additional tests after single reading with first reader only ($p=0.004$). Of the 205 more women who underwent additional tests after double reading with consensus and arbitration, 181 (88 %) had a false-positive result and 24 (12 %) were diagnosed of cancer. The 181 women with a false-positive result underwent 1,008 additional tests of which 201 (20 %) were invasive tests. From the 24 screen-detected cancers, seven (29 %) were DCIS and 17 (71 %) were I-II stages (Table 4).

Discussion

The results of this cost-consequence analysis showed that double reading with consensus and arbitration was more expensive than single reading with first reader only and had a higher proportion of false-positive results. The two strategies differed only slightly in terms of positive predictive values and cancer detection rates. In the absence of information about interval cancer rates, these results are not conclusive and should be confirmed through a cost-effectiveness evaluation.

In our study, double reading with consensus and arbitration was more expensive due to higher staff costs and costs of additional tests. Regarding staff costs, our results are in agreement with previous studies that assessed costs of double versus single reading [23–25]. The cost of additional tests is controversial. Although several studies reported high costs of additional tests at double reading [23, 25], Brown et al., did not [26]. Contrary to our results, these authors found a significant reduction in recall rate and number of additional tests when double reading was followed by consensus and

Table 4 Health-related outcomes of double reading with consensus and arbitration and single reading with first reader only

	False-positive results				P value	Screen detected cancers				
	Double reading with consensus and arbitration		Single reading with first reader only			Double reading with consensus and arbitration		Single reading with first reader only		
	N°	(%)	N°	(%)		N°	(%)	N°	(%)	
Total	2,559		2,378			263		239		
Prevalent screening	1,093	(42.7)	1,029	(43.3)	0.708	71	(27.0)	68	(28.5)	0.765
Subsequent screening	1,466	(57.3)	1,349	(56.7)		192	(73.0)	171	(71.5)	
Age at screening										
50-54	1,198	(46.8)	1,120	(47.1)	0.951	73	(27.8)	68	(28.5)	0.982
55-59	542	(21.2)	512	(21.5)		62	(23.6)	53	(22.2)	
60-64	449	(17.5)	403	(16.9)		60	(22.8)	54	(22.6)	
65-69	370	(14.5)	343	(14.5)		68	(25.8)	64	(26.7)	
BI-RADS										
Incomplete	1,169	(45.7)	1,132	(47.6)	0.048	98	(37.3)	98	(41.0)	0.017
Benign/Probably benign	698	(27.3)	570	(24.0)		10	(3.8)	0	(0.0)	
Suspicious	670	(26.2)	660	(27.8)		19	(7.2)	13	(5.4)	
Highly suggestive	22	(0.9)	16	(0.7)		136	(51.7)	128	(53.6)	
Stage										
DCIS						42	(16.2)	35	(14.9)	0.904*
I-II						206	(79.5)	189	(80.4)	
III-IV						11	(4.2)	11	(4.7)	
Molecular subtype										
Luminal A						156	(61.2)	140	(60.3)	0.927*
Luminal B						70	(27.5)	67	(28.9)	
HER 2						15	(5.9)	11	(4.7)	
Triple negative						14	(5.5)	14	(6.0)	

BI-RADS = Breast Imaging Report and Database System. *Based on available data.

arbitration. The explanation for these differences may be that their study was undertaken during prevalent screening when previous mammograms are not available and recall rates are often higher. In several other studies, the costs of additional tests were not calculated, but the authors reported high recall rates at double reading [14, 27, 28], suggesting increases in costs of additional tests.

The costs of additional tests in our study were mainly due to false-positive results. We calculated that each false-positive result cost a mean of 511 Euros in additional tests. To our knowledge, this is the first study in Spain to report this finding. Although several European studies have assessed the consequences of false-positive results, they did not report their costs [9]. In a study in Washington State, Chubak et al., assessed the cost of breast-related care in women with false-positive results and found a mean cost of 527 dollars per year [20]. Although this finding is not comparable with our result due to the differences between screening programmes, the results of both studies show an increase in costs due to false-positive results. Interestingly, we found that 20 % of additional tests in women with false-positive results were invasive tests. Invasive tests are expensive and can produce more negative psychological effects than non-invasive tests [9]. Future economic evaluations of false-positive results should thus take women who undergo invasive tests into account.

Our results suggest that single reading could imply net savings for the national health system in European countries. The average cost of a screening examination in our study varied little between double reading with consensus and arbitration (€ 44.9) and single reading (€ 39.1). This small variation, however, represented savings of about € 0.4 million per 57,157 screening examinations. Considering that thousands of women are screened every two years, changing to single reading could largely reduce the costs of breast cancer screening programmes.

Regarding the health-related outcomes, our results suggest that single reading may have a better interpretative accuracy than double reading. In our study, double reading with consensus and arbitration had more false-positive results than single reading while the positive predictive value was similar in both. Our results are in agreement with the U.S. National Cancer Institute that reported that the area under the ROC curve was higher among screening-units that did not perform double reading [29]. An examination of single reading using a better characterization of its diagnostic accuracy is needed to identify whether the area under the ROC curve of single reading is superior to double reading in our context.

The relation between overdiagnosis and reading strategies in breast cancer screening programmes has not been evaluated. Double reading has classically been considered beneficial because it detects more cancers than single reading, but it may increase the detection of ductal carcinoma in situ (DCIS), carrying a potential risk of overdiagnosis. Single reading, in

contrast, could reduce the probability of overdiagnosis due to a better interpretative accuracy. In our study, double reading with consensus and arbitration detected 24 more cancers than single reading; seven of these (29 %) were DCIS. Although the extent of overdiagnosis attributable to DCIS remains controversial [30], it is not recommended to increase its detection in breast cancer screening programmes [3]. Further prospective long-term studies are needed to evaluate the relation between overdiagnosis and reading strategies in breast cancer screening programmes.

Several studies have assessed the costs and cost-effectiveness of double versus single reading but in contrast with our results, they reported that double reading can be more cost-effective than single reading [10]. There are several reasons for this discrepancy. First, previous studies were performed in the context of screen-film mammography. Digital mammography can be more sensitive than screen-film mammography, reducing the possibility of missed lesions. Second, previous studies did not assess the false-positive results of reading strategies. Finally, previous studies were performed in the first or second round of screening programmes while our results came from the fifth and sixth rounds of a well-established screening programme.

The experience of readers can affect the effectiveness of reading strategies. All five radiologists in our study met the European guidelines recommending that over 5000 mammograms should be interpreted each year to assure high quality standards. The participation rate can also affect the effectiveness of reading strategies in certain conditions. For example, a very low participation rate can limit the opportunity of readers to gain more experience. The participation rate in our study was not as high as the 75 % recommended by the European guidelines [3]. However, in view of our readers' experience, it is unlikely that the participation rate directly affected their performance.

This study has several limitations. First, it is difficult to transfer costs of reading strategies from one country to another because breast cancer screening programmes differ considerably in unitary costs, working times, and protocol-reading variables. However, our results illustrate the relative increase in costs of double reading and may be similar to other European breast cancer screening programmes. Second, in this study we did not calculate the cost-effectiveness of reading strategies, but our results emphasize the need for future cost-effectiveness and cost-utility analyses. Third, we did not analyze indirect and non-health related costs because we performed the analysis from the health system perspective. Fourth, this study was conducted in a very specific context: four certified, highly trained, radiologists were involved in reading the mammograms. Therefore, the generalization of our results to other contexts with less trained radiologists is doubtful and additional data are needed. Fifth, we were unable to calculate sensitivity and specificity of the programme

because we cannot obtain information about interval cancers as data from population-based cancer registries are not available in our city. The absence of this information did not affect this cost-consequence analysis but should be taken into account in further cost-effectiveness evaluations. Finally, the major strengths of this analysis are that it presents context-specific real-world data and is the first cost-consequence analysis of double versus single reading in the era of digital mammography.

In conclusion, contrary to a previous systematic review [10], our results indicate that double reading is an expensive strategy that produces more false-positive results than single reading without significantly increasing the cancer detection rate. Our results are not conclusive as this study was conducted in a specific context and data regarding interval cancers was not available. However, in view of our results, further cost-effectiveness evaluations are needed to determine whether double reading is still necessary at digital screening mammography.

Acknowledgements The scientific guarantor of this publication is Margarita C. Posso Rivera. The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article. The authors state that this work has not received any funding. No complex statistical methods were necessary for this paper. Institutional Review Board approval was obtained. Written informed consent was waived by the Institutional Review Board. Methodology: retrospective, cross sectional study, performed at one institution.

References

1. Broeders M, Moss S, Nyström L et al (2012) The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen* 19:14–25
2. The Independent UK Panel on Breast Cancer Screening (2012) The benefits and harms of breast cancer screening: an independent review. *Lancet* 380:1778–1786
3. European Commission. Perry N, Broeders M, de Wolf C, et al. (2006) European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis, 4th edn. Office for Official Publications of the European Communities, Luxembourg
4. Gøtzsche PC, Jørgensen KJ (2013) Screening for breast cancer with mammography. *Cochrane Database Syst Rev*. doi:10.1002/14651858
5. Biller-Andorno N, Jüni P (2014) Abolishing Mammography Screening Programs? A View from the Swiss Medical Board. *N Engl J Med* 370:1965–1967
6. Jørgensen KJ (2013) Mammography screening. Benefits, harms, and informed choice 60:B4614
7. Paci E, Broeders M, Hofvind S, Puliti D, Duffy SW, EUROSCREEN Working Group (2014) European breast cancer service screening outcomes: a first balance sheet of the benefits and harms. *Cancer Epidemiol Biomarkers Prev* 23:1159–1163
8. World Health Organisation WHO (2002) International Agency for Research on Cancer. Breast cancer screening. World Health Organization, Geneva. Available via: http://www.iarc.fr/en/publications/pdfs-online/prev/handbook7/Handbook7_Breast-5.pdf. Accessed 26 Jul 2015.
9. Bond M, Pavey T, Welch K et al (2013) Systematic review of the psychological consequences of false-positive screening mammograms. *Health Technol Assess*. doi:10.3310/hta17130
10. Dinnes J, Moss S, Melia J, Blanks R, Song F, Kleijnen J (2001) Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: findings of a systematic review. *Breast* 10:455–463
11. Taylor P, Potts HW (2008) Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. *Eur J Cancer* 44:798–807
12. van den Biggelaar F, Kessels A, van Engelshoven J, Flobbe K (2009) Strategies for digital mammography interpretation in a clinical patient population. *Int J Cancer* 125:2923–2929
13. Román R, Sala M, Salas D et al (2012) Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. *Ann Oncol* 23:104–111
14. Taplin SH, Rutter CM, Elmore JG, Seger D, White D, Brenner RJ (2000) Accuracy of screening mammography using single versus independent double interpretation. *AJR Am J Roentgenol* 174:1257–1262
15. 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* 250:354–362
16. Duijm LEM, Groenewoud JH, Hendriks JHCL, de Koning HJ (2004) Independent double reading of screening mammograms in The Netherlands: effect of arbitration following reader disagreements. *Radiology* 231:564–570
17. Hofvind S, Geller BM, Rosenberg RD, Skaane P (2009) Screening-detected breast cancers: discordant independent double reading in a population-based screening program. *Radiology* 253:652–660
18. Klompenhouwer EG, Voogd AC, den Heeten GJ et al (2015) Discrepant screening mammography assessments at blinded and non-blinded double reading: impact of arbitration by a third reader on screening outcome. *Eur Radiol* 25:2821–2829
19. Gray AM, Clarke PM, Wolstenholme JL, Wordsworth S (2010) Applied Methods of Cost-effectiveness Analysis in Healthcare, 1st ed. Oxford University Press
20. Chubak J, Boudreau DM, Fishman PA, Elmore JG (2010) Cost of Breast-Related Care in the Year Following False Positive Screening Mammograms. *Med Care* 48:815–820
21. Payne JI, Martin T, Caines JS, Duggan R (2014) The Burden of False-Positive Results in Analog and Digital Screening Mammography: Experience of the Nova Scotia Breast Screening Program. *Can Assoc Radiol J* 65:315–320
22. Sala M, Domingo L, Macià F et al (2015) Does digital mammography suppose an advance in early diagnosis? Trends in performance indicators 6 years after digitalization. *Eur Radiol* 25:850–859
23. Leivo T, Salminen T, Sintonen H et al (1999) Incremental cost-effectiveness of double-reading mammograms. *Breast Cancer Res Treat* 54:261–267
24. Ciatto S, Del Turco MR, Morrone D et al (1995) Independent double reading of screening mammograms. *J Med Screen* 2:99–101
25. Seradour B, Wait S, Jacquemier J, Dubuc M, Piana L (1997) Double reading of mammograms in the Bouches-du-Rhône screening programme. Results and costs, 1990–1995. *J Radiol* 78:49–54

26. Brown J, Bryan S, Warren R (1996) Mammography screening: an incremental cost effectiveness analysis of double versus single reading of mammograms. *BMJ* 312:809–812
27. Ciatto S, Ambrogetti D, Bonardi R et al (2005) Second reading of screening mammograms increases cancer detection and recall rates Results in the Florence screening programme. *J Med Screen* 12: 103–106
28. Pauli R, Hammond S, Cooke J, Ansell J (1996) Comparison of radiographer/radiologist double film reading with single reading in breast cancer screening. *J Med Screen* 3:18–22
29. Taplin S, Abraham L, Barlow WE et al (2008) Mammography facility characteristics associated with interpretive accuracy of screening mammography. *J Natl Cancer Inst* 100:876–887
30. Ripping TM, Verbeek ALM, Fracheboud J, de Koning HJ, van Ravesteyn NT, Broeders MJM (2015) Overdiagnosis by mammographic screening for breast cancer studied in birth cohorts in The Netherlands. *Int J Cancer* 137:921–929

5.2. ARTÍCULO 2. ANÁLISIS DE COSTE-EFECTIVIDAD DE LA DOBLE LECTURA MAMOGRÁFICA

5.2.1. RESUMEN DE RESULTADOS

CARACTERÍSTICAS DE LA POBLACIÓN DE ESTUDIO

Se analizaron 28.636 mujeres con mamografía digital. Del total de mamografías evaluadas, 5.978 (20,9%) fueron prevalentes y 22.636 (79,1%) fueron incidentes. En un 5,6% (N = 1.614) de las mamografías se observó desacuerdo entre los radiólogos. El 93,1% (N = 26.661) de las mamografías no presentó hallazgos sospechosos y las mujeres fueron invitadas a continuar el cribado bienal. Un 4.8% (N = 1,366) de mujeres fueron recitadas para realizar alguna prueba complementaria y un 2.1% (N = 609) fueron invitadas a un control anual con mamografía.

RESULTADOS DE EFECTIVIDAD

La tasa de detección de cáncer varió de acuerdo a las tres estrategias evaluadas: i) doble lectura; ii) doble lectura en el cribado prevalente y lectura simple en el cribado incidente; y iii) lectura simple. El número de cánceres detectados fue de 148 (5,17 ‰) con la doble lectura, 138 (4,82 ‰) con la doble lectura en el cribado prevalente y de lectura simple en el cribado incidente y 137 (4,78 ‰) con la lectura simple (P = 0,77).

USO DE RECURSOS Y COSTES

Se estimaron los costes comunes y variables para las tres estrategias evaluadas. El coste común correspondiente a obtener 28.636 mamografías fue de 855.297 euros (€). Los costes variables consideraron el coste del personal y las pruebas adicionales.

El coste de la doble lectura fue de 1.318.968 €, seguido de 1.188.756 € con la doble lectura el cribado prevalente y lectura simple y de 1.135.445 € con la lectura simple. El coste de la doble lectura fue mayor debido a la mayor frecuencia de pruebas complementarias realizadas, mayor frecuencia de controles avanzados y al coste de la segunda lectura.

ANÁLISIS DE COSTE-EFECTIVIDAD

El coste medio correspondiente a la detección de un cáncer fue de 8.912 € con la doble lectura, 8.614 € con la doble lectura en el cribado prevalente y la lectura simple en el cribado incidente y 8.287 € con la lectura simple.

La lectura simple fue la menos costosa y detectó un menor número de cánceres. En segundo lugar se ubicó la doble lectura en el cribado prevalente y lectura simple en el cribado incidente y en tercer lugar la lectura simple. No se identificó ninguna estrategia dominada.

La relación de coste-efectividad incremental (RCEI) se calculó comparando una estrategia con la inmediatamente menos costosa. La RCEI de la doble lectura en el cribado prevalente y lectura simple en el cribado incidente frente a la lectura simple fue de 53.312 € por cáncer detectado. Esta estrategia fue descartada debido a que fue dominada por extensión (mayor coste por RCEI que la lectura doble). La RCEI de la lectura doble frente a la lectura simple fue de 16.684 €. Este monto fue más del doble del coste medio por cáncer detectado con la lectura simple (Tabla 5).

Tabla 5. Cálculo de la relación de coste-efectividad incremental (RCEI) de las estrategias incluidas en el análisis.

Estrategia de lectura	Coste estimado €	Efecto estimado		Coste incremental €	Efecto incremental	Coste promedio €	RCEI €
		Nº de cánceres	Tasa de detección				
Lectura simple	1.135.445	137	4,78%	-	-	8.287	-
Doble lectura prevalente y lectura simple incidente	1.188.756	138	4,82%	53.312	1	8.614	53.312
Doble lectura	1.318.968	148	5,17%	183.523	11	8.912	¶16.684

RCEI = relación de coste-efectividad incremental.

¶La estrategia con doble lectura en el cribado prevalente y lectura simple en el cribado incidente fue dominada por extensión por lo que la RCEI se calculó para la doble lectura frente a la lectura simple.

PUBLICACIÓN 2

POSSO M, CARLES M, RUÉ M, PUIG T, BONFILL X. COST-EFFECTIVENESS OF DOUBLE READING VERSUS SINGLE READING OF MAMMOGRAMS IN A BREAST CANCER SCREENING PROGRAMME. PLOS ONE. 2016; 11(7): e0159806.

PLoS One. 2016.

Factor de impacto 2,806. Q1 (15/64). Multidisciplinary sciences.

RESEARCH ARTICLE

Cost-Effectiveness of Double Reading versus Single Reading of Mammograms in a Breast Cancer Screening Programme

Margarita Posso^{1*}, Misericòrdia Carles², Montserrat Rué³, Teresa Puig^{1,4}, Xavier Bonfill^{1,4,5}

1 Service of Clinical Epidemiology and Public Health, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain, **2** Economics Department, Universitat Rovira i Virgili, Reus, Spain, **3** Basic Medical Sciences Department, Biomedical Research Institut of Lleida (IRBLLEIDA), Universitat de Lleida, Lleida, Spain, **4** Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, **5** CIBER of Epidemiology and Public Health (CIBERESP), Barcelona, Spain

* mposso@cochrane.es



OPEN ACCESS

Citation: Posso M, Carles M, Rué M, Puig T, Bonfill X (2016) Cost-Effectiveness of Double Reading versus Single Reading of Mammograms in a Breast Cancer Screening Programme. PLoS ONE 11(7): e0159806. doi:10.1371/journal.pone.0159806

Editor: Marcia Edilaine Lopes Consolaro, State University of Maringá/Universidade Estadual de Maringá, BRAZIL

Received: March 31, 2016

Accepted: July 10, 2016

Published: July 26, 2016

Copyright: © 2016 Posso et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. Additional data are available upon request from the authors.

Funding: The authors have no support or funding to report.

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

Abstract

Objectives

The usual practice in breast cancer screening programmes for mammogram interpretation is to perform double reading. However, little is known about its cost-effectiveness in the context of digital mammography. Our purpose was to evaluate the cost-effectiveness of double reading versus single reading of digital mammograms in a population-based breast cancer screening programme.

Methods

Data from 28,636 screened women was used to establish a decision-tree model and to compare three strategies: 1) double reading; 2) double reading for women in their first participation and single reading for women in their subsequent participations; and 3) single reading. We calculated the incremental cost-effectiveness ratio (ICER), which was defined as the expected cost per one additionally detected cancer. We performed a deterministic sensitivity analysis to test the robustness of the ICER.

Results

The detection rate of double reading (5.17%) was similar to that of single reading (4.78%; $P = .768$). The mean cost of each detected cancer was €8,912 for double reading and €8,287 for single reading. The ICER of double reading versus single reading was €16,684. The sensitivity analysis showed variations in the ICER according to the sensitivity of reading strategies. The strategy that combines double reading in first participation with single reading in subsequent participations was ruled out due to extended dominance.

Conclusions

From our results, double reading appears not to be a cost-effective strategy in the context of digital mammography. Double reading would eventually be challenged in screening programmes,

as single reading might entail important net savings without significantly changing the cancer detection rate. These results are not conclusive and should be confirmed in prospective studies that investigate long-term outcomes like quality adjusted life years (QALYs).

Introduction

Mammogram is the test of choice in European breast cancer screening programmes since it can detect breast cancer at an early stage [1–3]. Whereas digital mammography is a technology that can reduce false-positive results, no significant differences in the cancer detection rate were stated when it was compared to screen-film mammography [4]. In addition, an evaluation of its costs showed that screening with digital mammography can save long-term budget expense in breast cancer screening programmes [5]. Screening with digital mammography, therefore, has been widely implemented.

As two readers are unaware of each other's interpretation, double reading can increase sensitivity reducing the chance of missed lesions [6–12]. Thus, double reading of digital mammograms became the usual practice in European programmes [1]. However, the following reasons might bring into question its cost-effectiveness.

The effectiveness of double reading may be less important in situations where a high level of agreement between radiologists exists [13]. The benefit of double reading may be restricted to particular settings in which cancer detection is difficult, i.e. mammograms of women in their first participation (prevalent screening) when no previous images are available, women with small lesions that are not easy to find, or when the readers are less experienced [6,14–17]. In addition, having two readers may significantly increase the time, staff costs and resources used in the reading process [18].

Information obtained from cost-effectiveness analyses is useful to decision makers when deciding to implement breast cancer screening programmes and evaluate its benefits and potential harms. One previous cost-effectiveness analysis based on European data, reported that risk-based strategies could reduce harms and costs [19]. Shifting from double reading to single reading was not analysed in this study. However, it is reasonable to hypothesize that in some contexts more benefits can be obtained from single reading as it may reduce costs and false-positives without significantly reducing the cancer detection rate [20]. Conversely, other cost-effectiveness analyses performed in European countries reported double reading as a cost-effective strategy in programmes that used screen-film mammography [1,6,21,22],

Recently, economic evaluations have focused on the cost-effectiveness of double reading versus the combination of single reading and CAD (computer-aid detection) [23–26], whilst studies of single reading without CAD have not yet been published in the context of digital mammography [23,27]. In fact, as far as we know, little is known about whether double reading is a cost-effective strategy in digital screening. Therefore, the main purpose of this study was to assess the cost-effectiveness of double reading versus single reading of digital mammograms in a breast cancer screening programme.

Materials and Methods

Study population

This study was performed in women participating in a population-based breast cancer screening programme of the Hospital Sant Pau in an area of 390,000 inhabitants in Barcelona, Spain. The programme was funded by the Public Health Insurance to invite women of 50–69 years of

age for biennial screening. In this study, we included all digital mammograms performed from June 2009 until May 2011 and followed up until May 2013. We excluded two mammograms because the data of the reader’s interpretation was incomplete. A total of 28,636 mammograms were analysed, 5,978 (20.9%) corresponding to women participating for the first time in the breast cancer screening programme (prevalent screening), and 22,658 (79.1%) corresponding to subsequent participations (subsequent screening) (Table 1).

Two projections (mediolateral-oblique and craniocaudal) were taken per breast in all screening examinations. Four certified screening radiologists, who read at least 5,000 mammograms per year, read the mammograms in the breast cancer screening programme. Two radiologists separately read each mammogram (independent double reading). The radiologists classified the results of each mammogram as follows: (1) recall, in which case additional tests were requested to confirm or ruled out malignancy; (2) early recall, in which case the woman was programmed for a further screening mammography in 12 months; or (3) two-year screening, in which case the woman was programmed for a further screening mammography in two years. In case of disagreement between radiologists, the result of the mammogram was determined by consensus or by arbitration. Fig 1 shows the algorithm of the decisions made in a screening round.

Screen-detected cancers were pathologically confirmed breast cancers, both ductal carcinoma in situ (DCIS) and invasive carcinomas. We did not include breast malignancies other than primary breast cancers. The cancer detection rate was calculated as the number of screen-detected cancers divided by the number of participants. The Ethics Committee of the Hospital Sant Pau approved the study. Informed consent was not required, since the data were retrospectively collected, and records were anonymized before receipt and analysis.

Model design

A decision-tree model was used because of its appropriateness to reflect the immediate effects of decisions taken during a screening round. In this model, the main outcome of effectiveness was the screen-detected cancers; other performance and diagnostic accuracy outcomes were also analysed (S1 Table). The time horizon was four years—from June 2009 to May 2013—, which included a biennial screening round (2009–2011) plus two-year follow-up to confirm negative results. This model assumes that participants do not suffer other conditions that prevent them from successfully completing the entire time horizon.

The cost-effectiveness of double versus single reading was evaluated comparing the following strategies (Fig 2):

Table 1. Characteristics of the women included in the analysis.

	Participants in one screening round (2009–2011)	
	No.	%
Study population[†]	28,636	100.0
Prevalent screening	5,978	20.9
Incident screening	22,658	79.1
Age at screening		
50–54	8,181	28.6
55–59	6,947	24.3
60–64	7,047	24.6
65–69	6,461	22.6

[†]The information of these women was included in the decision-tree model as common parameters for all reading strategies.

doi:10.1371/journal.pone.0159806.t001

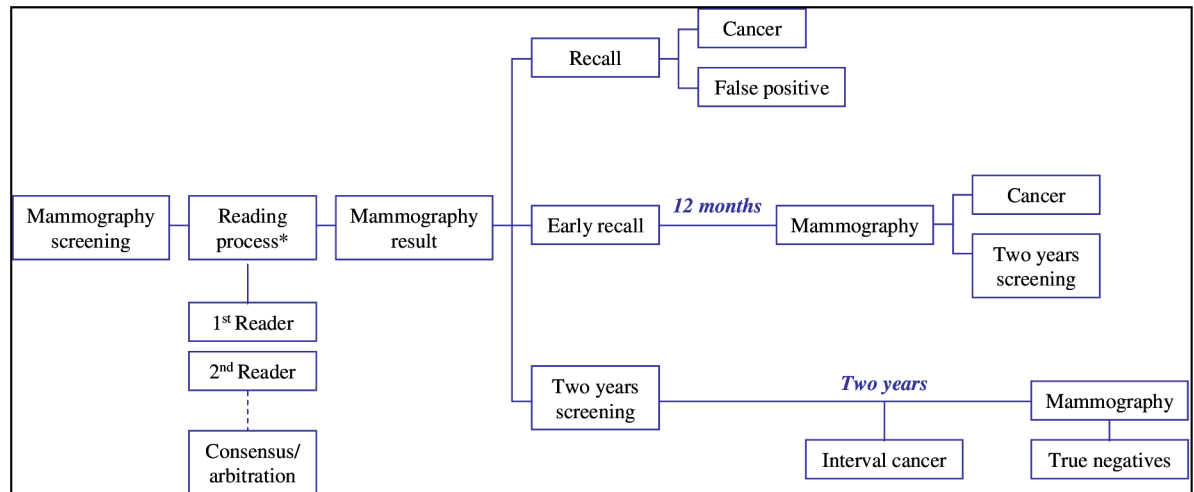


Fig 1. Algorithm followed during a biennial screening round in the programme. *The reading process included independent double reading followed by consensus and arbitration in case of disagreement.

doi:10.1371/journal.pone.0159806.g001

Double reading. Double reading with consensus and arbitration was the procedure performed in the breast cancer screening programme. For this strategy, we included the results of mammograms (recall, early recall or two-year screening) in the model, as well as the number of cancers in the way they were observed in a real setting.

Double reading in prevalent screening and single reading in incident screening. This strategy was based on the assumption that double reading was performed only for mammograms of women in their first participation, and single reading was performed for mammograms of women in their subsequent participations.

Single reading. This strategy was based on the assumption that the first radiologist alone determined the result of mammograms. In order to evaluate the effect of switching from the first to the second reader, we included the results of the second radiologist in the sensitivity analysis.

Costs data

As this study was conducted from the perspective of the public health system, we included direct healthcare costs of practitioner and diagnostic tests. Total costs were calculated using a bottom-up costing method in two stages: mammography screening and additional tests. Unitary costs for the first stage were obtained from the programme database. The financial department at the hospital provided costs for the second stage. All costs were calculated in Euro (€) (2010 value) (S2 Table). Since long-term health benefits are expected in a preventive programme, NICE’s recommendation carries a lower discount rate that reflects society’s preferences for benefits in the future [28–30]. In this study, we used a short-term horizon that prevents important variations associated with discounting. For these two reasons, no discount rate has been applied.

Determining cost-effectiveness

We compared reading strategies using incremental cost-effectiveness ratio (ICER) that indicates the additional cost of obtaining one additionally screen detected cancer. To construct the hierarchy of cost-effectiveness of the reading strategies we did the following: firstly, we calculated the average cost-effectiveness ratio (ACER) by dividing the total costs of each strategy by

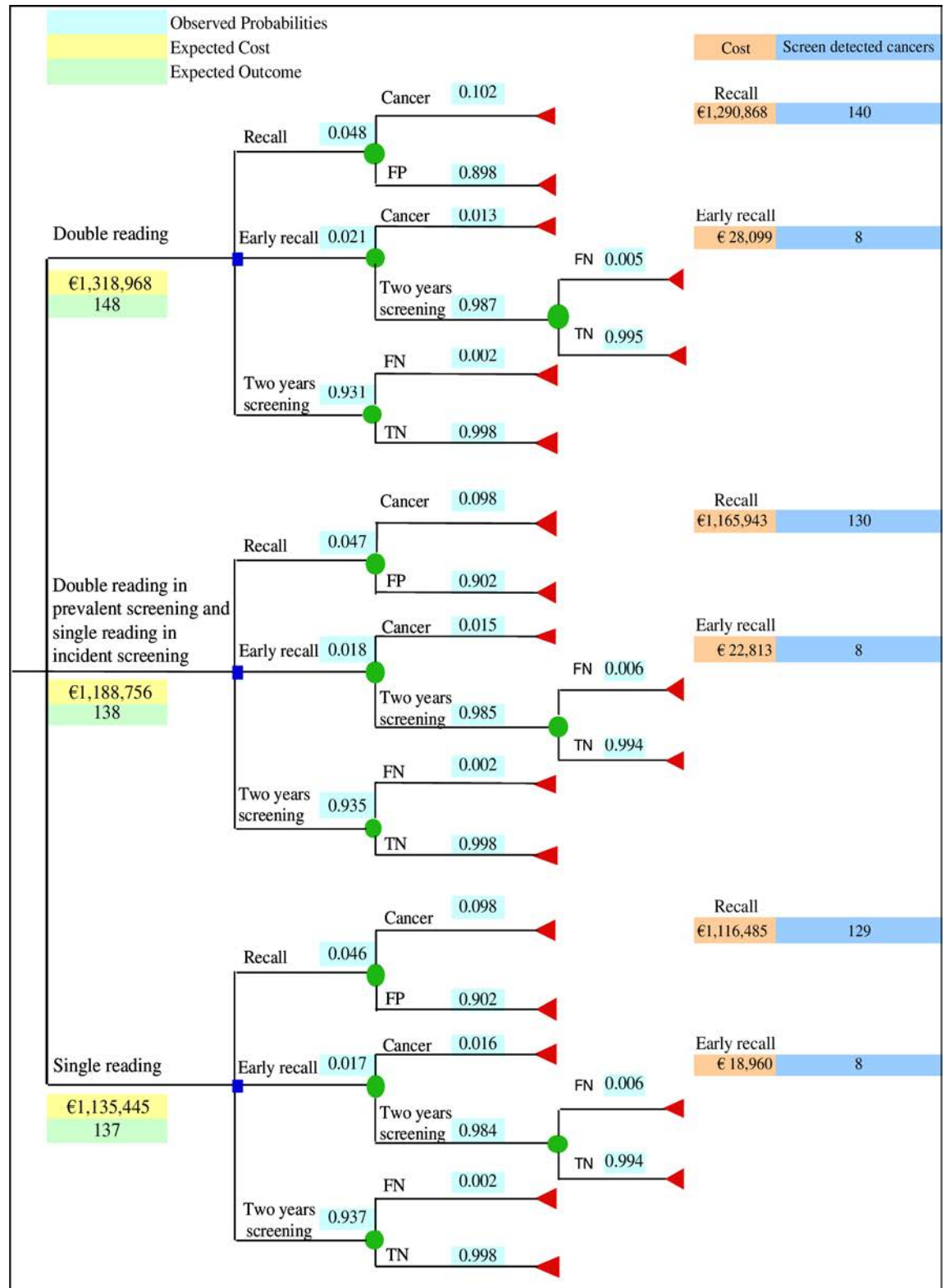


Fig 2. Decision-tree model used to evaluate the cost-effectiveness of the three reading strategies. FP = False Positive. FN = False Negative. TN = True Negative.

doi:10.1371/journal.pone.0159806.g002

its corresponding number of screen detected cancers. This allowed us to rule out, when necessary, a dominated strategy (more costly and less effective than alternatives). Secondly, we calculated ICERs and reported them starting with the lowest additional cost per additionally detected cancer. This allowed us to identify a strategy with lower effectiveness and higher ICER and to rule it out for extended dominance. Finally, ICERs for the non-ruled out strategies were compared. The most cost-effective strategy showed the lowest ICER.

Sensitivity analysis

A deterministic sensitivity analysis was carried out to assess the robustness of ICERs according to the variation in the following variables: participation rate, breast cancer prevalence, sensitivity of the reading strategy, positive-predictive value (PPV) of recall, early recall rate and staff costs (S3 Table).

Statistical tests were two-sided and all P values of less than 0.05 were considered statistically significant. We used Microsoft Excel, Redmond, Washington (2011) in all analyses.

Results

Effectiveness of reading strategies

The characteristics of the 28,636 women included in the study are presented in Table 1. The most relevant data for comparing the three reading strategies are the following. The number of screen detected cancers of the three strategies was: 148 (5.17‰) for double reading, 138 (4.82‰) for double reading in prevalent screening and single reading in incident screening, and 137 (4.78‰) for single reading ($P = 0.768$). Sixteen interval cancers (0.56‰) occurred after double reading (Table 2).

Characteristics of screen detected cancers. The number of screen detected cancers at double reading was higher in incident screening ($n = 104$) than in prevalent screening ($n = 44$). However, compared to incident screening (4.59‰), the cancer detection rate was higher in prevalent screening (7.36‰; $P = 0.008$). Of the 148 screen detected cancers, 117 (79.1%) were detected in concordance between both radiologists, while 31 (20.9%) were detected after consensus or arbitration. Thirty (20.5%) were carcinomas in situ (CIS) and 116 (79.5%) were invasive carcinomas. Compared to single reading, double reading increased by 25% ($n = 6$) and 4.5% ($n = 5$) the number of CIS and invasive carcinomas, respectively. No other relevant differences were observed in the characteristics of screen detected cancers according to reading strategies (S3 Table).

Cost of reading strategies and additional diagnostic tests

The cost of the three strategies was the following: €1,318,968 for double reading; €1,188,756 for double reading in prevalent screening and single reading in incident screening; and €1,135,445 for single reading. The final amount of common costs was €855,298, which was 64.9% and 75.3% of the total costs at double and single reading, respectively. Differences in non-common costs were mostly attributed to the number of readings and additional diagnostic tests (Table 3).

Cost effectiveness analysis

Average cost-effectiveness ratio. The average cost of each screen detected cancer for the three strategies was €8,912 for double reading; €8,614 for double reading in prevalent screening and single reading in incident screening; and €8,287 for single reading. Compared to double

Table 2. Effectiveness-outcomes of reading strategies included in the model.

	Double reading		Double reading in prevalent screening and single reading in incident screening		Single reading		P value
	No.	%	No.	%	No.	%	
Mammogram results							
Recall	1,366	4.8	1,333	4.7	1,322	4.6	.004
Early recall [†]	609	2.1	523	1.8	490	1.7	
Two-year screening	26,661	93.1	26,780	93.5	26,824	93.7	
Performance measure							
Readers' agreement	27,022	94.4	27,982	97.7	NA	NA	< .001
Consensus or arbitration	1,614	5.6	654	2.3	NA	NA	
PPV of recall	140	10.2	130	9.8	129	9.8	.884
Diagnostic accuracy							
Sensitivity	148	94.8%	138	88.5%	137	87.8%	.027 ^Σ
Specificity	27,214	95.5%	27,242	95.6%	27,256	95.7%	.389 ^Σ
Cancers							
Cancer detection rate [¶]	148	5.17‰	138	4.82‰	137	4.78‰	.768
Interval cancers [§]	16	0.56‰	26	0.91‰	27	0.94‰	.200-.093 ^Σ
Histologic type							
Invasive	116	79.5	111	81.6	111	82.2	.822
In situ	30	20.5	25	18.4	24	17.8	
Unknown	2	-	2	-	2	-	

NA = not applicable. PPV = positive predictive value.

[†]Women referred to mammography-control in 12 months.

^ΣChi square test of double versus single reading.

[¶]Detection rate per 1000 screened women.

[§]Interval cancers detected in the subsequent two years after a negative screening episode.

doi:10.1371/journal.pone.0159806.t002

Table 3. Costs estimation for a biennial period (2009–2011) according to reading strategy.

	Common parameters for all three reading strategies					
	No.			Cost €		
Mammogram	28,636			164,787		
Staff	NA			690,510		
Non-common costs	Double reading		Double reading in prevalent screening and single reading in incident screening		Single reading	
	No.	Cost €	No.	Cost €	No.	Cost €
^β Staff	NA	255,363	NA	156,874	NA	123,052
[¶] Additional tests	1,414	205,170	1376	173,890	1361	154,570
Supplies in early recall	609	2,331	523	2,002	490	1,876
*Maintenance in early recall	NA	807	NA	693	NA	649
Total		1,318,968		1,188,756		1,135,445

NA = not apply.

^βRadiologist and administrative staff involved in the reading process, consensus or arbitration; also administrative and technical staff involved in the early recall process.

[¶]Ultrasound, additional mammography, fine-needle aspiration cytology, core biopsy, open surgical biopsy, other minimal procedures.

*Including depreciation of mammography machine.

doi:10.1371/journal.pone.0159806.t003

Table 4. Process to calculate the Incremental Cost-Effectiveness Ratio (ICER) of the reading strategies.

Reading strategy	Expected cost €	Expected outcome		Incremental cost €	Incremental effect	Average cost €	ICER €
		No. of cancers	Detection rate				
Single reading	1,135,445	137	4.78‰			8,287	
Double reading in prevalent screening and single reading in incident screening	1,188,756	138	4.82‰	53,312	1	8,614	53,312
Double reading	1,318,968	148	5.17‰	183,523	11	8,912	[†] 16,684

[†]The strategy that combines double reading in prevalent screening with single reading in incident screening was ruled out by extended dominance; therefore, the value of the ICER represents the comparison between double and single reading.

doi:10.1371/journal.pone.0159806.t004

reading, the other two strategies were less expensive and less effective; therefore, none of the strategies was ruled out because of dominance.

Incremental cost-effectiveness ratio (ICER). Table 4 shows the process to calculate ICERs. The ICER of double reading in prevalent screening and single reading in incident screening versus single reading was €53,312. This strategy was ruled out of the analysis due to extended dominance. The ICER of double versus single reading was €16,684. This amount was more than twice the average cost per one cancer detected at single reading (€8,287).

In Fig 3, the ICERs are plotted over the cost-effectiveness plane where the continuous line represents the expected performance at single reading. The discontinuous line shows that shifting from single to double reading resulted in an increment of €183,523 in cost, and an increase of 0.39‰ in the cancer detection rate. In order to be as cost-effective as single reading, the required cancer detection rate and cost of double reading should shift to points A and B, respectively.

Sensitivity analysis

Our model was more sensitive to changes in terms of sensitivity of reading strategies for detecting cancers. Varying the current participation rate and observed breast cancer prevalence also affected our results, whereas changes in PPV of recall, staff costs or early recall proportion had less impact on the results (Fig 4 and S5 Table).

Discussion

We propose a decision-tree model to compare three reading strategies in a population-based breast cancer screening programme. The results of this cost-effectiveness analysis showed that the ICER per one additionally detected cancer was drastically higher (100% more expensive) for double reading than the average cost of each screen detected cancer at single reading. To our knowledge, this is one of the first economic evaluations of double reading versus single reading in the context of digital mammography [18], and the first one assessing single reading without CAD [27].

Our results might be useful for exploring those strategies that require less expensive ICER per one additionally screen detected cancer. Firstly, we found that the combination of double reading in prevalent screening and single reading in incident screening was not cost-effective due to extended dominance, which means that its ICER was higher than that of double reading. This result can be explained by the characteristics of cancers in incident screening that are frequently detected after discordance between readers [14]. Secondly, the effectiveness to detect cancers was similar between double reading and single reading. In fact, an agreement between readers' interpretations was observed in 94.4% of mammograms and the cancer detection rate

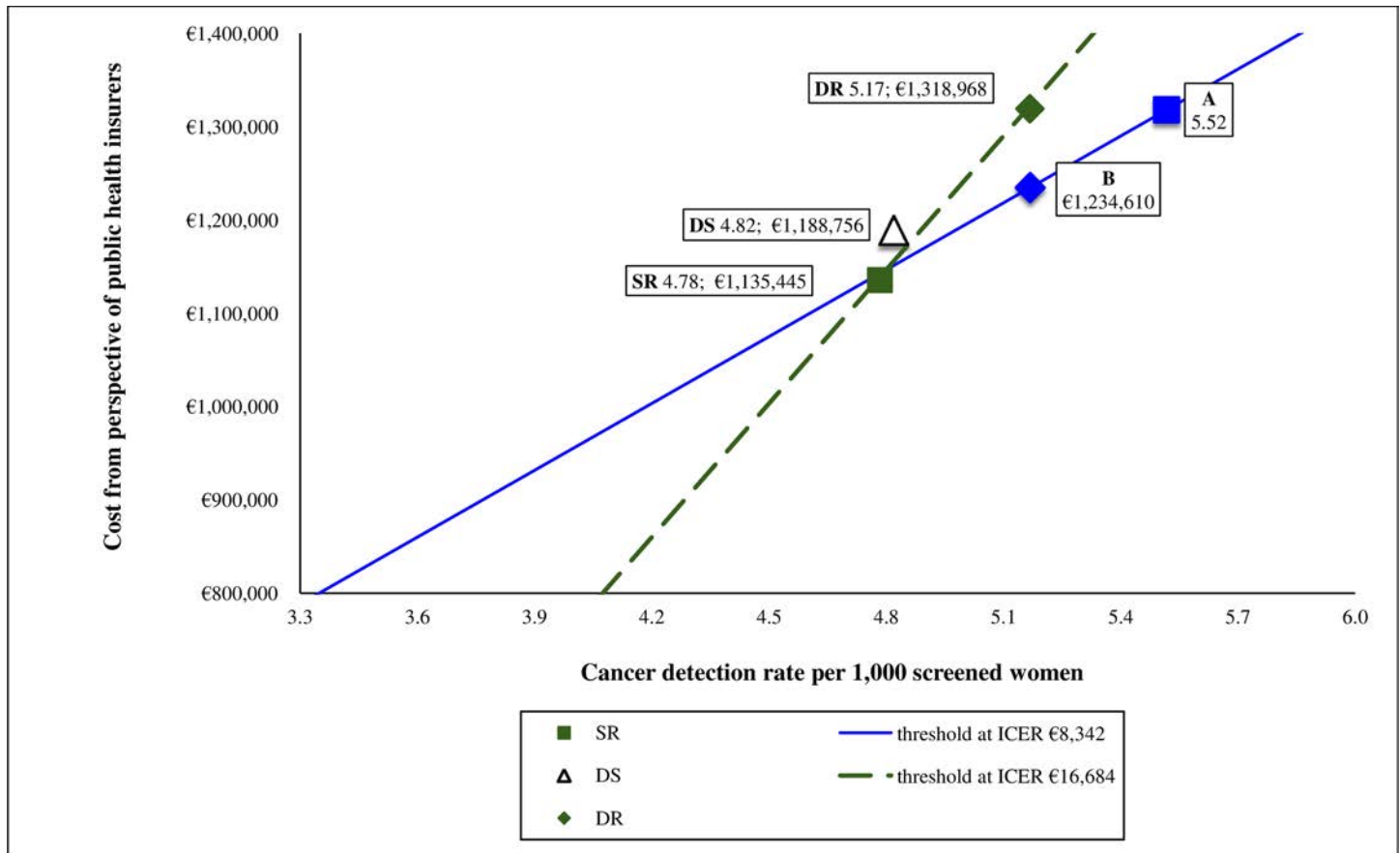


Fig 3. Cost-effectiveness plane illustrating differences in costs and cancer detection rates between the reading strategies. DR = Double reading. SR = Single reading. DS = Double reading in prevalent screening and single reading in incident screening. Continuous and dashed lines represent the thresholds if willingness to pay per one additionally detected cancer were €8,342 and €16,684, respectively. Point A represents the expected cancer detection rate at single reading if willingness to pay per single reading were equal to double reading. Point B represents the expected cost at single reading if the cancer detection rate at single reading were equal to that of double reading.

doi:10.1371/journal.pone.0159806.g003

of double reading increased only by 8%, as reported in a systematic review [6]. Thirdly, the ICER (€16,684) of double reading was approximately 100% more expensive than the average cost of one screen detected cancer at single reading. Therefore, double reading might not be as cost-effective as single reading.

Aside from the ICER, characteristics of additionally detected cancers after double reading might inform about the potential benefits of this reading strategy in breast cancer screening programmes. European guidelines recommend to limit the proportion of detected ductal carcinomas in-situ (DCIS) [1], due to its potential association with overdiagnosis. On the other hand, there is evidence that DCIS detection reduces interval cancer rates [31]. Based on a Spanish cohort, Blanch et al. [15] reported that double reading had a greater effect on detection of CIS. Similarly, we found that detection of CIS increased by 25% at double reading. Although currently there is not evidence that screening programmes should take measures to reduce DCIS detection, further studies will focus on the association between reading strategies and overdiagnosis.

In our study, both lower recall and early recall rates at single reading show important cost savings due to a reduction in additional diagnostic tests. This net benefit was not at the expense of decreasing its positive predictive value (PPV), which remained stable when comparing it

with double reading. Our results match with The US National Cancer Institute findings that showed a better interpretive accuracy and higher PPV at single reading [31].

The ICER at double reading remained rather stable in the sensitivity analysis. However, a small change (5%) in the sensitivity of double reading or single reading could carry important variations in the ICER. To be as cost-effective as single reading, detection rate at double reading should increase by 16%. On the contrary, a reduction in staff cost was not key determinant of ICER variations. Because the parameters we used to test uncertainties reflect the European guidelines' recommendations [1], we believe that our results might provide reasonable estimates to be extrapolated to other screening programmes with similar characteristics.

In contrast with our results, previous studies comparing double versus single reading reported that double reading was a cost-effective strategy [6,22]. This discrepancy can be explained by the fact that those studies were performed in the context of screen-film mammography. Digital mammography can be more sensitive than screen-film mammography, reducing the possibility of missed lesions. Furthermore, previous studies were performed decades ago in the firsts rounds of screening programmes when costs were probably lower than nowadays. Our results came from the fifth round of a well-established programme when the high experience of one single reader can be as good as the combination of two readers.

Although the scientific evidence is insufficient to determine the benefits of CAD on the readers interpretation [23,32,33]. Sato et al. [24] found that single reading + CAD was a cost-effective strategy increasing life years gained at a 'low' price. Our results cannot be comparable with those published by Sato et al. [24] because we did not evaluate CAD and we did not have

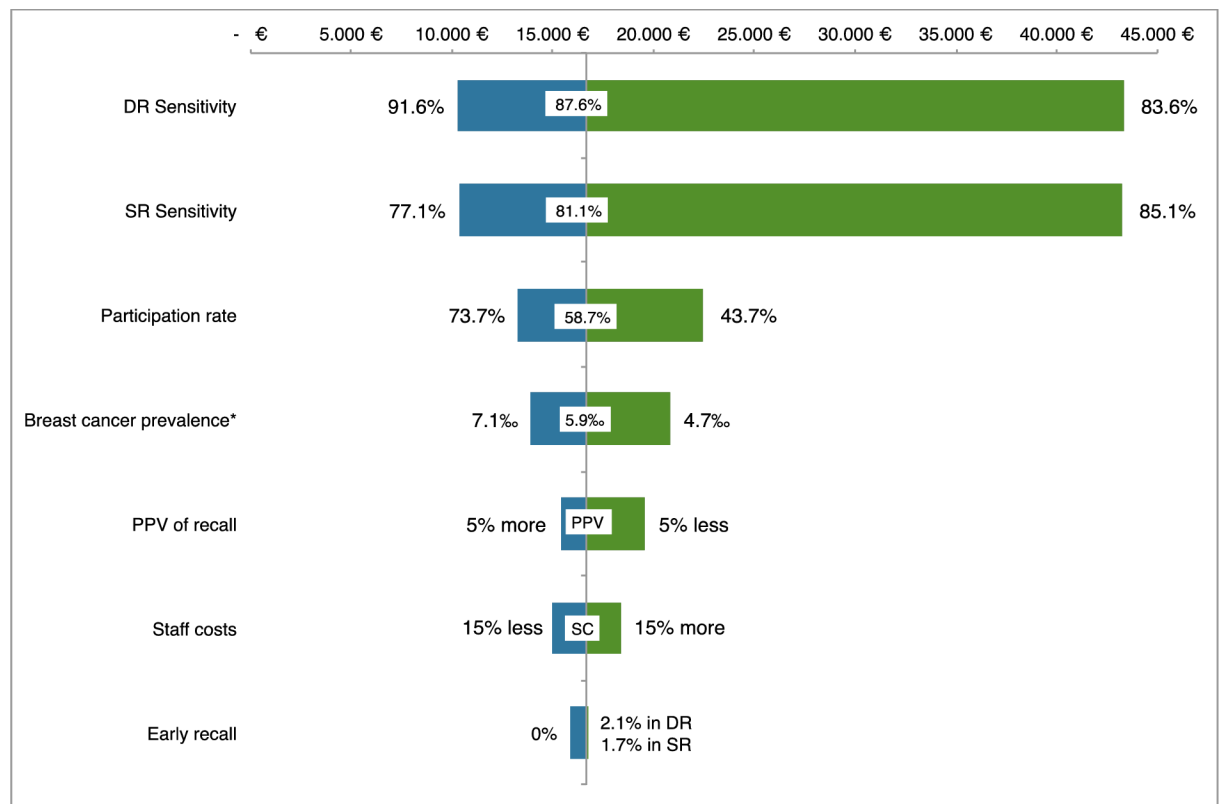


Fig 4. Sensitivity analysis of the incremental cost-effectiveness ratio (ICER) of double reading versus single reading. DR = Double reading. SR = Single reading. PPV = positive predictive value. SC = Staff costs. ER = early recall was 2.1% in double reading and 1.7% in single reading. *The prevalence of breast cancer was estimated as the number of true positives plus the number of false negatives.

doi:10.1371/journal.pone.0159806.g004

evidence of life-year gained with single reading. However, the results of both studies may stimulate rethinking of single reading alone or single reading + CAD as a feasible and efficient strategy in settings where shortage of radiologists exists.

The study has certain limitations. First, we did not assess the quality-adjusted life years (QALYs) for each strategy. Although there is no universal agreement, QALY is currently the most appropriate measure of health benefit, and interventions are usually considered cost-effective when the ICER placed on a QALY gained is up to €30,000 [28]. Nevertheless, our results might be useful for calculating expected costs per one additionally screen detected cancer and, therefore, can be taken into account when planning or evaluating breast cancer screening programmes. Second, it is difficult to compare our results with published cost-effectiveness evaluations [18] because of known differences in programmes between countries. However, our results may be similar to other European breast cancer screening programmes. Third, several economic evaluations have reported superiority of single reading with CAD over double reading [23,24]. We did not evaluate this new technology because it is not widely available in Spain. Fourth, although there is no statistical difference in detection rate between single and double reading, this study may not be adequately powered for the small expected difference. Finally, in this study we included information about five highly trained radiologists. The results can be different in other context with less trained radiologists or other professionals performing the reading of mammograms.

In conclusion, our results suggest that, in the context of digital mammography, double reading can be as effective as single reading but more expensive. Whereas double reading did not dramatically increase the cancer detection rate, the current question is how much the decision makers are willing to pay for the extra cancers detected. Further economic evaluation of randomised controlled trials may be crucial in determining whether the QALYs gained at double reading are comparable to those at single reading in the context of digital mammography.

Supporting Information

S1 Table. Definitions of performance measure and diagnostic accuracy outcomes.

(DOCX)

S2 Table. Unitary costs and cost estimation according to reading strategy. NA = not apply.

[†]Staff = medical, technician, and administrative. ^{*}Maintenance and repairs of the machine including depreciation. [‡]Administrative staff in early recall, as well as in consensus and arbitration. [§]In cancers cases suspected by one or both readers but with missing data, the average cost of additional tests was imputed.

(DOCX)

S3 Table. Value of parameters used in the sensitivity analysis.

(DOCX)

S4 Table. Characteristics of cancers detected at each reading strategy. [†]Detection rate per 1000 screened women. [‡]Linear test. [§]Interval cancers detected in the subsequent two years after a negative screening episode.

(DOCX)

S5 Table. Results of the sensitivity analysis. FN = false negatives.

(DOCX)

Acknowledgments

The authors thank Ms. Maria Victoria Leo Rosas, for her contributions to the language editing of the present article. The authors also thank members of the Screening Mammography

Program of Hospital Sant Pau, Barcelona, and participants in the programme. Margarita Posso is a Ph.D. candidate at the Methodology of Biomedical Research and Public Health program, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain.

Author Contributions

Conceived and designed the experiments: MP TP XB. Performed the experiments: MP MC MR TP XB. Analyzed the data: MP MC MR. Contributed reagents/materials/analysis tools: MP TP XB. Wrote the paper: MP MC MR TP XB.

References

1. Perry N, Puthaar E. European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis. European Communities; 2006.
2. Broeders M, Moss S, Nyström L, Njor S, Paap E, Massat N, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. 2012; 19 Suppl 1: 14–25. doi: [10.1258/jms.2012.012078](https://doi.org/10.1258/jms.2012.012078) PMID: [22972807](https://pubmed.ncbi.nlm.nih.gov/22972807/)
3. The Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. 2012; 380: 1778–1786. doi: [10.1016/S0140-6736\(12\)61611-0](https://doi.org/10.1016/S0140-6736(12)61611-0) PMID: [23117178](https://pubmed.ncbi.nlm.nih.gov/23117178/)
4. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med*. 2005; 353: 1773–1783. doi: [10.1056/NEJMoa052911](https://doi.org/10.1056/NEJMoa052911) PMID: [16169887](https://pubmed.ncbi.nlm.nih.gov/16169887/)
5. Comas M, Arrospe A, Mar J, Sala M, Vilapriño E, Hernández C, et al. Budget impact analysis of switching to digital mammography in a population-based breast cancer screening program: a discrete event simulation model. *PLoS ONE*. 2014; 9: e97459. doi: [10.1371/journal.pone.0097459](https://doi.org/10.1371/journal.pone.0097459) PMID: [24832200](https://pubmed.ncbi.nlm.nih.gov/24832200/)
6. Dinnes J, Moss S, Melia J, Blanks R, Song F, Kleijnen J. Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: findings of a systematic review. *The Breast*. 2001; 10: 455–463. doi: [10.1054/brst.2001.0350](https://doi.org/10.1054/brst.2001.0350) PMID: [14965624](https://pubmed.ncbi.nlm.nih.gov/14965624/)
7. Blanks RG, Wallis MG, Moss SM. 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*. 1998; 5: 195–201. doi: [10.1136/jms.5.4.195](https://doi.org/10.1136/jms.5.4.195) PMID: [9934650](https://pubmed.ncbi.nlm.nih.gov/9934650/)
8. Harvey SC, Geller B, Oppenheimer RG, Pinet M, Riddell L, Garra B. Increase in cancer detection and recall rates with independent double interpretation of screening mammography. *American Journal of Roentgenology*. 2003; 180: 1461–1467. PMID: [12704069](https://pubmed.ncbi.nlm.nih.gov/12704069/)
9. Matson M, Hibbert J, Field S. Double reporting of screening mammograms. *Clin Radiol*. 1997; 52: 567. PMID: [9240718](https://pubmed.ncbi.nlm.nih.gov/9240718/)
10. Thurffjell EL, Lernevall KA, Taube AA. Benefit of independent double reading in a population-based mammography screening program. *Radiology*. 1994; 191: 241–244. doi: [10.1148/radiology.191.1.8134580](https://doi.org/10.1148/radiology.191.1.8134580) PMID: [8134580](https://pubmed.ncbi.nlm.nih.gov/8134580/)
11. Thurffjell E. Mammography screening. One versus two views and independent double reading. *Acta Radiol*. 1994; 35: 345–350. PMID: [8011383](https://pubmed.ncbi.nlm.nih.gov/8011383/)
12. Klompenhouwer EG, Voogd AC, Heeten den GJ, Strobbe LJA, de Haan AFJ, Wauters CA, et al. Blinded double reading yields a higher programme sensitivity than non-blinded double reading at digital screening mammography: a prospected population based study in the south of The Netherlands. *Eur J Cancer*. 2015; 51: 391–399. doi: [10.1016/j.ejca.2014.12.008](https://doi.org/10.1016/j.ejca.2014.12.008) PMID: [25573788](https://pubmed.ncbi.nlm.nih.gov/25573788/)
13. Taplin SH, Rutter CM, Elmore JG, Seger D, White D, Brenner RJ. Accuracy of screening mammography using single versus independent double interpretation. *American Journal of Roentgenology*. 2000; 174: 1257–1262. doi: [10.2214/ajr.174.5.1741257](https://doi.org/10.2214/ajr.174.5.1741257) PMID: [10789773](https://pubmed.ncbi.nlm.nih.gov/10789773/)
14. Hofvind S, Geller BM, Rosenberg RD, Skaane P. Screening-detected breast cancers: discordant independent double reading in a population-based screening program. *Radiology*. 2009; 253: 652–660. doi: [10.1148/radiol.2533090210](https://doi.org/10.1148/radiol.2533090210) PMID: [19789229](https://pubmed.ncbi.nlm.nih.gov/19789229/)
15. Blanch J, Sala M, Román M, Ederra M, Salas D, Zubizarreta R, et al. Cumulative risk of cancer detection in breast cancer screening by protocol strategy. *Breast Cancer Res Treat*. 2013; 138: 869–877. doi: [10.1007/s10549-013-2458-5](https://doi.org/10.1007/s10549-013-2458-5) PMID: [23471648](https://pubmed.ncbi.nlm.nih.gov/23471648/)

16. Roelofs AAJ, Karssemeijer N, Wedekind N, Beck C, van Woudenberg S, Snoeren PR, et al. Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology*. 2007; 242: 70–77. doi: [10.1148/radiol.2421050684](https://doi.org/10.1148/radiol.2421050684) PMID: [17185661](https://pubmed.ncbi.nlm.nih.gov/17185661/)
17. Varela C, Karssemeijer N, Hendriks JHCL, Holland R. Use of prior mammograms in the classification of benign and malignant masses. *Eur J Radiol*. 2005; 56: 248–255. doi: [10.1016/j.ejrad.2005.04.007](https://doi.org/10.1016/j.ejrad.2005.04.007) PMID: [15890483](https://pubmed.ncbi.nlm.nih.gov/15890483/)
18. Leivo T, Salminen T, Sintonen H, Tuominen R, Auerma K, Partanen K, et al. Incremental cost-effectiveness of double-reading mammograms. *Breast Cancer Res Treat*. 1999; 54: 261–267. PMID: [10445425](https://pubmed.ncbi.nlm.nih.gov/10445425/)
19. Vilapriño E, Forné C, Carles M, Sala M, Pla R, Castells X, et al. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. *PLoS ONE*. 2014; 9: e86858. doi: [10.1371/journal.pone.0086858](https://doi.org/10.1371/journal.pone.0086858) PMID: [24498285](https://pubmed.ncbi.nlm.nih.gov/24498285/)
20. Posso MC, Puig T, Quintana MJ, Solà-Roca J, Bonfill X. Double versus single reading of mammograms in a breast cancer screening programme: a cost-consequence analysis. *Eur Radiol*. 2016. doi: [10.1007/s00330-015-4175-4](https://doi.org/10.1007/s00330-015-4175-4)
21. Ciatto S, Ambrogetti D, Bonardi R, Catarzi S, Risso G, Rosselli Del Turco M, et al. Second reading of screening mammograms increases cancer detection and recall rates. Results in the Florence screening programme. *J Med Screen*. 2005; 12: 103–106. doi: [10.1258/0969141053908285](https://doi.org/10.1258/0969141053908285) PMID: [15949122](https://pubmed.ncbi.nlm.nih.gov/15949122/)
22. Groenewoud JH, Otten JDM, Fracheboud J, Draisma G, van Ineveld BM, Holland R, et al. Cost-effectiveness of different reading and referral strategies in mammography screening in the Netherlands. *Breast Cancer Res Treat*. 2007; 102: 211–218. doi: [10.1007/s10549-006-9319-4](https://doi.org/10.1007/s10549-006-9319-4) PMID: [17004116](https://pubmed.ncbi.nlm.nih.gov/17004116/)
23. Azavedo E, Zackrisson S, Mejåre I, Heibert Arnlin M. Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. *BMC Med Imaging*. 2012; 12: 22. doi: [10.1186/1471-2342-12-22](https://doi.org/10.1186/1471-2342-12-22) PMID: [22827803](https://pubmed.ncbi.nlm.nih.gov/22827803/)
24. Sato M, Kawai M, Nishino Y, Shibuya D, Ohuchi N, Ishibashi T. Cost-effectiveness analysis for breast cancer screening: double reading versus single + CAD reading. *Breast Cancer*. Springer Japan; 2014; 21: 532–541. doi: [10.1007/s12282-012-0423-5](https://doi.org/10.1007/s12282-012-0423-5)
25. Guerriero C, Gillan MGC, Cairns J, Wallis MG, Gilbert FJ. Is computer aided detection (CAD) cost effective in screening mammography? A model based on the CADET II study. *BMC Health Serv Res*. 2011; 11: 11. doi: [10.1186/1472-6963-11-11](https://doi.org/10.1186/1472-6963-11-11) PMID: [21241473](https://pubmed.ncbi.nlm.nih.gov/21241473/)
26. Gilbert FJ, Astley SM, Gillan MGC, Agbaje OF, Wallis MG, James J, et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med*. 2008; 359: 1675–1684. doi: [10.1056/NEJMoa0803545](https://doi.org/10.1056/NEJMoa0803545) PMID: [18832239](https://pubmed.ncbi.nlm.nih.gov/18832239/)
27. Taylor P, Potts HWW. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. *Eur J Cancer*. 2008; 44: 798–807. doi: [10.1016/j.ejca.2008.02.016](https://doi.org/10.1016/j.ejca.2008.02.016) PMID: [18353630](https://pubmed.ncbi.nlm.nih.gov/18353630/)
28. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013 [Internet]. London: NICE; 2014. Available: <http://publications.nice.org.uk/pmg9>
29. O'Mahony JF, Paulden M. NICE's selective application of differential discounting: ambiguous, inconsistent, and unjustified. *Value Health*. 2014; 17: 493–496. doi: [10.1016/j.jval.2013.02.014](https://doi.org/10.1016/j.jval.2013.02.014) PMID: [25128041](https://pubmed.ncbi.nlm.nih.gov/25128041/)
30. Paulden M, O'Mahony JF, Culyer AJ, McCabe C. Some Inconsistencies in NICE's Consideration of Social Values. *Pharmacoeconomics*. 2014; 32: 1043–1053. doi: [10.1007/s40273-014-0204-4](https://doi.org/10.1007/s40273-014-0204-4) PMID: [25145802](https://pubmed.ncbi.nlm.nih.gov/25145802/)
31. Taplin S, Abraham L, Barlow WE, Fenton JJ, Berns EA, Carney PA, et al. Mammography facility characteristics associated with interpretive accuracy of screening mammography. *J Natl Cancer Inst*. 2008; 100: 876–887. doi: [10.1093/jnci/djn172](https://doi.org/10.1093/jnci/djn172) PMID: [18544742](https://pubmed.ncbi.nlm.nih.gov/18544742/)
32. Eadie LH, Taylor P, Gibson AP. A systematic review of computer-assisted diagnosis in diagnostic cancer imaging. *Eur J Radiol*. 2012; 81: e70–6. doi: [10.1016/j.ejrad.2011.01.098](https://doi.org/10.1016/j.ejrad.2011.01.098) PMID: [21345631](https://pubmed.ncbi.nlm.nih.gov/21345631/)
33. Bargalló X, Santamaría G, Del Amo M, Arguis P, Ríos J, Grau J, et al. Single reading with computer-aided detection performed by selected radiologists in a breast cancer screening program. *Eur J Radiol*. Elsevier; 2014;0. doi: [10.1016/j.ejrad.2014.08.010](https://doi.org/10.1016/j.ejrad.2014.08.010)

5.3. ARTÍCULO 3. REVISIÓN SISTEMÁTICA DE LA EFECTIVIDAD Y COSTE-EFECTIVIDAD DE LA DOBLE LECTURA MAMOGRÁFICA

5.3.1. RESUMEN DE RESULTADOS

CARACTERÍSTICAS DE LOS ESTUDIOS INCLUIDOS

A partir de 1.473 artículos que fueron identificados en la búsqueda, cuatro estudios se incluyeron en el análisis ¹⁴⁶⁻¹⁴⁹. Dos fueron estudios de coste-efectividad ^{147,148} y dos fueron observacionales ^{146,149}. Los cuatro fueron estudios diagnósticos con un bajo riesgo de sesgo y publicados durante o después del 2014. Tres de ellos fueron realizados en Europa y uno en Japón. Los cuatro estudios se basaron en programas bienales de cribado con mamografía digital en mujeres de 48 a 69 años. El número total de mamografías analizadas fue de 77.507, con tamaños de muestra que variaron entre 7.292 ¹⁴⁶ y 28.636 ¹⁴⁸.

ESTRATEGIAS DE LECTURA EVALUADAS

Los estudios evaluaron mamografías digitales valoradas por radiólogos o por médicos especializados en la lectura mamográfica. La doble lectura fue independiente en los cuatro estudios. En caso de desacuerdo entre los lectores, éste se resolvió por consenso ^{147,149}, arbitraje ¹⁴⁶ o una combinación de ellos ¹⁴⁸.

Los costes y efectos de la lectura simple fueron simuladas a partir de la primera lectura en los estudios europeos ^{146,148,149}. El estudio japonés ¹⁴⁷ simuló la lectura simple acompañada de un software de ayuda al diagnóstico (CAD). En el caso de los diseños utilizados en los estudios de coste-efectividad ambos utilizaron un árbol de decisión ^{147,148} y uno de ellos además del árbol incorporó un modelo de Markov ¹⁴⁷.

CALIDAD DE LOS ESTUDIOS INCLUIDOS

En general los estudios incluidos mostraron un bajo o moderado riesgo de sesgo evaluado con los instrumentos QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) ¹⁴¹ y CHEERS (Consolidated Health Economic Evaluation Reporting Standards) ¹⁴².

Se identificó una consideración respecto a la aplicabilidad de los resultados. Todos los lectores fueron altamente capacitados con más de 1.500 lecturas al año lo que puede limitar la aplicación de los resultados a otros contextos de cribado con lectores menos entrenados.

En relación a la calidad de las evaluaciones económicas, el riesgo de sesgo fue bajo en un estudio ¹⁴⁸ y moderado en otro ¹⁴⁷. Ambos estudios incluyeron sólo los costes directos de atención de la

salud en contextos específicos pero uno de ellos reportó una perspectiva social sin incorporar costes sociales ¹⁴⁷.

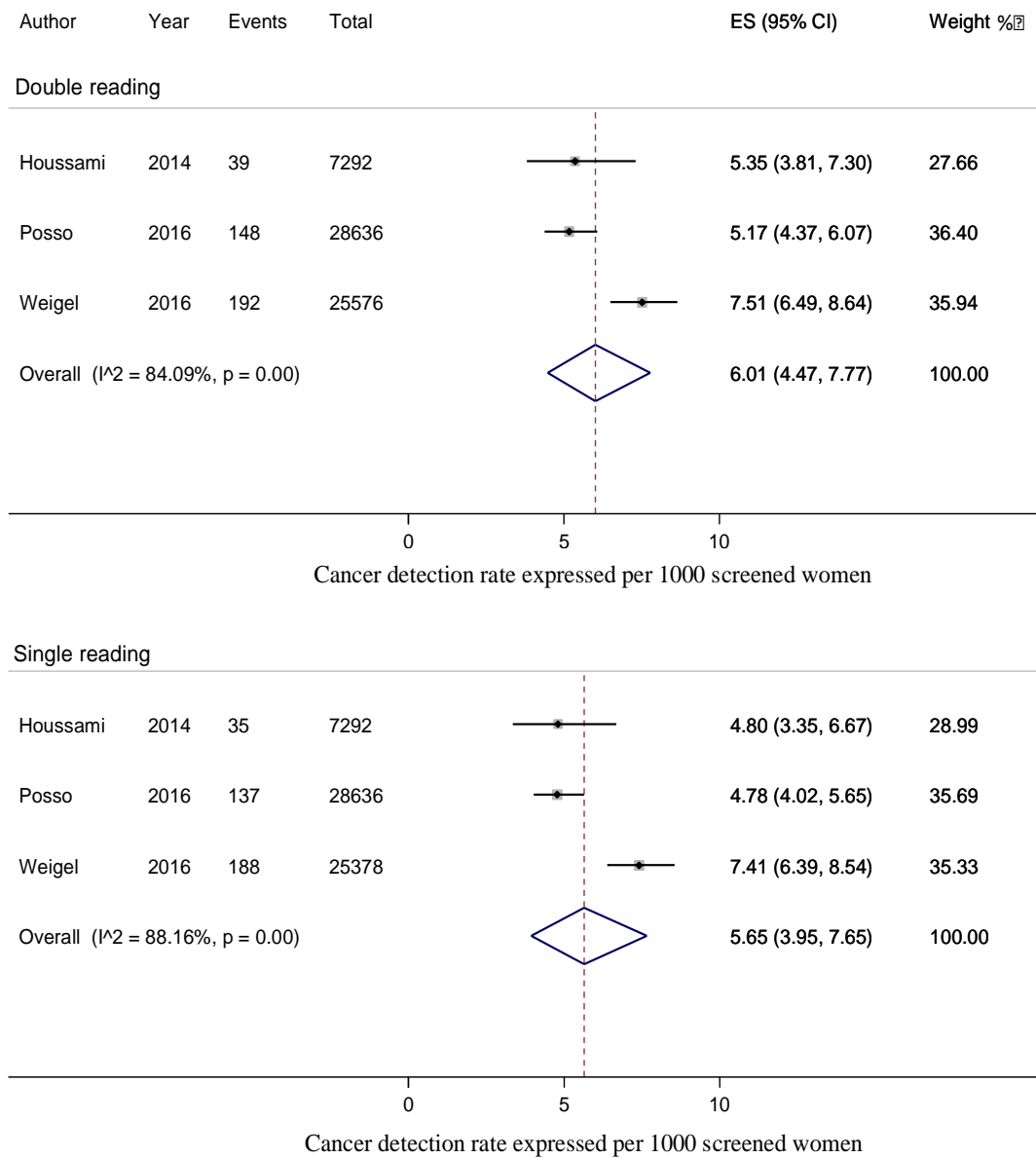
RESULTADOS DE EFECTIVIDAD

La tasa de detección de cáncer estimada a partir del análisis combinado de tres estudios fue de 6,2 por 1000 mujeres con la doble lectura y de 5,9 por 1000 mujeres con la lectura simple ($P = 0,76$) (Figura 7). La proporción de falsos positivos fue de 4,3% con la doble lectura y 4,1% con la lectura simple ($P = 0,12$).

RESULTADOS DE COSTE-EFECTIVIDAD

El estudio japonés reportó que la RCEI de la lectura simple más un sistema de ayuda al diagnóstico (CAD, Computer-Aided Detection) fue de 2.951 US \$ PPP por año de vida ganado lo que se consideró coste-efectivo para el umbral de disposición a pagar de 56.963 US \$ PPP. El otro estudio incluido reportó una RCEI de 24.717 US \$ PPP por cáncer detectado con la doble lectura comparado con la lectura simple.

Figura 3. Análisis combinado de la tasa de detección de cáncer con la doble lectura y con la lectura simple.



PUBLICACIÓN 3

POSSO M, PUIG T, CARLES M, RUÉ M, CANELO-AYBAR C, BONFILL X.
EFFECTIVENESS AND COST-EFFECTIVENESS OF DOUBLE READING IN DIGITAL
MAMMOGRAPHY SCREENING: A SYSTEMATIC REVIEW AND META-ANALYSIS.
EURJRAD. 2017; 96: 40-9.

European Journal of Radiology. 2017.

Factor de impacto 2,462. Q2 (45/126). Radiology, nuclear medicine & medical imaging.



Research article

Effectiveness and cost-effectiveness of double reading in digital mammography screening: A systematic review and meta-analysis



Margarita Posso^{a,e,*}, Teresa Puig^{a,b}, Misericòrdia Carles^c, Montserrat Rué^d,
Carlos Canelo-Aybar^{e,f}, Xavier Bonfill^{a,b,e,g}

^a Department of Clinical Epidemiology and Public Health, Hospital de la Santa Creu i Sant Pau (IIB Sant Pau), Barcelona, Spain

^b Universitat Autònoma de Barcelona (UAB), Barcelona, Spain

^c Economics Department and CREIP, Universitat Rovira i Virgili, Reus, Spain

^d Basic Medical Sciences Department, Biomedical Research Institut of Lleida (IRBLLEIDA), Universitat de Lleida, Lleida, Spain

^e Iberoamerican Cochrane Centre, Barcelona, Spain

^f School of Medicine, Peruvian University of Applied Sciences, Lima, Peru

^g CIBER of Epidemiology and Public Health (CIBERESP), Spain

ARTICLE INFO

Keywords:

Mammography

Mass screening

Breast neoplasms

Cost and cost analysis

Systematic review

ABSTRACT

Purpose: Double reading is the strategy of choice for mammogram interpretation in screening programmes. It remains, however, unknown whether double reading is still the strategy of choice in the context of digital mammography. Our aim was to determine the effectiveness and cost-effectiveness of double reading versus single reading of digital mammograms in screening programmes.

Methods: We performed a systematic review by searching the PubMed, Embase, and Cochrane Library databases up to April 2017. We used the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool and CHEERS (Consolidated Health Economic Evaluation Reporting Standards) checklist to assess the methodological quality of the diagnostic studies and economic evaluations, respectively. A proportion's meta-analysis approach, 95% Confidence Intervals (95% CI) and test of heterogeneity (P values) were used for pooled results. Costs are expressed US\$ PPP (United States Dollar purchasing power parities). The PROSPERO ID of this Systematic Review's protocol is CRD42014013804.

Results: Of 1473 potentially relevant hits, four high-quality studies were included. The pooled cancer detection rate of double reading was 6.01 per 1000 screens (CI: 4.47%–7.77%), and it was 5.65 per 1000 screens (CI: 3.95%–7.65%) for single reading (P = 0.76). The pooled proportion of false-positives of double reading was 47.03 per 1000 screens (CI: 39.13%–55.62%) and it was 40.60 per 1000 screens (CI: 38.58%–42.67%) for single reading (P = 0.12). One study reported, for double reading, an ICER (Incremental Cost-Effectiveness Ratio) of 16,684 Euros (24,717 US\$ PPP; 2015 value) per detected cancer. Single reading + CAD (computer-aided-detection) was cost-effective in Japan.

Conclusion: The evidence of benefit for double reading compared to single reading for digital mammography interpretation is scarce. Double reading seems to increase operational costs, have a not significantly higher false-positive rate, and a similar cancer detection rate.

1. Introduction

Mammography screening programmes have been widely implemented in Europe [1]. In these programmes, trained physicians, radiologists or radiographers can read mammograms following

different strategies; i.e. single reading with or without CAD (computer-aided-detection), blinded or unblinded double reading, and followed or not by consensus and arbitration [2–4]. When double reading is performed, each reader determines whether the result of the mammogram is normal -in which case the woman is invited for a further screening

Abbreviations: CAD, computer-aided-detection; DCIS, ductal carcinoma in situ; FNAC, fine-needle aspiration cytology; ICER, incremental cost-effectiveness ratio; NHS-EED, Economic Evaluation Database produced by the NIHR Centre for Reviews and Dissemination (CRD) at the University of York; NPV, net present value; US\$ PPP, United States Dollar purchasing power parities

* Corresponding author. Present address: Hospital de la Santa Creu i Sant Pau (IIB Sant Pau), C/Sant Antoni Maria Claret, 167, Pavelló 18, planta 0, CP: 08025, Barcelona, Spain.

E-mail addresses: mposso@cochrane.es (M. Posso), tpuig@santpau.cat (T. Puig), misericordia.carles@urv.cat (M. Carles), montse.rue@cmb.udl.cat (M. Rué), canelo@santpau.cat (C. Canelo-Aybar), xbonfill@santpau.cat (X. Bonfill).

<http://dx.doi.org/10.1016/j.ejrad.2017.09.013>

Received 24 April 2017; Received in revised form 14 September 2017; Accepted 19 September 2017

0720-048X/© 2017 Elsevier B.V. All rights reserved.

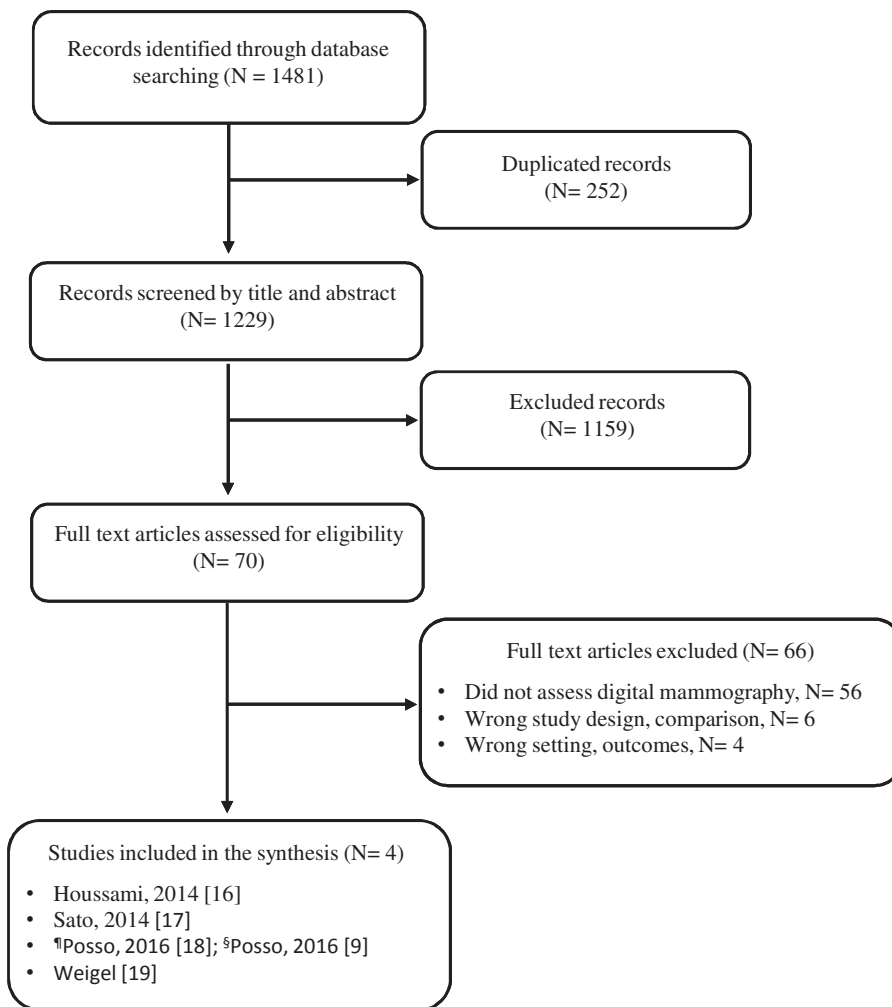


Fig. 1. PRISMA diagram of the study selection.

¶Main publication for the study. §Secondary publication not included in the meta-analyses.

round-, or abnormal -in which case additional tests are requested to confirm or rule out malignancy-. In case of disagreement between readers, the result of the mammogram can be determined by consensus and/or by arbitration [4].

The European screening guidelines published in 2006 [4], recommend double reading as the strategy of choice for mammogram interpretation. This recommendation, however, is based on studies performed in the context of screen-film mammography. These studies found that double reading was more effective than single reading increasing the sensitivity in 5–15% [4]. Regarding the cost-effectiveness, one systematic review published in 2001 suggested double reading as a cost-effective strategy [5]. More recently, in the same context of screen-film mammography, one study reported double reading with or without consensus/arbitration as a cost-effective strategy in the Netherlands [6].

Digital mammography is a relatively new technology that can be more sensitive for identifying small abnormalities [7,8]. The implementation of this technology, therefore, is rising very fast in European screening programmes [1]. Although there is little evidence on reading strategies for digital mammography, we have previously found that double reading can have more false-positives and similar detection rate than single reading [9]. Hence, we think that digital mammography may affect the accuracy for mammogram interpretation increasing the sensitivity of single reading.

As far as we know, there are no systematic reviews assessing double reading in the context of digital mammography. In order to support or eventually challenge the current recommendation of double reading in European screening programmes [4], a systematic review of the

effectiveness and cost-effectiveness of double reading of digital mammography has become relevant and is the aim of this systematic review.

2. Material and methods

2.1. Study design

We adhered to the guidance provided by the Cochrane Collaboration [10] and the PRISMA statement [11] for conducting and reporting this systematic review.

2.2. Search strategy and study selection

Databases were searched from 1st January 1990 to 20th February 2017, including Medline, EMBASE, and the Cochrane Library to obtain all relevant literature in the context of digital mammography. We searched for systematic reviews, experimental trials, and observational studies with a control group that compared double reading with single reading in the context of digital mammography. We also searched for cost-effectiveness, cost-utility and cost-benefit studies. We combined a series of terms for ‘Breast Neoplasms’, ‘Mass Screening’ and ‘double reading’. The search strategy is provided in Appendix A. We included studies we deemed as relevant based on our previous experience, and hand searched the bibliography of the included studies. Two reviewers independently screened the articles identified in the searches. Disagreements were resolved by consensus.

Table 1
Characteristics of included studies.

Reference	Country	Study design	Population	Readers	Index test	Comparator test	Reference standard
Houssami [16]	Italy	Analysis based on the prospective population-based STORM trial [32].	7292 women. Age: median, 58 years (≥ 48 years). Biennial screening. 16,000 examinees annually.	Radiologists with a median of 1791 screen-reads (from 1315 to 2370).	Single reading algorithm was modelled using the first read performed for each screening examination.	Double reading. Positive and discordant reads were recalled for assessment.	Suspicious mammogram: Excision histology or complete outcome assessment with and without needle biopsy. Negative mammogram: A minimum follow-up of 13 months to identify first year interval cancers.
Sato [17]	Japan	Cost-effectiveness analysis with a decision tree and a Markov model	Women aged 50–69 years.	Certified physicians	One physician using CAD (single + CAD).	Independent double-reading with consensus to solve disagreements	Suspicious examinee: Detailed examination including fine-needle aspiration cytology and core needle biopsy. Negative examinee: Two years period between screening examinations to identify interval cancers
Posso [18]	Spain	Cost-effectiveness analysis with a decision tree model	Biennial screening. 28,636 mammograms. Women aged 50–69 years.	Radiologists with > 5000 readings per year.	Single reading was modelled using the first read performed for each screening examination	Independent double-reading with consensus and/or arbitration to solve disagreements	Suspicious mammogram: Fine-needle aspiration, core-needle biopsy, and surgical biopsy. Negative mammogram: Two years period between screening examinations to identify interval cancers
Weigel [19]	Germany	Analysis based on the screening unit of the National Mammography Screening Programme.	Biennial screening. 25,579 mammograms. Women aged 50–69 years.	Radiologists with > 5000 readings per year and each had more than 5 years of experience in breast radiology.	Single reading algorithm was modelled using the first read performed for each screening examination	Independent double-reading with consensus to solve disagreements	Suspicious mammogram: Fine-needle aspiration, core-needle biopsy, and surgical biopsy. Negative mammogram: Two years period between screening examinations to identify interval cancers

STORM trial = 'screening with tomosynthesis or standard mammography' trial.

Table 2
Risk of Bias and Applicability Judgments of included studies according to QUADAS-2 [17] tool and CHEERS checklist [18].

Reference	QUADAS-2 Risk of bias				QUADAS-2 Applicability concerns				CHEERS checklist
	Patient selection	Index tests	Reference standard	Flow and timing	Patient selection	Index tests	Reference standard	Comments about risk of bias according to the information reported in the economic evaluations	
Houssami [16] Sato [17]	☺☺	☺☺	☺☺	☺☺	☺☺	☺☺	☺☺	Not applicable. Moderate risk of bias. The perspective is unclear. Design: Cost-effectiveness analysis. The analysis was performed from a social perspective but included only direct medical costs. Measurement of effectiveness was provided from one single city in Japan. Uncertainty in results was assessed by a probabilistic sensitivity analysis.	
Posso [18]	☺☺	☺☺	☺☺	☺☺	☺☺	☺☺	☺☺	Low risk of bias. Design: Cost-effectiveness analysis performed from the health system perspective. Measurement of effectiveness and estimating costs were provided from one single institution. Uncertainty in results was assessed by a deterministic sensitivity analysis. Variations between incident and subsequent screen-examinations were explored. Not applicable.	
Weigel [19]	☺☺	☺☺	☺☺	☺☺	☺☺	☺☺	☺☺	Not applicable.	

☺☺ = low risk; ☹☹ = high risk.

2.3. Inclusion criteria

Studies that compared single reading with double reading for digital mammogram interpretation were eligible for inclusion. Included studies must have presented data for at least one of the following outcomes: sensitivity, specificity, cancer detection rate, false-positive rate, interval cancers, mortality, overdiagnosis, quality of life, incremental cost-effectiveness ratios (ICERs), and net present values (NPVs). Restrictions of professionals performing screen-readings (radiologists, radiographers or other professionals), views (e.g. one or more, caudal or oblique), time between first and second reading, or whether double reading was followed or not by consensus and/or arbitration were not considered. Single-arm studies, studies that did not include digital mammograms, studies conducted out of the context of population-based screening programmes or studies that did not contain confirmation of neoplasms by pathology were excluded. Both ductal carcinoma in situ (DCIS) and invasive carcinomas were considered. No language restriction was applied.

2.4. Quality assessment and data extraction

Key study characteristics and outcomes were extracted and studies were quality assessed by one reviewer and checked for comparing the results by a second. Study quality was determined using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool [12] for diagnostic studies and CHEERS (Consolidated Health Economic Evaluation Reporting Standards) checklist [13] for economic evaluations.

2.5. Analysis and synthesis of the results

The study findings were interpreted, and similarities and differences between the studies described and tabulated. These considerations included study design, methodological quality, whether the study was appropriately powered, and characteristics of reading strategies, participants and outcome measures. Costs are converted and expressed in 2015 US\$ PPP (United States Dollar Purchasing Power Parities) [14]. We pooled the effectiveness results from individual studies by a proportion's meta-analysis approach when the studies were methodologically and clinically similar. Briefly, we implemented the *metaprop* command in Stata, which allows estimation of 95% Confidence Intervals (95% CI) using the exact binomial method and incorporates the Freeman-Tukey double arcsine transformation for computation of pooled proportions. Test of heterogeneity P values were informed [15]. The meta-analysis was conducted under the random-effect approach. The random-effects meta-analysis model involves an assumption that the effects being estimated in the different studies are not identical, but follow a distribution [10]. We performed the meta-analysis in order to find the range of effects that are obtained with double and single reading. In order to evaluate the effect of one study that used tomosynthesis, in addition to digital mammography, we performed a sensitivity analysis.

3. Results

A flow chart of the study selection is shown in Fig. 1 and the list of excluded studies is provided in Appendix B. No papers were identified that required translation. There were no controlled trials that compared the effect of double reading with single reading on digital mammography screening. The four studies that met our inclusion criteria were observational studies performed in the screening context [16–19], three of them were based in European data [16,18,19] and one was from Japan [17], with all four studies published during or after 2014. The total number of screening mammograms was 77,507, with sample sizes from 7292 [16] to 28,636 [18] for individual studies. The included studies involved women aged 48–69 years. All studies assessed women who underwent biennial mammography screening. Summary details of

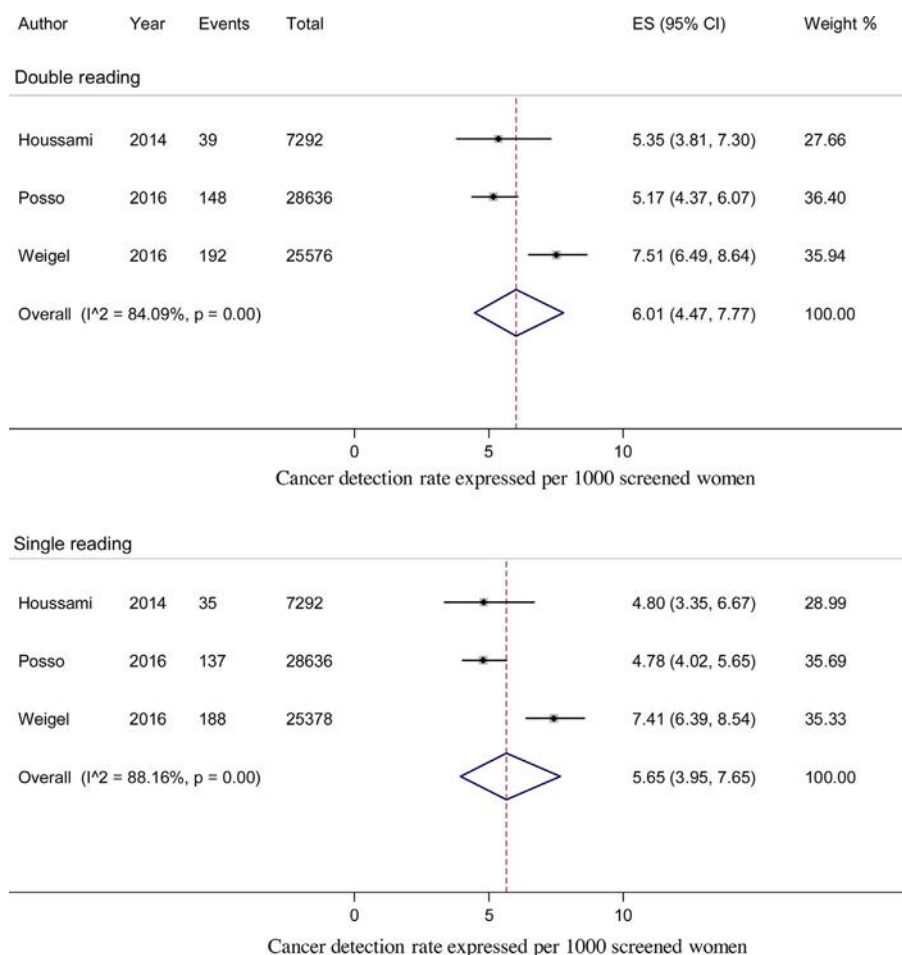


Fig. 2. Cancer detection rate pooled meta-analysis of double reading and single reading of digital mammography.

the included studies are provided in Table 1.

3.1. Reading strategies

Two studies used modelling techniques to assess the effectiveness of the reading strategies [16,19], and the other two studies modelled cost-effectiveness [17,18]. All studies used data from population based screening programmes and assessed readings performed by radiologists or by certified physicians. In all studies if the readers state that the findings of the mammogram were abnormal, a combination of additional tests was assigned to the woman in order to confirm or ruled out malignancy. The combination included additional mammographic projections, ultrasound, fine-needle aspiration cytology (FNAC), core biopsy, and open surgical biopsy.

Double reading was independently performed in the four studies; however, the studies did not inform whether the second reader was aware or not to the first reader’s interpretation. Disagreements between radiologists were solved by consensus [17,19], arbitration [16] or a combination of them [18]. Single reading was modelled in the studies. Whereas the European studies modelled single reading based on the first read performed for each screening examination [16,18,19], the Japanese study [17] modelled single reading using a decision tree and a Markov model.

3.2. Quality of included studies

Table 2 shows the risk of bias and applicability judgments ratings according to the QUADAS-2 tool [12], and the CHEERS checklist [13]. In relation to the risk of bias, the methodological quality is generally high in all included studies. The studies did not show limitations in the

patient selection. Modelling was based in women consecutively enrolled, a case-control design was avoided, and exclusions were justified. Double reading was the index test and single reading was comparator in three studies [16,18,19]. In one study the index test was single reading + CAD and the comparator was double reading [17]. In all studies single reading interpretations were reported without knowledge of the results of the double reading. The reference standard for positive cases was pathologic result of breast cancer and for negative cases was at least one-year follow-up.

In relation to the applicability judgments, we identified low concerns about patient selection, and high concerns about index test and reference standard. All studies included highly trained radiologists with more than 1500 readings per year or certified physicians; therefore, the results may not be applicable to other breast cancer screening contexts with less trained radiologists.

In relation to the quality assessment of the economic evaluations, we observed low and moderate risk of bias in the two included studies. The studies included only direct health-care costs provided from specific contexts (Table 2). The studies included in the pooled analysis were methodologically and clinically similar. These studies performed in European well-established breast screening programmes, with highly trained radiologists who performed independent double reading. Thus, the pooled effect represents the range of effects that can be obtained with double or single reading in a similar context.

3.3. Effectiveness of double reading

3.3.1. Cancer detection rate

Three studies reported the number of cancers detected after double reading compared to those detected after single reading [16,18,19].

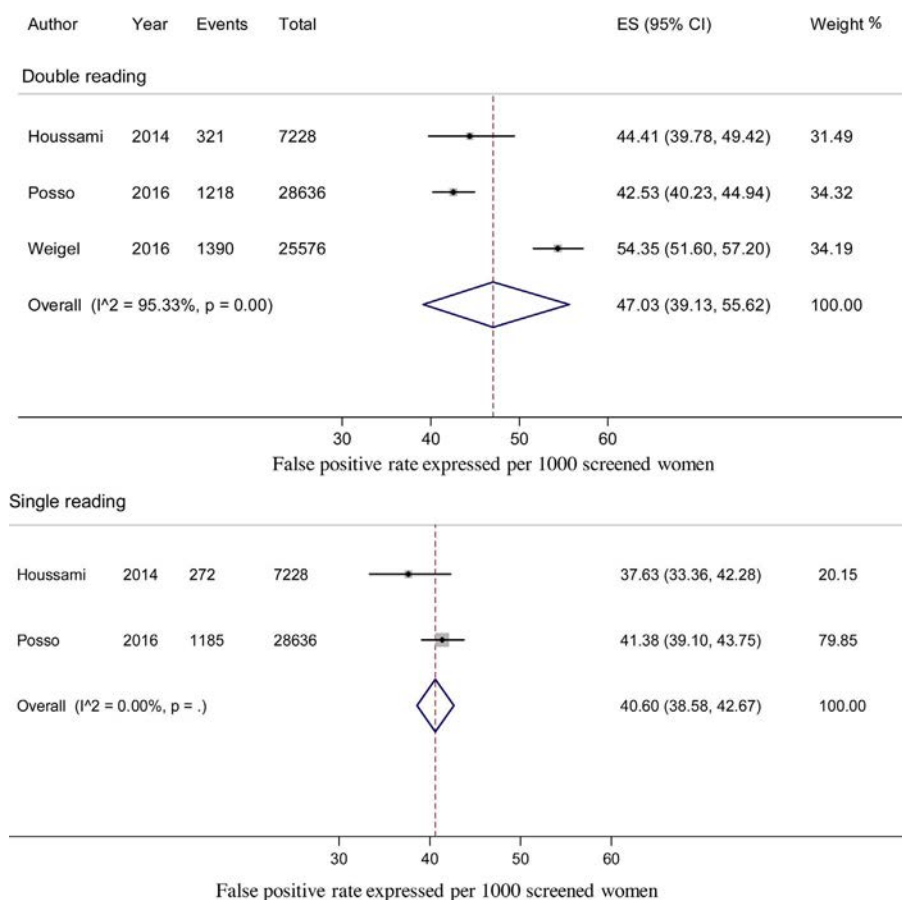


Fig. 3. False positive rate pooled meta-analysis of double reading and single reading of digital mammography.

The pooled cancer detection rate of double reading was 6.01 per 1000 screens (CI: 4.47–7.77 per 1000), that was not statistically different from 5.65 per 1000 screens (CI: 3.95–7.65 per 1000) observed in single reading ($P = 0.76$) (Fig. 2).

3.3.2. False-positive results

Two studies reported the proportion of false-positives results observed after double reading compared to those observed after single reading [16,18]. The pooled proportion of false-positives results of double reading was 47.03 per 1000 screens (CI: 39.13–55.62 per 1000) and it was 40.60 per 1000 (CI: 38.58–42.67 per 1000) for single reading ($P = 0.12$) (Fig. 3).

3.3.3. Interval cancers

Two studies reported data related with interval cancers [16,18]. One of them summed the false-negatives plus the interval cancers observed during the first year after screening. The sum for double reading was 3.6 per 1000 screens (CI: 2.2–4.9 per 1000) and it was 4.1 per 1000 screens (CI: 2.6–5.6 per 1000) for single reading ($P = 0.59$) [16]. The other study reported interval cancers observed during the two-year follow-up after screening. The interval cancer rate for double reading was 0.6 per 1000 (CI: 0.3–0.8 per 1000) screens and it was 0.9 per 1000 screens (CI: 0.6–1.3 per 1000) for single reading ($P = 0.09$) [18].

3.3.4. Sensitivity

Three studies reported the sensitivity of double reading compared to single reading [16,18,19]. The pooled sensitivity of double reading was 85.7% (CI: 77.7%–92.2%) and it was 80.4% (CI: 76.5%–84.0%) for single reading ($P = 0.60$) (Fig. 4). The study of Houssami et al. [16] reported the number of false negative results with tomosynthesis. Overall, 24 false negative results were reported, 6 interval cancers and 18 cancers identified only with tomosynthesis. Considering these false

negatives in a sensitivity analysis, the sensitivity of double reading in this study was 61.9% and the pooled result was 79.1%. Similarly, the sensitivity of single reading in this study was 55.6% and the pooled result was 74.1%.

3.3.5. Specificity

Three studies reported the specificity of double reading [16,18,19] and two of them [16,18] the specificity of single reading. The pooled specificity of double reading was 95.3% (CI: 94.4%–96.1%) and it was 95.9% (CI: 95.7%–96.1%) for single reading ($P = 0.13$) (Fig. 5). Considering the effect of including the false negatives identified with tomosynthesis in the study of Houssami et al. [16], the pooled specificity remains equal for both double and single reading.

3.4. Cost-Effectiveness of double reading compared to single reading

Sato et al. [17] performed a cost-effectiveness analysis from the societal perspective but only direct medical costs related to breast cancer screening and treatment were considered. Single reading + CAD extended in 0.0087 the number of life years gained compared with double reading. The ICER of single reading + CAD was 2951 US\$ PPP per life year gained that was considered cost-effective at a willingness to pay of 56,963 US\$ PPP. We published a cost-effectiveness analysis from the healthcare system perspective and assessed the programme’s direct costs for mammography interpretation and direct medical costs for diagnosis confirmation [18]. The ICER for double reading was 16,684 Euros (24,717 US\$ PPP) per cancer detected (Table 3).

4. Discussion

This systematic review included four studies comparing single reading with double reading for digital mammogram interpretation in

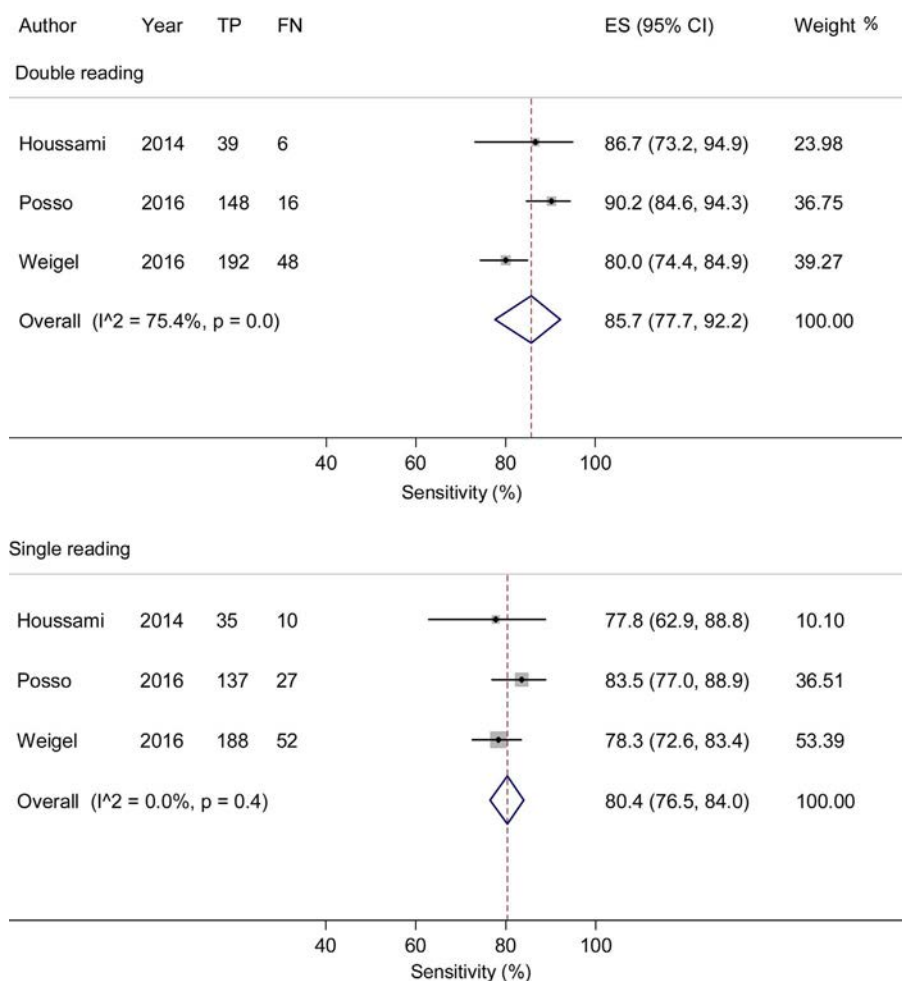


Fig. 4. Sensitivity (%) pooled meta-analysis of double reading and single reading of digital mammography.

breast cancer screening programmes. Three European studies [16,18,19] reported that double reading increases cancer detection rate and also the frequency of false positives. In one cost-effectiveness analyses, double reading showed an ICER of 16,684 Euros (24,717 US\$ PPP) per cancer detected [18] whereas single reading with CAD increased life years gained and may be cost-effective in Japan [17]. Firm conclusions could not be drawn due to the small number and heterogeneity of identified studies. Further, we did not identify any studies assessing long-term effects such as mortality, overdiagnosis, or quality of life.

Whereas the methodology of the included studies was generally good, we have serious concerns about the poor generalizability of the results of the included studies. The European studies [16,18] included only highly trained radiologists who can achieve high diagnostic accuracy because of their large experience in mammogram interpretation. Otherwise, less trained radiologists may not have the same diagnostic accuracy.

We found that the great majority of studies comparing double reading with single reading were conducted in the context of screen-film mammography, specifically 56 out of the 66 full text screened articles. Although the majority of European screening programmes have changed from screen-film to digital mammography over the last decade [1], scarce studies have assessed double reading in this new context. As digital mammography is associated with increases in detection rates [7], it seems that the supposed benefit of double reading attempting to increase the detection rate may be diminished. We think that further investigation is required to assess the cost-effectiveness of different reading strategies in the context of digital mammography.

Single reading is more affordable and easy to implement in all

screening programmes. On the contrary, the shortage of radiologists may bring difficulties for double reading all mammograms. Sato et al. [17] found that single reading + CAD was a cost-effective strategy. Although the scientific evidence is insufficient to determine whether the accuracy of single reading with or without CAD is equivalent to double reading [20–22], one of the reasons for conducting this systematic review is to stimulate rethinking of single reading alone or single reading + CAD as a feasible and efficient strategy.

As CAD is becoming a usual technique, knowledge about its effectiveness and cost-effectiveness becomes of key clinical and budgetary significance. We identified several studies assessing single reading + CAD for reading digitalized screen-film mammography and suggesting benefits of CAD on the radiologists interpretation [20–22]. However, factors affecting the cancer detection rate or false-positive rate of CAD can be different in the context of digital mammography [23]. With digital mammography the CAD system receives more accurate information from the image that improves the computer's ability to discriminate between true suspicious lesions and false-positive results [24]. Therefore, the results of Sato et al. [17] showing single reading + CAD as a cost-effective strategy should be replicated in European screening programmes.

A comparison of single reading with double reading in screening with digital breast tomosynthesis was recently performed by Houssami et al. [16]. Digital breast tomosynthesis is an improvement on conventional 2D imaging providing analysis of 3D mammographic data through a series of image slices allowing reconstruction in thin slices [25]. Two recently published systematic reviews have reported the potential for digital breast tomosynthesis to decrease recall rates and increase cancer detection rates [25,26]. In fact, Houssami et al.

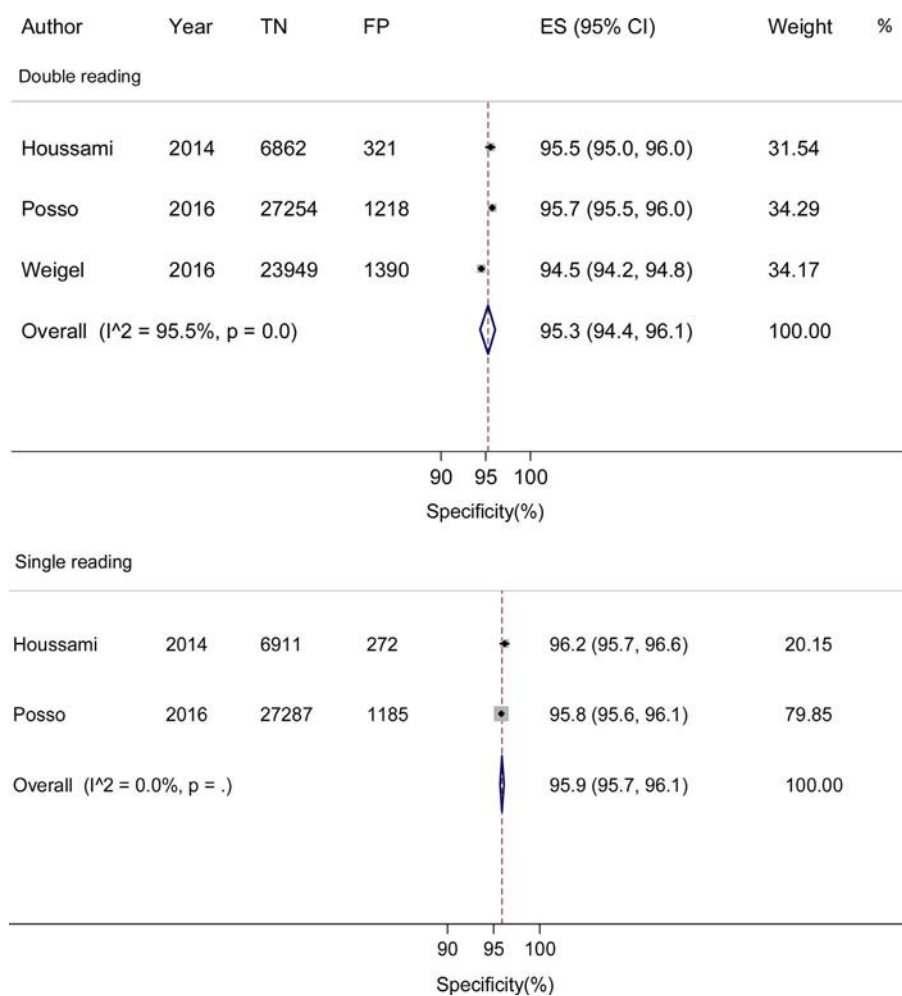


Fig. 5. Specificity (%) pooled meta-analysis of double reading and single reading of digital mammography.

reported that the main increase in cancer detection rate was reached by shifting to digital breast tomosynthesis regardless of the single or double reading strategy [16] and suggested that single reading of tomosynthesis may be better than double reading of digital mammography [27].

Given the well-recognised complexity of screening mammography interpretation, the radiologist’s reading volume and their experience have an effect on the reading performance. Variations in experience, reading volume and using consensus or arbitration in double reading have, therefore, an influence on tumour detection and false-positive rates. The studies included in this review were performed with highly trained professionals and we could not evaluate the association between experience or reading policies and results. Other studies based on film reading, however, showed a positive association between annual reading volume and false-positive rates. These studies showed a trend to a lower false-positive rate with increasing screening reading volume and no association between annual reading volume and cancer detection rate [28–31]. Further evaluation of the experience and reading volume in the digital context is needed.

The main limitation of this systematic review is the small number of studies included. Although it is possible that we were not able to identify all published studies comparing double reading with single reading, the literature search was comprehensive, with more than 1600 articles published between 1990 and 2016 that have been examined. Unfortunately, there are limited data comparing double reading with single reading in the context of digital mammography, digital mammography with CAD or digital breast tomosynthesis.

Our results are useful for decision makers as they inform about the effectiveness and cost-effectiveness of single reading with and without

CAD in the context of digital mammography. This information is not conclusive and should be confirmed with further research. Randomised controlled trials can be the most valid design to investigate whether or not the additional information from double reading confers incremental benefit in the context of digital mammography. Also, a cost-utility study comparing the cost per QALY gained at single and double digital reading should be performed within a randomised controlled trial. We sought to systematically review for long-term outcomes since they are critical outcomes for patients and are gaining wide acceptance in most fields of clinical research. However, these outcomes are difficult to evaluate and therefore they are not frequently reported. We did not find the effect of reading strategies on these outcomes but we think that future studies should focus on them. Our hypothesis points out that double reading may not be the most cost-effective way for digital mammography interpretation and should be tested through further research.

5. Conclusions

In conclusion, too little is currently known about the effectiveness and cost-effectiveness of double reading compared with single reading in the context of digital mammography. The uncertainty on the immediate effects such as cancer detection and false positives of double reading remains large and there are no publications on long-term health outcomes. Double reading seems to increase operational costs, have a not significantly higher false-positive rate, and a similar cancer detection rate. Substituting a second reader by CAD may be a cost-effective strategy. The emerging technologies such as digital mammography or tomosynthesis necessitate innovative screen-reading strategies to

Table 3
Cost-effectiveness results of studies assessing double reading and single reading of digital mammograms.

Reference	Incremental cost (double vs. single reading)	Incremental effect (double vs. single reading)	ICER	Threshold for cost-effective interventions	Reading strategy reported as cost-effective
Sato [17]	25.7 US\$ PPP	0.0087 life year gained	2950.7 US\$ PPP per life year gained	56,962.8 US\$ PPP	Single reading + CAD is cost-effective at a willingness to pay of 56,9962.8 US\$ PPP
Posso [18]	271,885.9 US\$ PPP	11 detected cancers	16,684 Euros (24,717.0 US\$ PPP) per detected cancer	Not reported	With an increase of 16,684 Euros (24,717.0 US\$ PPP) per detected cancer, double reading appears not to be a cost-effective strategy in the context of digital mammography

US\$ PPP = United States Dollar purchasing power parities. ICER: Incremental Cost-Effectiveness Ratio.

improve its efficiency.

Competing interests

The authors have no competing interests to declare. This protocol is registered in PROSPERO (2014:CRD42014013804).

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors thank Ms. María Victoria Leo Rosas for her contributions to the language editing of the present article, and Mr. Victor Velásquez Rimachi for his contributions to the review process.

Margarita Posso is a Ph.D. candidate at the Methodology of Biomedical Research and Public Health program, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ejrad.2017.09.013>.

References

- [1] L. Giordano, L. von Karsa, M. Tomatis, O. Majek, C. de Wolf, L. Lancucki, et al., Mammographic screening programmes in Europe: organization, coverage and participation, *J. Med. Screen.* 19 (Suppl. 1) (2012) 72–82, <http://dx.doi.org/10.1258/jms.2012.012085>.
- [2] E.G. Klompenhouwer, A.C. Voogd, G.J. den Heeten, L.J.A. Strobbe, A.F.J. de Haan, C.A. Wauters, et al., Blinded double reading yields a higher programme sensitivity than non-blinded double reading at digital screening mammography: a prospectively population based study in the south of The Netherlands, *Eur. J. Cancer* 51 (2015) 391–399, <http://dx.doi.org/10.1016/j.ejca.2014.12.008>.
- [3] F.J. Gilbert, S.M. Astley, M.G.C. Gillan, O.F. Agbaje, M.G. Wallis, J. James, et al., Single reading with computer-aided detection for screening mammography, *N. Engl. J. Med.* 359 (2008) 1675–1684, <http://dx.doi.org/10.1056/NEJMoa0803545>.
- [4] N. Perry, M. Broeders, C. de Wolf, S. Tömberg, R. Holland, L. von Karsa, *European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis*, 4th ed., Office for Official Publications of the European Communities, Luxembourg, 2006.
- [5] J. Dinnes, S. Moss, J. Melia, R. Blanks, F. Song, J. Kleijnen, Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: findings of a systematic review, *Breast* 10 (2001) 455–463, <http://dx.doi.org/10.1054/brst.2001.0350>.
- [6] J.H. Groenewoud, J.D.M. Otten, J. Fracheboud, G. Draisma, B.M. van Ineveld, R. Holland, et al., Cost-effectiveness of different reading and referral strategies in mammography screening in the Netherlands, *Breast Cancer Res. Treat.* 102 (2007) 211–218, <http://dx.doi.org/10.1007/s10549-006-9319-4>.
- [7] S. Vinnicombe, S.M. Pinto Pereira, V.A. McCormack, S. Shiel, N. Perry, I.M. Dos Santos Silva, Full-field digital versus screen-film mammography: comparison within the UK breast screening program and systematic review of published data, *Radiology* 251 (2009) 347–358, <http://dx.doi.org/10.1148/radiol.2512081235>.
- [8] P. Skaane, S. Hofvind, A. Skjennald, Randomized trial of screen-film versus full-field digital mammography with soft-copy reading in population-based screening program: follow-up and final results of Oslo II Study 1, *Radiology* 244 (2007) 708–717, <http://dx.doi.org/10.1148/radiol.2443061478>.
- [9] M.C. Posso, T. Puig, M.J. Quintana, J. Solà-Roca, X. Bonfill, Double versus single reading of mammograms in a breast cancer screening programme: a cost-consequence analysis, *Eur. Radiol.* 26 (2016) 3262–3271, <http://dx.doi.org/10.1007/s00330-015-4175-4>.
- [10] In: J.J. Deeks, P. Bossuyt, C. Gatsonis, (Eds.) *Cochrane Handbook for DTA Reviews*. Available from: < <http://methods.cochrane.org/sdt/handbook-dta-reviews/> > , Cochrane Collaboration; 2013. (Accessed September 19th 2017).
- [11] D. Moher, L. Shamseer, M. Clarke, D. Ghera, A. Liberati, M. Petticrew, et al., Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement, *Syst. Rev.* 4 (2015) 1, <http://dx.doi.org/10.1186/2046-4053-4-1>.
- [12] P.F. Whiting, A.W.S. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, et al., QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, *Ann. Intern. Med.* 155 (2011) 529–536, <http://dx.doi.org/10.7326/0003-4819-155-8-201110180-00009>.
- [13] D. Huserau, M. Drummond, S. Petrou, C. Carswell, D. Moher, D. Greenberg, et al., Consolidated health economic evaluation reporting standards (CHEERS) statement, *Value Health* 16 (2013) e1–e5, <http://dx.doi.org/10.1016/j.jval.2013.02.010>.

- [14] Organization for Economic Co-operation, Development (OECD), Purchasing Power Parities for GDP and Related Indicators, OECD, 2017 <https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm>. (Accessed 23 April 2017).
- [15] V.N. Nyaga, M. Arbyn, M. Aerts, Metaprop: a Stata command to perform meta-analysis of binomial data, *Arch. Public Health* 72 (39) (2014), <http://dx.doi.org/10.1186/2049-3258-72-39>.
- [16] N. Houssami, P. Macaskill, D. Bernardi, F. Caumo, M. Pellegrini, S. Brunelli, et al., Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading-evidence to guide future screening strategies, *Eur. J. Cancer* 50 (2014) 1799–1807, <http://dx.doi.org/10.1016/j.ejca.2014.03.017>.
- [17] M. Sato, M. Kawai, Y. Nishino, D. Shibuya, N. Ohuchi, T. Ishibashi, Cost-effectiveness analysis for breast cancer screening: double reading versus single + CAD reading, *Breast Cancer* 21 (2014) 532–541, <http://dx.doi.org/10.1007/s12282-012-0423-5>.
- [18] M. Posso, M. Carles, M. Rue, T. Puig, X. Bonfill, Cost-effectiveness of double reading versus single reading of mammograms in a breast cancer screening programme, *PLoS One* 11 (2016) e0159806, <http://dx.doi.org/10.1371/journal.pone.0159806>.
- [19] S. Weigel, W. Heindel, J. Heidrich, H.W. Hense, O. Heidinger, Digital mammography screening: sensitivity of the programme dependent on breast density, *Eur. Radiol.* (2016) 1–8, <http://dx.doi.org/10.1007/s00330-016-4636-4>.
- [20] E. Azavedo, S. Zackrisson, I. Mejère, M. Heibert Arnlind, Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review, *BMC Med. Imaging* 12 (2012) 22, <http://dx.doi.org/10.1186/1471-2342-12-22>.
- [21] X. Bargalló, G. Santamaría, M. Del Amo, P. Arguis, J. Ríos, J. Grau, et al., Single reading with computer-aided detection performed by selected radiologists in a breast cancer screening program, *Eur. J. Radiol.* (2014), <http://dx.doi.org/10.1016/j.ejrad.2014.08.010>.
- [22] L.H. Eadie, P. Taylor, A.P. Gibson, A systematic review of computer-assisted diagnosis in diagnostic cancer imaging, *Eur. J. Radiol.* 81 (2012) e70–e76, <http://dx.doi.org/10.1016/j.ejrad.2011.01.098>.
- [23] F. Engelken, R. Bremme, U. Bick, S. Hammann-Kloss, E.M. Fallenberg, Factors affecting the rate of false positive marks in CAD in full-field digital mammography, *Eur. J. Radiol.* 81 (2012) e844–e848, <http://dx.doi.org/10.1016/j.ejrad.2012.02.017>.
- [24] C. Dromain, B. Boyer, R. Ferré, S. Canale, S. Delalogue, C. Balleyguier, Computed-aided diagnosis (CAD) in the detection of breast cancer, *Eur. J. Radiol.* 82 (2013) 417–423, <http://dx.doi.org/10.1016/j.ejrad.2012.03.005>.
- [25] R. Hodgson, S.H. Heywang-Köbrunner, S.C. Harvey, Systematic review of 3D mammography for breast cancer screening, *Breast* 27 (2016) 52–61, <http://dx.doi.org/10.1016/j.breast.2016.01.002>.
- [26] F.J. Gilbert, L. Tucker, K.C. Young, Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool, *Clin. Radiol.* 71 (2016) 141–150, <http://dx.doi.org/10.1016/j.crad.2015.11.008>.
- [27] N. Houssami, D. Bernardi, M. Pellegrini, M. Valentini, C. Fantò, L. Ostillo, et al., Breast cancer detection using single-reading of breast tomosynthesis (3D-mammography) compared to double-reading of 2D-mammography: evidence from a population-based trial, *Cancer Epidemiol.* 47 (2017) 94–99, <http://dx.doi.org/10.1016/j.canep.2017.01.008>.
- [28] D.S. Buist, M.L. Anderson, S.J. Haneuse, E.A. Sickles, R.A. Smith, P.A. Carney, et al., Influence of annual interpretive volume on screening mammography performance in the United States, *Radiology* 259 (2011) 72–84, <http://dx.doi.org/10.1148/radiol.10101698>.
- [29] S. Haneuse, D.S. Buist, D.L. Miglioretti, M.L. Anderson, P.A. Carney, T. Onega, et al., Mammographic interpretive volume and diagnostic mammogram interpretation performance in community practice, *Radiology* 262 (2012) 69–79, <http://dx.doi.org/10.1148/radiol.11111026>.
- [30] I. Théberge, N. Hébert-Croteau, A. Langlois, D. Major, J. Brisson, Volume of screening mammography and performance in the Quebec population-based Breast Cancer Screening Program, *CMAJ* 172 (2005) 195–199, <http://dx.doi.org/10.1503/cmaj.1040485>.
- [31] R. Smith-Bindman, P. Chu, D.L. Miglioretti, C. Quale, R.D. Rosenberg, G. Cutter, et al., Physician predictors of mammographic accuracy, *J. Natl. Cancer Inst.* 97 (2005) 358–367, <http://dx.doi.org/10.1093/jnci/dji060>.
- [32] S. Ciatto, N. Houssami, D. Bernardi, F. Caumo, M. Pellegrini, S. Brunelli, et al., Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study, *Lancet Oncol.* 14 (2013) 583–589, [http://dx.doi.org/10.1016/S1470-2045\(13\)70134-7](http://dx.doi.org/10.1016/S1470-2045(13)70134-7).

5.4. ARTÍCULO 4. EXPRESIÓN DE BIOMARCADORES EN LESIONES BENIGNAS DE LA MAMA

5.4.1. RESUMEN DE RESULTADOS

CARACTERÍSTICAS DE LA POBLACIÓN DE ESTUDIO

Del total de 258 mujeres con lesiones benignas de la mama 86 fueron casos, es decir desarrollaron un cáncer y 172 fueron controles, es decir permanecieron libres de cáncer hasta el fin del seguimiento. El 65,1% del total (N = 168) presentaron una lesión benigna no proliferativa, el 30,2% (N = 78) una lesión proliferativa sin atipia y el 4,7% (N = 12) una lesión proliferativa con atipia.

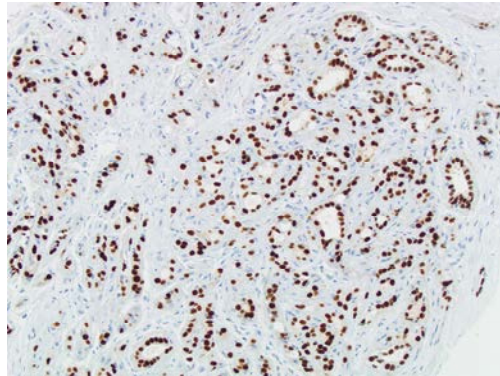
La edad media al momento del diagnóstico de lesión benigna fue de 56,4 años. La mediana de tiempo que transcurrió entre el diagnóstico de lesión benigna y el diagnóstico de cáncer en los casos fue de 71,4 meses, y la mediana de seguimiento en los controles fue de 69,2 meses (P = 0,69).

El 78% (N = 67/86) de los casos fueron detectados durante el cribado y el 22% (N = 19/86) fuera del cribado regular. El 77% (N = 66/86) fueron carcinomas invasivos, el 14% (N = 12/86) fueron in situ y la presencia de invasión fue desconocida en el 9% (N = 8/86) de los casos.

EXPRESIÓN DE REEPTORES DE ESTRÓGENO Y RIESGO DE CÁNCER

Comparado con las lesiones benignas que presentaron una expresión baja de receptores de estrógeno ($\leq 70\%$), la expresión moderada (71% a 90%) y alta ($> 90\%$) (Figura 8) mostró un riesgo mayor de cáncer de mama (OR = 1,98 y OR = 2,63; respectivamente).

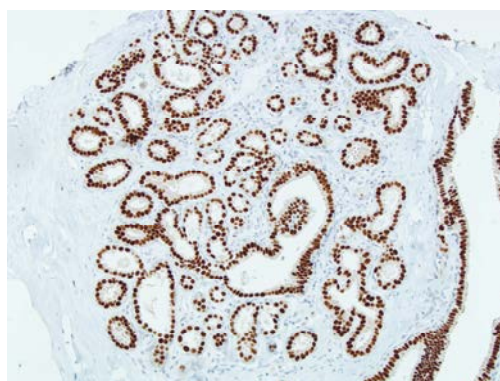
Figura 4. Expresión superior a 90% de receptores de estrógenos en una lesión no proliferativa de la mama.



EXPRESIÓN DE RECEPTORES DE PROGESTERONA Y RIESGO DE CÁNCER

En la línea de lo observado con la expresión de receptores de estrógenos, las lesiones con alta expresión de receptores de progesterona ($> 80\%$) (Figura 9) mostraron un mayor riesgo de cáncer de mama (OR = 2,22) comparado con las lesiones con baja expresión ($\leq 40\%$).

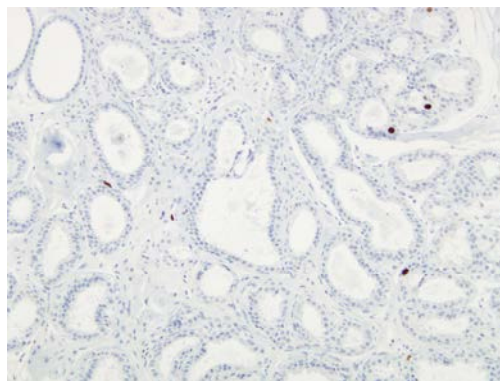
Figura 5. Expresión superior a 80% de receptores de progesterona en una lesión no proliferativa de la mama.



EXPRESIÓN DE KI67 EN LAS LESIONES PROLIFERATIVAS Y RIESGO DE CÁNCER

Casi todas las lesiones benignas mostraron algún grado de expresión de Ki67 (98,9%, N = 89/90). En comparación con las lesiones con una expresión menor al 1% de Ki67 (Figura 10), las mujeres con una expresión igual o superior al 1% mostraron un riesgo no significativamente superior de cáncer (OR = 1,16).

Figura 6. Expresión inferior a 1% de Ki67 en una lesión proliferativa de la mama.



PUBLICACIÓN 4


POSSO M, COROMINAS JM, SERRANO L, ROMÁN M, TORÁ-ROCAMORA I, ET AL.
BIOMARKERS EXPRESSION IN BENIGN BREAST DISEASES AND RISK OF SUBSEQUENT
BREAST CANCER: A CASE–CONTROL STUDY. *CANCER MED.* 2017; 6(6): 1482–89.

Cancer Medicine. 2017.

Factor de impacto 3,362. Q2 (90/217). *Oncology.*

ORIGINAL RESEARCH

Biomarkers expression in benign breast diseases and risk of subsequent breast cancer: a case–control study

Margarita Posso^{1,2} , Josep M. Corominas³, Laia Serrano³, Marta Román^{1,4}, Isabel Torá-Rocamora^{1,4}, Laia Domingo^{1,4,5}, Anabel Romero^{1,4}, María Jesús Quintana^{2,6}, María Vernet-Tomas⁷, Marisa Baré⁸, Carmen Vidal⁹, Mar Sánchez¹⁰, Francina Saladié¹¹, Carmen Natal¹², Joana Ferrer¹³, Sònia Servitja¹⁴, María Sala^{1,4}, Xavier Castells^{1,4} & on behalf of the BELE Study Group

¹Department of Epidemiology and Evaluation, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

²Department of Clinical Epidemiology and Public Health, Hospital de la Santa Creu i Sant Pau (IB Sant Pau), Barcelona, Spain

³Pathology Department, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

⁴Research Network on Health Services in Chronic Diseases (REDISSEC), Barcelona, Spain

⁵Agency for Health Quality and Assessment of Catalonia (AQuAS), Barcelona, Spain

⁶CIBER of Epidemiology and Public Health (CIBERESP), Barcelona, Spain

⁷Obstetrics and Gynecology Department, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

⁸Clinical Epidemiology and Cancer Screening, Parc Taulí University Hospital, Barcelona, Spain

⁹Cancer Prevention and Control Program, Catalan Institute of Oncology, Barcelona, Spain

¹⁰Direction General of Public Health, Department of Health, Government of Cantabria, Santander, Spain

¹¹Breast Cancer Screening Program of Tarragona, The Foundation League for the Research and Prevention of Cancer, Tarragona, Spain

¹²Principality of Asturias Breast Cancer Screening Program, Principality of Asturias, Oviedo, Spain

¹³Radiology Department, Hospital de Santa Caterina, Girona, Spain

¹⁴Oncology Department, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

Keywords

Benign breast disease, biomarkers, breast cancer, early detection, screening

Correspondence

Xavier Castells, Pg. Marítim 25–29, 08003, Barcelona, Catalonia, Spain.

Tel: (+34) 932483636;

Fax: (+34) 932483496;

E-mail: xcastells@parcdesalutmar.cat

Funding Information

This work was partly supported by Grants from Instituto de Salud Carlos III FEDER, (PI11/01296 and PI15/00098).

Received: 16 February 2017; Revised: 28

March 2017; Accepted: 29 March 2017

doi: 10.1002/cam4.1080

Abstract

Women with benign breast diseases (BBD) have a high risk of breast cancer. However, no biomarkers have been clearly established to predict cancer in these women. Our aim was to explore whether estrogen receptor (ER), progesterone receptor (PR), and Ki67 expression stratify risk of breast cancer in screened women with BBD. We conducted a nested case–control study. Women with breast cancer and prior BBDs (86 cases) were matched to women with prior BBDs who were free from breast cancer (172 controls). The matching factors were age at BBD diagnosis, type of BBD, and follow-up time since BBD diagnosis. ER, PR, and Ki67 expression were obtained from BBDs' specimens. Conditional logistic regression was used to estimate odds ratios (ORs), and 95% confidence intervals (CIs) of breast cancer risk according to ER, PR, and Ki67 expression. Women with >90% of ER expression had a higher risk of breast cancer (OR = 2.63; 95% CI: 1.26–5.51) than women with ≤70% of ER expression. Similarly, women with >80% of PR expression had a higher risk of breast cancer (OR = 2.22; 95% CI: 1.15–4.27) than women with ≤40% of PR expression. Women with proliferative disease and ≥1% of Ki67 expression had a nonsignificantly increased risk of breast cancer (OR = 1.16; 95% CI: 0.46–2.90) than women with <1% of Ki67 expression. A high expression of ER and PR in BBD is associated with an increased risk of subsequent breast cancer. In proliferative disease, high Ki67 expression may also have an increased risk. This information is helpful to better characterize BBD and is one more step toward personalizing the clinical management of these women.

Introduction

Benign breast diseases (BBD) are associated with an increased risk of subsequent breast cancer [1, 2]. The type of histological abnormality of BBD stratifies this

risk. Although proliferative disease with and without atypia could be associated with a two- to fourfold increased risk of developing breast cancer, nonproliferative diseases have a minimal increased risk of breast cancer [1].

The introduction of mammographic screening has led to a rise in the detection of BBD [2–4]. There are a number of mammographic features which, when considered suspicious, can lead to further diagnostic tests reaching a nonmalignant confirmation [5]. We have previously described that approximately 1.8% of screened women were found to have a BBD in our Spanish cohort [6]. Although it is possible that intensive screening can enhance the benefits for women with BBD, in fact, the majority of these women are still following routine screening recommendations [2, 4].

In order to better understand the biological characteristics of BBD, a few authors have focused on estrogen receptor (ER) [7, 8], progesterone receptor (PR) [9], and Ki67 [7, 10, 11] as biomarkers expressed in benign lesions that are not in proximity to concomitant breast cancer. The use of these biomarkers in the common clinical practice was born out by the fact that they have prognostic and response to treatment significance in women with breast cancer tumors [12, 13]. In women with BBD, however, the study of the expression of ER, PR, and Ki67 as predictors of subsequent breast cancer has shown inconsistent results [7–11]. Although some studies reported an increased risk of subsequent cancer in women with BBD that had high ER, PR, or Ki67 expression [7–9, 11], others reported different or contrary results [7, 14, 15].

To date, no biomarkers have been clearly established to predict cancer in women diagnosed with BBD. Moreover, as far as we know, studies that assessed these biomarkers have not been performed in breast cancer screening programs. Ideally, the identification of these biomarkers could improve both our ability to stratify an individual's risk for breast cancer and lead to more accurate follow-up or screening strategies. Therefore, the aim of our study was to explore whether ER and PR expression stratify risk of breast cancer in screened women with BBD. We also investigated the Ki67 expression in women with proliferative disease.

Material and Methods

Setting

The Spanish breast cancer screening program adheres to the recommendations of the European Guidelines for quality assurance in breast cancer screening and diagnosis [16]. Characteristics of the Spanish program have been previously described [6]. Briefly, women aged 50–69 years are routinely invited for a biennial screening. Biennial screening consists of two mammography views of each breast, a mediolateral oblique, and a cranio-caudal view. If the mammogram is negative, the woman is invited for further mammography screening in 2 years. On the contrary, if

suspicious mammographic findings are identified, the woman is recalled for further assessment to rule out malignancy. Further assessment often includes imaging procedures (i.e., additional mammography, ultrasonography, and magnetic resonance imaging) and/or biopsies (i.e., fine-needle aspiration, core needle biopsy, and open biopsy).

Study design

This study was designed as a nested case–control study among the subset of women with pathology-confirmed BBD within our Spanish screening cohort. The initial cohort subset included 10,262 women with BBD diagnosed in a breast cancer screening program between 1994 and 2011. Characteristics of our cohort have been previously reported [6]. Biopsies with indeterminate histological classification, for example, “negative for malignant cells” ($N = 4251$), were excluded from the analysis because they could not be classified in any of the BBD subtypes. Most of these biopsies with indeterminate classification came from fine-needle aspiration cytology. Therefore, 6011 women with BBD were followed up until December 31, 2015. Each BBD was classified into one of the following types: (1) nonproliferative disease; (2) proliferative disease without atypia; and (3) proliferative disease with atypia [17–19].

Cases and controls were selected from the study cohort subset. Cases were women with BBD who developed breast cancer during follow-up, whereas controls were women with BBD who remained cancer-free at least longer than the time to cancer in matched cases. From the 6011 women with BBD, 201 (3.3%) developed breast cancer during follow-up and were considered cases. We excluded 114 cases because they did not have sufficient tissue to assess biomarker's expression in the BBD sample and one case because it had evidence of mucinous carcinoma in the BBD sample. Thus, 86 cases with sufficient tissue to assess biomarkers' expression were included in the analysis. We randomly selected two controls per each case. Cases and controls were matched by age at the time of BBD diagnosis, type of BBD (nonproliferative disease, proliferative disease without atypia, or proliferative disease with atypia), and follow-up time between BBD diagnosis and end of follow-up. The sample size for the analysis included 258 women, 86 cases matched to 172 controls.

Both invasive and in situ carcinomas detected at regular screening or until the end of follow-up were analyzed. Age at BBD diagnosis, diagnosis of BBD, follow-up time between BBD diagnosis and the end of follow-up, and breast cancer events were retrieved from the breast cancer screening program database and hospital records. All protocol procedures and methods of ascertainment of BBD and breast cancer diagnosis were reviewed and

approved by the ethical committees at all participating institutions.

Pathology biopsies (ER, PR, Ki67 expression measurement)

Two pathologists independently reviewed the hematoxylin and eosin (H&E) slides from the biopsy blocks for the BBDs. The pathologists were blinded to each biopsy's case or control status. They completed detailed worksheets on the ER, PR, and Ki67 expressions when applied, and the subtype of BBD lesion (i.e., nonproliferative, proliferative without atypia, and proliferative with atypia) [17–19]. In the case of disagreement between pathologists, the quantitation was determined by consensus. If the woman had more than one biopsy during the study period, we used findings from the earliest biopsy performed. Individual tissue samples that exhibited more than one pathology diagnosis were classified by the most severe pathology finding.

Samples were fixed in 4% buffered formalin and embedded in paraffin. Four-micrometer sections of formalin-fixed, paraffin-embedded samples were deparaffinized and rehydrated. Staining H&E control and immunohistochemistry staining were performed for detecting ER, PR, and Ki67 expressions. Suitable positive tissue controls were used as quality control. Immunohistochemistry staining was performed using the Benchmark XT (VENTANA Roche).

To study ER and PR, we used the retrieval antigen protocol, endogenous peroxidase inhibition, detection system, DAB, and Hematoxylin from the UltraView system (VENTANA Roche). The primary antibodies used for ER and PR detection were Rabbit anti-human estrogen receptor alpha (clone SP1) prediluted (VENTANA Roche), and rabbit anti-human progesterone receptor (clone 1E2) prediluted (VENTANA Roche) and incubated for 20 min at 37°C.

To study Ki67, we used the retrieval antigen protocol, endogenous peroxidase inhibition, detection system, DAB, and Hematoxylin from the OptiView system (VENTANA Roche). The primary antibody used for Ki67 expression detection was rabbit anti-human Ki67 (clon 30-9) prediluted (VENTANA Roche) and incubated for 12 min at 37°C.

The percentage of ER, PR, and Ki67 staining was scored based on the number of positively stained nucleus per 100 studied cells observed with microscopic examination. At first, the entire sample was visualized in order to find the areas corresponding to the BBD lesion. If there were more than one BBD lesion, we evaluated the more severe one (i.e., proliferative disease with atypia followed by proliferative disease without atypia and by nonproliferative disease). In the case of proliferative disease, we evaluated the biomarker's expression in the areas with the highest proliferative activity. In the case of nonproliferative disease, we evaluated the biomarker's expression in the entire area

of the BBD lesion. We did not evaluate the background normal lobules.

Pathologic classification of cancer such as invasive carcinoma and carcinoma in situ were ascertained from hospital records.

Statistical analysis

The percentage of ER-, PR-, and Ki67-positive cells in each sample was obtained by the pathologists' microscopic assessment. We attempted to categorize the distribution of the ER and PR percentages into tertiles and we obtained the following groups: For ER (1) low expression: 0–70%; (2) moderate expression: 71–90%; and (3) high expression: >90%. For PR (1) low expression: 0–40%; (2) moderate expression: 41–80%; and (3) high expression: >80%. Regarding women with proliferative disease, Ki67 was categorized into two groups: (1) low expression: 0% to <1%; and (2) high expression: ≥1%. We calculated the median and range for continuous variables (age at BBD diagnosis, follow-up time between BBD diagnosis and the end of follow-up) and t-tests were used for comparisons. Chi-square tests were used to compare the distribution of ER, PR, and Ki67 expression across types of BBD and case-control status. Conditional logistic regression was used to estimate the odds ratios (ORs) and 95% confidence intervals (CIs) of breast cancer risk according to ER, PR, and Ki67 expression. We classified women according to the type of BBD and calculated ORs per each group. *P*-Value < 0.05 was considered statistically significant. Statistical analyses were performed using SPSS (v.19).

Results

Patients

Of the 201 subsequent breast cancer cases diagnosed from 1994 to 2015 in the study population, 86 had a previous BBD sample with sufficient tissue to assess ER and PR expression and were included in the analysis. We randomly selected two controls per each case. Thus, the sample size for the analysis included 258 women, 86 cases matched to 172 controls. Of the 258 women with BBD, 65.1% (*N* = 168) were classified as nonproliferative disease, 30.2% (*N* = 78) as proliferative disease without atypia, and 4.7% (*N* = 12) as proliferative disease with atypia. Overall, the mean age at BBD diagnosis was 56.4 years (56.3 and 56.5 years for cases and controls, respectively). Median time to cancer diagnosis among cases was 71.4 months (range, 18–193 months), and median length of cancer-free follow-up among controls was very similar (69.2 months, range, 12–183 months; *P* = 0.69) (Table 1). Breast cancer cases were screen-detected in 78% of all cases (*N* = 67/86), whereas the 22% (*N* = 19/86) was

Table 1. Characteristics of the nested case-control sample.

	Total (N = 258)		Case (N = 86)		Control (N = 172)		P-value
Benign breast disease (BBD)							
Nonproliferative (N, %)	168	65.1%	56	65.1%	112	65.1%	1.00 ²
Proliferative without atypia (N, %)	78	30.2%	26	30.2%	52	30.2%	
Proliferative with atypia (N, %)	12	4.7%	4	4.7%	8	4.7%	
Age at BBD diagnosis, years							
Mean, SD	56.4	0.3	56.3	0.5	56.5	0.4	0.70 ³
Median (Range)	56	(49–68)	55	(50–68)	56	(49–67)	
Age groups							
50–54 (N, %)	103	39.9%	36	41.9%	67	39.0%	0.50 ²
55–59 (N, %)	78	30.2%	29	33.7%	49	28.5%	
60–64 (N, %)	66	25.6%	19	22.1%	47	27.3%	
65–68 (N, %)	11	4.3%	2	2.3%	9	5.2%	
Years since BBD ¹ , months							
Mean, SD	70.0	2.5	71.4	4.6	69.2	3.0	0.69 ³
Median (Range)	63.5	(12–193)	67.0	(18–193)	62.0	(12–183)	

¹Follow-up time between year of BBD diagnosis and the end of follow-up.

²Chi-squared test comparing distributions of characteristics across case-control status.

³Two-sample t-tests comparing means across case-control status.

identified out of the regular screening. Seventy-seven percent ($N = 66/86$) of cases were invasive carcinomas, 14% ($N = 12/86$) were in situ, and the presence of invasion was unknown in 9% ($N = 8/86$) of cases.

ER expression

Nearly all BBD samples (99.2%; $N = 256/258$) showed $\geq 1\%$ of ER immunostaining expression. The highest ER expression was 98% (Fig. 1A), observed in 30.6% ($N = 79/258$) of all BBD samples. The ER expression mean scores were 83% and 74% ($P < 0.01$) in cases and controls, respectively. A higher proportion of BBD samples with $>90\%$ positively stained cells were observed in cases than in controls (Table 2). Compared to women with low ER expression ($\leq 70\%$), women with moderate (71% to 90%) and high expression ($>90\%$) had an elevated risk of breast cancer (OR = 1.98; 95% CI: 1.01–3.89 and OR = 2.63; 95% CI: 1.26–5.51, respectively) (Table 3). Restricting analyses to nonproliferative disease and to proliferative disease, we observed a similar tendency. For nonproliferative disease, ORs were nonsignificantly higher in women with moderate and high ER expression compared to women with low ER expression (OR = 1.85; 95% CI: 0.86–4.01 and OR = 2.42; 95% CI: 0.99–5.91, respectively). For proliferative disease, ORs were nonsignificantly higher in women with moderate and high ER expression compared to women with low ER expression (OR = 2.46; 95% CI: 0.59–10.29 and OR = 3.28; 95% CI: 0.82–13.13, respectively) (Table 3).

PR expression

Two hundred and fifty-two of 258 (97.7%) BBD samples showed $\geq 1\%$ of PR immunostaining expression. The

highest PR expression was 98% (Fig. 2A), observed in 12.4% ($N = 32/258$) of all BBD samples. The PR expression mean scores were 68% and 58% ($P < 0.01$) in cases and controls, respectively. A higher proportion of BBD samples with $>80\%$ positively stained cells were observed in cases than in controls (Table 2). Compared to women with low PR expression ($\leq 40\%$), women with high PR expression ($>80\%$) had an elevated risk of breast cancer (OR = 2.22; 95% CI: 0.15–4.27) (Table 3). Restricting analyses to nonproliferative disease and to proliferative disease, we observed a similar tendency. For nonproliferative disease, the OR was higher in women with high PR expression compared with women with low PR expression (OR = 2.48; 95% CI: 1.13–5.44). For proliferative disease, the OR was higher in women with high PR expression compared with women with low PR expression (OR = 1.81; 95% CI: 0.56–5.90) (Table 3).

Ki67 expression in women with proliferative disease

Almost all (98.9%; $N = 89/90$) proliferative disease samples showed some degree of Ki67 immunostaining expression; 28.9% ($N = 26/90$) of samples showed $<1\%$ of Ki67 expression (Fig. 3 and Table 2). The highest Ki67 expression was 10%, observed in one control. A similar proportion of BBD samples with $\geq 1\%$ positively stained cells were observed in cases (73.3%; $N = 22/30$) than in controls (70%; $N = 42/60$; $P = 0.11$) (Table 2). Compared to women with $<1\%$ of Ki67 expression, women with $\geq 1\%$ of Ki67 expression had a no significant increase in risk of breast cancer (OR = 1.16; 95% CI: 0.46–2.90) (Table 3).

Discussion

In this case–control study, we assessed whether the expression of estrogen receptors (ER) and progesterone

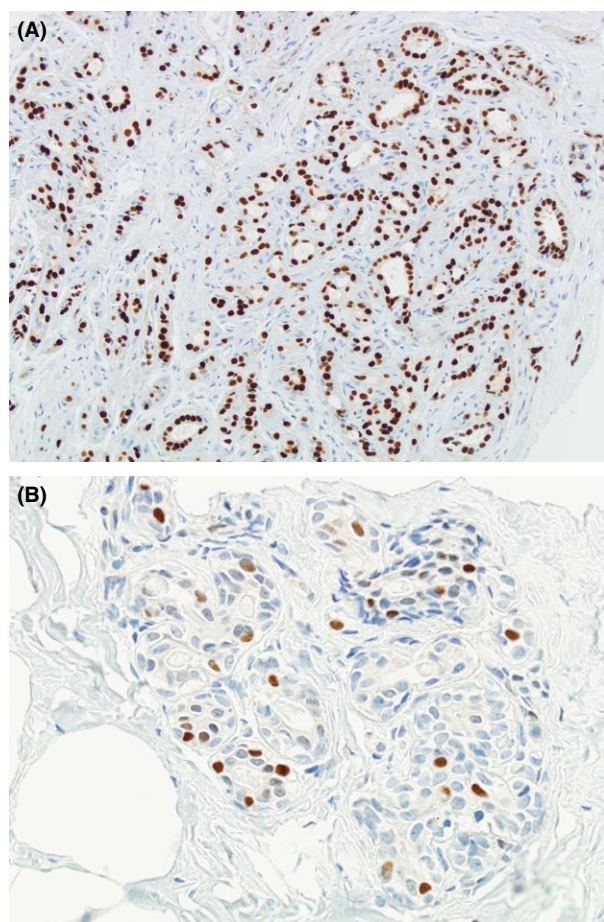


Figure 1. Estrogen receptor expression. (A) High expression (>90%). (B) Low expression (0–70%).

receptors (PR) stratifies the risk of subsequent cancer in women with benign breast diseases (BBD). We found both that women with high expression of ER or PR in BBD had an increased risk of subsequent breast cancer. This increased risk remains present when we analyzed women with nonproliferative disease. To the best of our knowledge, there is scarce evidence about the biomarkers' expression in these women. Therefore, our results may be useful to improve the clinical management of women with BBD identified in the screening context since they are mostly nonproliferative diseases.

We found that women with an ER expression higher than 90% in BBD had double the risk of subsequent breast cancer compared with an ER expression lower than 70%. This finding is in agreement with the hypothesis that estrogens promote normal growth of breast epithelium but it can also have an important role in the pathogenesis of breast cancer. Our results are consistent with two previous studies that reported that ER expression in nonatypical hyperplasia is associated with an increased risk of breast cancer [8, 20]. Other authors, however, have observed that ER expression does not increase the risk [7, 15]. The type of included benign lesions can explain the differences with our results. Whereas we included nonproliferative and proliferative diseases, Huh et al. [7] assessed normal breast tissue selected from biopsies with confirmed BBDs and Barr et al. [15] included only atypical hyperplasia.

We observed that high PR expression in BBD was associated with increased risk of breast cancer. Increased PR expression has also been observed in hyperplastic enlarged lobular units that are considered as the earliest histologically identifiable lesion with premalignant potential [9]. We did not find any other studies assessing PR expression in women with BBD. We found, however, one study that reported a decreasing trend for PR expression along

Table 2. Distribution of estrogen receptor (ER), progesterone receptor (PR), and Ki-67 expression according to the case-control status.

	Total (N = 258)		Case (N = 86)		Control (N = 172)		P-value ²
	N	%	N	%	N	%	
ER expression							
≤70%	80	31.0	18	20.9	62	36.1	0.04
71–90%	98	38.0	35	40.7	63	36.6	
>90%	80	31.0	33	38.4	47	27.3	
PR expression							
≤40%	83	32.2	23	26.7	60	34.9	0.02
41–80%	83	32.2	22	25.6	61	35.5	
>80%	92	35.7	41	47.7	51	29.7	
Proliferative disease	Total (N = 90)		Case (N = 30)		Control (N = 60)		
Ki67 expression ¹							
0% to <1%	26	28.3	8	26.7	18	30.0	0.74
≥1%	64	71.7	22	73.3	42	70.0	

¹Ki67 expression was evaluated only for proliferative disease.

²Chi-squared test comparing distributions of characteristics across case-control status.

Table 3. Odds ratios (OR) of subsequent breast cancer according to the percentage of estrogen receptor (ER), progesterone receptor (ER), and Ki67 expression.

	All benign breast diseases ¹			Nonproliferative disease			Proliferative disease ²		
	N	OR	(IC 95%)	N	OR	(IC 95%)	N	OR	(IC 95%)
ER expression									
≤70%	80	1		56	1		24	1	
71–90%	98	1.98	(1.01–3.89)	66	1.85	(0.86–4.01)	32	2.46	(0.59–10.29)
>90%	80	2.63	(1.26–5.51)	46	2.42	(0.99–5.91)	34	3.28	(0.82–13.13)
PR expression									
≤40%	83	1		60	1		23	1	
41–80%	83	0.87	(0.42–1.78)	56	0.68	(0.28–1.66)	27	1.43	(0.38–5.31)
>80%	92	2.22	(1.15–4.27)	52	2.48	(1.13–5.44)	40	1.81	(0.56–5.90)
Ki67 expression ³									
0% to <1%	–	–	–	–	–	–	26	1	
≥1%	–	–	–	–	–	–	64	1.16	(0.46–2.90)

OR, Conditional logistic regression was used to estimate odds ratios, and 95% confidence intervals of breast cancer risk according to ER, PR, and Ki67 expression. IC 95%, 95% confidence interval.

¹Women with nonproliferative disease, proliferative disease with atypia, and proliferative disease without atypia were included in the analyses.

²Women with proliferative disease with and without atypia were included in the analyses.

³Ki67 expression was evaluated only for proliferative disease.

with progression to malignancy [14]. This finding seems contrary to our results. Therefore, we consider that the PR expression should be further evaluated in larger studies.

Regarding Ki67, we found a nonsignificant association between high expression and subsequent breast cancer. Conversely, previous studies reported significant associations in women with sclerosis adenosis [10] or during the first 10 years post atypical hyperplasia [11]. In another study, Huh et al. [7] included women with BBD, mostly with proliferative disease, and observed a statistically significant increased risk of breast cancer in premenopausal women with >0.5% of Ki67 expression in BBD. In our study, we did not find a statistically significant increased risk. Discrepancies with our results can be explained by the sample size of our study, characteristics of included women, or the method used for quantifying the Ki67 expression and the obtained cut-offs. Although Huh et al. [7] obtained 0.28% as cut-off from digitalized methods, in our study, the great majority of specimens were classified in the group of ≤1% because pathologist's microscopic assessment did not provide lower cut-offs.

A better understanding of the characteristics of BBD identified in the context of screening will advance our clinical management of these increased risk lesions. Associations between genetic mutations and ER and PR expression were not addressed in our study. However, we think that further investigation should focus on ER and PR expression as biologically plausible biomarkers of

breast cancer risk [21]. On the other hand, clinical trials of selective estrogen receptor modulators and aromatase inhibitors have been proven to be effective at reducing breast cancer risk in women with atypical benign diseases [2, 22]. In agreement with these trials, we attempted to provide more accurate risk estimates that may lead to the utilization of preventive treatment in women at high risk who are in a position to benefit from them. Furthermore, we think that a more intensive screening, with a shorter interval time between examinations, should be explored as a suitable modality for women with BBD and high ER or PR expression.

The main limitation of this study is the relatively small sample size. We were not able to analyze the subset of women with atypical lesions neither differences between in situ and invasive carcinomas because of the small sample size. Further studies with larger number of cases should focus on these women. In addition, the generalizability of our findings is limited to women with BBD who have undergone a biopsy with sufficient tissue to examine ER, PR, and Ki67 expression. On the other hand, this study is strengthened by the fact that two breast pathologists, who were blinded to later cancer outcomes, reviewed all samples. Reproducibility of pathologists' microscopic examination compared with digitalized measures could be more useful for routine clinical practice in hospitals. Other strengths of this study are the use of a nested case-control design in the context of a screening program and the inclusion of nonproliferative diseases. Previous data came from studies assessing biologic characteristics

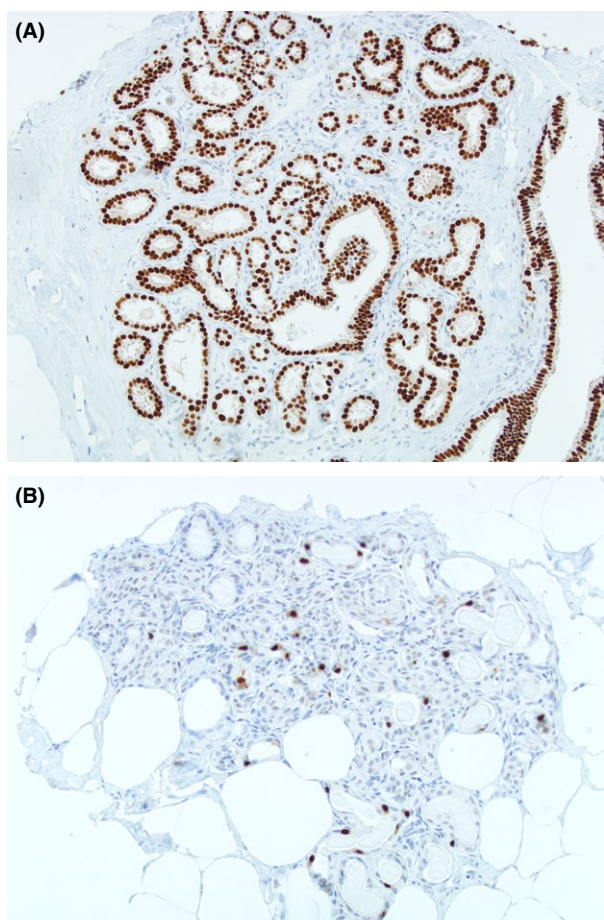


Figure 2. Progesterone receptor expression. (A) High expression (>80%). (B) Low expression (0–40%).

of BBDs mostly identified in younger women and out of the context of population-based screening programs. The current findings provide further evidence of the importance of nonproliferative diseases which are the most frequently detected BBDs in our setting [5].

In conclusion, a high expression of ER and PR in BBD is associated with increased risk of breast cancer. Although this increased risk was not demonstrated for the subset of nonproliferative or proliferative diseases, we found a tendency in both groups that should be confirmed in larger studies. Particularly in women with proliferative diseases, Ki67 may be associated with an increased risk. We believe that this information is helpful to better-characterize BBD, and is one more step toward the possibility of personalizing the clinical management of these women in the screening context.

Acknowledgments

Funding: This work was partly supported by Grants from Instituto de Salud Carlos III FEDER, (PI11/01296 and

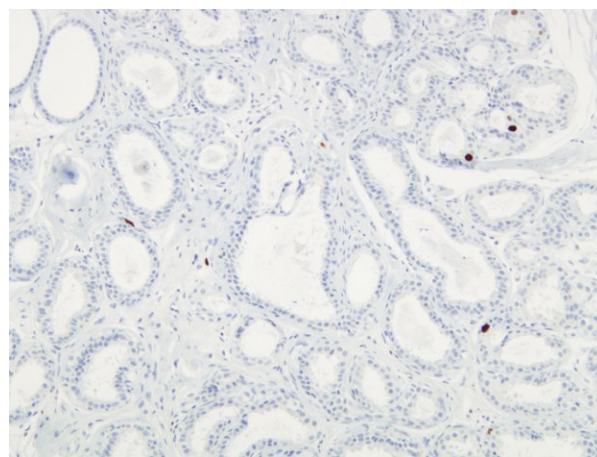


Figure 3. Ki67 < 1% of positive cells expression.

PI15/00098). The authors acknowledge the dedication and support of the entire Benign Lesion (BELE) Study Group (alphabetical order): IMIM (Hospital del Mar Medical Research Institute), Barcelona: Andrea Burón, Xavier Castells, Josep Maria Corominas, Laia Domingo, Margarita Posso, Ana Rodríguez-Arana, Marta Roman, Ana Romero, Sònia Servitja, Maria Sala, Laia Serrano, Isabel Torá-Rocamora, and Mar Vernet; Corporación Sanitària Parc Taulí, Sabadell: Marisa Baré and Xavier Andreu; Catalan Institute of Oncology, Barcelona: Núria Milà and Carmen Vidal; Hospital de la Santa Creu i Sant Pau, Barcelona: Laura López-Vilaró, Margarita Posso, and Maria Jesús Quintana; General Directorate of Public Health, Government of Cantabria: Mar Sánchez; Principality of Asturias Health Service: Carmen Natal; CIBERESP University of Oviedo: Ana Fernández-Somoano. Fundació Lliga per a la Investigació i Prevenció del Càncer, Reus, Universitat Rovira i Virgili: Jaume Galceran and Francina Saladié; Hospital Santa Caterina, Girona: Joana Ferrer; and Catalan Cancer Plan, Catalan Health Government, Barcelona: Josep Alfons Espinàs. Margarita Posso is a Ph.D. candidate at the Methodology of Biomedical Research and Public Health program, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain.

Conflict of Interest

None declared.

References

1. Dyrstad, S. W., Y. Yan, A. M. Fowler, and G. A. Colditz. 2015. Breast cancer risk associated with benign breast disease: systematic review and meta-analysis. *Breast Cancer Res. Treat.* 149:569–575.

2. Hartmann, L. C., A. C. Degnim, R. J. Santen, W. D. Dupont, and K. Ghosh. 2015. Atypical hyperplasia of the breast—risk assessment and management options. *N. Engl. J. Med.* 372:78–89.
3. Said, S. M., D. W. Visscher, A. Nassar, R. D. Frank, R. A. Vierkant, M. H. Frost, et al. 2015. Flat epithelial atypia and risk of breast cancer: a Mayo cohort study. *Cancer* 121:1548–1555.
4. Manfrin, E., R. Mariotto, A. Remo, D. Reghellin, F. Falsirollo, D. Dalfior, et al. 2009. Benign breast lesions at risk of developing cancer—A challenging problem in breast cancer screening programs. *Cancer* 115:499–507.
5. Castells, X., I. Torá-Rocamora, M. Posso, M. Román, M. Vernet-Tomas, A. Rodríguez-Arana, et al. February 2016. Risk of Breast Cancer in Women with False-Positive Results according to Mammographic Features. *Radiology* 151174.
6. Castells, X., L. Domingo, J. M. Corominas, I. Torá-Rocamora, M. J. Quintana, M. Baré, et al. 2015. Breast cancer risk after diagnosis by screening mammography of nonproliferative or proliferative benign breast disease: a study from a population-based screening program. *Breast Cancer Res. Treat.* 149:237–244.
7. Huh, S. J., H. Oh, M. A. Peterson, V. Almendro, R. Hu, M. Bowden, et al. 2016. The proliferative activity of mammary epithelial cells in normal tissue predicts breast cancer risk in premenopausal women. *Cancer Res.* 76:1926–1934.
8. Gobbi, H., W. D. Dupont, F. F. Parl, P. A. Schuyler, W. D. Plummer, S. J. Olson, et al. 2005. Breast cancer risk associated with estrogen receptor expression in epithelial hyperplasia lacking atypia and adjacent lobular units. *Int. J. Cancer* 113:857–859.
9. Lee, S., S. K. Mohsin, S. Mao, S. G. Hilsenbeck, D. Medina, and D. C. Allred. 2006. Hormones, receptors, and growth in hyperplastic enlarged lobular units: early potential precursors of breast cancer. *Breast Cancer Res.* 8:R6.
10. Nassar, A., T. L. Hoskin, M. L. Stallings-Mann, A. C. Degnim, D. C. Radisky, M. H. Frost, et al. 2015. Ki-67 expression in sclerosing adenosis and adjacent normal breast terminal ductal lobular units: a nested case-control study from the Mayo Benign Breast Disease Cohort. *Breast Cancer Res. Treat.* 151:89–97.
11. Santisteban, M., C. Reynolds, F. E. G. Barr, M. H. Frost, R. A. Vierkant, S. S. Anderson, et al. 2010. Ki67: a time-varying biomarker of risk of breast cancer in atypical hyperplasia. *Breast Cancer Res. Treat.* 121:431–437.
12. Perou, C. M., T. Sørlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, et al. 2000. Molecular portraits of human breast tumours. *Nature* 406:747–752.
13. Petrelli, F., G. Viale, M. Cabiddu, and S. Barni. 2015. Prognostic value of different cut-off levels of Ki-67 in breast cancer: a systematic review and meta-analysis of 64,196 patients. *Breast Cancer Res. Treat.* 153:477–491.
14. Zagouri, F., T. N. Sergentanis, and G. C. Zografos. 2007. Precursors and preinvasive lesions of the breast: the role of molecular prognostic markers in the diagnostic and therapeutic dilemma. *World J. Surg. Oncol.* 5:57.
15. Barr, F. E. G., A. C. Degnim, L. C. Hartmann, D. C. Radisky, J. C. Boughey, S. S. Anderson, et al. 2011. Estrogen receptor expression in atypical hyperplasia: lack of association with breast cancer. *Cancer Prev. Res.* 4:435–444.
16. Perry, N., M. Broeders, C. de Wolf, S. Tömborg, R. Holland, and L. von Karsa. 2006. European guidelines for quality assurance in breast cancer screening and diagnosis. 4th edn. Luxembourg: Office for Official Publications of the European Communities.
17. Dupont, W. D., and D. L. Page. 1985. Risk factors for breast cancer in women with proliferative breast disease. *N. Engl. J. Med.* 312:146–151.
18. Page, D. L., W. D. Dupont, L. W. Rogers, and M. S. Rados. 1985. Atypical hyperplastic lesions of the female breast. A long-term follow-up study. *Cancer* 55:2698–2708.
19. Fitzgibbons, P. L., D. E. Henson, and R. V. Hutter. 1998. Benign breast changes and the risk for subsequent breast cancer: an update of the 1985 consensus statement. *Cancer Committee of the College of American Pathologists. Arch. Pathol. Lab. Med.* 122:1053–1055.
20. Shaaban, A. M., J. P. Sloane, C. R. West, and C. S. Foster. 2002. Breast Cancer Risk in Usual Ductal Hyperplasia Is Defined by Estrogen Receptor- α and Ki-67 Expression. *Am. J. Pathol.* 160:597–604.
21. Allred, D. C., S. K. Mohsin, and S. A. Fuqua. 2001. Histological and biological evolution of human premalignant breast disease. *Endocr. Relat. Cancer* 8:47–61.
22. Clauser, P., M. A. Marino, P. A. Baltzer, M. Bazzocchi, and C. Zuiani. 2016. Management of atypical lobular hyperplasia, atypical ductal hyperplasia, and lobular carcinoma in situ. *Expert Rev. Anticancer Ther.* 16:335–346.



DISCUSIÓN

DISCUSIÓN

La discusión conjunta de esta tesis doctoral se articula en base a la evaluación de la efectividad y coste-efectividad del cribado con mamografía digital. Alrededor de este eje temático se evalúan dos aspectos clave del cribado como son la doble lectura mamográfica y las lesiones benignas de la mama. Para determinar la efectividad y el coste-efectividad de la doble lectura, se han desarrollado tres estudios, uno de coste-consecuencia, uno de coste-efectividad y una revisión sistemática y para caracterizar las lesiones benignas se ha desarrollado un estudio de casos y controles.

6.1. EFECTIVIDAD DE LA DOBLE LECTURA MAMOGRÁFICA

Durante la década de los 80, cuando se implementaron los programas de cribado, la mamografía analógica era la técnica de elección. Inicialmente se buscaba incrementar la detección de cáncer y se conocía poco sobre los efectos adversos del cribado. En ese contexto, la doble lectura demostró un incremento en la tasa de detección de cáncer¹⁰³ y los programas europeos realizaron esfuerzos para implementarla como estrategia de cribado poblacional⁶⁰.

Más recientemente, a partir de los años 2000, la mamografía digital ha desplazado a la analógica¹⁵⁰. Actualmente, incrementar la tasa de detección no constituye el único objetivo del cribado, sino que más bien se intenta mejorar el equilibrio entre los riesgos y beneficios¹⁵¹⁻¹⁵³. En este nuevo contexto, se dispone de escasa evidencia sobre la efectividad de las estrategias de lectura y sobre su efecto sobre los riesgos y beneficios. A pesar de esto, los programas de cribado con mamografía digital continúan utilizando la doble lectura sin realizar cambios significativos derivados de la tecnología digital que les acerque más a sus objetivos.

La evaluación de la efectividad de la doble lectura supone necesariamente una evaluación de su impacto sobre la frecuencia de falsos positivos y sobre la tasa de detección de cáncer. Por ello, el desarrollo de un estudio de coste-consecuencia y una revisión sistemática en este trabajo de tesis

procura aportar información relevante sobre estos dos desenlaces fundamentales para evaluar el equilibrio de riesgos y beneficios. Los resultados de estos dos estudios muestran que en el contexto de la mamografía digital, la doble lectura incrementa significativamente la frecuencia de falsos positivos y de forma más modesta la tasa de detección de cáncer. Se trata de unos resultados que consideramos relevantes dadas las consecuencias que podrían tener en la población de cribado.

Concretamente, en el estudio de coste-consecuencia se identificó un incremento del 0,3% en la proporción de resultados falsos positivos atribuible a la doble lectura, siendo de 4,2% con la lectura simple y 4,5% con la doble lectura ¹⁵⁴. Esta mayor frecuencia de falsos positivos con la doble lectura fue corroborada con la revisión sistemática. El metanálisis de los resultados de tres estudios europeos ^{146,148,149} mostró un incremento del 0,6% en la proporción de falsos positivos con la doble lectura. En ambos estudios, la proporción de falsos positivos ha sido calculada a partir de una o dos rondas de cribado. Estas cifras son considerablemente inferiores a las que se observarían de forma acumulada durante 10 rondas de cribado bienal. Se ha estimado que el riesgo acumulado de falsos positivos en mujeres de 50 a 69 años después de 10 rondas de mamografía es del 20% al 32% en Europa ^{89,155,156}. Por lo tanto, desde una perspectiva poblacional, considerando los 20 años de cribado con mamografía, la proporción de falsos positivos generados por la doble lectura podría ser mucho mayor.

Respecto a la tasa de detección de cáncer, el estudio de coste-consecuencia mostró que esta tasa con la doble lectura no fue significativamente superior a la lectura simple (4,6 por 1000 y 4,2 por 1000 mamografías, respectivamente). Un resultado similar se confirmó con el metanálisis (6,0 por 1000 y 5,7 por 1000, respectivamente). No obstante la similitud en las tasas de detección, existe un número de casos que de no ser por la doble lectura no se hubieran detectado durante el cribado. Algunos de estos cánceres detectados gracias a la doble lectura podrían representar un beneficio real y otros podrían contribuir al sobrediagnóstico. Un mejor conocimiento de las características radiológicas y clínicas de estos casos difíciles de detectar, o que sólo se detectan gracias a la doble lectura, permite valorar mejor el beneficio de esta estrategia.

En el estudio de coste-consecuencia se pudo evaluar las características de los casos más frecuentemente detectados con la doble lectura. Estos casos tuvieron imágenes radiológicas categorizadas como benignas o probablemente benignas según BI-RADS (Breast Imaging Reporting and Data System), fueron carcinomas *in situ* o no avanzados y de subtipo molecular

diferente al triple negativo. Otros autores, en base a resultados de mamografía analógica, informaron que los casos detectados con la doble lectura presentaron microcalcificaciones ¹⁰⁹. Parece ser, por tanto, que el beneficio de la doble lectura no es el mismo en toda la población cribada sino que se restringe a la detección de lesiones benignas o cánceres de comportamiento no invasivo.

Existe gran controversia sobre la magnitud de los casos que serían sobrediagnosticados con el cribado mamográfico, de modo que las estimaciones varían entre un 1% y un 80% ^{3,157,158}. De hecho, la doble lectura y las nuevas tecnologías diagnósticas, al incrementar la sensibilidad en la detección de lesiones benignas o carcinomas *in situ*, podrían incrementar también el sobrediagnóstico. Actualmente no se cuenta con evidencia suficiente sobre la relación de las estrategias de lectura y el sobrediagnóstico. Esta falta de evidencia dificulta valorar si el aparente beneficio de la doble lectura al incrementar la tasa de detección queda diluido al demostrar que contribuye con un incremento en el sobrediagnóstico.

Más allá de los resultados inmediatos como son la tasa de detección y la frecuencia de falsos positivos, la efectividad de la doble lectura se ha de valorar a la luz de los resultados a largo plazo. La evaluación de resultados a largo plazo es ampliamente aceptada debido a que permite valorar más objetivamente los efectos relevantes para los pacientes ¹⁵⁹. Sin embargo, estos resultados son difíciles de objetivar y por lo tanto no se informan con frecuencia. En este sentido, la revisión sistemática desarrollada en este trabajo de tesis demuestra que los estudios disponibles no han evaluado los efectos a largo plazo de la doble lectura como son la mortalidad, calidad de vida o sobrediagnóstico, evidenciando así la necesidad de futuros estudios que aborden este tema.

6.2. COSTE-EFECTIVIDAD DE LA DOBLE LECTURA MAMOGRÁFICA

Si bien la evidencia sobre la efectividad de la doble lectura en el contexto de la mamografía digital es escasa, la evidencia sobre su coste-efectividad lo es aún más. Esto sorprende si se tiene en cuenta que evaluar el coste que representa una estrategia de cribado frente a su efectividad es fundamental para apoyar su implementación y continuidad. En este trabajo de tesis, además de la información económica que aporta el estudio de coste-consecuencia, se ha desarrollado un estudio

de coste-efectividad cuyos resultados muestran que la doble lectura no parece ser eficiente en el contexto de la mamografía digital, mucho menos cuando se aplica solo en el cribado prevalente.

Como era de esperar, los recursos empleados para realizar la segunda lectura, el consenso y el arbitraje generan un coste superior al coste de realizar la lectura simple. Este incremento se produce a expensas de realizar un mayor número de pruebas complementarias. De hecho, en el estudio de coste-consecuencia se evidenció que cada resultado falso positivo tuvo un coste promedio de 511 € atribuible a las pruebas adicionales¹⁵⁴. Al incrementar en un 12,5% el coste total del personal y un 30% el coste de las pruebas complementarias¹⁵⁴, nuestros resultados indican que la doble lectura conlleva un mayor uso de recursos profesionales que no se ve compensado por una reducción en las pruebas complementarias. Más aún, dado que el incremento de pruebas complementarias no fue de utilidad para incrementar de forma significativa la tasa de detección de neoplasias, el coste de oportunidad de la doble lectura no parece estar justificado desde el punto de vista del cribado ni del uso de recursos a nivel hospitalario asistencial.

El coste medio por cáncer detectado con la doble lectura (8,912 €) parece similar al de la lectura simple (8,288 €), según hemos estimado en el estudio de coste-efectividad¹⁴⁸. No obstante su parecido, estas cifras no se traducen en que ambas estrategias sean igual de eficientes. Para valorar la eficiencia de la doble lectura, se calculó la relación de coste-efectividad (RCEI) que permite cuantificar el coste que representa cada cáncer detectado de forma adicional. La RCEI de la doble lectura fue de 16,684 € por cáncer detectado. Este monto duplica el coste medio de la detección de un cáncer con la lectura simple (8,288 €). Si bien no existe un consenso sobre la disponibilidad a pagar por cáncer detectado, estos resultados permiten cuestionar la eficiencia de la doble lectura. En todo caso, resulta difícil justificar una disponibilidad a pagar que representa el doble del coste medio por cada cáncer detectado de forma adicional.

La RCEI de 16,684 € por cáncer detectado, es susceptible de ser variable. Para estimar estas variaciones se realizó un análisis de sensibilidad determinístico en el que se identificó que un cambio del 5% en la sensibilidad de la doble lectura o de la lectura simple es el parámetro que más afecta la RCEI. Por el contrario, el efecto de otros parámetros de calidad estipulados en las guías europeas⁶⁰ no mostraron un efecto sobre la RCEI. Así, el porcentaje de participación, la prevalencia de cáncer, el valor predictivo positivo de la recitación y el porcentaje de controles avanzados no afectaron la RCEI. Por tanto, la cifra de 16,684 € por cáncer detectado con la doble lectura podría ser parcialmente aplicable a otros programas de cribado en el contexto europeo.

Otras evaluaciones económicas europeas se han centrado en evaluar y comprobar que el cribado con mamografía es una estrategia coste-efectiva frente a no realizar cribado ¹⁶⁰⁻¹⁶². También se han publicado modelos para simular el efecto de realizar un cribado basado en los factores de riesgo ^{163,164}. Sin embargo, estos estudios persisten en considerar la doble lectura como parámetro constante y no evalúan la diferencia que supondría aplicarla solo en el cribado prevalente. Por el contrario, en este trabajo de tesis si se realizó una evaluación del coste-efectividad de la doble lectura en el cribado prevalente.

Con una RCEI de 53.312 € por cáncer detectado, realizar doble lectura en el cribado prevalente y lectura simple en el cribado incidente demuestra no ser coste-efectivo. De hecho, esta estrategia está dominada por extensión. La dominancia por extensión se presenta cuando una estrategia aunque más costosa a nivel global es a la vez la que presenta una menor RCEI. En este caso, realizar doble lectura solo en el cribado prevalente es menos costoso que realizarla en ambos cribados, prevalente e incidente. Sin embargo, la RCEI de realizar la doble lectura en ambos cribados es menor que la RCEI de realizarla solo en el cribado prevalente. Esto se explica porque los cánceres que se presentan en el cribado incidente son más difíciles de detectar y, por tanto, el uso de la doble lectura es más eficiente en el cribado incidente.

6.3. NUEVAS TECNOLOGÍAS Y OTRAS VARIABLES RELACIONADAS CON LA EFECTIVIDAD Y COSTE-EFECTIVIDAD DEL CRIBADO

Aunque la mamografía sigue siendo recomendada como prueba de cribado tiene una limitación importante al ser una técnica que produce imágenes en dos dimensiones de un órgano que tiene tres dimensiones. Inevitablemente, esto implica que las lesiones puedan quedar ocultas por la superposición de tejido mamario. De hecho, la superposición de tejido denso puede conducir a falsos positivos ¹⁶⁵. En este sentido, investigar sobre mejoras en la mamografía digital, como el uso de los sistemas de ayuda al diagnóstico (CAD, Computer-Aided Detection) o la tomosíntesis cobra relevancia.

A medida que los sistemas de ayuda al diagnóstico pasan a ser una tecnología de uso habitual, el conocimiento sobre su eficacia y rentabilidad adquiere importancia clínica y presupuestaria. Es por esto que el sistema CAD ha sido ampliamente evaluado en contextos clínicos y de cribado con

mamografía analógica ¹⁶⁶, aunque existe escasa evidencia en el contexto de la mamografía digital. Tal como se demostró en la revisión sistemática desarrollada en este trabajo de tesis, solo se cuenta con un estudio sobre CAD en mamografía digital de cribado, en el cual la lectura simple acompañada de CAD fue más eficiente que la doble lectura ¹⁴⁷.

La evidencia científica aún es insuficiente para determinar los beneficios del CAD en la interpretación de las mamografías y por tanto, actualmente no se considera que el CAD pueda reemplazar a la segunda lectura en el cribado con mamografía. Sin embargo, se ha de tener en cuenta que estos sistemas han mejorado continuamente. Ahora, a diferencia de las imágenes analógicas, tienen la posibilidad de obtener información de mayor calidad de las imágenes digitales y, por tanto, podrían discriminar con mayor precisión las lesiones sospechosas. En consecuencia, sigue siendo necesario realizar estudios que clarifiquen esta cuestión.

La tomosíntesis es una tecnología de imagen digital que tiene el potencial de superar el problema de superposición de tejidos que se observa con la mamografía digital ^{167,168}. Por ello, la doble lectura en el contexto de la tomosíntesis podría no ser de utilidad pues por si sola, la tomosíntesis ya representa una mejora en la sensibilidad y especificidad. De hecho, uno de los estudios incluidos en la revisión sistemática informó que el mayor aumento en la tasa de detección de cáncer se alcanzó gracias a la incorporación de la tomosíntesis, independientemente de la estrategia de lectura utilizada ¹⁴⁶. Más recientemente, los mismos autores han reportado que una lectura simple de las imágenes producidas por la tomosíntesis detecta más cánceres y reduce la frecuencia de falsos positivos en comparación con la práctica habitual de doble lectura mamográfica ¹⁶⁹.

Los resultados de este trabajo de tesis se basan en el desempeño de profesionales experimentados y con gran volumen de lecturas al año ^{148,154} por lo que no ha sido posible valorar la efectividad y coste-efectividad de la doble lectura con volúmenes menores de lectura. Sin embargo, se reconoce que la experiencia de los radiólogos y el número anual de mamografías leídas influyen en la interpretación de la mamografía de cribado y por tanto en su efectividad y coste-efectividad. Debido a esto, muchos países han adoptado requisitos mínimos de volumen de lectura. En la mayoría de los países europeos, los programas siguen las guías europeas que estipulan que los radiólogos deben interpretar un mínimo de 5000 mamografías al año ⁶⁰. Por el contrario, en Australia ¹⁷⁰ o Canadá ¹⁷¹ los requisitos de volumen interpretativo son mucho más bajos con rangos que varían entre 480 a 2000 mamografías por año.

La experiencia de los radiólogos obtenida mediante una formación especializada o por los años de trabajo en unidades específicas de manejo de patología mamaria afecta también la variabilidad en la interpretación de las imágenes ¹⁷². Algunos estudios han evaluado el efecto de intervenciones educativas sobre la efectividad de la lectura mamográfica con resultados contradictorios y variables en cuanto a la sensibilidad y la tasa de detección del cáncer de mama ¹⁷³⁻¹⁷⁵.

Por último, la tasa de participación también puede afectar la efectividad de las estrategias de lectura en ciertas condiciones. Por ejemplo, una tasa de participación muy baja puede limitar la oportunidad de los lectores para ganar más experiencia. Por tanto, todos estos aspectos han de tenerse en cuenta para garantizar un correcto desempeño de las estrategias de lectura mejorando el perfil de efectividad y coste-efectividad.

6.4. LESIONES BENIGNAS Y PERSONALIZACIÓN EN EL CRIBADO

Las mujeres con lesiones benignas ¹¹⁰, densidad mamaria elevada, antecedentes familiares de cáncer de mama, o pruebas invasivas previas, entre otros, han sido identificadas por algunos autores como población susceptible para la personalización del cribado, aplicando intervalos de cribado más cortos, comenzando el cribado a edad más temprana ¹⁶³, o utilizando una prueba de cribado distinta a la mamografía ¹⁷⁶. De esta manera, sería posible mejorar los beneficios del cribado en este grupo de mujeres, a la vez que sería posible reducir los efectos adversos en el grupo de mujeres sin estas características.

La personalización en el cribado probablemente represente una adecuación en la utilización de los recursos. Algunas mujeres, las que más se benefician, recibirían un seguimiento más intenso y con pruebas más sofisticadas, pero la mayoría de la población probablemente se beneficiaría con un cribado más espaciado en el tiempo.

Particularmente en las mujeres con mamas densas, la mamografía tiene una menor sensibilidad y especificidad ¹⁶⁵. El tejido mamario denso se compone principalmente de estructuras ductales y tejido conectivo, mientras que el tejido mamario no denso es en su mayoría grasa ¹⁶⁵. En base a la distribución de los porcentajes de áreas densas y grasas se categoriza la densidad de las mamas. Se cuenta con diferentes sistemas de clasificación de la densidad mamaria que en todos los casos pasa

por una valoración inicial de la mamografía ¹⁷⁷. Una vez identificadas las mujeres con mamas densas, el uso de diferentes estrategias de detección podría mejorar la detección temprana del cáncer de mama.

Las tecnologías que más comúnmente se utilizan en mujeres con mamas densas son el ultrasonido de mano, el ultrasonido automatizado, la tomosíntesis y la resonancia magnética. De estas cuatro tecnologías, sólo la tomosíntesis podría ser utilizada en el cribado, tal como consta en las recientemente publicadas recomendaciones europeas para el cribado de cáncer de mama ⁵¹. Estas recomendaciones están formuladas siguiendo los criterios del sistema GRADE (Grading of Recommendations Assessment, Development and Evaluation) ¹⁷⁸ por lo que el balance entre riesgos y beneficios y el uso de recursos y costes son dos criterios que, entre otros, han sido evaluados de forma rigurosa y explícita.

En el caso de las lesiones benignas, la caracterización de las mismas por medio de un análisis detallado de los patrones mamográficos con los que se manifiestan ¹¹³ o los biomarcadores que expresan podría ser de utilidad para identificar aquellas con mayor potencial maligno. Otras variables de la historia hormonal y reproductiva de la mujer también podrían ser de interés en este grupo de mujeres. Se ha reportado que en mujeres con lesiones benignas el estado postmenopáusico se asocia con menor riesgo de cáncer, mientras que el uso de terapia hormonal estaría asociado con un mayor riesgo ¹⁷⁹.

Los resultados del estudio de casos y controles ¹⁸⁰ desarrollado en este trabajo de tesis señalan que las lesiones benignas con alta expresión de receptores de estrógeno (RE) o de progesterona (RP) tienen un mayor riesgo de desarrollar cáncer de mama posterior. Concretamente, si la expresión de RE es superior al 70% o la de RP es superior al 80% la posibilidad de desarrollar cáncer podría ser hasta dos veces superior comparado con tener una baja expresión de estos biomarcadores.

Al incluir tanto a mujeres con lesiones proliferativas y no proliferativas, se pudo evaluar cada subgrupo de forma individual. Particularmente para el grupo de mujeres con lesiones no proliferativas, el incremento de riesgo asociado a la expresión de RE se mantuvo presente ¹⁸⁰. Otros autores han reportado resultados discordantes pero todos en base a lesiones detectadas fuera del cribado. Por un lado, la expresión de RE en la hiperplasia sin atipia tendría un mayor riesgo de cáncer ^{119,181} y por otro lado la expresión de RE en el tejido mamario normal ¹¹⁸ o en las lesiones hiperplásicas con atipia ¹⁸² no aumentaría el riesgo de cáncer.

Definir estrategias de cribado basadas en la personalización es una tarea pendiente. Esta tesis doctoral apunta la evaluación de biomarcadores como un criterio más a tener en cuenta. Particularmente, la valoración de la expresión de los RE y RP podría beneficiar el manejo y seguimiento de las mujeres con lesiones benignas.

6.5. LIMITACIONES Y FORTALEZAS

LIMITACIONES

Las principales limitaciones de este trabajo de tesis son las propias de los diseños metodológicos de cada uno de los estudios que la componen. Algunas limitaciones son comunes. Entre ellas, están las fuentes de información en los análisis de coste-consecuencia y coste-efectividad, que son representativas del contexto en el cual el uso de recursos y costes se desarrolló. Aunque se basan en cohortes grandes de mujeres participantes en el cribado, los costes estimados para cada una de las estrategias evaluadas pueden no ser transferible a otros contextos.

Además no se analizaron los costes indirectos y el análisis es prioritariamente de interés desde la perspectiva del sistema de salud. Sin embargo, se informa sobre aumentos relativos en los costes y el análisis de sensibilidad incorpora escenarios similares a otros programas europeos por lo que los resultados pueden ser parcialmente aplicables a otros contextos.

De forma similar a los costes, los resultados de efectividad observados pueden ser aplicables a programas de cribado que cuenten con radiólogos muy experimentados en la lectura mamográfica. La generalización de los resultados de este trabajo de tesis a otros contextos con radiólogos menos entrenados podría ser limitada.

En el estudio de coste-efectividad no se evaluaron los años de vida ajustados a la calidad para cada estrategia. Los resultados pueden ser útiles para calcular los costes esperados por cada cáncer detectado. Por lo tanto, pueden tenerse en cuenta al planificar o evaluar programas de cribado y para futuros estudios de coste-utilidad.

La revisión sistemática intenta proporcionar una información completa, fiable y sintética sobre la efectividad y coste-efectividad de la doble lectura. Sin embargo, los pocos estudios disponibles no permiten que esta revisión clarifique las incertidumbres que existen sobre el tema. Si bien los

estudios incluidos mostraron un bajo riesgo de sesgo, la valoración global de la evidencia disponible no permite proponer respuestas concluyentes.

Finalmente, el número relativamente pequeño de lesiones benignas evaluadas en este trabajo de tesis también es una potencial limitación. Esto no permitió analizar el subgrupo de mujeres con lesiones atípicas ni diferenciar entre carcinomas *in situ* e invasivos. Por tanto, es necesario continuar evaluando la expresión de biomarcadores en estos grupos específicos.

FORTALEZAS

Los resultados de este trabajo de tesis son novedosos y relevantes para los programas de cribado, especialmente en este momento en que la mamografía digital está ampliamente extendida en Europa. Aportan información útil para ayudar a definir estrategias de cribado que mejoren el balance entre riesgos y beneficios y a su vez mejoren la eficiencia.

Los estudios realizados en este trabajo de tesis confirman la utilidad de las evaluaciones económicas en la toma de decisiones en salud. El estudio de casos y controles es útil para evaluar eventos poco frecuentes, como son las lesiones benignas, y la revisión sistemática para la gestión de la evidencia y conocimientos actuales. Si bien el diseño óptimo para evaluar la efectividad y coste-efectividad de diferentes estrategias de cribado es el ensayo clínico aleatorizado ¹⁸³, los recursos y el esfuerzo que se requieren para realizarlo son considerables y se han de sopesar con el valor de la evidencia que se espera obtener. Como alternativa se cuenta con otros diseños metodológicos como son los utilizados en esta tesis doctoral.

Particularmente, los modelos de decisión permiten evaluar escenarios hipotéticos a partir de ciertas asunciones. En este trabajo de tesis, el modelo que se plantea tanto en el estudio de coste-consecuencia como en el estudio de coste-efectividad es sencillo y preciso para la evaluación del coste y beneficio inmediatos que representa cada una de las estrategias de lectura evaluadas. Además, tal como se confirmó con la revisión sistemática, el estudio de coste-efectividad es una de las primeras evaluaciones económicas de la doble lectura en el contexto digital. Finalmente, en el estudio de casos y controles se dispone de información exhaustiva de variables clínicas y expresión de biomarcadores de las lesiones benignas que hasta aquel momento no habían sido evaluadas en el contexto del cribado.

En conjunto, los tres estudios originales se basan en información de cohortes importantes de mujeres cribadas y con periodos de seguimiento suficientemente largos para garantizar una adecuada evaluación de los resultados. El manejo de la información y la combinación estadística de los resultados a través del metanálisis en la revisión sistemática proporciona un resumen cuantitativo de los resultados que facilita su interpretación. Por tanto, los resultados de los cuatro estudios ofrecen evidencia novedosa explorando áreas del cribado en las que existe gran controversia.

6.6. IMPLICACIONES PARA LOS PROGRAMAS DE CRIBADO

De los cuatro estudios se puede concluir que resulta necesario y oportuno realizar una evaluación permanente de la efectividad y coste efectividad de las estrategias de lectura, sobre todo teniendo en cuenta que las nuevas tecnologías diagnósticas afectan sustancialmente los resultados del cribado.

Aunque se ha sugerido que para incrementar las tasas de detección de cáncer es inevitable asumir un incremento en los falsos positivos, los programas de cribado deberían investigar y preferir las estrategias que promueven una reducción en los falsos positivos sin afectar de forma significativa las tasas de detección. En este sentido, la lectura simple parece ser una estrategia más adecuada en el contexto de la mamografía digital.

Otro beneficio de la lectura simple es que podría implicar ahorros netos para el sistema sanitario. El coste promedio de un episodio de cribado varió poco entre la lectura doble con el consenso y el arbitraje (44,9 €) y la lectura simple (39,1 €). Esta pequeña variación, sin embargo, representó un ahorro de aproximadamente 0,4 millones de euros por 57.157 episodios de cribado. Teniendo en cuenta que miles de mujeres son examinadas cada dos años, el cambio a la lectura simple podría reducir en gran medida los costes de los programas de cribado.

La mejora en la adecuación de la asignación de los recursos permite invertir en nuevas tecnologías y estrategias más orientadas a la personalización del cribado. En el caso concreto de las mujeres con lesiones benignas, los biomarcadores de expresión de receptores de estrógenos y progesterona

mejoran el manejo clínico y la planificación de estrategias de cribado que les ofrezcan mayores beneficios.

A nivel europeo, la Iniciativa de la Comisión Europea para el control del cáncer de mama (ECIBC) ha iniciado un proyecto para desarrollar las nuevas Guías europeas para la detección y el diagnóstico del cáncer de mama. Se prevé que durante los próximos dos años, se desarrollen y publiquen aproximadamente 90 recomendaciones basadas en evidencia y entre ellas se tendrán en cuenta las estrategias de lectura mamográfica. Particularmente, los resultados de la revisión sistemática desarrollada en este trabajo de tesis podrían ser de interés para el desarrollo de esta recomendación.

6.7. IMPLICACIONES PARA LA INVESTIGACIÓN

Los resultados de los estudios de coste-consecuencia y coste-efectividad constituyen una pieza importante para planificar y desarrollar estudios de coste-utilidad en los que se valoren resultados a largo plazo, como es el cálculo de los años de vida ajustados por calidad. Es importante mencionar que idealmente este estudio debería desarrollarse mediante un ensayo clínico controlado, amplio, bien diseñado e incorporando la perspectiva de la sociedad.

Además, los resultados de este trabajo de tesis dan pie a cuestionar la doble lectura como posible factor asociado al sobrediagnóstico. Por ello, se necesitan más estudios prospectivos que permitan evaluar el efecto de las estrategias de lectura y las nuevas tecnologías sobre la reducción o el incremento del sobrediagnóstico.

El beneficio de la doble lectura podría quedar restringido al cribado incidente o a situaciones en las que exista mayor variabilidad. Queda pendiente evaluar el valor de la doble lectura en contextos de mamografía digital con radiólogos menos experimentados. Así también queda pendiente evaluar el efecto de la doble lectura sobre la detección de lesiones benignas.

Sin duda se ha avanzado mucho desde la época en que el cribado se realizaba con mamografía analógica y con la lectura única de una sola proyección. Se ha evolucionado tanto en las tecnologías de detección como en las estrategias de lectura. Los esfuerzos apuntan a utilizar tecnologías más sofisticadas y muchas veces más costosas, como la tomosíntesis o incluso la

biopsia líquida. El uso de estas tecnologías en aquellas mujeres que potencialmente se beneficiarían más no representaría propiamente un gasto sino una adecuada inversión de los recursos. Para garantizar esto, la evaluación de la efectividad y coste-efectividad debe ser una tarea permanente en los programas de cribado.

Los esfuerzos de la investigación futura deberían continuar en la dirección de personalizar las estrategias y tecnologías del cribado con el objetivo de garantizar el mejor balance de riesgo-beneficio y de coste-efectividad.




CONCLUSIONES

CONCLUSIONES

1. Los resultados de esta tesis doctoral muestran que, en el contexto de la mamografía digital, la doble lectura incrementa significativamente la frecuencia de falsos positivos y de forma más modesta la tasa de detección del cáncer de mama.
2. El beneficio que muestra la doble lectura al incrementar la sensibilidad del cribado se restringe a la detección de cánceres de comportamiento no invasivo o lesiones benignas identificadas como falsos positivos. Este aparente beneficio de la doble lectura quedaría diluido si se demuestra que contribuye a incrementar el sobrediagnóstico.
3. La doble lectura de la mamografía digital no es coste-efectiva cuando se aplica solo en el cribado prevalente y tampoco parece serlo en el cribado incidente.
4. Si bien no existe un consenso sobre la disponibilidad a pagar por cáncer detectado, los resultados observados permiten cuestionar la eficiencia de la doble lectura. En todo caso, resulta difícil justificar una disponibilidad a pagar significativamente mayor (>100%) por cada cáncer detectado de forma adicional con la doble lectura en comparación con el coste medio de cada cáncer detectado con la lectura simple.
5. Dado que no se cuenta con evidencia sobre la efectividad a largo plazo de la doble lectura en el contexto de la mamografía digital, es necesario que futuros estudios evalúen su efecto sobre la mortalidad, la calidad de vida y el sobrediagnóstico.
6. Las lesiones benignas con alta expresión de receptores de estrógeno (RE) o de progesterona (RP) tienen un mayor riesgo de desarrollar cáncer de mama posteriormente. Concretamente, si la expresión de RE es superior al 70% o la de RP es superior al 80% la posibilidad de desarrollar cáncer podría ser hasta dos veces superior comparado con tener una baja expresión de estos biomarcadores.

7. Definir estrategias de cribado basadas en el riesgo mejora el uso de recursos y la eficiencia del cribado. Este trabajo de tesis doctoral apunta la evaluación de biomarcadores como un criterio más a tener en cuenta en la personalización del cribado.



CONCLUSIONS

CONCLUSIONS

1. This doctoral thesis shows that in the context of digital mammography double reading significantly increases the frequency of false positive results, while the increase in the cancer detection rate is moderate.
2. Double reading has the benefit of increasing screening sensitivity. However, this benefit is restricted to detecting non-invasive tumours or benign lesions that are identified as false positive results. This apparent benefit will be dismissed, if double reading increases overdiagnosis.
3. Double reading of digital mammography is not cost-effective for prevalent screening, and it appears that neither for incident screening.
4. Although there is no consensus on the willingness to pay per one screen-detected cancer, double reading's efficiency is challenged at the light of the aforementioned results. It is difficult to justify a drastically higher willingness to pay (100% more expensive) per one additionally detected cancer with double reading compared to the average cost of each cancer detected with single reading.
5. It is necessary to further evaluate the effect of double reading on mortality, quality of life and overdiagnosis since there is not evidence on the long-term effectiveness of double reading in the context of digital mammography.
6. Benign lesions with high expression of estrogen receptors (ER) or progesterone receptors (PR) have a higher risk of subsequent breast cancer. Specifically, if the RE expression is higher than 70% or the RP expression is higher than 80% the risk of developing cancer is two times higher compared to having a low expression of these biomarkers.
7. Defining risk-based screening strategies improves the use of resources and the efficiency of screening. This doctoral thesis pointed out that is important to evaluate biomarkers as one additional factor to take into account for individualized screening.



REFERENCIAS

REFERENCIAS

1. International Agency for Research on Cancer. *GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012*. 2012. doi:10.3322/caac.20107/full.
2. DeSantis CE, Bray F, Ferlay J, Lortet-Tieulent J, Anderson BO, Jemal A. International Variation in Female Breast Cancer Incidence and Mortality Rates. *Cancer Epidemiol Biomarkers Prev*. 2015;24(10):1495-1506. doi:10.1158/1055-9965.EPI-15-0535.
3. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. *N Engl J Med*. 2016;375(15):1438-1447. doi:10.1056/NEJMoa1600249.
4. López-Abente G, Nuñez O, Pérez-Gómez B, Aragonés N, Pollán M. La Situación Del Cáncer En España: Informe 2015 (The Situation of Cancer in Spain: Report 2015). *SSRN Journal*. March 2015. doi:10.2139/ssrn.2863882.
5. Tsilidis KK, Papadimitriou N, Capothanassi D, et al. Burden of Cancer in a Large Consortium of Prospective Cohorts in Europe. *J Natl Cancer Inst*. 2016;108(10). doi:10.1093/jnci/djw127.
6. Galceran J, Amejjide A, Carulla M, et al. Cancer incidence in Spain, 2015. *Clin Transl Oncol*. 2017;19(7):799-825. doi:10.1007/s12094-016-1607-9.
7. Borrás J, Amejjide A, Vilardell L, Valls J. Evolución de la incidencia del cáncer en Cataluña, 1985-2002. *Medicina Clínica*. 2008;131:11-18. doi:10.1016/S0025-7753(08)76428-5.
8. Pollán M, Michelena MJ, Ardanaz E, et al. Breast cancer incidence in Spain before, during and after the implementation of screening programmes. *Ann Oncol*. 2010;21 Suppl 3:iii97-iii102. doi:10.1093/annonc/mdq095.
9. King M-C, Marks JH, Mandell JB, New York Breast Cancer Study Group. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science*. 2003;302(5645):643-646. doi:10.1126/science.1088759.
10. Lux MP, Bani MR, Fasching PA, Beckmann MW. Hereditary Breast and Ovarian Cancer. In: *The Molecular Basis of Human Cancer*. Vol 52. New York, NY: Humana Press, New York, NY; 2017:401-421. doi:10.1007/978-1-59745-458-2_26.
11. Wooster R, Weber BL. Breast and ovarian cancer. *N Engl J Med*. 2003;348(23):2339-2347. doi:10.1056/NEJMra012284.
12. Allred DC, Mohsin SK, Fuqua SA. Histological and biological evolution of human premalignant breast disease. *Endocr Relat Cancer*. 2001;8(1):47-61.
13. Hartmann LC, Sellers TA, Frost MH, et al. Benign breast disease and the risk of breast cancer. *N Engl J Med*. 2005;353(3):229-237. doi:10.1056/NEJMoa044383.
14. Dyrstad SW, Yan Y, Fowler AM, Colditz GA. Breast cancer risk associated with benign breast disease: systematic review and meta-analysis. *Breast Cancer Res Treat*. 2015;149(3):569-575. doi:10.1007/s10549-014-3254-6.
15. Bradley CJ, Given CW, Roberts C. Race, socioeconomic status, and breast cancer treatment and survival. *JNCI Journal of the National Cancer Institute*. 2002;94(7):490-496.
16. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. *The Lancet*. 2002;360(9328):187-195. doi:10.1016/S0140-6736(02)09454-0.
17. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol*. 2012;13(11):1141-1151. doi:10.1016/S1470-2045(12)70425-4.
18. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. *The Lancet*. 1997;350(9084):1047-1059.
19. Glass AG, Lacey JV, Carreon JD, Hoover RN. Breast cancer incidence, 1980-2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. *J Natl Cancer Inst*. 2007;99(15):1152-1161. doi:10.1093/jnci/djm059.

20. Robson M, Offit K. Clinical practice. Management of an inherited predisposition to breast cancer. *N Engl J Med*. 2007;357(2):154-162. doi:10.1056/NEJMcp071286.
21. Foulkes WD. Inherited susceptibility to common cancers. *N Engl J Med*. 2008;359(20):2143-2153. doi:10.1056/NEJMra0802968.
22. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *The Lancet*. 2001;358(9291):1389-1399. doi:10.1016/S0140-6736(01)06524-2.
23. National Cancer Institute (NCI). *SEER Stat Fact Sheets: Breast*. 2012.
24. Henderson TO, Amsterdam A, Bhatia S, et al. Systematic review: surveillance for breast cancer in women treated with chest radiation for childhood, adolescent, or young adult cancer. *Ann Intern Med*. 2010;152(7):444-55-W144-54. doi:10.7326/0003-4819-152-7-201004060-00009.
25. Clemons M, Loijens L, Goss P. Breast cancer risk following irradiation for Hodgkin's disease. *Cancer Treatment Reviews*. 2000;26(4):291-302.
26. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med*. 2007;356(3):227-236. doi:10.1056/NEJMoa062790.
27. Nelson HD, Zakher B, Cantor A, et al. Risk Factors for Breast Cancer for Women Aged 40 to 49 Years A Systematic Review and Meta-analysis. *Ann Intern Med*. 2012;156(9):635-648. doi:10.7326/0003-4819-156-9-201205010-00006.
28. Dupont WD, Page DL. Breast cancer risk associated with proliferative disease, age at first birth, and a family history of breast cancer. *Am J Epidemiol*. 1987;125(5):769-779.
29. Colditz GA, Willett WC, Hunter DJ, et al. Family history, age, and risk of breast cancer. Prospective data from the Nurses' Health Study. *JAMA*. 1993;270(3):338-343.
30. Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA. Family history and the risk of breast cancer: a systematic review and meta-analysis. *Int J Cancer*. 1997;71(5):800-809.
31. Hulka BS. Epidemiologic analysis of breast and gynecologic cancers. *Prog Clin Biol Res*. 1997;396:17-29.
32. Zhang Y, Kiel DP, Kreger BE, et al. Bone mass and the risk of breast cancer among postmenopausal women. *N Engl J Med*. 1997;336(9):611-617. doi:10.1056/NEJM199702273360903.
33. Qu X, Zhang X, Qin A, et al. Bone mineral density and risk of breast cancer in postmenopausal women. *Breast Cancer Res Treat*. 2013;138(1):261-271. doi:10.1007/s10549-013-2431-3.
34. Brinton LA, Schairer C, Hoover RN, Fraumeni JF. Menstrual factors and risk of breast cancer. *Cancer Invest*. 1988;6(3):245-254.
35. Harper S, Lynch J, Meersman SC, Breen N, Davis WW, Reichman MC. Trends in area-socioeconomic and race-ethnic disparities in breast cancer incidence, stage at diagnosis, screening, mortality, and survival among women ages 50 years and over (1987-2005). *Cancer Epidemiol Biomarkers Prev*. 2009;18(1):121-131. doi:10.1158/1055-9965.EPI-08-0679.
36. Peairs KS, Choi Y, Stewart RW, Sateia HF. Screening for breast cancer. *Seminars in Oncology*. 2017;44(1):60-72. doi:10.1053/j.seminoncol.2017.02.004.
37. Morimoto LM, White E, Chen Z, et al. Obesity, body size, and risk of postmenopausal breast cancer: the Women's Health Initiative (United States). *Cancer Causes Control*. 2002;13(8):741-751.
38. Boyd NF, Stone J, Vogt KN, Connelly BS, Martin LJ, Minkin S. Dietary fat and breast cancer risk revisited: a meta-analysis of the published literature. *British Journal of Cancer*. 2003;89(9):1672-1685. doi:10.1038/sj.bjc.6601314.
39. Nelson HD, Humphrey LL, Nygren P, Teutsch SM, Allan JD. Postmenopausal hormone replacement therapy: scientific review. *JAMA*. 2002;288(7):872-881.
40. Enger SM, Ross RK, Paganini-Hill A, Bernstein L. Breastfeeding experience and breast cancer risk among postmenopausal women. *Cancer Epidemiol Biomarkers Prev*. 1998;7(5):365-369.
41. Ellison RC, Zhang Y, McLennan CE, Rothman KJ. Exploring the relation of alcohol consumption to risk of breast cancer. *Am J Epidemiol*. 2001;154(8):740-747.
42. Tretli S. Height and weight in relation to breast cancer morbidity and mortality. A prospective study of 570,000 women in Norway. *Int J Cancer*. 1989;44(1):23-30.
43. Gaudet MM, Gapstur SM, Sun J, Diver WR, Hannan LM, Thun MJ. Active smoking and breast cancer risk: original cohort data and meta-analysis. *J Natl Cancer Inst*. 2013;105(8):515-525. doi:10.1093/jnci/djt023.

44. Gram IT, Park S-Y, Kolonel LN, et al. Smoking and Risk of Breast Cancer in a Racially/Ethnically Diverse Population of Mainly Women Who Do Not Drink Alcohol: The MEC Study. *Am J Epidemiol*. 2015;182(11):917-925. doi:10.1093/aje/kwv092.
45. Pizot C, Boniol M, Mullie P, et al. Physical activity, hormone replacement therapy and breast cancer risk: A meta-analysis of prospective studies. *Eur J Cancer*. 2016;52:138-154. doi:10.1016/j.ejca.2015.10.063.
46. Wu Y, Zhang D, Kang S. Physical activity and risk of breast cancer: a meta-analysis of prospective studies. *Breast Cancer Res Treat*. 2013;137(3):869-882. doi:10.1007/s10549-012-2396-7.
47. Wilson J, Jungner G. *Principles and Practice of Screening for Disease*. World Health Organization Public Health Papers, No. 34. Geneva; 1968.
48. Andermann A, Blancquaert I, Beauchamp S, Déry V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ*. 2008;86(4):317-319. doi:10.2471/BLT.07.050112.
49. Harris R, Sawaya GF, Moyer VA, Calonge N. Reconsidering the criteria for evaluating proposed screening programs: reflections from 4 current and former members of the U.S. Preventive services task force. *Epidemiol Rev*. 2011;33(1):20-35. doi:10.1093/epirev/mxr005.
50. Walter SD, Day NE. Estimation of the duration of a pre-clinical disease state using screening data. *Am J Epidemiol*. 1983;118(6):865-886.
51. **Recommendations on Breast Cancer Screening.** <http://ecibc.jrc.ec.europa.eu/recommendations/list/3>.
52. McCann J, Stockton D. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer*. 2002. doi:10.1186/bcr455.
53. Wald N, Chamberlain J, Hackshaw A. Report of the European Society of Mastology Breast Cancer Screening Evaluation Committee. In: Vol 79. 1993:371-379.
54. Carro-Alonso B, Sainz-Martínez J, Lamata-Hernández F, Fernández-Gómez JA, Villavieja-Atance JL. Rendimiento diagnóstico de la mamografía en pacientes con lesiones clínicamente ocultas. Análisis de 1005 lesiones. *Rev Med Univ Navarra*. 2006;50:21-25.
55. Giordano L, Karsa Von L, Tomatis M, et al. Mammographic screening programmes in Europe: organization, coverage and participation. *J Med Screen*. 2012;19 Suppl 1:72-82. doi:10.1258/jms.2012.012085.
56. Bonfill X, Marzo MM, Medina C, Roura P, Rue M. L'EFECTMTAT DEL CRIBATGE DEL CÁNCER DE MAMA EN EL NOSTRE ENTORN. *Gaceta Sanitaria*. 1992;6(30):128-142. doi:10.1016/S0213-9111(92)71104-2.
57. Castells X, Sala M, Ascunce N, Salas D, Zubizarreta R, Casamitjana M. Descripción del cribado del cáncer en España. Proyecto DESCRIC. September 2008:1-311.
58. Ascunce N, Delfrade J, Salas D, Zubizarreta R, Ederria M, en nombre de Red de Programas de Cribado de Cáncer. [Breast cancer screening: characteristics and results of the Spanish programs]. *Medicina Clinica*. 2013;141(1):13-23. doi:10.1016/j.medcli.2012.03.030.
59. Departament de Salut. Generalitat de Catalunya, (null) PDD. *Pla Contra El Càncer a Catalunya 2015-2020*. 2015. http://cancer.gencat.cat/web/.content/mery/documents/Pla-contra-el-cancer-a-catalunya_vs-14-abril-2015.pdf.
60. Perry N, Puthaar E. *European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis*. European Communities; 2006.
61. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. *JNCI Journal of the National Cancer Institute*. 2000;92(18):1490-1499.
62. Miller VA, Cousino M, Leek AC, Kodish ED. Hope and Persuasion by Physicians During Informed Consent. *J Clin Oncol*. September 2014;JCO.2014.55.2588. doi:10.1200/JCO.2014.55.2588.
63. Bjurstam N, Björnelid L, Warwick J, et al. The Gothenburg Breast Screening Trial. *Cancer*. 2003;97(10):2387-2396. doi:10.1002/cncr.11361.
64. Bjurstam NG, Björnelid LM, Duffy SW. Updated results of the Gothenburg Trial of Mammographic Screening. *Cancer*. 2016;122(12):1832-1835. doi:10.1002/cncr.29975.
65. Chu KC, Smart CR, Tarone RE. Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan clinical trial. *JNCI Journal of the National Cancer Institute*. 1988;80(14):1125-1132.
66. Habbema JD, van Oortmarssen GJ, van Putten DJ, Lubbe JT, van der Maas PJ. Age-specific

- reduction in breast cancer mortality by screening: an analysis of the results of the Health Insurance Plan of Greater New York study. *JNCI Journal of the National Cancer Institute*. 1986;77(2):317-320.
67. Shapiro S, Strax P, Venet L, Venet W. Proceedings: Changes in 5-year breast cancer mortality in a breast cancer screening program. *Proc Natl Cancer Conf*. 1972;7:663-678.
68. Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. *BMJ*. 1988;297(6654):943-948. doi:10.1136/bmj.297.6654.943.
69. Zackrisson S. Rate of over-diagnosis of breast cancer 15 years after end of Malmo mammographic screening trial: follow-up study. *BMJ*. 2006;332(7543):689-692. doi:10.1136/bmj.38764.572569.7C.
70. Frisell J, Lidbrink E, Hellström L, Rutqvist LE. Followup after 11 years--update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Res Treat*. 1997;45(3):263-270.
71. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *The Lancet*. 2002;359(9310):909-919. doi:10.1016/S0140-6736(02)08020-0.
72. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med*. 2002;137(5 Part 1):305-312.
73. Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ*. 2014;348:g366. doi:10.1136/bmj.g366.
74. Tabár L, Gad A, Holmberg L, Ljungquist U. Significant reduction in advanced breast cancer. Results of the first seven years of mammography screening in Kopparberg, Sweden. *Diagn Imaging Clin Med*. 1985;54(3-4):158-164.
75. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. *JNCI Journal of the National Cancer Institute*. 1982;69(2):349-355.
76. Fagerberg G, Baldetorp L, Gröntoft O, Lundström B, Månson JC, Nordenskjöld B. Effects of repeated mammographic screening on breast cancer stage distribution. Results from a randomised study of 92 934 women in a Swedish county. *Acta Radiol Oncol*. 1985;24(6):465-473.
77. Moss SM, Cuckle H, Evans A, et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet*. 2006;368(9552):2053-2060. doi:10.1016/S0140-6736(06)69834-6.
78. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev*. 2013;(6):CD001877. doi:10.1002/14651858.CD001877.pub5.
79. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. *British Journal of Cancer*. 2013;108(11):2205-2240. doi:10.1038/bjc.2013.177.
80. Canadian Task Force on Preventive Health Care, Tonelli M, Connor Gorber S, et al. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. *CMAJ*. 2011;183(17):1991-2001. doi:10.1503/cmaj.110334.
81. Siu AL, U.S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2016;164(4):279-296. doi:10.7326/M15-2886.
82. World Health Organization. *WHO Position Paper on Mammography Screening*. 2014.
83. James TA, Wade JE, Sprague BL. The impact of mammographic screening on the surgical management of breast cancer. *J Surg Oncol*. 2016;113(5):496-500. doi:10.1002/jso.24184.
84. Bond M, Pavey T, Welch K, et al. Systematic review of the psychological consequences of false-positive screening mammograms. *Health Technol Assess*. 2013;17(13):1-170-vi. doi:10.3310/hta17130.
85. Jørgensen KJ. Mammography screening Benefits, harms and informed choice. *Dan Med J*. 2013;60(4):B4614.
86. Musolino A, Michiara M, Conti GM, et al. Human epidermal growth factor receptor 2 status and interval breast cancer in a population-based cancer registry study. *Journal of Clinical Oncology*. 2012;30(19):2362-2368. doi:10.1200/JCO.2011.37.6434.
87. Domingo L, Salas D, Zubizarreta R, et al. Tumor phenotype and breast density in distinct categories of interval cancer: results of population-based mammography screening in Spain. *Breast Cancer Res*.

- 2014;16(1):R3. doi:10.1186/bcr3595.
88. Warner E. Clinical practice. Breast-cancer screening. *N Engl J Med*. 2011;365(11):1025-1032. doi:10.1056/NEJMcp1101540.
89. Roman R, Sala M, Salas D, et al. Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. *Ann Oncol*. 2012;23(1):104-111. doi:10.1093/annonc/mdr032.
90. Hofvind S, Thoresen S, Tretli S. The cumulative risk of a false-positive recall in the Norwegian Breast Cancer Screening Program. *Cancer*. 2004;101(7):1501-1507. doi:10.1002/cncr.20528.
91. Elmore JG, Reisch LM, Barton MB, et al. Efficacy of breast cancer screening in the community according to risk level. *J Natl Cancer Inst*. 2005;97(14):1035-1043. doi:10.1093/jnci/dji183.
92. Roman R, Sala M, La Vega De M, et al. Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening. *Breast Cancer Res Treat*. 2011;130(2):543-552. doi:10.1007/s10549-011-1581-4.
93. Pisano ED, Yaffe MJ. Digital mammography. *Radiology*. 2005;234(2):353-362. doi:10.1148/radiol.2342030897.
94. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med*. 2005;353(17):1773-1783. doi:10.1056/NEJMoa052911.
95. Karssemeijer N, Bluekens AM, Beijerinck D, et al. Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. *Radiology*. 2009;253(2):353-358. doi:10.1148/radiol.2532090225.
96. Vinnicombe S, Pinto Pereira SM, McCormack VA, Shiel S, Perry N, Santos Silva dos IM. Full-Field Digital versus Screen-Film Mammography: Comparison within the UK Breast Screening Program and Systematic Review of Published Data. *Radiology*. 2009;251(2):347-358. doi:10.1148/radiol.2512081235.
97. Domingo L, Romero A, Belvis F, 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):2020-2028. doi:10.1007/s00330-011-2143-1.
98. de Munck L, De Bock GH, Otter R, et al. Digital vs screen-film mammography in population-based breast cancer screening: performance indicators and tumour characteristics of screen-detected and interval cancers. *British Journal of Cancer*. 2016;115(5):517-524. doi:10.1038/bjc.2016.226.
99. Sala M, Domingo L, Macià F, Comas M, Burón A, Castells X. Does digital mammography suppose an advance in early diagnosis? Trends in performance indicators 6 years after digitalization. *Eur Radiol*. 2014;25(3):850-859. doi:10.1007/s00330-014-3431-3.
100. Sala M, Salas D, Belvis F, 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):388-395. doi:10.1148/radiol.10100874.
101. Comas M, Arrospide A, Mar J, et al. Budget Impact Analysis of Switching to Digital Mammography in a Population-Based Breast Cancer Screening Program: A Discrete Event Simulation Model. Gorlova OY, ed. *PLoS ONE*. 2014;9(5):e97459-10. doi:10.1371/journal.pone.0097459.
102. Klompenhouwer EG, Voogd AC, Heeten den GJ, et al. Blinded double reading yields a higher programme sensitivity than non-blinded double reading at digital screening mammography: A prospecting population based study in the south of The Netherlands. *European Journal of Cancer*. January 2015;1-9. doi:10.1016/j.ejca.2014.12.008.
103. Dinnes J, Moss S, Melia J, Blanks R, Song F, Kleijnen J. Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: findings of a systematic review. *The Breast*. 2001;10(6):455-463. doi:10.1054/brst.2001.0350.
104. Shaw CM, Flanagan FL, Fenlon HM, McNicholas MM. Consensus Review of Discordant Findings Maximizes Cancer Detection Rate in Double-Reader Screening Mammography: Irish National Breast Screening Program Experience. *Radiology*. 2009;250(2):354-362. doi:10.1148/radiol.2502080224.
105. Caumo F, Brunelli S, Tosi E, et al. On the role of arbitration of discordant double readings of screening mammography: experience from two Italian programmes. *Radiol med*. 2011;116(1):84-91. doi:10.1007/s11547-010-0606-0.
106. Ciatto S, Ambrogetti D, Rizzo G, et al. The role of arbitration of discordant reports at double reading of screening mammograms. *J Med Screen*. 2005;12(3):125-127. doi:10.1258/0969141054855337.

107. Klompenhouwer EG, Weber RJP, Voogd AC, et al. Arbitration of discrepant BI-RADS 0 recalls by a third reader at screening mammography lowers recall rate but not the cancer detection rate and sensitivity at blinded and non-blinded double reading. *The Breast*. 2015;24(5):601-607. doi:10.1016/j.breast.2015.06.004.
108. Duijm LEM, Groenewoud JH, Hendriks JHCL, de Koning HJ. Independent Double Reading of Screening Mammograms in the Netherlands: Effect of Arbitration Following Reader Disagreements. *Radiology*. 2004;231(2):564-570. doi:10.1148/radiol.2312030665.
109. Hofvind S, Geller BM, Rosenberg RD, Skaane P. Screening-detected Breast Cancers: Discordant Independent Double Reading in a Population-based Screening Program. *Radiology*. 2009;253(3):652-660. doi:10.1148/radiol.2533090210.
110. Hartmann LC, Degnim AC, Santen RJ, Dupont WD, Ghosh K. Atypical hyperplasia of the breast--risk assessment and management options. *N Engl J Med*. 2015;372(1):78-89. doi:10.1056/NEJMs1407164.
111. Said SM, Visscher DW, Nassar A, et al. Flat epithelial atypia and risk of breast cancer: A Mayo cohort study. *Cancer*. 2015;121(10):1548-1555. doi:10.1002/cncr.29243.
112. Manfrin E, Mariotto R, Remo A, et al. Benign breast lesions at risk of developing cancer--a challenging problem in breast cancer screening programs: five years' experience of the Breast Cancer Screening Program in Verona (1999-2004). *Cancer*. 2009;115(3):499-507. doi:10.1002/cncr.24038.
113. Castells X, Torá-Rocamora I, Posso M, et al. Risk of Breast Cancer in Women with False-Positive Results according to Mammographic Features. *Radiology*. 2016;280(2):379-386. doi:10.1148/radiol.2016151174.
114. Santen RJ, Mansel R. Benign breast disorders. *N Engl J Med*. 2005;353(3):275-285. doi:10.1056/NEJMra035692.
115. Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. *N Engl J Med*. 1985;312(3):146-151. doi:10.1056/NEJM198501173120303.
116. Page DL, Dupont WD, Rogers LW, Rados MS. Atypical hyperplastic lesions of the female breast. A long-term follow-up study. *Cancer*. 1985;55(11):2698-2708.
117. Sakorafas GH, Tsiotou AG. Ductal carcinoma in situ (DCIS) of the breast: evolving perspectives. *Cancer Treatment Reviews*. 2000;26(2):103-125. doi:10.1053/ctrv.1999.0149.
118. Huh SJ, Oh H, Peterson MA, et al. The Proliferative Activity of Mammary Epithelial Cells in Normal Tissue Predicts Breast Cancer Risk in Premenopausal Women. *Cancer Research*. 2016;76(7):1926-1934. doi:10.1158/0008-5472.CAN-15-1927.
119. Gobbi H, Dupont WD, Parl FF, et al. Breast cancer risk associated with estrogen receptor expression in epithelial hyperplasia lacking atypia and adjacent lobular units. *Int J Cancer*. 2005;113(5):857-859. doi:10.1002/ijc.20632.
120. Lee S, Mohsin SK, Mao S, Hilsenbeck SG, Medina D, Allred DC. Hormones, receptors, and growth in hyperplastic enlarged lobular units: early potential precursors of breast cancer. *Breast Cancer Res*. 2006;8(1):R6. doi:10.1186/bcr1367.
121. Nassar A, Hoskin TL, Stallings-Mann ML, et al. Ki-67 expression in sclerosing adenosis and adjacent normal breast terminal ductal lobular units: a nested case-control study from the Mayo Benign Breast Disease Cohort. *Breast Cancer Res Treat*. April 2015:1-9. doi:10.1007/s10549-015-3370-y.
122. Santisteban M, Reynolds C, Barr Fritcher EG, et al. Ki67: a time-varying biomarker of risk of breast cancer in atypical hyperplasia. *Breast Cancer Res Treat*. 2009;121(2):431-437. doi:10.1007/s10549-009-0534-7.
123. Perou CM, Sørbye T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747-752. doi:10.1038/35021093.
124. Petrelli F, Viale G, Cabiddu M, Barni S. Prognostic value of different cut-off levels of Ki-67 in breast cancer: a systematic review and meta-analysis of 64,196 patients. *Breast Cancer Res Treat*. 2015;153(3):477-491. doi:10.1007/s10549-015-3559-0.
125. Allred DC. Molecular Biomarkers of Risk in Premalignancy and Breast Cancer Prevention. *Cancer Prevention Research*. 2011;4(12):1947-1952. doi:10.1158/1940-6207.CAPR-11-0478.
126. Zagouri F, Sergentanis TN, Zografos GC. Precursors and preinvasive lesions of the breast: the role of molecular prognostic markers in the diagnostic and therapeutic dilemma. *World J Surg Onc*. 2007;5(1):57-11. doi:10.1186/1477-7819-5-57.
127. Barr FEG, Degnim AC, Hartmann LC, et al. Estrogen Receptor Expression in Atypical Hyperplasia:

- Lack of Association with Breast Cancer. *Cancer Prevention Research*. 2011;4(3):435-444. doi:10.1158/1940-6207.CAPR-10-0242.
128. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press, USA; 2015.
129. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal 2013*. London: National Institute for Health and Care Excellence (NICE); 2013.
130. López Bastida J, Oliva J, Antoñanzas F, et al. Propuesta de guía para la evaluación económica aplicada a las tecnologías sanitarias. *Gaceta Sanitaria*. 2010;24(2):154-170. doi:10.1016/j.gaceta.2009.07.011.
131. Kobelt G. *Health Economics: an Introduction to Economic Evaluation*. Third Edition. London: London: BSC Print Ltd Google Scholar; 2013.
132. Cleemput I, Neyt M, Thiry N, De Laet C, Leys M. *Threshold Values for Cost-Effectiveness in Health Care*. Belgian Health Care Knowledge Centre; 2008.
133. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Intervention. Version 5.1. 0*. The Cochrane Collaboration; 2011. doi:10.1002/14651858.CD004959.pub4/full.
134. PRISMA-P Group, Moher D, Shamseer L, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):e1000326-e1000329. doi:10.1186/2046-4053-4-1.
135. Joanna Briggs Institute. *Joanna Briggs Institute Reviewers' Manual: 2014 Edition/Supplement. 2014*. The Joanna Briggs Institute
136. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal 2013*. London: National Institute for Health and Care Excellence (NICE); 2013. <https://www.nice.org.uk/process/pmg9/chapter/foreword>.
137. Shemilt I, Mugford M, Byford S, et al. *Incorporating Economics Evidence*. Chichester, UK: John Wiley & Sons, Ltd; 2008:449-479. doi:10.1002/9780470712184.ch15.
138. Anderson R. Systematic reviews of economic evaluations: utility or futility? *Health Econ*. 2010;19(3):350-364. doi:10.1002/hec.1486.
139. O'Mahony JF, Paulden M. NICE's Selective Application of Differential Discounting_ Ambiguous, Inconsistent, and Unjustified. *Value in Health*. 2014;17(5):493-496. doi:10.1016/j.jval.2013.02.014.
140. Paulden M, O Mahony JF, Culyer AJ, McCabe C. Some Inconsistencies in NICE's Consideration of Social Values. *Pharmacoeconomics*. 2014;32(11):1043-1053. doi:10.1007/s40273-014-0204-4.
141. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536. doi:10.7326/0003-4819-155-8-201110180-00009.
142. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) Statement. *Value in Health*. 2013;16(2):e1-e5. doi:10.1016/j.jval.2013.02.010.
143. Conversion rates - Purchasing power parities (PPP) - OECD Data. <https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm>.
144. Nyaga VN, Arbyn M, Aerts M. Metaprop: a Stata command to perform meta-analysis of binomial data. *Archives of Public Health 2014 72:1*. 2014;72(1):39. doi:10.1186/2049-3258-72-39.
145. Castells X, Domingo L, Corominas JMA, et al. Breast cancer risk after diagnosis by screening mammography of nonproliferative or proliferative benign breast disease: a study from a population-based screening program. *Breast Cancer Res Treat*. 2014;149(1):237-244. doi:10.1007/s10549-014-3208-z.
146. Houssami N, Macaskill P, Bernardi D, et al. Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading â€“ Evidence to guide future screening strategies. *European Journal of Cancer*. 2014;50(10):1799-1807. doi:10.1016/j.ejca.2014.03.017.
147. Sato M, Kawai M, Nishino Y, Shibuya D, Ohuchi N, Ishibashi T. Cost-effectiveness analysis for breast cancer screening: double reading versus single + CAD reading. *Breast Cancer*. 2014;21(5):532-541. doi:10.1007/s12282-012-0423-5.
148. Posso M, Carles M, Rue M, Puig T, Bonfill X. Cost-Effectiveness of Double Reading versus Single Reading of Mammograms in a Breast Cancer Screening Programme. *PLoS ONE*. 2016;11(7):e0159806. doi:10.1371/journal.pone.0159806.
149. Weigel S, Heindel W, Heidrich J, Hense HW, Heidinger O. Digital mammography screening: sensitivity of the programme dependent on breast density. *Eur Radiol*. November 2016:1-8.

- doi:10.1007/s00330-016-4636-4.
150. Giordano L, Karsa Von L, Tomatis M, et al. Mammographic Screening Programmes in Europe: Organization and Participation. *J Med Screen.* 2012;19(1_suppl):72-82. doi:10.1258/jms.2012.012085.
 151. Paci E, Broeders M, Hofvind S, Puliti D, Duffy SW, EUROSCREEN Working Group. European breast cancer service screening outcomes: a first balance sheet of the benefits and harms. *Cancer Epidemiol Biomarkers Prev.* 2014;23(7):1159-1163. doi:10.1158/1055-9965.EPI-13-0320.
 152. Screening IUPOBC. The benefits and harms of breast cancer screening: an independent review. *The Lancet.* 2012;380(9855):1778-1786. doi:10.1016/S0140-6736(12)61611-0.
 153. Biller-Andorno N, Jüni P. Abolishing mammography screening programs? A view from the Swiss Medical Board. *N Engl J Med.* 2014;370(21):1965-1967. doi:10.1056/NEJMp1401875.
 154. Posso MC, Puig T, Quintana MJ, Solà-Roca J, Bonfill X. Double versus single reading of mammograms in a breast cancer screening programme: a cost-consequence analysis. *Eur Radiol.* 2016;26(9):3262-3271. doi:10.1007/s00330-015-4175-4.
 155. Hofvind S, Thoresen S, Tretli S. The cumulative risk of a false-positive recall in the Norwegian Breast Cancer Screening Program. *Cancer.* 2004;101(7):1501-1507. doi:10.1002/cncr.20528.
 156. Castells X. Cumulative false positive recall rate and association with participant related factors in a population based breast cancer screening programme. *Journal of Epidemiology & Community Health.* 2006;60(4):316-321. doi:10.1136/jech.2005.042119.
 157. Lynge E, Beau A-B, Christiansen P, et al. Overdiagnosis in breast cancer screening: The impact of study design and calculations. *Eur J Cancer.* 2017;80:26-29. doi:10.1016/j.ejca.2017.04.018.
 158. Puliti D, Duffy SW, Miccinesi G, et al. Overdiagnosis in Mammographic Screening for Breast Cancer in Europe: A Literature Review. *J Med Screen.* 2012;19(1_suppl):42-56. doi:10.1258/jms.2012.012082.
 159. Pardo-Hernández H, Alonso-Coello P. Patient-important outcomes in decision-making: a point of no return. *Journal of Clinical Epidemiology.* 2017;88:4-6. doi:10.1016/j.jclinepi.2017.05.014.
 160. Arrospide A, Rue M, van Ravesteyn NT, et al. Economic evaluation of the breast cancer screening programme in the Basque Country: retrospective cost-effectiveness and budget impact analysis. *BMC Cancer.* June 2016;1-9. doi:10.1186/s12885-016-2386-y.
 161. Sankatsing VRDV, Heijnsdijk EAM, van Luijt PA, van Ravesteyn NT, Fracheboud J, de Koning HJ. Cost-effectiveness of digital mammography screening before the age of 50 in The Netherlands. *Int J Cancer.* 2015;137(8):1990-1999. doi:10.1002/ijc.29572.
 162. Carles M, Vilaprinyo E, Cots F, et al. Cost-effectiveness of early detection of breast cancer in Catalonia (Spain). *BMC Cancer.* 2011;11(1):192. doi:10.1186/1471-2407-11-192.
 163. Vilaprinyo E, Forn C, Carles M, et al. Cost-Effectiveness and Harm-Benefit Analyses of Risk-Based Screening Strategies for Breast Cancer. Sapino A, ed. *PLoS ONE.* 2014;9(2):e86858-10. doi:10.1371/journal.pone.0086858.
 164. Evans DG, Astley S, Stavrinou P, et al. Improvement in risk prediction, early detection and prevention of breast cancer in the NHS Breast Screening Programme and family history clinics: a dual cohort study. *Programme Grants Appl Res.* 2016;4(11):1-210. doi:10.3310/pgfar04110.
 165. Lee CI, Chen LE, Elmore JG. Risk-based Breast Cancer Screening: Implications of Breast Density. *Med Clin North Am.* 2017;101(4):725-741. doi:10.1016/j.mcna.2017.03.005.
 166. Azavedo E, Zackrisson S, re IM, Arnlind MH. Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. *BMC Medical Imaging.* 2012;12(1):1-1. doi:10.1186/1471-2342-12-22.
 167. Gur D, Zuley ML, Anello MI, et al. Dose reduction in digital breast tomosynthesis (DBT) screening using synthetically reconstructed projection images: an observer performance study. *Academic Radiology.* 2012;19(2):166-171. doi:10.1016/j.acra.2011.10.003.
 168. Gilbert FJ, Tucker L, Gillan MG, et al. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme—a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. *Health Technol Assess.* 2015;19(4):i-xxv-1-136. doi:10.3310/hta19040.
 169. Houssami N, Bernardi D, Pellegrini M, et al. Breast cancer detection using single-reading of breast tomosynthesis (3D-mammography) compared to double-reading of 2D-mammography: Evidence

- from a population-based trial. *Cancer Epidemiology*. 2017;47:94-99. doi:10.1016/j.canep.2017.01.008.
170. Roder DM, Ward GH, Farshid G, Gill PG, BreastScreen Australia National Quality Management Committee. Influence of service characteristics on high priority performance indicators and standards in the BreastScreen Australia program. *Asian Pac J Cancer Prev*. 2014;15(14):5901-5908.
171. Coldman AJ, Major D, Doyle GP, et al. Organized breast screening programs in Canada: effect of radiologist reading volumes on outcomes. *Radiology*. 2006;238(3):809-815. doi:10.1148/radiol.2382041684.
172. Nodine CF, Kundel HL, Mello-Thoms C, et al. How experience and training influence mammography expertise. *Academic Radiology*. 1999;6(10):575-585.
173. Geller BM, Bogart A, Carney PA, et al. Educational interventions to improve screening mammography interpretation: a randomized controlled trial. *AJR Am J Roentgenol*. 2014;202(6):W586-W596. doi:10.2214/AJR.13.11147.
174. Carney PA, Abraham L, Cook A, et al. Impact of an educational intervention designed to reduce unnecessary recall during screening mammography. *Academic Radiology*. 2012;19(9):1114-1120. doi:10.1016/j.acra.2012.05.003.
175. Bassett LW, Hollatz-Brown AJ, Bastani R, Pearce JG, Hirji K, Chen L. Effects of a program to train radiologic technologists to identify abnormalities on mammograms. *Radiology*. 1995;194(1):189-192. doi:10.1148/radiology.194.1.7997550.
176. Lee CI, Cevik M, Alagoz O, et al. Comparative Effectiveness of Combined Digital Mammography and Tomosynthesis Screening for Women with Dense Breasts. *Radiology*. 2015;274(3):772-780. doi:10.1148/radiol.14141237.
177. Sickles EA, D'Orsi C, Bassett L, et al. *Acr Bi-Rads® Mammography*. in: *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. Reston, VA, American College of Radiology.; 2013.
178. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*. 2011;64(4):383-394. doi:10.1016/j.jclinepi.2010.04.026.
179. Arthur R, Wang Y, Ye K, et al. Association between lifestyle, menstrual/reproductive history, and histological factors and risk of breast cancer in women biopsied for benign breast disease. *Breast Cancer Res Treat*. June 2017. doi:10.1007/s10549-017-4347-9.
180. Posso M, Corominas JM, Serrano L, et al. Biomarkers expression in benign breast diseases and risk of subsequent breast cancer: a case-control study. *Cancer Med*. 2017;6(6):1482-1489. doi:10.1002/cam4.1080.
181. Shaaban AM, Sloane JP, West CR, Foster CS. Breast cancer risk in usual ductal hyperplasia is defined by estrogen receptor-alpha and Ki-67 expression. *Am J Pathol*. 2002;160(2):597-604. doi:10.1016/S0002-9440(10)64879-1.
182. Barr Fritcher EG, Degenim AC, Hartmann LC, et al. Estrogen receptor expression in atypical hyperplasia: lack of association with breast cancer. *Cancer Prev Res (Phila)*. 2011;4(3):435-444. doi:10.1158/1940-6207.CAPR-10-0242.
183. Bell KJL, Bossuyt P, Glasziou P, Irwig L. Assessment of changes to screening programmes: why randomisation is important. *BMJ*. 2015;350(mar30 14):h1566-h1566. doi:10.1136/bmj.h1566.



ANEXOS

ABREVIATURAS

AVAC	Años de vida ajustados por calidad
AVG	Años de vida ganados
BI-RADS	American College of Radiology Breast Imaging Reporting and Data System
ECA	Ensayo clínico aleatorizado
FI	Factor de impacto
GRADE	Grading of Recommendations Assessment, Development and Evaluation
IC	Intervalo de confianza
MeSH	Medical Subject Headings
Q1	Primer cuartil
Q2	Segundo cuartil
RCEI	Relación de coste-efectividad incremental
RS	Revisión sistemática
VPP	Valor predictivo positivo

OTRAS PUBLICACIONES RELACIONADAS CON EL TEMA DE LA TESIS

ARTICULO 1: Castells X, Torá-Rocamora I, Posso M, Román M, Vernet-Tomas M, Rodríguez-Arana A, et al. Risk of Breast Cancer in Women with False-Positive Results according to Mammographic Features. *Radiology*. 2016; 280 (2): 379-86. FI 7,896.

ARTICULO 2: Posso M, Puig T, Bonfill X. [Balancing risks and benefits of mammography screening for breast cancer: would you support its recommendation in Peruvian women?]. *Rev Peru Med Exp Salud Pública*. 2015; 32 (1): 117-28.

Risk of Breast Cancer in Women with False-Positive Results according to Mammographic Features¹

Xavier Castells, PhD
 Isabel Torá-Rocamora, PhD
 Margarita Posso, MPH
 Marta Román, PhD
 María Vernet-Tomas, PhD
 Ana Rodríguez-Arana, MD
 Laia Domingo, PhD
 Carmen Vidal, MPH
 Marisa Baré, PhD
 Joana Ferrer, MD
 María Jesús Quintana, PhD
 Mar Sánchez, MD
 Carmen Natal, MPH
 Josep A. Espinàs, PhD
 Francina Saladié, MD
 María Sala, PhD
 For the BELE Study Group

¹ From the Departments of Epidemiology and Evaluation (X.C., I.T., M.P., L.D., M. Sala), Obstetrics and Gynecology (M.V.), and Image Diagnostics (A.R.), IMIM (Hospital del Mar Medical Research Institute), Pg. Marítim 25-29, 08003 Barcelona, Catalonia, Spain; Research Network on Health Services in Chronic Diseases (REDISSEC), Barcelona, Spain (X.C., I.T. L.D., M.B., M. Sala); Cancer Registry of Norway, Oslo, Norway (M.R.); Cancer Prevention and Monitoring Program, Catalan Institute of Oncology, Barcelona, Catalonia, Spain (C.V.); Clinical Epidemiology and Cancer Screening, Parc Taulí University Hospital, Sabadell, Catalonia, Spain (M.B.); Radiology Department, Hospital de Santa Caterina, Girona, Catalonia, Spain (J.F.); Epidemiology Department, Hospital de la Santa Creu i Sant Pau, Barcelona, Catalonia, Spain (M.J.Q.); Direction General of Public Health, Department of Health, Government of Cantabria, Santander, Cantabria, Spain (M. Sánchez); Principality of Asturias Breast Cancer Screening Program, Principality of Asturias, Oviedo, Spain (C.N.); Catalan Cancer Plan, Department of Health, Catalonia, Spain (J.A.E.); Spanish Cancer Research Network (RTICC), Carlos III Health Institute, Madrid, Spain (J.A.E.); and Breast Cancer Screening Program of Tarragona, The Foundation League for the Research and Prevention of Cancer, Tarragona, Catalonia, Spain (F.S.). Received May 26, 2015; revision requested July 18; revision received October 13; accepted October 30; final version accepted November 25. Supported by the Instituto de Salud Carlos III FEDER (grant P111/01296). Address correspondence to X.C. (e-mail: xcastells@hospitaldelmar.cat).

The members of the BELE Study Group are listed in Appendix E1 (online).

© RSNA, 2016

Purpose:

To assess the risk of breast cancer in women with false-positive screening results according to radiologic classification of mammographic features.

Materials and Methods:

Review board approval was obtained, with waiver of informed consent. This retrospective cohort study included 521 200 women aged 50–69 years who underwent screening as part of the Spanish Breast Cancer Screening Program between 1994 and 2010 and who were observed until December 2012. Cox proportional hazards regression analysis was used to estimate the age-adjusted hazard ratio (HR) of breast cancer and the 95% confidence interval (CI) in women with false-positive mammograms as compared with women with negative mammograms. Separate models were adjusted for screen-detected and interval cancers and for screen-film and digital mammography. Time without a breast cancer diagnosis was plotted by using Kaplan-Meier curves.

Results:

When compared with women with negative mammograms, the age-adjusted HR of cancer in women with false-positive results was 1.84 (95% CI: 1.73, 1.95; $P < .001$). The risk was higher in women who had calcifications, whether they were (HR, 2.73; 95% CI: 2.28, 3.28; $P < .001$) or were not (HR, 2.24; 95% CI: 2.02, 2.48; $P < .001$) associated with masses. Women in whom mammographic features showed changes in subsequent false-positive results were those who had the highest risk (HR, 9.13; 95% CI: 8.28, 10.07; $P < .001$).

Conclusion:

Women with false-positive results had an increased risk of breast cancer, particularly women who had calcifications at mammography. Women who had more than one examination with false-positive findings and in whom the mammographic features changed over time had a highly increased risk of breast cancer. Previous mammographic features might yield useful information for further risk-prediction models and personalized follow-up screening protocols.

© RSNA, 2016

Online supplemental material is available for this article.

False-positive results are one of the major concerns in breast cancer screening. False-positive results require additional assessment, including imaging or invasive procedures such as surgical biopsy, to rule out malignancy and can lead to patient anxiety (1,2). Furthermore, it has been estimated that the cumulative risk of false-positive findings in women aged 50–69 years after 10 screening mammography rounds is 20%–32% in Europe (3–5) and 49%–63% in the United States (6,7).

Previous studies have shown that having had a false-positive result is associated with an increased risk that breast cancer will be detected in subsequent mammographic examinations (8,9). A more recent study stated that this risk appears to be higher in women who underwent fine-needle aspiration cytology or biopsy rather than imaging procedures alone (10).

Calcifications, asymmetric density, and distortion are the most frequent

mammographic features associated with breast cancer at the time of diagnosis (11). Moreover, these features are also described in the baseline mammography screening of interval cancers (12–14). On the contrary, finding a mass appears to be frequently described in mammography screening and associated with nonmalignant lesions (11). Nevertheless, to our knowledge, the risk of breast cancer associated with the mammographic features of false-positive results has not been evaluated. The aim of this study was to assess the risk of breast cancer in women with false-positive results at screening according to the radiologic classification of the mammographic features.

Materials and Methods

The study protocol was approved by the institutional review boards at all participating institutions. Informed consent was not required, since the data were retrospectively collected and anonymized.

Setting and Study Population

The retrospective cohort study included 521 200 women aged 50–69 years who underwent screening at least once between 1994 and 2010 in a population-based breast cancer screening program in Spain. Follow-up was conducted until December 2012 and included only those women who had at least 2 years of follow-up data to enable identification of interval cancers. Women in whom breast cancer was detected the first time they underwent screening in this screening program were excluded from analysis.

Screening Procedures

In accordance with the recommendations of the European Guidelines

for Quality Assurance in Breast Cancer Screening and Diagnosis (15), the Spanish Breast Cancer Screening Program provides free biennial screening to women aged 50–69 years. The standard procedure for radiologic performance was double-view imaging (mediolateral-oblique and craniocaudal images were acquired) and double reading with consensus or arbitration. Mammograms were read by trained radiologists who interpreted at least 1500 screening mammograms per year.

The Breast Imaging Reporting and Data System (BI-RADS) scale was used to classify mammograms. Women with a BI-RADS score of 3, 4, 5, or 0 were recalled for further assessment within a maximum of 2 months after screening mammography. The results of these further assessments to confirm the presence or absence of malignancy were obtained within 1 month after the most recent diagnostic test. Further assessments may include imaging procedures (mammography, ultrasonography [US], or magnetic resonance [MR] imaging), invasive procedures (fine-needle aspiration, core-needle biopsy, surgical biopsy, vacuum-assisted biopsy, or percutaneous excisional biopsy), or a combination thereof. A false-positive result was defined as any screening

Advances in Knowledge

- The excess risk of breast cancer in women with false-positive results was higher in women who had calcifications, regardless of whether calcifications were (age-adjusted hazard ratio [HR], 2.73; 95% confidence interval [CI]: 2.28, 3.28; $P < .001$) or were not (HR, 2.24; 95% CI: 2.02, 2.48; $P < .001$) associated with a mass, as compared with a reference population with negative mammograms.
- Women with two or more false-positive results in which the mammographic features changed over time had a large increase in risk of developing breast cancer (HR, 9.13; 95% CI: 8.28, 10.07; $P < .001$).
- Our data suggest that the risk of breast cancer in women with false-positive results was similar between screen-detected (HR, 1.77; 95% CI: 1.65, 1.90) and interval (HR, 2.01; 95% CI: 1.81, 2.24) cancers.

Implication for Patient Care

- The mammographic features in women with false-positive results might be useful for further risk prediction models and when planning follow-up protocols for these women.

Published online before print

10.1148/radiol.2016151174 Content code: BR

Radiology 2016; 000:1–8

Abbreviations:

BI-RADS = Breast Imaging Reporting and Data System

CI = confidence interval

HR = hazard ratio

Author contributions:

Guarantors of integrity of entire study, X.C.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, X.C., I.T., M.P., M.V., A.R., L.D., C.V., J.F., M.J.Q., M. Sánchez, C.N., J.A.E., F.S., M. Sala; clinical studies, X.C., I.T., M.V., A.R., C.V., M.B., J.F., M.J.Q., M. Sánchez, J.A.E., F.S.; statistical analysis, X.C., I.T., M.R., M.V., A.R., C.V., M.J.Q., M. Sánchez, F.S., M. Sala; and manuscript editing, all authors

Conflicts of interest are listed at the end of this article.

mammogram with a BI-RADS score of 3, 4, 5, or 0, which thereby required further diagnostic tests (results in the subsequent 3 months) and for which a diagnosis of invasive breast cancer or ductal carcinoma in situ was not made in the immediate recall assessment. Negative screening mammograms were defined as those without recommendation for further assessment. In accordance with the European Guidelines (15), in our screening program, women with negative mammograms or false-positive results were invited to continue with mammography screening every 2 years.

Mammographic Features

The mammographic features in the false-positive group were classified as mass (circumscribed or not), distortion, asymmetric density, and calcifications with or without associated mass. Other multiple features were classified as calcifications with either distortion or asymmetric density, mass with distortion, and asymmetric density with distortion.

We also analyzed a subgroup of women who had two or more false-positive examinations and in whom the mammographic features changed over time. Mammographic features were considered to have changed when a woman with a false-positive result at prior screening had another false-positive result at any subsequent screening but the mammographic features changed over time and were different between examinations. We did not consider changes in appearance, nor did we consider changes in the location of a similarly described finding. For instance, we did not consider a woman who had a mass as her first false-positive result and calcifications as any subsequent false-positive result, regardless of the location of mammographic features.

Breast Cancer Cases

All screen-detected cancers are routinely recorded in the screening program database. Interval cancers (primary breast cancers diagnosed after a negative screening test and before the

next screening invitation) were identified by merging data from participants who underwent screening with data from population-based cancer registries, the regional minimum basic data set, and hospital-based cancer registries. Both invasive carcinoma and in situ carcinoma were histopathologically confirmed in each screening setting. We assessed the location of breast cancer with respect to the side of the false-positive result. Then, two possible situations were considered: cancer was deemed ipsilateral when the false-positive result and the cancer were in the same breast and contralateral when the false-positive result and the cancer were in different breasts. We also compared the proportion of ductal carcinoma in situ diagnosed in the negative group versus that in the false-positive group.

Statistical Analyses

We compared the time at risk in women with negative mammograms with the time at risk in women with false-positive results. Women in the negative group had only negative mammograms during the study period, while women in the false-positive group had at least one false-positive result. The index mammogram was the first negative mammogram in the negative group and the first false-positive result in the false-positive group. The time at risk in both groups was estimated from the time of the index mammogram until the occurrence of a breast cancer diagnosis, the end of follow-up, or censoring, whichever came first. Time at risk was measured as median time, and the corresponding 25th and 75th percentiles are also listed.

To compare the risk of breast cancer among women with false-positive results and those without false-positive results, we estimated the age-adjusted hazard ratios (HRs) and 95% confidence intervals (95% CIs) of breast cancers using Cox proportional hazards regression analysis. We analyzed the risk of breast cancer separately for screen-detected and interval cancers and according to the type of mammography technique used (screen-film vs

digital). We estimated the risk of breast cancer among women who showed a change in mammographic features. The proportionality of hazards assumption was considered to be justified after examination of the Schoenfeld residuals (16). We tested the time without breast cancer diagnosis in women with false-positive findings according to the mammographic features by means of Kaplan-Meier curves and log-rank test. Additionally, we analyzed the median time between the screen-detected cancer diagnosis and previous negative mammography.

Statistical tests were two sided, and a *P* value of less than .05 indicated a significant difference. Analyses were performed by using Stata (version 11; Stata, College Station, Tex) and R (version 3.1.2; R Foundation, Vienna, Austria) statistical packages.

Results

We analyzed 521 200 women who underwent 1 626 873 screening mammographic examinations. Breast cancer was more frequently detected at screening ($n = 4608$ [71.5%]) than during the interval between screening rounds ($n = 1839$ [28.5%]). Among the 445 125 women in the negative group, 353 722 (79.5%) had screen-film mammograms and 91 403 (20.5%) had digital mammograms. Among the 76 075 women in the false-positive group, 55 902 (71.9%) had screen-film mammograms and 21 878 (28.1%) had digital mammograms ($P = .410$). Women in the negative group had a median follow-up time of 72 months (25th–75th percentile range, 45–113 months) and women in the false-positive group had a median follow-up time of 68 months (25th–75th percentile range, 24–100 months) (Table 1).

In women with false-positive results, 52 334 (65.5%) underwent mammography in additional views, 35 317 (44.2%) underwent US, and 324 (0.4%) underwent MR imaging. In regard to invasive tests, 6757 (8.4%) women underwent fine-needle aspiration; 3582 (4.5%), core-needle biopsy; 861 (1.1%), surgical biopsy; 576 (0.7%), vacuum-assisted

Table 1

Characteristics of Women Included in the Study

Characteristic	Total	Negative Mammograms	Mammographic Features of False-Positive Results
No. of women who underwent screening	521 200	445 125	76 075
No. of screening mammograms	1 626 873	1 345 431	281 442
No. of breast cancers after a false-positive mammographic finding	6 447	5 004	1 443
No. of screen-detected cancers	4 608 (71.5)	3 609 (72.1)	999 (69.2)
No. of interval cancers	1 839 (28.5)	1 395 (27.9)	444 (30.8)
Type of mammography of the mammogram index			
Screen film	409 624 (78.6)	353 722 (79.5)	55 902 (71.9)
Digital	111 576 (21.4)	91 403 (20.5)	21 878 (28.1)
Women's age of the mammogram index			
50–54 y	274 759 (52.7)	238 906 (53.7)	35 853 (47.1)
55–59 y	114 900 (22.0)	96 966 (21.8)	17 934 (23.6)
60–64 y	97 664 (18.7)	82 705 (18.6)	14 959 (19.7)
65–69 y	33 877 (6.5)	26 548 (6.0)	7 329 (9.6)
Time at risk (mo)*	72 (44–111)	72 (45–113)	68 (24–100)

Note.—Unless otherwise indicated, data in parentheses are percentages.

* Data are median time at risk, and data in parentheses are the 25th–75th percentile range.

biopsy; and 144 (0.2%), percutaneous excisional biopsy.

Risk of Breast Cancer

A higher breast cancer risk was observed in women with false-positive results than in women with negative mammograms (HR, 1.84; 95% CI: 1.73, 1.95; $P < .001$). The most prevalent mammographic features were mass ($n = 39\,501$ [51.9%]), calcifications ($n = 15\,379$ [20.2%]), and asymmetric density ($n = 13\,156$ [17.3%]). Multiple mammographic features were found in 5108 (6.7%) of 76 075 mammograms in the false-positive group, and calcifications associated with mass were the most prevalent ($n = 4000$ [5.3%]). All subtypes of mammographic features had a significant risk of breast cancer. The highest risk was found among women who had calcifications associated with a mass (HR, 2.73; 95% CI: 2.28, 3.28; $P < .001$) or not associated with a mass (HR, 2.24; 95% CI: 2.02, 2.48; $P < .001$). The lowest breast cancer risk was found for mass (HR, 1.59; 95% CI: 1.47, 1.73; $P < .001$) and

asymmetric density (HR, 1.58; 95% CI: 1.36, 1.83; $P < .001$) (Table 2).

The proportion of ductal carcinoma in situ in the entire cohort was 13.5% ($n = 778$), with proportions of 15.7% ($n = 207$) and 12.9% ($n = 571$) in the false-positive and negative groups, respectively ($P = .601$). The location of breast cancer with respect to the location of a false-positive result was not available in 271 (18.8%) cases. In 1172 women with available information, 851 (72.6%) tumors were ipsilateral to the location of false-positive results.

Women with false-positive results had a higher risk of screen-detected cancers (HR, 1.77; 95% CI: 1.65, 1.90; $P < .001$) and interval cancers (HR, 2.01; 95% CI: 1.81, 2.24; $P < .001$) than did women with negative mammograms. The highest risk of screen-detected cancer was observed in women with calcifications (HR, 2.32; 95% CI: 2.06, 2.61; $P < .001$), and the highest risk of interval cancer was found in women with calcifications associated with mass (HR, 3.93; 95% CI: 2.95, 5.24; $P < .001$) (Table 2). The risk of

breast cancer in the group of women who underwent screen-film mammography (HR, 1.86; 95% CI: 1.74, 1.98; $P < .001$) was similar to the risk in the group of women who underwent digital mammography (HR, 1.72; 95% CI: 1.47, 2.01; $P < .001$) (Table 3).

The Kaplan-Meier curves and log-rank test showed differences in time without breast cancer diagnosis among the subtypes of mammographic features (Figure). The lowest median time was observed in asymmetric density (median time, 35 months; 25th–75th percentile range, 23–51 months), followed by distortion (median time, 48 months; 25th–75th percentile range, 24–84 months), calcifications (median time, 50 months; 25th–75th percentile range, 31–78 months), mass (median time, 54 months; 25th–75th percentile range, 25–93 months), calcifications associated with mass (median time, 61 months; 25th–75th percentile range, 36–92 months), and other multiple features (median time, 61 months; 25th–75th percentile range, 25–73 months).

Risk of Breast Cancer in Women in Whom the Mammographic Features Changed over Time

A total of 3292 (4.1%) of women with false-positive results had a change in their mammographic features on subsequent false-positive mammograms. These women had a high increase in the risk of breast cancer (HR, 9.13; 95% CI: 8.28, 10.07; $P < .001$) (Table 4). Cancers detected in women with changes in mammographic features were 30.2% (436/1443) of all cancers detected in women with false-positive results. Most cancers detected in women with changes in mammographic features were screen-detected cancers (94.5%); however, in our overall study population, the screen-cancer detection rate was 71.5% ($P = .074$). The median time between the screen-cancer diagnosis and previous mammography was 1 year in this subgroup of women, whereas in the rest of women with or without previous false-positive examinations, the same median time was 2 years.

Table 2
Age-adjusted Risk of Breast Cancer in Women with False-Positive Results according to the Mammographic Features and by Screen-detected and Interval Cancers

Finding	Overall Breast Cancer			Screen-detected Breast Cancer			Interval Breast Cancer				
	No. of Mammograms	Time at Risk (mo)*	No. of Breast Cancers	HR†	PValue	No. of Breast Cancers	HR†	PValue	No. of Breast Cancers	HR†	PValue
Negative mammograms	445 (125)	72 (45–113)	5004	1.00 (Referent)	...	3609	1.00 (Referent)	...	1395	1.00 (Referent)	...
Mammographic features of false-positive results‡	76 075 (100)	68 (24–100)	1443	1.84 (1.73–1.95)	<.001	999	1.77 (1.65–1.90)	<.001	444	2.01 (1.81–2.24)	<.001
Mass§	39 501 (51.9)	70 (24–107)	663	1.59 (1.47–1.73)	<.001	541	1.51 (1.37–1.66)	<.001	212	1.82 (1.57–2.09)	<.001
Calcifications§	15 379 (20.2)	72 (44–108)	388	2.24 (2.02–2.48)	<.001	290	2.32 (2.06–2.61)	<.001	98	2.03 (1.66–2.49)	<.001
Asymmetric density§	13 156 (17.3)	49 (24–92)	183	1.58 (1.36–1.83)	<.001	119	1.44 (1.20–1.73)	<.001	64	1.92 (1.49–2.47)	<.001
Distortion§	2931 (3.9)	48 (24–82)	55	2.09 (1.61–2.73)	<.001	42	2.24 (1.64–3.04)	<.001	13	1.73 (1.00–2.99)	<.05
Calcification associated with mass§	4000 (5.3)	70 (24–109)	120	2.73 (2.28–3.28)	<.001	72	2.27 (1.80–2.87)	<.001	48	3.93 (2.95–5.24)	<.001
Other multiple features§	1108 (1.4)	61 (24–96)	34	2.21 (2.29–4.50)	<.001	25	3.29 (2.22–4.87)	<.001	9	3.02 (1.57–5.82)	<.001
Calcification with asymmetric density	410 (0.5)	68 (24–107)	8	1.94 (0.97–3.87)	.062	8	2.69 (1.34–5.38)	<.01	0
Mass with distortion	346 (0.5)	48 (24–77)	12	4.08 (2.31–7.18)	<.001	8	3.81 (1.90–7.62)	<.001	4	4.75 (1.78–12.67)	<.01
Calcification with distortion	265 (0.3)	72 (24–96)	12	4.20 (2.38–7.40)	<.001	7	3.38 (1.61–7.10)	<.01	5	6.34 (2.63–15.25)	<.001
Asymmetric density with distortion	87 (0.1)	49 (24–72)	2	3.09 (0.77–12.36)	.111	2	4.40 (1.10–17.60)	<.05	0

Note.—Unless otherwise indicated, data in parentheses are percentages.

* Data are medians, and data in parentheses are 25th and 75th percentile range.

† Data in parentheses are 95% CIs.

‡ Schoenfeld residuals test for model with all mammographic features of false-positive results: $\chi^2 = 0.442$, $P = .517$ for overall breast cancers; $\chi^2 = 0.27$, $P = .602$ for screen-detected breast cancers; and $\chi^2 = 0.40$, $P = .528$ for interval breast cancers.

§ Schoenfeld residuals global test for model with subtypes of mammographic features of false-positive results: $\chi^2 = 1.39$, $P = .284$ for overall breast cancers; $\chi^2 = 0.33$, $P = .603$ for screen-detected breast cancers; and $\chi^2 = 2.00$, $P = .158$ for interval breast cancers.

Table 3

Age-adjusted Risk of Breast Cancer in Women with False-Positive Results according to the Type of Mammography

Type of Mammography and Finding	Total	No. of Breast Cancers	HR*	P Value
Screen film	409 624			
Negative mammograms	353 722	4324	1.00 (Referent)	...
Mammographic features of false-positive results [†]	55 902 (100)	1236	1.86 (1.74, 1.98)	<.001
Mass [‡]	28 731 (51.4)	568	1.63 (1.50, 1.78)	<.001
Calcification [†]	12 540 (22.4)	335	2.16 (1.94, 2.42)	<.001
Asymmetric density [†]	9834 (17.6)	163	1.68 (1.44, 1.97)	<.001
Distortion [†]	1346 (2.4)	36	2.11 (1.52, 2.93)	<.001
Calcification associated with mass [‡]	2730 (4.9)	100	2.73 (2.24, 3.33)	<.001
Other multiple features [‡]	721 (1.3)	26	3.12 (2.12, 4.59)	<.001
Calcification associated with asymmetric density	298 (0.5)	6	1.73 (0.78, 3.85)	.179
Mass associated with distortion	205 (0.4)	10	4.55 (2.45, 8.46)	<.001
Calcification associated with distortion	187 (0.3)	9	3.83 (1.99, 7.37)	<.001
Asymmetric density associated with distortion	31 (0.1)	1	3.19 (0.45, 22.68)	.246
Digital (n = 111,576)				
Negative mammograms	91 403	680	1.00 (Referent)	...
Mammographic features of false-positive results [†]	21 878 (100)	207	1.72 (1.47, 2.01)	<.001
Mass [‡]	10 770 (49.2)	87	1.35 (1.08, 1.70)	<.05
Calcification [†]	2839 (13.0)	53	2.96 (2.23, 3.91)	<.001
Asymmetric density [†]	3322 (15.2)	20	1.05 (0.67, 1.64)	.827
Distortion [†]	1585 (7.2)	19	1.98 (1.26, 3.13)	<.01
Calcification associated with mass [‡]	1270 (5.8)	20	2.69 (1.72, 4.19)	<.001
Other multiple features [‡]	387 (1.8)	8	3.42 (1.70, 6.87)	<.01
Calcification associated with asymmetric density	112 (0.5)	2	2.95 (1.74, 11.84)	.126
Mass associated with distortion	141 (0.6)	2	2.59 (0.65, 10.37)	.179
Calcification associated with distortion	78 (0.4)	3	5.68 (1.83, 17.67)	<.01
Asymmetric density associated with distortion	56 (0.3)	1	2.77 (0.39, 19.69)	.309

Note.—Unless otherwise indicated, data in parentheses are percentages.

* Data in parentheses are 95% CIs.

[†] Schoenfeld residuals test for model with all mammographic features of false-positive results: $\chi^2 = 0.06$, $P = .813$ for screen-film mammography; $\chi^2 = 1.42$, $P = .491$ for digital mammography.

[‡] Schoenfeld residuals global test for model with subtypes of mammographic features of false-positive results: $\chi^2 = 2.55$, $P = .110$ for screen-film mammography; $\chi^2 = 5.14$, $P = .527$ for digital mammography.

Discussion

Women with false-positive results had an increased risk of breast cancer; this finding is in agreement with the findings of other studies (8,10). The excess risk found was higher in women with calcifications than in women with other mammographic features. In addition,

women with two or more false-positive examinations in whom the mammographic features changed over time had the highest increase in risk of developing breast cancer.

In our study, masses were the most frequent mammographic feature observed on false-positive images. However, the risk of breast cancer in

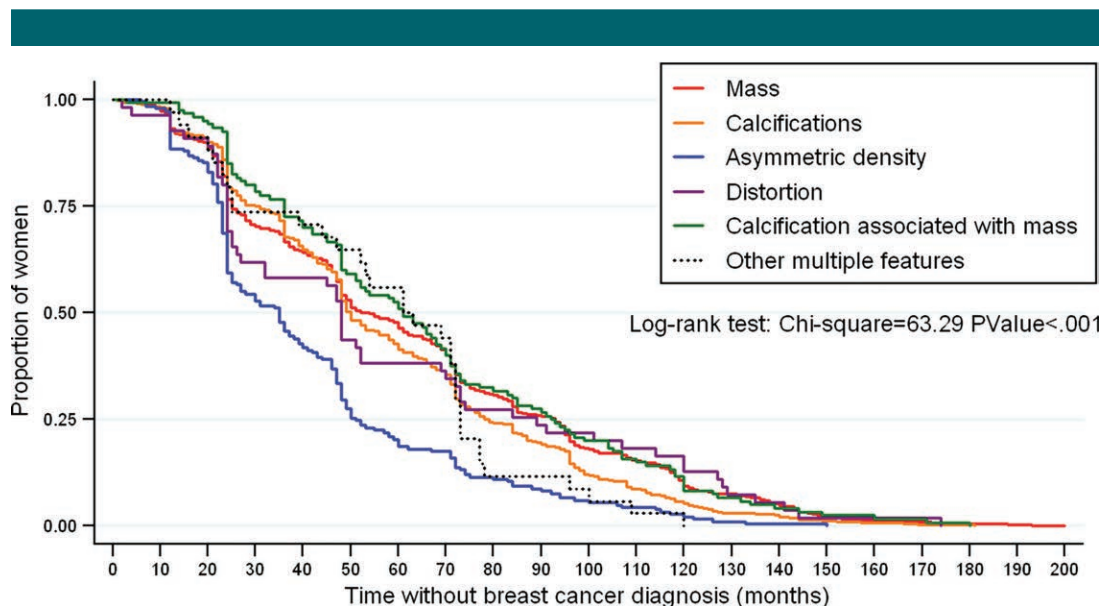
women with masses was lower than that in women with calcifications. It is a widely known fact that calcifications are frequently observed on breast cancer mammograms (17) and on baseline screening mammograms in women with interval cancers (14).

Our data suggest that the risk of breast cancer among women with false-positive results was no different regarding screen-detected and interval cancers. Moreover, we found no differences in the risk of breast cancer according to the type of mammography (screen-film vs digital mammography).

Only a small percentage (4.1%) of women had two or more false-positive studies in which mammographic features changed over time; however, these women had 30.2% of all breast cancers diagnosed in women with false-positive results. In addition, this group had the highest risk (HR, 9.13) of breast cancer in the entire cohort. From our results, it seems reasonable to offer different screening pathways to the women who have changes in mammographic features of false-positive results. In fact, the exploratory analysis in which we assessed the time between the screen-cancer diagnosis and previous mammography revealed that yearly screening may be the preferred pathway in this group, as an exception to the 2-year standard interval recommended by the European guidelines (15). Thus, women with changes in their mammographic features would probably gain a maximum benefit from intensive screening.

We observed that 72.6% of cancers were ipsilateral to the false-positive result. False-positive results may be associated with the presence of benign breast diseases, which have higher risk of breast cancer (18–21). Another explanation is that some of the false-positive findings actually harbored cancer that was not diagnosed until a later date.

This study was strengthened by the long follow-up period and large sample size. A limitation of this study was the lack of information on the quadrant in which we found mammographic features that did not enable us to study if tumors were found in the same location.



Kaplan-Meier curve shows time without a breast cancer diagnosis in women with false-positive results according to the mammographic features. Median time to breast cancer diagnosis was 54 months (25th–75th percentile range, 25–93 months) for mass, 50 months (25th–75th percentile range, 31–78 months) for calcifications, 35 months (25th–75th percentile range, 23–51 months) for asymmetric density, 48 months (25th–75th percentile range, 24–84 months) for distortion, 61 months (25th–75th percentile range, 36–92 months) for calcification associated with mass, and 61 months (25th–75th percentile range, 25–73 months) for other multiple features.

Table 4

Age-adjusted Risk of Breast Cancer in Women with Two or More False-Positive Results in Which Mammographic Features Changed over Time

Finding	No. of Women	No. of Breast Cancers	HR*	PValue
Negative mammogram	445 125	5004	1.00 (Referent)	...
Mammographic features of false-positive results†	76 075 (100.0)	1443	1.84 (1.73, 1.95)	<.001
Mammographic features without change‡	72 783 (95.7)	1007	1.36 (1.27, 1.46)	<.001
Mammographic features with change‡	3292 (4.1)	436	9.13 (8.28, 10.07)	<.001

Note.—Unless otherwise indicated, data in parentheses are percentages.

* Data in parentheses are the 95% CI.

† Schoenfeld residuals test for model with all mammographic features of false-positive results: $\chi^2 = 0.42, P = .517$.

‡ Schoenfeld residuals global test for model with subtypes of mammographic features of false-positive results: $\chi^2 = 1.71, P = .191$.

We found that when compared with women with negative mammograms, women with false-positive results have a higher risk of breast cancer detection, particularly those whose mammograms showed calcifications. Furthermore, women with two or more false-positive results in which the mammographic features changed over time had a

greatly increased risk. As with other variables, such as family history, age, and breast density, previous screening results may be useful to improve the accuracy of risk prediction models. Population-based breast cancer screening programs should consider the antecedent of having a false-positive result for recommending yearly follow-up.

Disclosures of Conflicts of Interest: X.C. disclosed no relevant relationships. I.T. disclosed no relevant relationships. M.P. disclosed no relevant relationships. M.R. disclosed no relevant relationships. M.V. disclosed no relevant relationships. A.R. disclosed no relevant relationships. L.D. disclosed no relevant relationships. C.V. disclosed no relevant relationships. M.B. disclosed no relevant relationships. J.F. disclosed no relevant relationships. M.J.Q. disclosed no relevant relationships. M. Sánchez disclosed no relevant relationships. C.N. disclosed no relevant relationships. J.A.E. disclosed no relevant relationships. F.S. disclosed no relevant relationships. M. Sala disclosed no relevant relationships.

References

1. Bond M, Pavey T, Welch K, et al. Systematic review of the psychological consequences of false-positive screening mammograms. *Health Technol Assess* 2013;17(13):1–170, v–vi.
2. Román R, Sala M, De La Vega M, et al. Effect of false-positives and women’s characteristics on long-term adherence to breast cancer screening. *Breast Cancer Res Treat* 2011; 130(2):543–552.
3. Hofvind S, Thoresen S, Tretli S. The cumulative risk of a false-positive recall in the

- Norwegian Breast Cancer Screening Program. *Cancer* 2004;101(7):1501–1507.
4. Román R, Sala M, Salas D, et al. Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. *Ann Oncol* 2012;23(1):104–111.
 5. Castells X, Molins E, Macià F. Cumulative false positive recall rate and association with participant related factors in a population based breast cancer screening programme. *J Epidemiol Community Health* 2006;60(4):316–321.
 6. Elmore JG, Barton MB, Mocerri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998;338(16):1089–1096.
 7. Hubbard RA, Miglioretti DL, Smith RA. Modelling the cumulative risk of a false-positive screening test. *Stat Methods Med Res* 2010;19(5):429–449.
 8. von Euler-Chelpin M, Risør LM, Thorsted BL, Vejborg I. Risk of breast cancer after false-positive test results in screening mammography. *J Natl Cancer Inst* 2012;104(9):682–689.
 9. Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst* 2006;98(17):1204–1214.
 10. Castells X, Román M, Romero A, et al. Breast cancer detection risk in screening mammography after a false-positive result. *Cancer Epidemiol* 2013;37(1):85–90.
 11. Georgian-Smith D, Lawton T. Basic principles. In: *Breast imaging and pathologic correlations: a pattern-based approach*. Philadelphia, Pa: Walters Kluwer, 2014; 18–105.
 12. Hoff SR, Samset JH, Abrahamsen AL, Vigeland E, Klepp O, Hofvind S. Missed and true interval and screen-detected breast cancers in a population based screening program. *Acad Radiol* 2011;18(4):454–460.
 13. Broeders MJ, Onland-Moret NC, Rijken HJ, Hendriks JH, Verbeek AL, Holland R. Use of previous screening mammograms to identify features indicating cases that would have a possible gain in prognosis following earlier detection. *Eur J Cancer* 2003;39(12):1770–1775.
 14. Hofvind S, Geller B, Skaane P. Mammographic features and histopathological findings of interval breast cancers. *Acta Radiol* 2008;49(9):975–981.
 15. Perry N, Broeders M, de Wolf C, Tömberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Luxembourg: Office for Official Publications of the European Communities, 2006.
 16. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994;81(3):515–526.
 17. Vinnicombe S, Pinto Pereira SM, McCormack VA, Shiel S, Perry N, Dos Santos Silva IM. Full-field digital versus screen-film mammography: comparison within the UK breast screening program and systematic review of published data. *Radiology* 2009;251(2):347–358.
 18. Hartmann LC, Sellers TA, Frost MH, et al. Benign breast disease and the risk of breast cancer. *N Engl J Med* 2005;353(3):229–237.
 19. Kabat GC, Jones JG, Olson N, et al. A multicenter prospective cohort study of benign breast disease and risk of subsequent breast cancer. *Cancer Causes Control* 2010;21(6):821–828.
 20. Tice JA, O'Meara ES, Weaver DL, Vachon C, Ballard-Barbash R, Kerlikowske K. Benign breast disease, mammographic breast density, and the risk of breast cancer. *J Natl Cancer Inst* 2013;105(14):1043–1049.
 21. Castells X, Domingo L, Corominas JM, et al. Breast cancer risk after diagnosis by screening mammography of nonproliferative or proliferative benign breast disease: a study from a population-based screening program. *Breast Cancer Res Treat* 2015;149(1):237–244.

BALANCE ENTRE RIESGOS Y BENEFICIOS DEL TAMIZAJE MAMOGRÁFICO DE CÁNCER DE MAMA ¿APOYARÍA SU RECOMENDACIÓN EN MUJERES PERUANAS?

Margarita Posso^{1,2,a}, Teresa Puig^{1,3,b}, Xavier Bonfill^{1,2,3,c}

RESUMEN

El objetivo de este estudio es valorar, en base a un análisis crítico y estructurado de la evidencia, si el tamizaje mamográfico de cáncer de mama en Perú es una intervención recomendable. El análisis se realizó utilizando los criterios propuestos por el sistema GRADE (Grading of Recommendations Assessment, Development and Evaluation). Se realizó una búsqueda de la literatura en Medline y en otras fuentes de información nacional (literatura gris). Para el desarrollo de la recomendación, a favor o en contra del tamizaje, se utilizó la mejor evidencia disponible sobre los beneficios, riesgos, costos y preferencias de la población. Los resultados incluyeron cinco revisiones sistemáticas (RS) que evaluaron los riesgos y beneficios del tamizaje; dos evaluaciones económicas realizadas en Perú y un estudio que valoró indirectamente las preferencias de las mujeres peruanas. La calidad de la evidencia a partir de las RS fue moderada a favor del tamizaje en mujeres de 50 a 69 años. El balance entre riesgos y beneficios mostró una mayor probabilidad de sobrediagnóstico comparado con el beneficio en la reducción de la mortalidad. La estrategia más costo-efectiva fue la mamografía trienal. La percepción de barreras podría comprometer seriamente la participación de las mujeres. En conclusión, la recomendación a favor del tamizaje mamográfico en Perú es débil, más aun si se tienen en cuenta otras necesidades poblacionales más urgentes. En caso de llevarse a cabo, la estrategia de tamizaje más adecuada para la población peruana consistiría en realizar una mamografía trienal en mujeres de 50 a 69 años.

Palabras clave: Cribado; Neoplasias de la mama; Mamografía (fuente: DeCS BIREME).

BALANCING RISKS AND BENEFITS OF MAMMOGRAPHY SCREENING FOR BREAST CANCER: WOULD YOU SUPPORT ITS RECOMMENDATION IN PERUVIAN WOMEN?

ABSTRACT

The aim of this study is to assess the balance between benefits and risks of population-based mammography screening programs in Peruvian women. We followed the criteria proposed by the GRADE (Grading of Recommendations Assessment, Development and Evaluation) working group. A generic search strategy for published studies was performed using Medline and other sources of national data (gray literature). The evidence of benefits, risks, costs and preferences of the population was used in order to state a recommendation in favour, or against, screening. We found five systematic reviews (SR) that evaluated the balance between benefit and risks, two economic evaluations based on Peruvian data, and one study that reported the preferences of Peruvian women. The quality of evidence of the SR was moderate in favour of screening in women aged 50-69 years. The balance of risks and benefits showed a higher probability of overdiagnosis compared with the reduction in mortality. The most cost-effective strategy was the triennial mammography. Perceived barriers could seriously compromise the participation of women. In conclusion, the recommendation of mammography screening for Peruvian women is weak, even more if we take into account other health necessities of the population. However, if implemented, triennial mammography in women aged 50-69 years could be the more suitable screening strategy

Key words: Straining; Breast Neoplasms; Mammography (source: MeSH NLM).

¹ Universidad Autónoma de Barcelona, Barcelona, España.

² Centro Cochrane Iberoamericano. Institut de Recerca Biomèdica Sant Pau (IIB Sant Pau), Barcelona, España.

³ Hospital de la Santa Creu i Sant Pau. Institut de Recerca Biomèdica Sant Pau (IIB Sant Pau), Barcelona, España.

^a Médica especialista en Medicina Preventiva y Salud Pública, MPH; ^b Médica especialista en Medicina Preventiva y Salud Pública, MPH, PhD; ^c médico especialista en Oncología Médica, MPH, PhD

El presente artículo forma parte de la tesis doctoral titulada: "Cribado poblacional del cáncer de mama: aspectos relacionados con la efectividad y coste-efectividad", de la autoría de Margarita Posso. Departamento de Pediatría, Obstetricia y Ginecología y Medicina Preventiva. Universitat Autònoma de Barcelona. Barcelona, España.

Recibido: 07-08-14 Aprobado: 28-01-15

Citar como: Posso M, Puig T, Bonfill X. Balance entre riesgos y beneficios del tamizaje mamográfico de cáncer de mama ¿apoyaría su recomendación en mujeres peruanas?. Rev Peru Med Exp Salud Publica. 2015;32(1):117-28.

INTRODUCCIÓN

A nivel mundial, con una tasa ajustada por edad de 43,3/100 000 habitantes, según cifras publicadas en 2012 por la Agencia Internacional de Investigación en Cáncer, el cáncer de mama ocupa el segundo lugar en frecuencia en mujeres ⁽¹⁾. En Perú, según el Sistema Nacional de Vigilancia Epidemiológica, durante el periodo 2006-2011, el 16,6% (11 271) del total de casos de cáncer diagnosticados en mujeres correspondió a pacientes con cáncer de mama ⁽²⁾, afectando principalmente a las mayores de 50 años (80/100 000) ⁽¹⁾. Así, en el país esta neoplasia es la tercera causa de mortalidad por cáncer, contando 10 defunciones por 100 000 habitantes ⁽¹⁾ y la supervivencia es de 70% a los 5 años ⁽³⁾.

En Perú, aproximadamente un 40% de las pacientes son diagnosticadas en estadios avanzados lo que reduce las posibilidades terapéuticas y conlleva un peor pronóstico ⁽³⁾. La edad promedio al diagnóstico es 54 años y a la muerte 58,4 años ⁽³⁾. Esta rápida progresión de la enfermedad podría estar condicionada por el escaso acceso al tratamiento y su baja efectividad en los estadios avanzados pues solo un 69,5% de la población cuenta con un sistema de cobertura en salud ⁽⁴⁾. Se ha estimado que en la población femenina el 1,2% de los años de vida saludables perdidos debido a muerte prematura o discapacidad (AVISA) son atribuibles al cáncer de mama ⁽³⁾. En general, las enfermedades no transmisibles representarían el 66% de los AVISA y en particular los tumores malignos serían responsables del 7% de la carga de enfermedad ⁽⁵⁾.

Tal como lo establece la Organización Mundial de la Salud (OMS), las mujeres sintomáticas no se beneficiarían de un programa de tamizaje, sino más bien de un circuito de diagnóstico y tratamiento rápido ⁽⁶⁾. Para la población asintomática la mamografía es el único método de tamizaje que se ha revelado eficaz. Métodos de tamizaje como la autoexploración mamaria o la exploración clínica no han demostrado grandes beneficios ⁽⁵⁾. Los estudios disponibles muestran deficiencias en estas técnicas debido a su baja sensibilidad, alto porcentaje de falsos positivos, discontinuidad en la realización y variabilidad en la técnica ⁽⁷⁻⁹⁾. Según un estudio realizado en Chile, durante el 2004-2007 se alcanzó un 65% de participación en el tamizaje clínico. A pesar de esta gran adherencia no se evidenció un incremento en el porcentaje de diagnósticos tempranos ni se redujo el porcentaje de estadios avanzados ⁽¹⁰⁾.

Existen estudios cuyos resultados indican un efecto beneficioso del tamizaje mamográfico sobre la reducción en la mortalidad por cáncer de mama ^(11,12)

por lo que en muchos países se han implementado programas poblacionales ^(13,14). Sin embargo, desde el momento en que se establecieron estos programas ha existido un intenso debate en relación a sus posibles beneficios y perjuicios. Clásicamente se han descrito consecuencias negativas en las mujeres con resultados falsos positivos ^(15,16) y más recientemente el debate gira en torno al sobrediagnóstico. El sobrediagnóstico se ha definido como la detección de lesiones, incluidos algunos carcinomas, que debido a su historia natural nunca se habrían manifestado clínicamente ⁽¹⁷⁾. Los perjuicios derivados del sobrediagnóstico son frecuentes y conllevan un deterioro importante sobre la calidad de vida de las mujeres sometidas a otras pruebas diagnósticas invasivas o tratamientos quirúrgicos innecesarios ^(18,19).

El gran número de publicaciones científicas en este tema ha generado una mayor necesidad de sistematizar la revisión de la literatura para extraer conclusiones válidas. Si bien las RS proporcionan información esencial sobre la eficacia del tamizaje mamográfico de cáncer de mama, el juicio para recomendar una intervención a nivel poblacional requiere un enfoque más amplio. En este sentido, la utilización crítica de la evidencia para orientar la toma de decisiones es necesaria, más aún si tenemos en cuenta que los recursos son siempre limitados.

Durante los últimos años el sistema GRADE viene siendo ampliamente utilizado para el desarrollo de recomendaciones ⁽²⁰⁾. Las organizaciones que están adheridas al grupo de trabajo o que utilizan GRADE son referentes a nivel internacional como son la OMS, la Colaboración Cochrane, diferentes sociedades científicas norteamericanas y europeas, entre otros. GRADE propone clasificar la calidad de la evidencia y desarrollar recomendaciones de forma explícita, rigurosa y consistente. Bajo este enfoque las recomendaciones a favor o en contra de una intervención deben desarrollarse con base en la mejor evidencia disponible y siguiendo principalmente los siguientes criterios: 1) los beneficios o la eficacia, 2) los riesgos o seguridad, 3) los costos o utilización de recursos, y 4) las preferencias de la población ⁽²¹⁾.

En Perú, mediante el denominado "Plan Esperanza" el Estado ha priorizado el manejo del cáncer ⁽²²⁾. La Coalición Multisectorial liderada por el Ministerio de Salud ha planteado como objetivo para el año 2016 realizar como mínimo una mamografía al 30% de las mujeres mayores de 40 años ⁽²³⁾. Con el objetivo de contribuir de manera cuantitativa a esta iniciativa, en este trabajo se han utilizado los criterios propuestos por el sistema GRADE para elaborar recomendaciones útiles para la toma de decisiones y adaptadas al contexto del país.

MÉTODO

ESTRATEGIA DE BÚSQUEDA

Para identificar estudios sobre la eficacia y seguridad de la intervención se realizó una búsqueda de la literatura en Medline. Se combinaron los términos relacionados con cáncer de mama, programas poblacionales de tamizaje y eficacia (Tabla 1). Las evaluaciones económicas y los estudios de preferencias también se buscaron en Medline incorporando en la estrategia anterior los términos relacionados con Perú, costos y valoraciones. La búsqueda manual de literatura gris se realizó principalmente para identificar estudios de utilización de recursos y preferencias de la población. Se revisaron las páginas web del Ministerio de Salud, la Dirección General de Epidemiología, el Instituto Nacional de Salud, el Instituto Nacional de Enfermedades Neoplásicas y en el repositorio nacional de la Red Peruana de Tesis Digitales.

ELECCIÓN DE LOS DESENLACES O RESULTADOS CRÍTICOS. EFICACIA Y SEGURIDAD

Eficacia: el propósito del tamizaje es detectar precozmente las neoplasias y evitar que alcancen estadios más avanzados. Este adelantamiento en el momento del diagnóstico permitiría una intervención temprana y una disminución de la mortalidad por esta causa. Es por esto que la mejor forma de evaluar el impacto positivo del tamizaje mamográfico es midiendo la disminución en la mortalidad ⁽¹⁷⁾.

Seguridad: el sobrediagnóstico es el principal efecto indeseable en las mujeres participantes de un programa de tamizaje poblacional. Un cáncer sobrediagnosticado es aquel que de no ser por el tamizaje no hubiera alterado la vida de la mujer ⁽¹⁷⁾. El impacto negativo de

las pruebas adicionales y los tratamientos innecesarios en las mujeres sobrediagnosticadas supera el impacto negativo de otros efectos indeseables como son los resultados falsos negativos o falsos positivos, la radiación en el tórax, etc. ⁽¹⁷⁾.

EVALUACIÓN DE LA CALIDAD DE LA EVIDENCIA

GRADE propone evaluar la calidad de la evidencia en su conjunto, independientemente de la calidad individual de cada uno de los estudios sobre la eficacia y seguridad. Así, para el desarrollo de este apartado, se valoraron de forma explícita y estructurada el riesgo de sesgo (limitaciones en los estudios), la inconsistencia, la dirección de la evidencia y la imprecisión de los desenlaces de interés. Se utilizaron las tablas GRADE para el resumen de la evidencia ⁽²¹⁾.

ANÁLISIS DEL USO DE RECURSOS

Se respondió a la pregunta sobre si el uso de recursos compensa el beneficio neto esperado. Por tanto, para el desarrollo de este apartado se han incluido solo evaluaciones económicas completas en las cuales se valore de forma simultánea el uso de los recursos respecto al efecto incremental ganado con una intervención frente a otra. Las evaluaciones económicas varían respecto al contexto en el que se desarrollan por lo que en este estudio solo se han incluido evaluaciones económicas desarrolladas en Perú.

ANÁLISIS DE LAS PREFERENCIAS DE LA POBLACIÓN

GRADE propone evaluar si se puede asumir que las preferencias son similares en la población que va a recibir la intervención. En caso de ser homogénea a favor la recomendación sería fuertemente aconsejable.

Tabla 1. Estrategia de búsqueda en Ovid (<http://gateway.ovid.com/>).

1. Breast Neoplasms/
2. (Breast adj2 (neoplasm\$ or carcinoma\$ or tumour\$ or tumor\$ or cancer\$)).ti,ab.
3. 1 or 2
4. Mass Screening/
5. Mammography/
6. (breast adj2 (screen\$ or program\$ or service\$)).ti,ab.
7. (mammogra\$ or mamogra\$).ti,ab.
8. (Breast adj2 (radiograph\$ or imaging or visualize or visualise or exam\$ or test\$)).ti,ab.
9. 4 or 5 or 6 or 7 or 8
10. (over-diagnosis OR over diagnosis OR overdiagnosis OR underdiagnosis OR underdiagnosis OR under diagnosis OR mis-diagnosis OR mis-diagnosis OR over-treat\$ OR over treat\$ OR overtreat\$ OR under-treat\$ OR under treat\$ OR false positive OR false-positive OR harm\$ OR wound OR injur\$). ti,ab.
11. (life year gained OR QALY OR quality adjusted life year OR quality of life OR quality OR better life OR healthy life OR happy life OR mortal\$ OR death\$ OR dead\$ OR deceased\$ OR defunct\$ OR fallen OR departed OR breast cancer mortality OR breast-cancer mortality OR cancer mortality OR preventable death\$ OR preventable defunct\$). ti,ab.
12. 10 or 11
13. 3 and 9 and 12

Tabla 2. Características de los ensayos clínicos aleatorizados (ECA) incluidos en las principales revisiones sistemáticas (RS) que evalúan los riesgos y beneficios del cribado mamográfico de cáncer de mama.

Revisión sistemática	Modelo de metanálisis	Características de los estudios incluidos en el metanálisis						
		ECA (año de inicio) ^{ref}	Población (n)	Edad (años)	Intervención	Control	Intervalo ¹ (meses)	Duración ^d (años)
Cochrane Reviews ²⁴	Efecto fijo	Canadá I (1980) ²⁸	50 430	40-49	M + EC + AE	AE	12	5
		Canadá II (1980) ²⁹	39 405	50-59	M + EC + AE	EC + AE	12	5
		Göteborg (1982) ³⁰	52 222	39-59	M	NC	18	7
		Kopparberg (1977) ³¹	57 897	38-75	M + AE	NC	24-33	7
		Malmö (1976) ³²	42 238	45-69	M	NC	18-24	12
		New York (1963) ³³	62 000	40-64	M + EC	NC	12	3
		Ostergötland (1978) ³⁴	76 970	38-75	M + AE	NC	24-33	7
		Stockholm (1981) ³⁵	60 800	39-65	M	NC	24-28	4
UK age trial (1991) ³⁶	160 921	39-41	M	NC	12	8		
UK National Health Service ²⁵	Efecto aleatorio	Igual al anterior*						
Canadian TFPHC ²⁶	Efecto aleatorio	Igual al anterior*						
US PSTF 2009 ²⁷	Efecto aleatorio	Igual al anterior*						

ECA= ensayo clínico aleatorizado, M= mamografía, EC= examen clínico, AE= autoexploración, NC= no cribado.*Todas las RS incluyeron en el metanálisis los mismos ECA. ¹Intervalo de tiempo entre una prueba de cribado y la siguiente. ^dDuración del estudio.

Para el desarrollo de este apartado se incorporaron referencias nacionales en las que se valoren de la forma más directa posible las preferencias de la población afectada.

DESARROLLO DE LA RECOMENDACIÓN FINAL

Considerando el balance entre riesgos y beneficios del uso de los recursos y las preferencias de la población, se desarrolló una recomendación a favor o en contra. La fuerza de la recomendación se basó en la incertidumbre de los resultados que afectan el sentido de la recomendación. Todo el proceso se estructuró utilizando la tabla propuesta por GRADE para la generación de recomendaciones ⁽²¹⁾.

RESULTADOS

Se identificaron cinco RS, elaboradas por reconocidos referentes científicos como son la RS Cochrane ⁽²⁴⁾, la del Servicio Nacional de Salud del Reino Unido ⁽²⁵⁾, las guías de práctica clínica de la Task Force canadiense ⁽²⁶⁾ y estadounidense ⁽²⁷⁾ y la RS de la Red Europea de Programas de Tamizaje de Cáncer de Mama ⁽¹¹⁾. En todas se ha realizado un análisis combinado de los resultados de nueve ensayos clínicos aleatorizados (ECA) en los que se incluyeron más de 600 000 mujeres entre 39 y 74 años ⁽²⁸⁻³⁶⁾ (Tabla 2).

EFICACIA DEL TAMIZAJE MAMOGRÁFICO SOBRE LA MORTALIDAD

La RS de Gøtzsche ⁽²⁴⁾ reportó una disminución relativa de la mortalidad del 10 y 25% en los ECA, con una aleatorización óptima y subóptima, respectivamente. Esta RS concluye que una estimación más rigurosa del efecto sería del 15%, punto intermedio entre 10 y 25%. Por tanto, en términos absolutos, se evitaría una muerte de cada 2000 mujeres invitadas al tamizaje durante 10 años (Tabla 3).

A diferencia de la RS de Gøtzsche que utilizó para el metanálisis un modelo de efecto fijo, la RS publicada por el Grupo del Reino Unido ⁽²⁵⁾ utilizó uno de efecto aleatorio. La principal diferencia consiste en que el modelo de efectos aleatorios incorpora en el análisis la variabilidad entre los estudios, lo que mejoraría el nivel de confianza y, en cierto modo, la validez externa de los resultados ⁽³⁷⁾. Con este modelo se reportó una disminución relativa de la mortalidad del 20%. Aplicando este porcentaje a la realidad británica, donde las mujeres entre 50 y 69 años participan en el tamizaje de forma trienal, los autores estiman que para evitar una muerte sería necesario invitar a 235 mujeres durante 20 años.

Por su parte, los grupos canadiense y estadounidense ^(26,38) identificaron el mayor beneficio en el grupo de 50 a 69 años (21%) reportando que para evitar una muerte sería

Tabla 3. Beneficios y riesgos del cribado mamográfico. Resultados de las principales RS publicadas entre el 2009 y 2013.

Revisión Sistemática (referencia)	Disminución de la mortalidad				Riesgos				
	Número de estudios	Análisis de subgrupo	Relativa* (IC 95%)	Absoluta¶ (intervalo del cribado)	Número de estudios	Periodo evaluado	Falsos positivos	Exceso de procedimientos invasivos	Sobrediagnóstico
Cochrane Reviews ²³	8	Estimación final	~15%						
	3	Adecuada aleatorización	10% (-2 a 21)	1/2000 invitadas durante 10 años (B/T)	3	acumulado 13 años	49%	19%	30%
	5	Inadecuada aleatorización	25% (17 a 33)						
UK National Health Service ²⁴	8		20% (11 a 27)	1/235 invitadas durante 20 años (T)	3	tras una mamografía	3,2%	1%	11-19%
Canadian TFPHC ²⁵	8	39 a 49 años	15% (4 a 25)	1/2108 invitadas durante 11 años (B/T)	8	acumulado 11 años	33%	3,6%	30-52%
	7	50 a 69 años	21% (10 a 32)	1/720 invitadas durante 11 años (B/T)	7	acumulado 11 años	28%	3,7%	NR
	2	70 a 74 años	32% (-1 a 55)	No reportado	2	acumulado 11 años	21%	2,6%	NR
US PSTF 2009 ²⁶	8	39 a 49 años	15% (4 a 25)	1/1 904 invitadas durante 12 años (B/T)	8	tras una mamografía	10%	NR	1-10%
	6	50 a 59 años	24% (1 a 25)	1/1 339 invitadas durante 12 años (B/T)	6	tras una mamografía	11%	NR	1-10%
	2	60 a 69 años	32% (10 a 32)	1/337 invitadas durante 7 años (B/T)	2	tras una mamografía	14%	NR	1-10%
	1	70 a 74 años	-12% (-72 a 27)	No es posible realizar el calculo	1	tras una mamografía	15%	NR	1-10%
Euroscreen Working Group ¹⁰	7	IBM	25% (19 a 31)	7-9/1 000 cribadas durante 20 años y seguidas hasta los 79 años (B)	con ajuste adecuado	acumulado 20 años	20%	3%	0-10%
	8	Casos y controles	31% (27 a 43)		estudios sin ajuste				1-58%

RS= revisiones sistemáticas, NR= no reportado, IBM= incidence-based mortality (estudios que midieron las mortalidad en mujeres que con certeza se realizaron una o más pruebas de cribado mamográfico).

*¶ La disminución relativa y la absoluta se obtienen dimensionando la diferencia entre el número de muertes que se producen en el grupo de mujeres invitadas al cribado frente al grupo control. Mientras que la disminución relativa, es la proporción o porcentaje en que se reduce el riesgo relativo de muerte por cáncer de mama, la disminución absoluta se expresa mediante el número total de muertes evitadas.

necesario invitar a participar durante 11 años a 720 mujeres (Tabla 3).

Si bien la mejor forma de medir la eficacia de una intervención es mediante los estudios experimentales, una aproximación más cercana a la realidad podría obtenerse a partir de otros diseños pragmáticos⁽³⁹⁾. Bajo esta premisa, el grupo europeo Euroscreen realizó un metanálisis de estudios observacionales. Además de los estudios de tipo caso control, se incluyeron aquellos que siguieron una metodología *incidence-based mortality* (IBM), lo que permitiría un manejo más adecuado del factor exposición⁽⁴⁰⁾. La disminución en la mortalidad estimada a partir de los estudios IBM fue del 25% y del 31% en los estudios de tipo caso control. Así, los autores reportaron que realizando el tamizaje bienal a 1000 mujeres durante 20 años se evitarían 7 a 9 muertes⁽¹¹⁾ (Tabla 3).

SEGURIDAD DEL TAMIZAJE. EL SOBREDIAGNÓSTICO

Los estudios disponibles dimensionaron el sobrediagnóstico como el exceso de neoplasias

detectadas en el grupo tamizaje frente al número total de neoplasias clínicamente diagnosticadas en el grupo control. Dos perspectivas, la poblacional y la individual, han sido utilizadas para reportar el sobrediagnóstico. En ambas, el numerador está determinado por el exceso de neoplasias. Bajo una perspectiva individual, en el denominador se incluirían solo las neoplasias diagnosticadas durante el periodo del tamizaje mientras que bajo una perspectiva poblacional se incluirían también las neoplasias diagnosticadas durante el periodo posterior al tamizaje.

Tanto la RS Cochrane⁽²⁴⁾ como la del Reino Unido⁽²⁵⁾ consideraron adecuados para el estudio del sobrediagnóstico a tres ECA^(28,29,41). En estos, las mujeres aleatorizadas al grupo control no fueron invitadas a participar de ningún tamizaje posterior a la intervención y tuvieron un periodo de seguimiento adecuado. La frecuencia de sobrediagnóstico fue del orden del 11% desde una perspectiva poblacional y alrededor del 19% desde la individual⁽²⁵⁾. Según la RS Cochrane esta cifra sería mayor, cercana al 30%,

debido al tamizaje oportunista en el grupo control. Cifras menores (1 a 10%) se reportaron a partir de los estudios observacionales (38,42). Con estos datos, los autores estimaron que de cada 1000 mujeres invitadas a participar del tamizaje bienal durante 20 años, entre 4 y 20 mujeres sufrirían un sobrediagnóstico (11,24). Dicho de otra manera, si una mujer cribada recibe el diagnóstico de cáncer, la probabilidad de que se trate de un sobrediagnóstico sería de un 19% (25) (Tabla 3).

EVALUACIÓN DE LA CALIDAD DE LA EVIDENCIA

La mayoría de los EC incluidos en las RS mostraron limitaciones serias con posibles distorsiones en la aleatorización, ocultamiento de la secuencia, y cegamiento. Además, todos estos ECA fueron realizados hace aproximadamente 20 o 30 años por lo que su validez hoy en día podría quedar en entredicho, más aun considerando que el riesgo basal de la población femenina peruana es diferente a la población europea. Sin embargo, los resultados fueron consistentes, se midieron de forma directa y mostraron una gran precisión, por lo que finalmente la calidad de la evidencia fue moderada (Tabla 4).

El escenario es diferente para los estudios observacionales. En este caso, a pesar de tratarse de estudios contemporáneos, la calidad de la evidencia podría ser de baja o muy baja calidad debido a las limitaciones propias de su diseño, medición errónea entre la exposición y el desenlace, problemas con el manejo de las variables de confusión y grupos de comparación inadecuados (Tabla 5).

BALANCE ENTRE LOS RIESGOS Y BENEFICIOS

La incertidumbre en torno a las estimaciones del beneficio y el sobrediagnóstico fueron importantes. Por cada 1000 mujeres que participen del tamizaje durante 20 años, entre 2 y 9 muertes serían evitadas y de 4 a 20 mujeres sufrirían un sobrediagnóstico. Los principales determinantes de esta gran variabilidad fueron el tipo de estudio, la edad de las mujeres, la duración del tamizaje y del seguimiento. En todo caso, aunque el número de casos sobrediagnosticados parece ser más frecuente que el número absoluto de muertes evitadas, una clave para considerar un balance negativo sería determinar si las consecuencias del sobrediagnóstico alcanzan una magnitud cercana a la muerte.

Tabla 4. Resumen de la evidencia y evaluación de la calidad de los ensayos clínicos aleatorizados (ECA) que evaluaron el efecto del cribado mamográfico sobre la mortalidad por cáncer de mama. Adaptado de la Guía de Práctica Clínica (GPC) Nacional Canadiense²⁶.

Número de estudios	Evaluación de la calidad						Resumen de los resultados				Calidad de la evidencia
	Diseño	Limitaciones en el estudio	Inconsistencia	Evidencia indirecta	Imprecisión	Otros	Mortalidad por cáncer de mama		Efecto		
							Cribado n (%)	Control n (%)	RR (IC 95%)	Absoluto (NNC)	
Mortalidad por cáncer de mama en mujeres de 40 a 49 años											
8	ECA	Serio†	No serio§	No serio¶	No serio**	No††	448 (0,29)	625 (0,32)	0,85 (0,75-0,96)	1 por cada 2108	Moderada
Mortalidad por cáncer de mama en mujeres de 50 a 69 años											
7	ECA	Serio‡‡	No serio§§	No serio¶¶	No serio**	No††	639 (0,47)	743 (0,64)	0,79 (0,68-0,90)	1 por cada 721	Moderada
Mortalidad por cáncer de mama en mujeres de 70 a 74 años											
2	ECA	Serio***	No serio†††	No serio¶¶¶	Serio¶¶¶	No††	49 (0,47)	50 (0,68)	0,68 (0,45-1,01)	NE	Baja

Nota: Las estimaciones de RR (riesgo relativo) se basan en el meta-análisis con modelo de efectos aleatorios.
 ECA= ensayo clínico aleatorizado. RR= riesgo relativo. IC = intervalo de confianza. NNC = número necesario a cribar (mujeres que necesitan ser cribadas cada 2 años durante un promedio de 11 años para prevenir 1 muerte por cáncer de mama). NE= No estimable debido a la falta de precisión del RR, la intervención podría incrementar la mortalidad en lugar de reducirla.
 † Cinco estudios cuasi-aleatorizados y tres ECA verdaderos. El cegamiento y el ocultamiento de la secuencia no están claros en los cinco primeros.
 § No existe heterogeneidad, p = 0,48; I² = 0%.
 ¶ La evidencia que se evalúa es directa. El efecto del cribado sobre la reducción de la mortalidad es una pregunta que se aborda de forma similar en todos los estudios en relación a la población, intervención, comparación y resultados.
 ** El tamaño total de la muestra es grande y IC95% de los resultados es estrecho.
 †† No se cuenta con estudios suficientes para evaluar el sesgo de publicación.
 ‡‡ El cegamiento y el ocultamiento de la secuencia no están claros en los cinco estudios y sólo dos son ECA verdaderos.
 §§ No existe heterogeneidad, p = 0,12, I² = 41%.
 *** El cegamiento y ocultamiento no están claros.
 ††† No existe heterogeneidad, p = 0,75; I² = 0%.
 ¶¶¶ El tamaño total de la muestra es grande, pero el número total de eventos es pequeño lo que condiciona un IC95% amplio.

Tabla 5. Tabla GRADE de los estudios observacionales que evaluaron el efecto del cribado mamográfico sobre la mortalidad por cáncer de mama. Elaboración propia a partir de los estudios incluidos en la Revisión Sistemática (RS) del Grupo Europeo Euroscreen¹¹.

Evaluación de la calidad							Resumen de los resultados				
Número de estudios	Diseño	Limitaciones en el estudio	Inconsistencia	Evidencia indirecta	Imprecisión	Otros	Mortalidad por cáncer de mama		Efecto		Calidad de la evidencia
							Invitadas al cribado n (%)	No invitadas n (%)	Relativo (IC 95%)	Absoluto (NNC)	
Mortalidad por cáncer de mama en mujeres de 45 a 69 años											
20	IBM	No serio‡	No serio§	No serio¶	No serio*	No†	ND	ND	RR 0,75 (0,69-0,81)	ND***	Baja
Mortalidad por cáncer de mama en mujeres de 45 a 69 años (cribado) frente a otros grupos de edad (25 a 85, control)											
12	Estudio de tendencia	Serio‡‡	Serio§§	Serio¶¶	No serio**	No†	ND	ND	Reducción anual‡‡ 1-9% Reducción al final‡‡‡ 28-36%	ND	Muy baja
Mortalidad por cáncer de mama en mujeres de 40 a 79 años											
7	CC	Serio‡‡	No serio§§§	No serio¶¶	No serio*	No†	ND	ND	OR 0,69 (0,57-0,73)	ND***	Muy baja

Nota: Los estudios observacionales presentan serias limitaciones para estimar el efecto de una intervención. De acuerdo al sistema GRADE la calidad de la evidencia derivada de estudios observacionales es baja. Podría mejorar si se demuestra un adecuado manejo de las variables de confusión, una gran magnitud del efecto y un gradiente dosis-respuesta.

IBM= estudios que siguen la metodología *incidence-based mortality*. Los eventos se miden solo en aquellas mujeres que fueron diagnosticadas de cáncer de mama después de su primera invitación al cribado. CC= estudio caso control. IC = intervalo de confianza. RR= riesgo relativo. OR= *odds ratio*. NNC = número necesario a cribar ND= no disponible.

‡ Manejo adecuado de las variables de confusión más relevantes. El diseño permite reducir el riesgo de errores en la medición de la exposición a la intervención y el desenlace. El grupo control mostró características similares al grupo intervención en edad y localización geográfica. La magnitud del efecto es importante. No se cuenta con información sobre el gradiente dosis-respuesta.

§ No existe heterogeneidad, p = 0,40, I² = no reportado

¶ El efecto del cribado sobre la mortalidad se mide de forma directa. El resultado se mide posterior a la intervención, no se evalúan resultados intermedios, la población utilizada y los instrumentos de medición son apropiados para evaluar el efecto.

* El tamaño de la muestra es adecuado y el IC 95% es estrecho.

† No se evaluó el sesgo de publicación.

‡‡ Reducción de la mortalidad reportada de forma anual.

‡‡‡ Reducción de la mortalidad reportada al final del periodo del cribado.

‡‡ No se reportó un manejo adecuado de las variables de confusión. Sólo se ajustó el sesgo de auto-selección. El riesgo de sesgo es alto.

§§ Gran heterogeneidad. Los autores no realizaron un análisis combinado de los datos debido a la gran variabilidad entre los estudios (instrumentos de medición, grupos de comparación).

¶¶ Importantes diferencias epidemiológicas entre los grupos de comparación. La incidencia de cáncer de mama varía de acuerdo a la edad siendo más frecuente en mujeres de 45 a 69 años.

** El tamaño total de la muestra es grande. No se cuenta con el IC95% de las estimaciones.

§§§ No existe heterogeneidad, p = 0,17; I² = no reportado.

*** En un estudio posterior, considerando solo los datos de los estudios IBM y CC, los autores estiman que para reducir la mortalidad en 7 a 9 casos es necesario cribar a 1000 mujeres¹⁸

La magnitud del efecto del sobrediagnóstico es difícil de calcular, en parte, porque no se ha podido cuantificar claramente su frecuencia. Es difícil determinar si una imagen sospechosa representa realmente un peligro para la mujer o, por el contrario, se trata de una lesión que nunca afectará su vida. Como ejemplo, los autores han considerado al carcinoma ductal *in situ*. Debido a su progresión lenta o estacionaria, en muchos casos este carcinoma no progresaría a estadios avanzados. Si fuera identificado en el tamizaje, sería innecesariamente tratado afectando negativamente la calidad de vida y el bienestar psicológico y económico de la mujer⁽¹⁵⁾.

Otra clave para orientar bien el balance entre los riesgos y beneficios es identificar los grupos que más

se beneficiarían de la intervención. Consistentemente el grupo de edad en que se observan los mayores beneficios es el de 50 a 69 años. Mujeres de otras edades podrían estar igual de expuestas a un sobrediagnóstico y obtendrían menos beneficios del tamizaje mamográfico.

USO DE LOS RECURSOS

Se han identificado dos evaluaciones económicas que comparan diferentes estrategias para el control del cáncer de mama en el Perú^(43,44). Ambos estudios coinciden en reportar que solo brindar tratamiento a los casos sintomáticos sin incorporar ningún tipo de prevención sería la estrategia menos recomendable.

En el estudio de costo-efectividad de Zelle *et al.* ⁽⁴³⁾, se comparan 94 posibles escenarios que combinan las actividades de prevención y tratamiento más relevantes en Perú. El análisis se realizó desde la perspectiva del Sistema de Salud y siguiendo la metodología recomendada por la OMS para el desarrollo de evaluaciones económicas de países en vías de desarrollo (WHO-CHOICE) ⁽⁴⁵⁾. Se consideraron muy costo-efectivas a las intervenciones con una relación de costo-efectividad incremental (RCE) por debajo de \$ 4608 para evitar un AVISA. El tamizaje con mamografía trienal en mujeres de 45 a 69 años, 60% fija (urbana) y 40% móvil (rural) combinado con el tratamiento estándar, sin trastuzumab, en todos los estadios de la enfermedad mostró una RCE de \$ 4125 por AVISA, siendo así la estrategia más costo-efectiva. Ampliar el rango del tamizaje trienal hasta los 40 años y realizarlo de forma bienal incrementaría la RCE a \$ 5659 y \$27 477, respectivamente. Otras estrategias, incluidas las combinaciones con examen clínico y punción inmediata con aguja fina (PAF), mostraron RCE por encima del umbral recomendado ⁽⁴³⁾.

Desde la misma perspectiva del sistema de salud, un estudio de costo-utilidad comparó los años de vida ajustados por calidad (AVAC) que se obtendrían a partir de cinco diferentes estrategias. La mamografía, con un promedio en costo y AVAC de \$ 35,93 y \$ 0,3963, respectivamente, se reportó como la estrategia más recomendable. Tal como refiere el autor, la incertidumbre alrededor de esta estimación es bastante amplia, debido principalmente a la falta de datos representativos de la población peruana ⁽⁴⁴⁾.

VALORES Y PREFERENCIAS

En Perú los estudios sobre conocimientos, actitudes, valores o preferencias de las mujeres frente al tamizaje de cáncer de mama son escasos ⁽⁴⁶⁻⁴⁸⁾. Recientemente Huaman *et al.* realizaron un estudio de validación del cuestionario de susceptibilidad, beneficios y barreras ante el tamizaje mamográfico en 296 mujeres entre 40 y 65 años. El 57,2% de las participantes refirió haberse realizado una mamografía alguna vez en su vida. En el análisis multivariado, los únicos factores asociados a realizarse una mamografía fueron la edad entre 50 y 60 años y tener conocimientos medios o altos sobre la mamografía y el cáncer de mama. Por el contrario, la percepción de las barreras (miedo, vergüenza, olvido, priorización de otras necesidades, etc.) estuvo fuertemente asociada a la no realización de la mamografía ⁽⁴⁶⁾.

Aunque la utilidad del tamizaje parece ser reconocida en la población femenina peruana esto no queda reflejado en la realización habitual de este tipo de pruebas. Un

estudio nacional en base a la Encuesta Demográfica de Salud Familiar 2008 (ENDES) reportó que un 34,6% de las mujeres entre 20 y 49 años se había realizado alguna vez el autoexamen de mamas. La realización de la mamografía sería mucho menos frecuente y con importantes diferencias geográficas a nivel nacional ⁽⁴⁹⁾.

RECOMENDACIÓN FINAL

Como se muestra en la Tabla 6, basados en el análisis del equilibrio entre beneficios, riesgos, costos y preferencias, la implementación del tamizaje mamográfico en Perú sería débilmente recomendable en el grupo de mujeres de 50 a 69 años. Por otro lado, esta intervención no sería recomendable en las mujeres menores de 50 y mayores de 69 años debido a que los beneficios esperados no superarían los riesgos del tamizaje.

DISCUSIÓN

En el Perú las estrategias que no incorporan medidas preventivas y solo se basan en el tratamiento, han demostrado ser las menos recomendables ^(43,44). Es innegable el impacto negativo que esta enfermedad tiene en la sociedad por lo que de forma muy acertada se han iniciado importantes esfuerzos para mejorar el manejo del cáncer a nivel nacional. En todos los contextos los recursos son siempre limitados y aquellos que se dediquen al control de una enfermedad no pueden dedicarse a otras actividades. Este costo de oportunidad nos obliga a evaluar muy rigurosamente las estrategias disponibles y elegir las más eficientes.

Particularmente, en relación con el cáncer de mama, la propuesta de la Coalición Multisectorial consiste en realizar el tamizaje a las mujeres entre los 40 y 69 años, de preferencia se realizará examen clínico asociado a PAF en el grupo de 40 a 49 y mamográfico a partir de los 50 ⁽¹⁵⁾. Algunas consideraciones alrededor de esta propuesta podrían mejorar los resultados de una intervención a gran escala.

Sobre el tamizaje en el grupo de 40 a 49 años. No existe evidencia a favor del examen clínico como prueba de tamizaje a nivel poblacional, no ha demostrado reducir la mortalidad ni mejorar la detección del cáncer en estadios tempranos. De hecho, independientemente de la edad de la mujer, la OMS solo recomienda el tamizaje clínico en lugares en los que la mamografía no está disponible ⁽⁵⁰⁾. Además, en mujeres jóvenes la frecuencia de la enfermedad no es alta, lo que reduciría aun más el valor predictivo positivo del examen incrementando el número de lesiones benignas detectadas. En este

Tabla 6. Generación de las recomendaciones sobre el cribado mamográfico a partir de la evidencia disponible (Tablas 3 y 4). El sentido y la fortaleza de la recomendación se presentan bajo la perspectiva del sistema GRADE^{20,21}.

Pregunta: ¿Se recomienda la implementación del cribado mamográfico de cáncer de mama en el Perú?		Especificaciones
Factor	Juicio	Juicio razonado
Calidad de la evidencia ¿La confianza con los resultados es alta o moderada?	Sí <input type="checkbox"/> No <input checked="" type="checkbox"/>	La confianza en los resultados es moderada. Los desenlaces evaluados provienen de 8 ECA y 39 estudios observacionales. Aunque el riesgo de sesgo está presente, en la mayoría de los estudios los resultados muestran una gran consistencia y precisión lo que favorece una mayor confianza en los resultados.
A más alta confianza con los resultados más probable es realizar recomendaciones fuertes.	X <input type="checkbox"/>	
Balance entre los efectos beneficiosos y los perjudiciales ¿Al aplicar esta intervención, los efectos beneficiosos van a superar los perjudiciales, o a la inversa?	Sí <input type="checkbox"/> No <input checked="" type="checkbox"/>	El balance entre beneficios y perjuicios varía de acuerdo a los estudios siendo menos favorable en los ECA. Se ha estimado que de cada 2000 mujeres que participan en el cribado bienal durante 10 años, sólo se reduciría la mortalidad en 1 caso mientras que 10 mujeres sufrirían las consecuencias del sobrediagnóstico. Con estos datos, la frecuencia de perjuicios supera a los beneficios. Sin embargo, sería razonable otorgarle un mayor peso a evitar una muerte frente a la magnitud del daño del sobrediagnóstico en 10 mujeres.
A más seguridad que el balance es claramente beneficioso o perjudicial, más probable es realizar recomendaciones fuertes.	X <input type="checkbox"/>	
Uso de recursos ¿El uso de recursos compensa el beneficio neto esperado?	Sí <input checked="" type="checkbox"/> No <input type="checkbox"/>	El estudio de coste-efectividad de Zelle <i>et al</i> ³ demuestra que la práctica habitual para el manejo de cáncer de mama en el Perú no es coste-efectiva. Después de evaluar 96 posibles escenarios, la recomendación que se deriva del estudio es realizar cribado trienal a mujeres entre 45 y 69 años.
Cuanto más bajo sea el coste de una intervención, en relación a las alternativas, más probable es realizar recomendaciones fuertes.	X <input type="checkbox"/>	
Valores y preferencias ¿Se puede asumir que los valores y preferencias son similares en la población que va a recibir la intervención?	Sí <input type="checkbox"/> No <input checked="" type="checkbox"/>	En Perú no se cuenta con estudios sobre valores y preferencias de las mujeres que podrían beneficiarse del cribado mamográfico de cáncer de mama. Algunos autores han apuntado que la participación en el cribado podría ser baja debido, entre otros, a la falta de conocimiento del cribado y a la priorización de otras necesidades ⁴⁶⁻⁴⁹ .
A más seguridad que los valores y preferencias son similares entre la población probable es realizar recomendaciones fuertes.	X <input type="checkbox"/>	
Fuerza de la recomendación: DEBIL a FAVOR en mujeres de 50 a 69 años y débil en contra en mujeres menores de 50 y mayores de 70 años.		

ECA= ensayo clínico aleatorizado

contexto, realizar una prueba invasiva, como es el PAF, para determinar la histología del tumor podría mejorar la precisión en los resultados. En contraposición, la PAF incrementaría los costos y la necesidad de recursos humanos adecuadamente entrenados. Por tanto, si incorporamos todos estos factores bajo la perspectiva GRADE, la recomendación sería débil en contra de este tipo de tamizaje en mujeres de 40 a 59 años.

En las mujeres a partir de los 50 años el tamizaje mamográfico trienal parece ser la prueba de elección. Reducir el intervalo a bienal o anual incrementaría los costos sin mejorar sensiblemente los beneficios, lo que representaría una mala utilización de los recursos ⁽⁴³⁾. No solo es suficiente conocer el perfil de costo-efectividad o costo-utilidad de una intervención antes de implementarla, también es necesario valorar el costo de la información que será necesaria para evaluar los resultados de la misma ⁽⁵¹⁾. En Perú, este valor de la información podría ser alto. Sería necesario, por ejemplo, implementar un registro que permita evaluar cambios en los estadios al diagnóstico y su repercusión en la mortalidad. Así también tendrán que generarse otras fuentes de información para valorar el cumplimiento de los estándares del tamizaje.

Por otro lado, la participación y adherencia de las mujeres depende en gran medida de sus conocimientos y actitudes. En México, la baja participación reportada (22%) ⁽⁵²⁾ sería parcialmente atribuible a la escasa información, la incomodidad y el miedo que sienten las mujeres frente al tamizaje ⁽⁵³⁾. En Perú, una gran área de trabajo consistirá en educar a la población y mejorar la confianza de las mujeres en los servicios de salud. Según los estudios disponibles si los porcentajes de participación no acercan al 70% la intervención reduciría su nivel de costo-efectividad y los recursos asignados serían difícilmente justificables ⁽⁴³⁾.

Finalmente, una barrera frente a la implementación de un programa coordinado a nivel nacional es la actual organización segmentada del sistema sanitario. El 85% de los médicos oncólogos ejercen en Lima donde, además, se concentran la mayoría de los equipos de prevención, diagnóstico y tratamiento del cáncer ⁽⁵⁴⁾. Del mismo modo, la disponibilidad al tratamiento, que es un componente crucial en el manejo del cáncer, no está garantizada en todo el territorio nacional. Las

poblaciones rurales estarían especialmente vulnerables debido a que los centros locales no pueden ofrecer un adecuado cuidado de las pacientes con cáncer. Por tanto, un reto importante consistirá en el equipamiento de los servicios de salud priorizando la máxima utilización de los recursos humanos y logísticos disponibles.

CONCLUSIONES

El tamizaje de cáncer de mama no solo supone beneficios sino también riesgos para la población participante. Si bien la mamografía es la única estrategia que ha demostrado una reducción en la mortalidad, el balance entre los riesgos y beneficios podría ser negativo si la intervención se implementa en la población equivocada. En Perú la recomendación es débil a favor del tamizaje mamográfico en las mujeres de 50 a 69 años y débil en contra en las mujeres fuera de este rango de edad. La evidencia disponible indica que los resultados más beneficiosos y costo-efectivos se presentarían mediante un tamizaje trienal en mujeres de 50 a 69 años. En lugares donde la mamografía está disponible, el tamizaje clínico no parece justificarse, mucho menos en mujeres jóvenes en quienes la prevalencia de la enfermedad no es alta.

Importantes retos en materia de equidad, educación, redistribución de los recursos, acceso y mejora de la confianza en los sistemas sanitarios podrían condicionar los resultados del tamizaje a nivel nacional. Frente a otras posibles áreas en las que invertir los recursos disponibles, solo mediante la evaluación rigurosa de los resultados del programa se podrá valorar si es justificable continuar con la intervención elegida. La toma de decisiones basadas en la evidencia es indispensable, en este sentido el sistema GRADE es una herramienta de gran utilidad.

Contribuciones de autoría: XBC y MPR han participado en la concepción del artículo, y su redacción. La búsqueda sistemática de la literatura, así como el análisis e interpretación de los datos fueron realizados por MPR y supervisados por XBC y TPR. TPR y XBC realizaron la revisión crítica del artículo y todos los autores han aprobado su versión final.

Conflictos de interés: los autores declaran no tener conflictos de interés.

Fuentes de financiamiento: autofinanciado.

REFERENCIAS BIBLIOGRÁFICAS

- Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, *et al.* GLOBOCAN 2012: Estimated cancer incidence, mortality and prevalence worldwide in 2012 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013 [citado el 22 de mayo del 2014]. Disponible en: <http://globocan.iarc.fr>
- Ramos W, Venegas D. Análisis de la situación del cáncer en el Perú 2013. Lima: MINSA/DGE;2013.
- Justo N, Wilking N, Jönsson B, Luciani S, Cazap E. A review of breast cancer care and outcomes in Latin America. *Oncologist*. 2013;18(3):248-56. doi: 10.1634/theoncologist.2012-0373.
- Instituto Nacional de Estadística e Informática. Condiciones de vida en el Perú. Informe Técnico N° 2. Lima: INEI; 2014.
- Yip C-H, Cazap E, Anderson BO, Bright KL, Caleffi M, Cardoso F, *et al.* Breast cancer management in middle-resource countries (MRCs): consensus statement from the Breast Health Global Initiative. *Breast*. 2011 Apr;20 Suppl 2:S12-9. doi: 10.1016/j.breast.2011.02.015.
- World Health Organization. Cribado y detección precoz del cáncer [Internet]. Geneva: WHO; c2015 [citado el 14 de enero 14 del 2015]. Disponible en: <http://www.who.int/cancer/detection/es/>
- Thomas D, Gao D, Ray R, Wang W, Allison C, Chen F, *et al.* Randomized trial of breast self-examination in Shanghai: final results. *J Natl Cancer Inst*. 2002 Oct 2;94(19):1445-57.
- Nelson AL. Controversies regarding mammography, breast self-examination, and clinical breast examination. *Obstet Gynecol Clin North Am*. 2013 Sep;40(3):413-27. doi: 10.1016/j.ogc.2013.05.001.
- Spence D. Bad medicine: clinical breast examination. *BMJ*. 2012 Oct 11;345:e6789. doi: 10.1136/bmj.e6789.
- Puschel K, Thompson B. Mammogram screening in Chile: using mixed methods to implement health policy planning at the primary care level. *Breast*. 2011 Apr;20 Suppl 2:S40-5. doi: 10.1016/j.breast.2011.02.002.
- Paci E, EUROSCREEN Working Group. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen*. 2012;19 Suppl 1:5-13.
- Broeders M, Moss S, Nyström L, Njor S, Jonsson H, Paap E, *et al.* The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen*. 2012;19 Suppl 1:14-25.
- Giordano L, von Karsa L, Tomatis M, Majek O, de Wolf C, Lancucki L, *et al.* Mammographic screening programmes in Europe: organization, coverage and participation. *J Med Screen*. 2012;19 Suppl 1:72-82.
- 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. *Ann Oncol*. 2008 Apr;19(4):614-22.
- Bond M, Pavey T, Welch K, Cooper C, Garside R, Dean S, *et al.* Systematic review of the psychological consequences of false-positive screening mammograms. *Health Technol Assess*. 2013 Mar;17(13):1-170, v-vi. doi: 10.3310/hta17130.
- Salas D, Ibáñez J, Román R, Cuevas D, Sala M, Ascunce N, *et al.* Effect of start age of breast cancer screening mammography on the risk of false-positive results. *Prev Med*. 2011 Jul-Aug;53(1-2):76-81. doi: 10.1016/j.ypmed.2011.04.013.
- Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M, *et al.* The benefits and harms of breast cancer screening: an independent review. *Br J Cancer*. 2013 Jun 11;108(11):2205-40. doi: 10.1038/bjc.2013.177.
- Jørgensen K. Mammography screening. Benefits, harms, and informed choice. *Dan Med J*. 2013 Apr;60(4):B4614.
- Muir JA. Evidence-Based Healthcare: How to Make Health Policy and Management Decisions. Churchill Livingstone. New York; 2001.
- Guyatt G, Oxman A, Akl E, Kunz R, Vist G, Brozek J, *et al.* GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011 Apr;64(4):383-94. doi: 10.1016/j.jclinepi.2010.04.026.
- GRADE Working Group [Internet]. c2005-2009 [citado el 14 de enero del 2015]. Disponible en: <http://www.gradeworkinggroup.org/>
- Declaran de interés nacional la Atención Integral del Cáncer y Mejoramiento del Acceso a los Servicios Oncológicos en el Perú y dictan otras medidas. Decreto Supremo N° 009-2012-SA (3 de noviembre de 2012).
- Instituto Nacional de Enfermedades Neoplásicas. Coalición Multisectorial "Perú contra el Cáncer". Plan Nacional para el Fortalecimiento de la Prevención y Control del Cáncer en el Perú. Lima: INEN; 2006.
- Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev*. 2013 Jun 4;6:CD001877. doi: 10.1002/14651858.CD001877.pub5.
- Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet*. 2012 Nov 17;380(9855):1778-86. doi: 10.1016/S0140-6736(12)61611-0.
- The Canadian Task Force on Preventive Health Care. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. *CMAJ*. 2011 Nov 22;183(17):1991-2001. doi: 10.1503/cmaj.110334.
- US Preventive Services Task Force. Screening for breast cancer: US Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2009 Nov 17;151(10):716-26, W-236. doi: 10.7326/0003-4819-151-10-200911170-00008.
- Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med*. 2002 Sep 3;137(5 Part 1):305-12.
- Miller AB, To T, Baines C-J, Wall C. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-

- 59 years. *J Natl Cancer Inst.* 2000 Sep 20;92(18):1490-9.
30. Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet.* 2002 Mar 16;359(9310):909-19.
 31. Tabar L, Gad A, Holmberg L, Ljungquist U. Significant reduction in advanced breast cancer. Results of the first seven years of mammography screening in Kopparberg, Sweden. *Diagn Imaging Clin Med.* 1985;54(3-4):158-64.
 32. Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, et al. Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. *BMJ.* 1988 Oct 15;297(6654):943-8.
 33. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst.* 1982 Aug;69(2):349-55.
 34. Fagerberg G, Baldetorp L, Gröntoft O, Lundström B, Månson JC, Nordenskjöld B. Effects of repeated mammographic screening on breast cancer stage distribution. Results from a randomised study of 92 934 women in a Swedish county. *Acta Radiol Oncol.* 1985 Nov-Dec;24(6):465-73.
 35. Frisell J, Lidbrink E, Hellström L, Rutqvist LE. Followup after 11 years-update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Res Treat.* 1997 Sep;45(3):263-70.
 36. Moss S, Cuckle H, Evans A, Johns L, Waller M, Bobrow L, et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet.* 2006 Dec 9;368(9552):2053-60.
 37. Borenstein M, Hedges L, Higgins J, Rothstein H. *Introduction to Meta-Analysis.* Chichester: John Wiley & Sons.; 2009.
 38. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L; U.S. Preventive Services Task Force. Screening for breast cancer: Systematic Evidence Review Update for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2009 Nov 17;151(10):727-37, W237-42. doi: 10.7326/0003-4819-151-10-200911170-00009.
 39. Ioannidis J, Greenland S, Hlatky MA, Khoury MJ. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet.* 2014 Jan 11;383(9912):166-75. doi: 10.1016/S0140-6736(13)62227-8.
 40. Njor S, Nyström L, Moss S, Paci E, Broeders M, Segnan N, et al. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. *J Med Screen.* 2012;19 Suppl 1:33-41.
 41. Zackrisson S, Andersson I, Janzon L, Manjer J, Garne J. Rate of overdiagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ.* 2006 Mar 25;332(7543):689-92. Epub 2006 Mar 3.
 42. Puliti D, Duffy S, Miccinesi G, de Koning H, Lyng E, Zappa M, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen.* 2012;19 Suppl 1:42-56.
 43. Zelle SG, Vidaurre T, Abugattas JE, Manrique JE, Sarria G, Jeronimo J, et al. Cost-effectiveness analysis of breast cancer control interventions in Peru. *PLoS One.* 2013 Dec 10;8(12):e82575. doi: 10.1371/journal.pone.0082575.
 43. Gutiérrez-Aguado A. Costo utilidad de intervenciones preventivas para cáncer de mama en el Perú. *Rev Peru Ginecol Obstet.* 2012;58(4):253-61.
 45. World Health Organization. *Choosing interventions that are cost effective* [Internet]. Geneva: WHO; c2015 [citado el 22 de mayo del 2014]. Disponible en: <http://www.who.int/choice/en/>
 46. Huaman MA, Kamimura-Nishimura KI, Kanamori M, Siu A, Lescano AG. Validation of a susceptibility, benefits, and barrier scale for mammography screening among Peruvian women: a cross-sectional study. *BMC Womens Health.* 2011 Dec 7;11:54. doi: 10.1186/1472-6874-11-54.
 47. Sánchez Castro AC, Sáenz Orellana ADP. Conocimientos, actitudes y prácticas del autoexamen de mama en mujeres de 15 - 45 años que acuden a los consultorios de Planificación Familiar y Papanicolaou del Hospital Nacional Daniel Alcides Carrión, periodo agosto - diciembre del 2010. [tesis] Universidad Nacional Mayor de San Marcos, Facultad de Medicina. Lima, Perú; 2011.
 48. Vela KV, Alarcón-Rozas AE. Estudio piloto sobre actitudes preventivas para cáncer de mama en mujeres de Sullana. *Rev Med Hered.* 2002;13(4):131-4.
 49. Romani F, Gutiérrez C, Ramos-Castillo J. Autoexamen de mama en mujeres peruanas: prevalencia y factores sociodemográficos asociados. Análisis de la Encuesta Demográfica de Salud Familiar (ENDES). *An Fac Med.* 2011;72(1):23-31.
 50. Anderson BO, Yip CH, Smith RA, Shyyan R, Sener SF. Guideline implementation for breast healthcare in low-income and middle-income countries. *Cancer.* 2008 Oct 15;113(8 Suppl):2221-43. doi: 10.1002/ncr.23844.
 51. Claxton K. Centre for Health Economics. *A Pilot Study of Value of Information Analysis to Support Research Recommendations for the National Institute for Health and Clinical Excellence.* CHE Research Paper 4. York, UK: University of York 2005.
 52. Knaul FM, Nigenda G, Lozano R, Arreola-Ornelas H, Langer A, Frenk J. Breast cancer in Mexico: an urgent priority. *Salud Publica Mex.* 2009;51 Suppl 2:s335-44.
 53. Nigenda G, Caballero M, González-Robledo LM. Barreras de acceso al diagnóstico temprano del cáncer de mama en el Distrito Federal y en Oaxaca. *Salud Publica Mex.* 2008;51(2):s254-62.
 54. Goss P, Lee B, Badovinac-Crnjevic T, Strasser-Weippl K, Chavarri-Guerra Y, St Louis J, et al. Planning cancer control in Latin America and the Caribbean. *Lancet Oncol.* 2013 Apr;14(5):391-436. doi: 10.1016/S1470-2045(13)70048-2.

Correspondencia: Margarita Posso
 Dirección: : C/ St. Antoni M. Claret 167
 Centro Cochrane Iberoamericano - Hospital
 de la Santa Creu i Sant Pau, Pabellón 18, CP
 08025, Barcelona, España
 Teléfono: (+34) 935537814
 Correo electrónico: mposso@cochrane.es