

Rendimiento diagnóstico de la mamografía digital en el cribado del cáncer de mama

Diagnostic performance of
digital mammography in breast
cancer screening. *Executive
Summary*

**INFORMES DE EVALUACIÓN DE TECNOLOGÍAS SANITARIAS
AETSA 2007/09**

INFORMES, ESTUDIOS E INVESTIGACIÓN



MINISTERIO
DE CIENCIA
E INNOVACIÓN



Ministerio de Ciencia e Innovación
AIC Agencia de Evaluación
de Tecnologías Sanitarias
Instituto
de Salud
Carlos III



MINISTERIO
DE SANIDAD
Y POLÍTICA SOCIAL



JUNTA DE ANDALUCÍA
CONSEJERÍA DE SALUD

“Impreso en cartulina  y papel fabricado con pasta libre de madera”

Rendimiento diagnóstico de la mamografía digital en el cribado del cáncer de mama

Diagnostic performance of
digital mammography in breast
cancer screening. *Executive
summary*

Márquez Cruz, María Dolores

Rendimiento diagnóstico de la mamografía digital en el cribado del cáncer de mama. María Dolores Márquez Cruz, Soledad Márquez Calderón.— Sevilla: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía, 2009.

100 p; 24 cm. (Colección: Informes, estudios e investigación. Ministerio de Sanidad y Política Social. Serie: Informes de Evaluación de Tecnologías Sanitarias)

1. Neoplasias de la mama / diagnóstico 2. Cribado / métodos 3 Mama / radiografía I. Márquez Calderón, Soledad. II Andalucía. Agencia de Evaluación de Tecnologías Sanitarias III. España. Ministerio de Sanidad y Política Social.

Autoras: María Dolores Márquez Cruz, Soledad Márquez Calderón.

Edita: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía

Avda. de la Innovación s/n.

Edificio RENTA SEVILLA, 2ª planta.

41020 Sevilla

España – Spain

Este documento se ha realizado en el marco de colaboración previsto en el Plan de Calidad para el Sistema Nacional de Salud elaborado por el Ministerio de Sanidad y Política Social, al amparo del convenio de colaboración suscrito por el Instituto de Salud Carlos III, organismo autónomo del Ministerio de Ciencia e Innovación y la Agencia de Evaluación de Tecnologías Sanitarias de Andalucía

© de la presente edición: Ministerio de Sanidad y Política Social.

© de los contenidos: Consejería de Salud – JUNTA DE ANDALUCÍA

ISBN: 978-84-96990-40-1

NIPO: 477-10-051-2

Depósito Legal: SE-1784-2010

Imprime: Egea Impresores, Sevilla.

Este documento puede ser reproducido en todo o en parte, por cualquier medio, siempre que se cite explícitamente su procedencia.

<http://publicaciones.administración.es>

<http://www.juntadeandalucia.es/salud/aetsa>

Rendimiento diagnóstico de la mamografía digital en el cribado del cáncer de mama

Diagnostic performance of
digital mammography in breast
cancer screening. *Executive
summary*



MINISTERIO
DE CIENCIA
E INNOVACIÓN



Instituto
de Salud
Carlos III

Ministerio de Ciencia e Innovación

AIE Agencia de Evaluación
TIS de Tecnologías Sanitarias



MINISTERIO
DE SANIDAD
Y POLÍTICA SOCIAL



Plan de Calidad
para el Sistema Nacional
de Salud



JUNTA DE ANDALUCÍA
CONSEJERÍA DE SALUD

Conflicto de Interés

Las autoras declaran que no tienen intereses que puedan competir con el interés primario y los objetivos de este informe e influir en su juicio profesional al respecto.

Agradecimientos

Este trabajo se ha beneficiado de forma importante de las aportaciones de las siguientes revisoras externas:

- Dolores Salas Trejo. Jefa de la Oficina del Plan de Cáncer. Conselleria de Sanitat. Comunidad Valenciana.
- Mercedes Acebal Blanco. Servicio de Radiodiagnóstico. Hospital Costa del Sol. Marbella.

La Agencia de Evaluación de Tecnologías Sanitarias de Andalucía y las autoras agradecen a las revisoras de este texto el esfuerzo realizado, su dedicación y sus aportaciones.

Los contenidos del informe son responsabilidad de las autoras, procediendo el eximente habitual en el caso de las revisoras.

Índice

| | |
|---|----|
| Resumen ejecutivo | 13 |
| Executive summary | 17 |
| 1 Introducción | 21 |
| 1.1 Epidemiología y características del cáncer de mama | 21 |
| 1.2 Cribado del cáncer de mama con mamografía | 22 |
| 1.3 Descripción de la tecnología: mamografía digital | 24 |
| 1.4 Estado del conocimiento y justificación de la revisión | 25 |
| 2. Objetivos | 27 |
| 3 Material y Métodos | 29 |
| 3.1 Tipo de estudio | 29 |
| 3.2 Búsqueda bibliográfica | 29 |
| 3.3 Selección de los estudios: criterios de inclusión y exclusión | 31 |
| 3.4 Evaluación crítica de los estudios seleccionados | 32 |
| 3.5 Extracción de la información | 32 |
| 3.6 Síntesis de los resultados | 33 |
| 4 Resultados | 35 |
| 4.1. Selección de artículos según criterios de inclusión | 35 |
| 4.2. Descripción de los artículos | 36 |
| 4.3 Valoración crítica de los artículos | 48 |
| 4.4 Resultados principales de rendimiento diagnóstico | 51 |
| 4.5. Síntesis de los resultados | 63 |
| 5 Discusión | 69 |
| 5.1. Discusión de las limitaciones y aspectos metodológicos | 70 |
| 5.2 Discusión resultados | 71 |
| 5.3 Implicaciones para la práctica clínica y la investigación | 74 |
| 6 Conclusiones | 77 |

| | |
|------------------------------------|----|
| Referencias | 79 |
| Anexo 1: análisis económico | 85 |
| 1. Introducción | 85 |
| 2. Objetivos | 87 |
| 3. Método | 87 |
| 4. Resultados y Discusión | 90 |
| 5. Conclusiones | 95 |
| 6. Recomendaciones | 95 |
| Referencias | 97 |
| Anexo 2: QUADAS | 99 |

Índice de tablas y Figuras

| | |
|---|----|
| Tabla 1. Número de artículos hallados en cada base de datos..... | 31 |
| Tabla 2. Fases de la selección..... | 35 |
| Tabla 3. Artículos EXCLUIDOS..... | 35 |
| Tabla 4. Descripción de las características del estudio de cohorte realizado en dos centros de Estados Unidos por Lewin et al: Artículos derivados del estudio..... | 43 |
| Tabla 5. Descripción de las características de los tres estudios realizados en el marco del programa de cribado de Noruega: artículos derivados de estos estudios que aportan datos sin seguimiento de las mujeres..... | 44 |
| Tabla 6. Descripción de las características de los dos estudios realizados en el marco del programa de cribado de Noruega que aportan datos de seguimiento de las mujeres..... | 45 |
| Tabla 7. Descripción de las características del estudio DMIST (Digital Mammographic Imaging Screening Trial)..... | 46 |
| Tabla 8. Descripción de las características del estudio realizado en el marco del programa de cribado de Florencia..... | 47 |
| Tabla 9 . Evaluación de la calidad de los estudios seleccionados a través de la Herramienta QUADAS (19)..... | 50 |
| Tabla 10. Parámetros de rendimiento diagnóstico de la mamografía digital versus la mamografía convencional en el estudio de Lewin..... | 51 |
| Tabla 11. Parámetros de rendimiento diagnóstico de la mamografía digital versus la mamografía convencional en los estudios noruegos (sin considerar seguimiento)... | 55 |
| Tabla 12. Parámetros de rendimiento diagnóstico de la mamografía digital versus la mamografía convencional en los estudios noruegos (considerando el seguimiento)..... | 59 |
| Tabla 13. Área bajo la curva ROC para la mamografía digital y convencional basado en la escala de malignidad de 7 puntos de la mamografía..... | 60 |
| Tabla 14 . Exactitud diagnóstica de MC y MD con uso de la escala BI-RADS®* después de 365 días de seguimiento..... | 61 |
| Tabla 15 . Exactitud diagnóstica de MC y MD con uso de la escala de malignidad de 7 puntos* después de 455 días de seguimiento..... | 61 |
| Tabla 16. Parámetros de rendimiento diagnóstico de la mamografía digital versus la mamografía convencional en el estudio de Del Turco (43)..... | 63 |

Resumen ejecutivo

Título: Rendimiento diagnóstico de la mamografía digital en el cribado del cáncer de mama.

Autoras: María Dolores Márquez Cruz, Soledad Márquez Calderón.

ANTECEDENTES

El cáncer de mama constituye uno de los problemas de salud más importantes para las mujeres de los países industrializados, debido a su alta incidencia, a su mortalidad y a la repercusión personal y social que comporta. Hasta el momento, no se dispone de estrategias eficaces para la prevención primaria del cáncer de mama, siendo la prevención secundaria, mediante la detección precoz del cáncer de mama y el tratamiento adecuado, el instrumento fundamental para el control de la enfermedad hoy día. Durante más de tres décadas, la prueba utilizada en el cribado del cáncer de mama ha sido la mamografía analógica o convencional, pero en los últimos años ha surgido como alternativa la mamografía digital de campo completo. Este informe se plantea por el interés que tiene revisar el conocimiento sobre las ventajas y limitaciones de las nuevas alternativas de pruebas de cribado del cáncer de mama.

OBJETIVOS

El objetivo principal es revisar la evidencia científica disponible acerca del rendimiento diagnóstico de la mamografía digital (MD) comparada con la mamografía analógica o convencional (MC) en el cribado de cáncer de mama.

A partir de la bibliografía hallada en la revisión para dar respuesta al objetivo principal, se planteó el objetivo secundario de revisar la evidencia sobre la comparación de MD y MC en términos de resultados en salud.

METODOLOGÍA

Se ha realizado una revisión sistemática de la literatura. Se partió de una revisión sistemática previa que cubría la literatura publicada hasta 2005 y se recuperaron los artículos originales incluidos en ella. Posteriormente se realizó una búsqueda que cubrió el periodo 2005-2007 en las bases referenciales MEDLINE, PRE-MEDLINE y EMBASE. Se han explorado también otras fuentes como Cochrane Library Plus, Centre for Reviews and Dissemination, International Network of Agencies for Health Technolo-

gy Assessment, National Guidelines Clearinghouse, ECRI Institute, Sumsearch y Tripdatabase.

Los principales criterios de inclusión de artículos fueron: población (mujeres adultas asintomáticas), intervención y comparación (cribado de cáncer de mama con MD versus con MC) y medidas de resultados (cualquier parámetro relacionado con el rendimiento diagnóstico, resultados intermedios – como la tasa de rellamadas- o resultados finales –como mortalidad-).

Tanto la evaluación crítica de los artículos originales seleccionados como la extracción de resultados se realizó por dos evaluadoras de manera independiente, resolviéndose las discrepancias por consenso. Para evaluar la calidad se utilizó la herramienta QUADAS y se clasificaron los estudios según la validez del diseño y la existencia de seguimiento.

La síntesis de resultados se realizó de forma cualitativa.

A partir de la revisión se consideró la oportunidad de explorar los aspectos económicos relacionados con las dos alternativas. Para no distorsionar los objetivos del informe, todo lo relativo a esta aproximación económica se encuentra detallado en el Anexo I.

RESULTADOS

A partir de la revisión sistemática hallada en la búsqueda preliminar se incluyeron 6 artículos originales. En la búsqueda definitiva (período 2005-2007) se obtuvieron 271 referencias, quedando por tanto un cómputo final de 277 documentos. Tras eliminar duplicados y hacer una selección por lectura de título y resumen, quedaron 24 artículos. Una vez leídos a texto completo, se desecharon 13 por no cumplir los criterios de inclusión y quedaron finalmente 11 artículos originales (que procedían de 6 estudios).

Todos los estudios encontrados se centraron en la valoración del rendimiento diagnóstico de las mamografías y resultados intermedios del cribado. Para graduar la importancia de los resultados finales según la calidad de los estudios, éstos se han clasificado en tres grupos según la validez del diseño (comparabilidad de los grupos y existencia de seguimiento tras el cribado), ya que la herramienta QUADAS no discriminó mucho en este caso (todos los estudios estaban en un rango entre 7 y 9 respuestas afirmativas de las 14 preguntas de QUADAS).

Comparación del rendimiento diagnóstico de MD y MC en los estudios con mayor validez (cohortes comparables y seguimiento):

En este grupo se incluyeron resultados de 3 estudios: un ensayo clínico con asignación aleatoria (Oslo II, con 23.929 mujeres) y dos estudios de cohorte única (a cada mujer se le hacía una MC y una MD: Estudios Oslo I – 3.683 mujeres- y DMIST -49.258 mujeres-). La sensibilidad de la MD fue

superior a la de la MC en el único ensayo clínico, no hallándose diferencias estadísticamente significativas en los dos estudios de cohorte única con seguimiento. Asimismo, no se encontraron diferencias entre MD y MC en la tasa de detección de cánceres de intervalo, en la tasa de falsos negativos ni en el valor predictivo negativo.

En dos de los estudios, la especificidad de la MD fue menor que la de la MC y la tasa de falsos positivos fue mayor con la MD que con la MC, no hallándose diferencias en el tercer estudio (a pesar de ser el de mayor muestra). Los distintos resultados sobre especificidad podrían relacionarse con la distinta forma de interpretar las mamografías y con el contexto de los estudios (programas poblacionales en los dos con menor especificidad de la MD y cribado oportunista en el estudio que no halló diferencias).

Sólo el estudio de cribado oportunista dio resultados de rendimiento diagnóstico tras el seguimiento desagregados por edad y otras variables. En las mujeres menores de 50 años, perimenopaúsicas y con mamas densas, la MD tuvo mayor área bajo la curva ROC y sensibilidad que la MC; y no hubo diferencias en especificidad y valor predictivo positivo.

Comparación del rendimiento diagnóstico de MD y MC en las publicaciones que no incluyen datos de seguimiento tras el cribado:

Los dos estudios Oslo (I y II) aportaron también resultados del cribado sin tener en cuenta el seguimiento. Además, hay un estudio de cohorte única (con 4.489 mujeres) y dos estudios con cohortes elegidas de forma separada – una cribada con MD y otra con MC- (con 343.002 y 28.770 mujeres) que aportaron datos de rendimiento diagnóstico de tipo transversal. Todos estos estudios coinciden en que la tasa de detección de cáncer de mama en una ronda de cribado no es significativamente diferente cuando se usa MD o MC.

Sólo dos de los estudios aportaron datos sobre el porcentaje de carcinomas in situ, presentando resultados discordantes: uno no halló diferencias entre MD y MC y otro encontró un mayor porcentaje con MD.

De los cinco estudios que aportaron datos sobre el valor predictivo positivo, los tres con un diseño de mayor validez (grupos comparables) y uno de los estudios con cohortes separadas concluyeron que no había diferencias entre MD y MC. Sólo un estudio con cohortes separadas encontró mayor valor predictivo positivo con la MD.

Comparación de la MD y MC en términos de otros resultados:

No se han encontrado estudios que aporten datos de comparación entre la MD y la MC en términos de mortalidad u otros resultados en salud.

Los resultados sobre la tasa de rellamadas tras una mamografía de cribado fueron divergentes entre los distintos estudios. Los estudios más estrictos en la lectura de las mamografías no encontraron diferencias entre

MD y MC, excepto en el subgrupo de mujeres de 50-69 años del único ensayo clínico, donde la tasa de rellamadas fue mayor con la MD. En el resto de estudios los resultados fueron opuestos: mayor tasa de rellamadas con MD (1 estudio) y con MC (1 estudio).

El porcentaje de biopsias se estudió en dos trabajos de cohorte única, siendo diferentes los resultados entre ellos: el estudio de mayor puntuación en la escala de validez no encontró diferencias entre MD y MC, mientras que el de menor validez encontró mayor tasa de biopsias con MC.

CONCLUSIONES

No se han encontrado diferencias importantes entre la MD y la MC en términos de sensibilidad, tasa de detección de cáncer y valor predictivo positivo, independientemente de la validez del diseño. Los resultados sobre la especificidad, porcentaje de carcinomas in situ, tasa de rellamadas y porcentaje de biopsias presentan mayores discrepancias entre los estudios, por lo que no puede concluirse que haya una clara ventaja de un tipo de mamografía sobre otro. Por último, no se han encontrado datos sobre posibles diferencias de la MD y la MC en términos de mortalidad y otros resultados de salud.

Executive summary

Title: Diagnostic performance of digital mammography in breast cancer screening.

Authors: María Dolores Márquez Cruz, Soledad Márquez Calderón.

BACKGROUND

Breast cancer is one of the main health problems for Western countries' women due to its high incidence and mortality, and to the personal and social impact that it entails. Up to the moment, there are no effective strategies available for the primary prevention of breast cancer. For this reason, secondary prevention by early detecting breast cancer and delivering appropriate treatment is the main tool to control the disease nowadays. For more than three decades, the test employed to screen breast cancer has been analogical mammography. However, full field digital mammography has emerged as an alternative in the last few years. The present report has been drawn up due to the increasing interest in assessing the evidence on the advantages and limitations of new technologies for breast cancer screening.

OBJECTIVES

The main objective of the present report is to review the scientific evidence available on the diagnostic performance of digital mammography (DM) as compared to analogical or traditional mammography (TM) in breast cancer screening.

From the bibliography found in the review to approach the main objective, it was posed a secondary objective: to assess the scientific evidence on the comparison between screening with DM and TM in terms of health outcomes.

METHODS

A systematic review of the scientific literature was carried out. It was started from a previous systematic review covering the literature published up to 2005; and the original papers included in it were retrieved. Afterwards a search was run on the reference bases MEDLINE, PRE-MEDLINE and EMBASE. Other sources that have been explored are the following: Cochrane Library Plus, Centre for Reviews and Dissemination, INAHTA, National Guidelines Clearinghouse, ECRI Institute, Sumsearch and Tripdatabase.

The main inclusion criteria of the papers were: population (asymptomatic adult women), intervention and comparison (breast cancer screening

with DM vs. with TM) and outcome measures (any parameter related to the diagnostic performance, intermediate results -like recall rate-, or final outcome –as mortality).

Critical assessment of the selected original papers and data extraction were conducted by two assessors separately; and the disagreements were solved by consensus. To evaluate the quality, QUADAS tool was used. The studies were classified depending on the validity of the design and the existence of follow-up.

The data were summarised qualitatively.

RESULTS

Six original papers were obtained from the systematic review found in the preliminary search. In the definite search (period 2005-2007), 271 additional references were retrieved. After duplicates had been discarded and inclusion criteria were applied to all papers by means of reading the titles and abstracts, 24 documents remained. Once they were fully read, 13 were discarded as they did not meet the inclusion criteria, remaining 11 original papers. These 11 papers reported results from 6 studies.

All the studies found were focused on assessing the diagnostic performance of the mammographies and intermediate outcomes of screening. The studies were classified according to the design's validity (comparability of the groups and existence of follow-up after screening), since QUADAS tool did not discriminate much in this case (all the studies were in a range between 7 and 9 positive answers of the 14 questions of QUADAS). Results are reported separately for each group of studies.

Comparison of DM and TM in terms of diagnostic performance, based on the most valid studies (with comparable cohorts and follow-up after screening):

In this group, there were included results from 3 studies: a randomised clinical trial (Oslo II, with 23.929 women) and two single-cohort studies (each woman received a TM and a DM: Oslo I – 3.683 women- and DMIST -49.258 women). The sensitivity of DM was higher to that of TM in the randomised clinical trial, but no statistically significant differences were found in the two single-cohort studies with follow-up. Likewise no differences between DM and TM were found in terms of interval cancer detection rate, false-negative rate or negative predictive value.

In two of the studies the specificity of DM was lower than that of TM and the false-positive rate was higher with DM. However, no differences between both types of mammographies were found in the third study (despite it was the one with the highest sample size). These differences between the studies in terms of the results on specificity could be related to different

forms of interpreting the mammographies and to the studies' context (population programmes in the two studies which showed lower specificity of DM and opportunistic screening in the study that did not find any difference between mammographies).

Only one of the studies with follow-up data (DMIST, based on opportunistic screening of breast cancer) reported results segmented by age and other variables. In women aged less than 50, who were perimenopausal or had dense breasts, DM showed a bigger area under the ROC curve and a higher sensitivity than TM; and there were no differences in specificity and positive predictive value.

Comparison of DM and TM in terms of diagnostic performance, based on the papers that do not include follow-up data after screening:

The two Oslo studies (I and II) also provided results from the screening without taking into account the follow-up. Moreover there were a single-cohort study (with 4.489 women) and two double-cohort studies – cohorts selected separately: one of them was screened with DM and the other with TM - (with 343.002 and 28.770 women) that provided data on diagnostic performance without follow-up after screening. All these studies show that the rate of breast cancer in a screening round is not significantly different when either DM or TM is used.

Only two of the studies provided data on the percentage of carcinomas in situ, but results were different between them: one did not find differences between DM and TM, and the other one reported higher percentage with DM.

Of the five studies that provided data on positive predictive value, the three considered to have high validity (with comparable groups) and another one (with two separate cohorts) concluded that there were no differences between DM and TM. Just one study with separate cohorts reported higher positive predictive value with DM.

Comparison between DM and TM in terms of other outcomes:

No studies were found that provided comparison data between DM and TM in terms of either mortality or any other health outcomes.

The results on the comparison of DM and TM in terms of recall rate after a screening mammography were divergent among the studies. The strictest studies in the reading of the mammographies did not find differences between DM and TM, except for the subgroup of women aged 50-69 in the only one clinical trial (higher recall rate after DM). In the rest of studies, the results were different: the highest recall rate was after DM in one study and after TM in another one.

The percentage of biopsies was reported in two single-cohort studies. The results were different among them: the study that obtained a higher

score in the validity scale (QUADAS) did not find differences between DM and TM while the one with lower validity found a higher biopsy rate after TM.

CONCLUSIONS

No important differences were found between DM and TM in terms of sensitivity, cancer detection rate and positive predictive value, regardless of the design's validity. The results on specificity, percentage of carcinomas in situ, recall rate and percentage of biopsies showed greater discrepancies among the studies. Therefore, it cannot be concluded that there is a clear advantage of a type of mammography over the other one. Finally, there was not found data about the comparison of screening with DM and TM in terms of mortality and other health outcomes.

1 Introducción

1.1 Epidemiología y características del cáncer de mama

El cáncer de mama constituye uno de los problemas de salud más importantes para las mujeres de los países industrializados, debido a su alta incidencia, a su mortalidad y a la repercusión personal y social que comporta.

Se trata de la neoplasia más frecuente en la mujer y, considerando los dos sexos conjuntamente, es el cáncer más frecuente en el mundo después del de pulmón. Además, se trata de la primera causa de muerte por cáncer en la población femenina. En España, aunque la incidencia y mortalidad por cáncer de mama son menores que en otros países europeos, las cifras siguen siendo altas. Así, la incidencia ajustada por edad es de 70 por 100.000 mujeres al año, y anualmente fallecen unas 5.700 mujeres por cáncer de mama, lo que representa el 16,7% de todos los fallecimientos por cáncer del sexo femenino y una tasa ajustada de mortalidad de 21,42 por 100.000 mujeres. Asimismo, en España la supervivencia global de los casos de cáncer de mama (78% a los 5 años) es algo superior a la media europea (76%) (1, 2).

Hasta el momento, no se dispone de estrategias eficaces para la prevención primaria del cáncer de mama, ya que los principales factores de riesgo de esta enfermedad están relacionados claramente con el sexo femenino, la edad superior a 50 años y los antecedentes familiares, todos ellos factores difícilmente modificables. Por este motivo, es la prevención secundaria, mediante la detección precoz del cáncer de mama y el tratamiento adecuado, la que constituye hoy día el instrumento fundamental para el control de la enfermedad (2).

En estos últimos años, la tasa de mortalidad por cáncer de mama ha ido disminuyendo, atribuyéndose dicho beneficio a los programas de detección precoz y a los avances en el tratamiento sistémico de la enfermedad. (3).

1.2 Cribado del cáncer de mama con mamografía

La mamografía es la prueba que se utiliza habitualmente en la detección precoz del cáncer de mama. Su aceptabilidad, mínimos efectos adversos y bajo coste de aplicación, han facilitado la extensión de su uso en el cribado poblacional. El estudio puede hacerse con una única proyección oblicua medio lateral de ambas mamas o bien con dos proyecciones (oblicua medio lateral y cráneo caudal) bilaterales. Actualmente existe acuerdo en que la utilización conjunta de dos proyecciones como prueba de cribado incrementa la sensibilidad y la tasa de detección de cáncer, y puede reducir la tasa de repetición de mamografías (4).

Después de varios años en que el cribado del cáncer de mama venía haciéndose en casi todos los países, en el año 2001 se desató una polémica al respecto, tras la publicación de una revisión de la literatura sobre el cribado de cáncer de mama (5). En ella, los autores concluían que no existían evidencias suficientes de que éste produjese beneficios en la supervivencia de la mujer. Unos años más tarde, en 2006, los mismos autores publicaron una revisión actualizada, en la que añadían nuevas evidencias y concluían, a partir de ensayos clínicos realizados en Europa y Estados Unidos, que la mamografía de cribado reduce entre un 15% y un 20% la mortalidad por cáncer de mama con respecto a la población sin cribado rutinario (6). Aunque aún pueden existir controversias sobre la magnitud de los efectos del cribado, ahora la evidencia es suficiente para concluir que el cribado reduce la mortalidad por cáncer de mama. Con todo, es necesario señalar que de todos los métodos utilizados para la detección precoz de este cáncer, sólo se ha demostrado la eficacia para disminuir la mortalidad en el caso de la mamografía periódica utilizada en el contexto de programas poblacionales de cribado.

Como se ha comentado anteriormente, la mayoría de los países han introducido programas de cribado de cáncer de mama con mamografía. Entre los distintos países y regiones, así como entre las distintas guías de práctica clínica, existen diferencias sobre la edad a la que se debe comenzar el cribado y sobre la periodicidad de la mamografía. Según la U.S. Preventive Services Task Force (7) se recomienda realizar el cribado de cáncer de mama con mamografía, con o sin examen clínico de la mama, cada 1-2 años para mujeres de 40 años en adelante. En la revisión sistemática que soporta las recomendaciones de la U.S. Preventive Services Task Force se encuentran evidencias fuertes de que el cribado con mamografía cada 12-33 meses reduce

significativamente la mortalidad por cáncer de mama en mujeres de 50-69 años y evidencias débiles para las mujeres de 40-49 años (8). Sin embargo, la Canadian Task Force no recomienda la mamografía de cribado en mujeres de 40-49 años, por considerar que las evidencias sobre su efectividad no son suficientes (9). Asimismo, en España, el Programa de Actividades Preventivas y Promoción de la Salud de la Sociedad Española de Medicina Familiar y Comunitaria, recomienda iniciar el cribado de mama con mamografía en las mujeres de 50 años y continuarlo hasta los 70 (10).

El menor beneficio del cribado del cáncer de mama con mamografía en las mujeres de 40-49 años puede explicarse por diversos factores. Por una parte, las características de la mama en mujeres más jóvenes dificultan la interpretación de la mamografía. Por otra parte, el cáncer de mama tiene una incidencia menor y un carácter más agresivo en estas edades (11). En el caso de las mujeres de edad avanzada, no se dispone de suficientes datos para evaluar la efectividad del cribado, ya que la mayoría de los ensayos clínicos no incluyen a mujeres mayores de 64 años de edad (2), con lo cual la evidencia científica disponible es aún más débil que para las mujeres jóvenes (11).

Por último, hay que tener siempre en cuenta que los programas de cribado no solo tienen beneficios. Según la revisión sistemática de Gøtzsche (6), de cada 2.000 mujeres sanas que participan durante 10 años en un programa de cribado, a 1 se le prolongará la vida y a otras 10 se les diagnosticará una lesión intrascendente que no se hubiera diagnosticado sin cribado y que por tanto, recibirán un tratamiento innecesario. No pueden minimizarse las consecuencias de los falsos positivos y negativos del cribado (2). Los falsos positivos conllevan ansiedad y afectan a la calidad de vida de las mujeres, además de provocar exploraciones complementarias (mamografía de diagnóstico, ecografía o biopsia) y costes adicionales (12,13). Los falsos negativos pueden provocar un retraso en el diagnóstico, debido a la sensación de seguridad de la mujer ante una prueba negativa.

Para minimizar los efectos adversos del cribado, la mamografía debe garantizar una alta calidad. Actualmente, aunque la mamografía analógica presenta múltiples ventajas, también tiene limitaciones, ya que tanto su sensibilidad como su especificidad no son óptimas. La sensibilidad de la mamografía obtenida en los programas de cribado varía entre un 85% y 95% y la especificidad es superior al 90%. El valor predictivo de un resultado positivo en la mamografía de cribado varía entre un 5% y un 10%. Es decir, sólo entre 5 y 10 mujeres de cada 100 con mamografía positiva son después confirmadas tras pruebas adicionales, como portadoras de un cáncer de mama. Este valor está muy condicionado por la prevalencia del cáncer de mama en la población que participa en el programa (4,11,14).

1.3 Descripción de la tecnología: mamografía digital

Durante más de tres décadas, la prueba utilizada en el cribado del cáncer de mama ha sido la mamografía analógica. En los últimos años, con el desarrollo de las nuevas técnicas de diagnóstico por imagen, ha surgido la mamografía digital. Es una de las últimas técnicas de imagen que ha incorporado la tecnología digital y se ha propuesto como alternativa a la mamografía convencional en el cribado y diagnóstico del cáncer de mama. La digitalización de otras técnicas de imagen es un elemento de presión para el uso de la mamografía digital, ya que facilitaría la organización de los servicios de radiología.

La mamografía digital difiere de la análoga en que produce imágenes de rayos X de la mama usando detectores digitales en vez de película de radiografía. La adquisición de la imagen, el procesado posterior, la lectura radiológica y el archivado se realizan por medios digitales, lo que permite optimizar cada uno de estos procesos. Estas características hacen atractiva a la mamografía digital como prueba de cribado, ya que facilita la detección y diagnóstico asistido por ordenador, la mejora de las imágenes (aumentar el contraste, eliminar artefactos, etc.), la teleasistencia (las imágenes pueden almacenarse, localizarse y transmitirse electrónicamente, lo que hace posible la consulta a larga distancia con otros especialistas), el transporte de las mamografías y su disponibilidad en los programas de cribado para comparar con mamografías previas, etc. (13,15). Desde el punto de vista de la gestión sanitaria, una de las ventajas más importantes estaría en la posibilidad de integrar toda la información en las redes digitales de información radiológica y hospitalaria. Asimismo, permitiría la eliminación de algunos costes asociados a la mamografía analógica (chasis, películas, líquidos, transporte y archivo de mamografías, etc.).

Sin embargo, la implantación de la mamografía digital también tiene inconvenientes y costes asociados. De hecho, no solo requiere un cambio de los mamógrafos, sino que para rentabilizar su rendimiento necesitaría de otras inversiones importantes como el cambio de los negatoscopios por estaciones de trabajo que requieren mayor resolución que otras técnicas, sistemas de almacenamiento, etc.

Existen dos tipos principales de mamógrafos digitales: directos e indirectos. Los detectores de la mamografía digital directa convierten sus lecturas en carga eléctrica, y utilizan normalmente el selenio como fotodetector, por su afinidad con los rayos X. La mamografía digital indirecta

(sistema digital con conversión indirecta) utiliza una capa centelleante que absorbe los rayos X (tierras raras: yoduro de cesio), genera fotones de luz y éstos son detectados por fotodiodos (paneles de silicio amorfo) (15,16).

1.4 Estado del conocimiento y justificación de la revisión

Como ya se ha comentado, el cáncer de mama es un importante problema de salud en nuestro medio y, dado que hasta el momento no es posible evitar la enfermedad, los esfuerzos de prevención se orientan al diagnóstico precoz. Por ello tiene interés el estudio de las ventajas y limitaciones de las nuevas alternativas de pruebas de cribado, como es el caso de la mamografía digital.

La evaluación de la mamografía digital, en comparación con la analógica, en el cribado de cáncer de mama mediante una revisión sistemática de la literatura ha sido propuesta por un grupo de expertos consultados por la Agencia de Evaluación de Tecnologías Sanitarias de Andalucía. Esto ha tenido lugar en el contexto de un ejercicio de priorización dentro del Programa Nacional de Evaluación de Tecnologías Sanitarias del Ministerio de Sanidad y Consumo (actualmente, Ministerio de Sanidad y Política Social).

Antes de iniciar la revisión se exploró si existían otras revisiones sistemáticas sobre el tema, encontrando tres.

La revisión más reciente se había publicado en 2007 (13) e incluía artículos que abarcaban desde 1989 hasta 2005 (sin especificar el mes de 2005 donde acababa la búsqueda). Uno de los objetivos de la revisión fue comparar la evidencia científica disponible sobre los resultados de la mamografía digital versus la analógica en el contexto del cribado de cáncer de mama. No encontraron diferencias estadísticamente significativas entre ambas tecnologías en términos de tasa de detección de cáncer, si bien los autores concluían que la mamografía digital tenía mejores resultados de rendimiento diagnóstico en el subgrupo de mujeres con mamas radiológicamente densas, menores de 50 años, y peri o premenopáusicas. Sin embargo, esta revisión no matizó los resultados de los estudios originales incluidos según la calidad de los mismos y los propios autores advierten de la necesidad de mejores evidencias y de la importancia de estudios de evaluación económica si se pretende una sustitución de la técnica.

Las otras dos revisiones sistemáticas eran más antiguas. Una de ellas se había publicado en 2004 (17), si bien incluía artículos hasta diciembre de 2002. Se centraba en la precisión de las nuevas tecnologías en el cribado de cáncer de mama, llegando a la conclusión de que no hay evidencias suficientes para apoyar aún el uso de nuevas tecnologías de imagen, incluida la mamografía digital. Por último, la tercera revisión sistemática hallada se había publicado en 2002 (18), y versa sobre la comparación técnica, clínica y económica de la mamografía digital y la convencional en el contexto del sistema sanitario canadiense. Llegan a la conclusión de que la mamografía digital ofrece algunos beneficios técnicos y clínicos, pero que la evidencia de esos beneficios es débil y la magnitud de los mismos no compensa los costes diferenciales; por ello recomiendan la tecnología de menor coste (mamografía convencional).

La mamografía digital es una tecnología nueva en la que puede esperarse una cierta intensidad en el ritmo en que aparecen las publicaciones. Dado que la revisión más reciente incluye estudios hasta 2005 (13), se ha considerado que tenía interés actualizar el tema hasta finales de 2007. Además, parece necesario incorporar algunos aspectos metodológicos en el enfoque de la revisión sistemática, fundamentalmente la matización de las conclusiones según la calidad de los estudios originales.

2. Objetivos

PRINCIPAL

Revisar la evidencia científica disponible acerca del rendimiento diagnóstico de la mamografía digital comparada con la mamografía analógica en el cribado de cáncer de mama.

SECUNDARIO

A partir de la bibliografía hallada en la revisión para dar respuesta al objetivo principal, se planteó el siguiente objetivo secundario:

Revisar la evidencia científica disponible acerca de los resultados en salud del cribado basado en la mamografía digital comparado con el cribado basado en la mamografía analógica.

3 Material y Métodos

3.1 Tipo de estudio

Se ha realizado una revisión sistemática de la literatura sobre la mamografía digital en comparación con la analógica, en el cribado del cáncer de mama, abarcando el período 1995-2007.

3.2 Búsqueda bibliográfica

3.2.1 Bases de datos y estrategia de búsqueda

Se realizó una primera búsqueda bibliográfica exploratoria enfocada a identificar revisiones sistemáticas. Se utilizaron las palabras clave “*digital mammography*” y “*breast cancer screening*” en las bases de datos de la Cochrane Library Plus, Centre for Reviews and Dissemination (CRD), International Network of Agencies for Health Technology Assessment (INAHTA), National Guidelines Clearinghouse (NGC), ECRI Institute, Sumsearch y Tripdatabase. Se encontró una revisión sistemática (13) en la Cochrane Library Plus, sobre la evidencia científica disponible relativa al rendimiento diagnóstico de la mamografía digital comparándola con la analógica en el contexto clínico. Esta revisión incluía el período de 1989 a 2005, sin especificar el último mes de inclusión de estudios de 2005.

A partir de este hallazgo en la búsqueda preliminar, se decidió incluir los estudios originales de la revisión sistemática encontrada y actualizar la búsqueda de estudios originales. La búsqueda de estudios originales se realizó en las bases referenciales MEDLINE (incluido PRE-MEDLINE mediante PUBMED) y EMBASE. En ésta búsqueda no hubo limitación por idioma, pero si por año, limitándose a los artículos publicados desde el año 2005 (completo) hasta Octubre de 2007.

Las estrategias de búsqueda fueron las siguientes:

MEDLINE (2005-2007) realizada en Octubre de 2007
#1.(mammography or mammogram\$.m_titl.

- #2. digital.m_titl.
- #3. 1 AND 2
- #4. screening.ti,ab.
- #5. 3 AND 4
- #6. Mammography/
- #7. Radiographic Image Enhancement/
- #8. Radiographic Image Interpretation, Computer-Assisted/
- #9. Image Processing, Computer-Assisted/
- #10. 7 OR 8 OR 9
- #11. 6 AND 10
- #12. 12.Breast Neoplasms/pc, di, ra [Prevention & Control, Diagnosis, Radiography]
- #13. exp Mass Screening/
- #14. 12 OR 13
- #15. 11 AND 14
- #16. 5 OR 15
- #17. 18. limit 16 to yr="2005 – 2007"

EMBASE (2005-2007) realizada en Octubre de 2007

- #1. mammography or mammogram*
- #2. digital
- #3. screening
- #4. 1 AND 2 AND 3
- #5. 'digital mammography'/exp AND 'breast cancer'/exp AND ('screening'/exp OR 'mass screening'/exp OR 'cancer screening'/exp OR 'screening':ti,ab) AND [embase]/lim AND [2005-2007]/py
- #6. 4 OR 5 AND [2005-2007]/py

3.2.2 Resultados de la búsqueda

A partir de la revisión sistemática hallada en la búsqueda preliminar, se incluyeron 6 artículos originales (todos los relacionados con cribado de cáncer de mama que estaban incluidos en la revisión). Tras la búsqueda definitiva en Medline y Embase, se obtuvieron 270 referencias. Se añadió además un artículo que aparecía en el listado de bibliografía de uno de los artículos hallados. El cómputo final fue de 277 referencias. Se detectaron 24 duplicados en las distintas bases de datos y tras la eliminación de estos, quedaron 253 referencias bibliográficas.

En la Tabla 1 se muestra el número de artículos obtenidos tras realizar las búsquedas en las diferentes bases de datos.

| Tabla 1. Número de artículos hallados en cada base de datos. | | |
|---|------------------|---|
| FUENTE | ARTÍCULOS | TOTAL (CON DUPLICADOS/ SIN DUPLICADOS) |
| Revisión sistemática de la búsqueda preliminar | 6 | 277 – 24 Duplicados = 253 |
| MEDLINE | 221 | |
| EMBASE | 49 | |
| Listado de bibliografía de uno de los artículos | 1 | |

3.3 Selección de los estudios: criterios de inclusión y exclusión

Para la selección de los estudios se definieron los siguientes **criterios de inclusión:**

- Población: mujeres adultas asintomáticas con edad comprendida entre 40 y 70 años.
- Intervención y comparación: cribado de cáncer de mama con mamografía digital versus mamografía convencional. Se incluyeron los estudios que comparaban estas dos técnicas, bien en el mismo grupo o en dos grupos de pacientes.
- Resultados: se incluyeron estudios que daban información sobre al menos uno de los siguientes resultados:
 1. Rendimiento diagnóstico: Sensibilidad, especificidad, área bajo la curva ROC, valor predictivo positivo (VPP) y valor predictivo negativo (VPN).
 2. Intermedios: Tasa de rellamadas.
 3. Finales: Mortalidad, supervivencia, calidad de vida.
- Diseño: Se consideraron revisiones sistemáticas, ensayos clínicos y estudios observacionales (transversales y de cohortes).
- Se limitó por idiomas a inglés, francés y español.
- No se tuvieron en cuenta aspectos técnicos del equipo de mamografía convencional ni digital, ni se limitó por tamaño de muestra.

Criterios de exclusión:

- Estudios experimentales con simulación de imágenes.
- Estudios realizados en animales.
- Estudios en que las mamografías convencionales se digitalizan para la posterior lectura por ordenador.

3.4 Evaluación crítica de los estudios seleccionados

Se ha realizado la evaluación crítica de los artículos seleccionados originales mediante la herramienta QUADAS (Quality Assesment of Diagnostic Accuracy Studies) (19) (Anexo 2). Esta herramienta analiza las siguientes características :

- Adecuación del espectro de mujeres seleccionadas a la población que recibirá la prueba evaluada en la práctica clínica.
- Aspectos relacionados con posibles sesgos asociados a la prueba de referencia utilizada: adecuación de la prueba e independencia con respecto a las técnicas evaluadas.
- Comparabilidad de los grupos en cuanto a número y tipo de pruebas recibidas.
- Tiempo entre la realización de la técnica realizada y la prueba de referencia.
- Descripción de las técnicas utilizadas y de la prueba de referencia.
- Cegamiento de los evaluadores de la prueba de referencia con respecto a los resultados obtenidos en las técnicas evaluadas y viceversa.
- Resultados dudosos o indeterminados y pérdidas durante el seguimiento.

La evaluación de la calidad de los estudios se realizó de forma independiente por dos personas. Las dudas y discrepancias se resolvieron mediante discusión y consenso.

Aunque la herramienta no contempla una medida final de la calidad de estos estudios, a modo de resumen se presenta la suma de los ítem en los que se ha respondido “sí”, a fin de facilitar una ideal global del cumplimiento de los criterios de calidad.

Además del resultado de la valoración de cada estudio con QUADAS se ha tenido en cuenta la validez del diseño para la evaluación comparada de la mamografía digital y convencional.

3.5 Extracción de la información

La extracción de la información de cada uno de los artículos seleccionados fue realizada por dos personas de forma independiente y las discrepancias se resolvieron por discusión y consenso.

3.6 Síntesis de los resultados

Se ha realizado una síntesis cualitativa de los resultados. El metanálisis se ha descartado debido a la diversidad de los diseños y la heterogeneidad de las características de los estudios incluidos en la revisión (en términos de población, protocolo de la lectura de mamografías y criterios de rellamada, resultados, etc.). Para la síntesis cualitativa, se han agrupado los estudios en base a la validez del diseño (comparabilidad de los grupos y existencia o no de seguimiento) y matizando los resultados según la calidad de los estudios y las características del contexto.

4 Resultados

4.1. Selección de artículos según criterios de inclusión

La selección partió de los 253 artículos que quedaron tras la eliminación de duplicados. Se inició mediante la lectura del título y el resumen, excluyéndose en esta fase 229 artículos. En la Tabla 2 se muestran las fases de la selección.

| Tabla 2. Fases de la selección. | | |
|--|-----------|-----------|
| FASES | INCLUIDOS | EXCLUIDOS |
| DUPLICADOS | 253 | 24 |
| POR TITULO Y RESUMEN | 24 | 229 |
| POR LECTURA A TEXTO COMPLETO | 11 | 13 |

El motivo de exclusión más frecuente era que el objetivo del estudio no estaba relacionado con la pregunta de investigación (144 estudios), seguido por estudios no originales (63, entre ellos cartas, presentaciones a congresos, revisiones no sistemáticas, artículos de opinión, etc.) y otros motivos (22; entre ellos, estudios sobre simulaciones de mamografías, idioma, etc.).

De los 24 artículos incluidos para la lectura a texto completo, se seleccionaron 11 artículos originales. De los 13 artículos no seleccionados en esta fase, 7 se excluyeron por no responder al objetivo de la revisión y 6 por no tratarse de artículos originales (cartas, ficha de tecnología emergente, artículos de opinión, revisiones no sistemáticas, etc.). En la Tabla 3 se presentan los motivos detallados de exclusión de estos 13 artículos.

| Tabla 3. Artículos EXCLUIDOS | |
|-------------------------------------|---|
| AUTOR/ AÑO | MOTIVO DE EXCLUSIÓN |
| Glitin (20) 2007 | No responde al objetivo de la revisión. En este estudio sólo realizan a la población mamografías convencionales. |
| Berns (21) 2006 | No responde al objetivo de la revisión. En este estudio comparan el tiempo de interpretación y la adquisición de la imagen. |
| Dershaw D. (22) 2006 | No es un artículo original. Se trata de una revisión no sistemática. |

| Tabla 3. Artículos EXCLUIDOS (continuación) | |
|--|---|
| Skaane (23) 2006 | No responde al objetivo de la revisión. En este estudio, la población es la misma que en uno de los estudios incluidos, pero añaden la lectura por ordenador de las mamografías convencionales tras digitalizarlas. |
| Skaane (24) 2005 | No responde al objetivo de la revisión. Este estudio versa sobre la aportación de nuevos lectores de las mamografías digitales o convencionales. |
| Van Ongeval (25) 2006 | No es un artículo original. Se trata de una revisión no sistemática. |
| Hailey D (26) 2006 | No es un artículo original. Se trata de una ficha de tecnología emergente de la Canadian Agency for Drugs and Technologies in Health. |
| Bonardi R (27) 2005 | No responde al objetivo de la revisión. En este estudio no realizan a la población mamografías digitales, sino que digitalizan las mamografías convencionales. |
| Jong R.A. (28) 2005 | No es un artículo original. Se trata de una revisión no sistemática. |
| Ge J (29) 2007 | No responde al objetivo de la revisión. El objetivo de este estudio es detectar microcalcificaciones en la mama con la ayuda de la lectura por ordenador.. |
| Lewin (30) 2007 | No es un artículo original. Se trata de una revisión no sistemática. |
| Yaffe M.J (31) 2006 | No es un artículo original. Se trata de un artículo de opinión. |
| Ciatto S (32) 2006 | No responde al objetivo de la revisión. Se trata de un estudio sobre los costes diferenciales de ambas técnicas mamográficas. |

4.2. Descripción de los artículos

4.2.1. Grupos de artículos por estudio de procedencia y diseño

De los 11 artículos originales incluidos en la revisión:

- Tres de ellos (33-35) forman parte de un estudio de cohortes prospectivo realizado en dos centros de Estados Unidos, donde todas las mujeres reciben mamografía digital y convencional y son seguidas durante un año. El primero de los 3 artículos (33) cuenta

los resultados preliminares del estudio (antes de haber completado la inclusión de toda la muestra de mujeres) por lo que sólo se ha utilizado para completar la lectura crítica del artículo principal¹ de este grupo (34). El tercero de los artículos (35) es un re-análisis de los datos del estudio, realizado retrospectivamente, que pretende valorar qué aporta la realización de los dos tipos de mamografías (digital y convencional) respecto a uno sólo.

- Cinco de ellos (36-40) proceden de tres estudios realizados en el marco del programa nacional de cribado de cáncer de mama de Noruega (Norwegian Breast Cancer Screening Program: NBCSP):
 - **Estudio Oslo I.** Este trabajo, sobre un grupo de mujeres que reciben todas los dos tipos de mamografía (digital y convencional), da lugar a 2 artículos: uno aporta los datos iniciales del estudio y podría considerarse como un diseño transversal (resultados de rendimiento diagnóstico en el momento de realización de las mamografías) (36) y el otro aporta datos a los dos años de seguimiento de las mismas mujeres, por lo que podría considerarse como un diseño de cohorte prospectivo (37).
 - **Estudio Oslo II.** Se trata de un ensayo clínico con asignación aleatoria de parte de las mujeres del programa de cribado noruego a dos grupos: uno al que se hace mamografía digital y otro al que se hace mamografía convencional; dando lugar a dos artículos. El primero de los artículos aporta resultados del ensayo en un momento del tiempo (rendimiento diagnóstico en el momento de realización de las mamografías) (38) y el segundo artículo añade los resultados tras 1,5-2 años de seguimiento de las mismas mujeres (39).
 - **Estudio de Vigeland *et al.*** (40). Aprovecha el inicio del programa de cribado en uno de los condados de Noruega, donde se decide hacer mamografía digital a todas las mujeres. Compara los resultados de rendimiento diagnóstico de la mamografía digital en este condado (sin hacer seguimiento) con los de la mamografía convencional realizada en la primera ronda del programa en los demás condados de Noruega (sin considerar seguimiento). Se trata pues de un estudio transversal.
- Dos de los artículos, ambos de Pisano et al. (41,42) derivan de un estudio de cohorte prospectivo realizado en Estados Unidos y

¹ No se incluyen los resultados del estudio de Lewin 2001 para no dar datos repetidos. Se incluyen los resultados de los dos artículos que aportan información sobre la muestra completa de mujeres (Lewin 2002 y Glueck 2007).

Canadá (Digital Mammographic Imaging Screening Trial: DMIST), en el que se realiza mamografía digital y convencional a todas las mujeres incluidas. El primer artículo (41) se limita a contar los objetivos y metodología del estudio, por lo que sólo se ha utilizado para completar la lectura crítica del artículo principal de este grupo (42). Este segundo artículo aporta los resultados del estudio tras dos años de seguimiento.

- Por último, el artículo de Del Turco (43), que deriva de un programa de cribado realizado en Florencia (Italia). El estudio compara un grupo de mujeres seleccionadas a partir de una unidad móvil de cribado que sólo usaba mamografía digital y otro grupo seleccionado a partir de una unidad móvil que sólo usaba mamografía convencional. Se trata de un análisis retrospectivo de los datos del programa de cribado, sin información sobre seguimiento de las mujeres, por lo que a efectos de diseño se puede considerar como un estudio transversal.

A modo de resumen, de cara a la extracción de resultados se han utilizado 9 artículos, todos excepto los de Lewin (33) y Pisano a (41), que sólo se han usado para completar la cumplimentación del QUADAS en los artículos principales de su grupo. La valoración crítica mediante QUADAS se ha realizado en 8 artículos, todos excepto los dos anteriores y el estudio de Glueck (35), dado que sólo se trata de un reanálisis de los datos de otro estudio.

4.2.2. Descripción detallada de los estudios: población, intervenciones y medidas de resultados

Estudio realizado en dos centros de Estados Unidos por el grupo de Lewin *et al.* (Tabla 4)

Como se ha comentado, este estudio consta de 3 artículos (33-35). El artículo principal (34) incluye la totalidad de la muestra: 4.489 mujeres mayores de 40 años que acudieron a 2 centros médicos para hacerse una mamografía de cribado. A estas mujeres se le realizaron 6.736 exámenes (cada examen incluía una MC y una MD): 2.533 exámenes de mujeres que sólo participaron una vez en el cribado, 3.300 exámenes de 1665 mujeres que participaron dos veces y 873 exámenes de 291 mujeres que participaron tres veces. Por tanto, el estudio mezcla información de distintas rondas de cribado.

Aunque el estudio se presenta como si fuera una cohorte con seguimiento prospectivo, realmente se analizan los datos como si fuera un estudio transversal, ya que siempre se informa sobre exámenes realizados y no sobre mujeres. Tampoco se dan los resultados por separado para cada

grupo de mujeres según el número de rondas que incluyen, ni queda claro si las mujeres que sólo participaron una vez tuvieron seguimiento o no.

Cada una de las dos mamografías de un examen era leída por un radiólogo diferente que no conocía el resultado de la otra mamografía. Cada caso se clasificaba según el BI-RADS® (Breast Imaging Reporting and Data System²) y se le asociaba una recomendación y una probabilidad de padecer cáncer (entre 0 y 100%). Una mamografía se consideraba positiva si el BI-RADS® era de 0, 4, ó 5. Cuando había discrepancia entre el resultado de la MD y la MC se discutían los casos y –aunque no se hizo muy frecuentemente– cabía la posibilidad de no seguir realizando pruebas adicionales después de discutir el caso. Por tanto, hubo algunas mujeres con MD o MC inicialmente positiva en las que no se siguió investigando ni llegaron a tener una biopsia. Por último, en este estudio se consideraron falsos negativos los casos con mamografías negativas en los que se diagnosticó un cáncer en menos de un año.

Estudios realizados en el marco del programa nacional de cribado de cáncer de mama de Noruega (Tablas 5 y 6).

En el marco del programa nacional de cribado de cáncer de mama de Noruega, se han realizado tres estudios: dos observacionales (Oslo I (36, 37) y Estudio del condado de Vestfold (40)) y un ensayo clínico con asignación aleatoria (Oslo II (38,39)). Los estudios Oslo I y II son del mismo grupo de autores y tienen muchos puntos de la metodología en común, por lo que se cuentan juntos.

El estudio Oslo I incluyó a 3.683 mujeres de 50-69 años que acudían al programa de cribado en su tercera ronda y las siguió durante dos años. El estudio Oslo II incluyó a más de 20.000 mujeres (cuyos resultados muestra en dos grupos de edad: 45-49 años y 50-69 años) y las siguió un año y medio (a las de 45-49 años) o dos años (a las de 50-69 años). Estas mujeres fueron aleatorizadas a través de un código numérico de llamada con ajustes por edad y área de residencia³. La asignación aleatoria se realizó de forma que un 70% de las mujeres recibieran MC y un 30% MD.

Tanto en el estudio Oslo I como en el Oslo II, las mamografías digitales se leyeron con monitor (*soft copy*). Las imágenes de las MC y MD fueron

² Las categorías de BI-RADS® son las siguientes: 0 necesidad de otros estudios, 1 negativo, 2 hallazgo benigno, 3 probablemente benigno, 4 sospechoso, 5 altamente sugestivo de malignidad, 6 malignidad comprobada mediante diagnóstico histológico.

³ Esto se hizo para tener grupos comparables a pesar de que las mujeres procedían de 27 áreas de la región de Oslo, muy diferentes en nivel socioeconómico entre ellas.

interpretadas por 8 radiólogos con entrenamiento previo, o con más de 4 años de experiencia en cribado mamográfico. Estos 8 lectores se dividían en 2 equipos de 4 radiólogos cada uno: un equipo interpretaba las MC y otro las MD, siendo cada imagen leída de forma independiente por 2 radiólogos. Se utilizó una escala de 5 puntos de probabilidad de cáncer en las mamografías⁴, donde todos los casos clasificados como 2 ó mayor eran automáticamente seleccionados para discutirse en una reunión de consenso (donde intervenía el equipo completo de radiólogos del tipo de mamografía que se tratara). En las reuniones tenían disponibles todas las mamografías anteriores de la mujer. También se llevaban a discusión los casos donde había síntomas clínicos y los exámenes técnicamente incorrectos. Las rellamadas se hacían en los casos en que se obtenía una puntuación de 3 ó más puntos por al menos uno de los dos lectores.

En cuanto a las medidas de resultado en los estudios Oslo I y Oslo II, cabe destacar que ambos incluyen información sobre los resultados iniciales del cribado (cánceres iniciales) y del seguimiento (cánceres de intervalo, cánceres detectados en la siguiente ronda de cribado y falsos negativos de la mamografía inicial). El estudio Oslo II aporta los resultados desagregados por grupos de edad (45-49 y de 50- 69 años).

En el último estudio de los realizados en Noruega (40), participaron 18.239 mujeres de 50-69 años del Condado de Vestfold (incluidas entre 2004 y 2005), a las que se les realizó una MD de cribado. Para la comparación usaron los datos de las 324.763 mujeres de los otros 18 Condados de Noruega, a las que se les había realizado MC durante la primera ronda del programa nacional de cribado.

El equipo de MD que utilizaron en este estudio difiere de los de otros, ya que está basado en paneles de detección de selenio⁵. Las MD fueron leídas por 5 radiólogos, 4 de ellos con más de 5 años de experiencia en leer MC y 1 con un año de experiencia en MC, y todos con una experiencia de 5 meses en lectura de MD. Se llevó a cabo una doble lectura independiente por los radiólogos, utilizando la misma escala de 5 categorías para determinar la probabilidad de cáncer que se había usado en los estudios Oslo I y II. Además, se usó una reunión de consenso para decidir, con la misma metodología que los estudios Oslo I y II.

⁴ La escala tenía las siguientes categorías: 1 normal (definitivamente benigno), 2 probablemente benigno, 3 indeterminado, 4 probable malignidad y 5 malignidad.

⁵ Con los paneles de selenio los píxeles son menores y el procesamiento de la imagen es diferente, resultando una imagen de alto contraste.

Estudio DMIST (Digital Mammographic Imaging Screening Trial) realizado en Estados Unidos y Canadá (Tabla 7).

Este estudio de cohorte prospectivo fue realizado por el American College of Radiology Imaging Network y se inició en el año 2001. El artículo principal (42), publicado en 2005, aporta los resultados del cribado de 42760 mujeres que se presentaron en 32 centros diferentes de Estados Unidos y Canadá para una mamografía de cribado (sin síntomas en ese momento) y que fueron seguidas entre 11 y 15 meses. No hubo restricción por edad de las mujeres, siendo la edad media de 54,6 años y el rango intercuartílico de 47 a 61.

A todas las mujeres se les realizó tanto MC como MD. Se utilizaron 5 tipos de equipos de MD diferentes, según el que había disponible en cada centro.

La MD y MC de cada mujer fueron interpretadas de forma independiente por 2 radiólogos con experiencia previa en lectura de mamografías (un radiólogo valoró la MD y otro radiólogo la MC). Utilizaron 3 escalas para la clasificación de las lesiones: la escala BI-RADS®, una escala de 7 ítems de malignidad⁶ (para obtener el área bajo la curva ROC), y una recomendación acerca de la necesidad de re-llamada en una escala de 6 puntos. Para pasar a exámenes adicionales, incluida la biopsia, bastaba con que uno de los dos radiólogos lo recomendase.

El diagnóstico definitivo de cáncer al final del seguimiento se determinó si había una biopsia positiva entre el inicio del estudio y hasta los 15 meses siguientes, incluyendo una segunda mamografía de cribado al final de seguimiento. Los resultados indeterminados o sin información (biopsias con poca muestra no concluyentes, mujeres con segunda mamografía –al final del seguimiento– con BI-RADS® 3-4-5, personas que murieron durante el seguimiento) se excluyeron del estudio. Las biopsias se interpretaron por un patólogo local (del centro) y por un patólogo central del estudio. Las discrepancias entre ambos se resolvieron por consenso.

Los resultados fundamentales se refieren a la exactitud diagnóstica de la MC comparada con la MD mediante la medición del área bajo la curva ROC. Hacen además análisis de subgrupos según edad (mayores y menores de 50 años), patrones de densidad mamaria según escala BI-RADS®, equipo utilizado, riesgo de cáncer y estado menopáusico. Los resultados sobre rendimiento diagnóstico de las mamografías se dan a los 12 y 15 meses.

⁶ Las categorías de la escala eran: 1 definitivamente no maligno, 2 casi definitivamente no maligno, 3 probablemente no maligno, 4 posibilidad de ser maligno, 5 probablemente maligno, 6 muy probablemente maligno, 7 definitivamente maligno.

Estudio realizado en Florencia (Tabla 8)

Este estudio es un análisis realizado retrospectivamente con datos del programa de cribado de cáncer de mama en Florencia (43). Se incluyeron los resultados de 2 grupos de mujeres atendidas en dos unidades móviles diferentes (una con MD y otra con MC), que se seleccionaron de forma que pudieran ser apareadas por década de edad y par de radiólogos que interpretaron las mamografías. Se incluyeron 28.770 mujeres de 50-69 años, sin aportar datos de seguimiento. Los resultados corresponden a distintas rondas de cribado.

Las mamografías fueron interpretadas por 4 radiólogos con más de 5 años de experiencia en mamografías de cribado. La lectura la realizaban 2 radiólogos de forma no independiente, es decir, el segundo radiólogo conocía los resultados del primero. Utilizaron para la clasificación de la lesión la escala de BI-RADS® y otra escala de cinco categorías⁷.

La exactitud de la MD se compara con la de la MC (tasa de rellamadas, tasa de detección de cancer y VPP), aportando los datos desagregados según edad (50-59 y 60-69 años), densidad mamaria, tipo de ronda de cribado (prevalente o incidente), y tipo de anomalías radiológicas.

⁷ 1 negativo, 2 benigno, 3 dudoso, 4 sospechoso y 5 maligno.

| TABLA 4. Descripción de las características del estudio de cohorte realizado en dos centros de Estados Unidos por Lewin et al: Artículos derivados del estudio. | | | | | | |
|--|--|--|---|---|---|--|
| AUTOR- AÑO - N | DISEÑO | POBLACIÓN | INTERVENCIÓN (prueba índice) | COMPARACIÓN | GOLD ESTÁNDAR | MEDIDAS DE RESULTADO* |
| LEWIN et al (2001) (33) N=3890 Estudio preliminar. Se usa fundamentalmente para completar el QUADAS de Lewin 2002 | Aunque se presenta como cohorte con seguimiento de un año, realmente se analiza como un estudio transversal. Todas mujeres reciben MC y MD, leídas por 2 radiólogos (1 MC y otro MD). | - 3890 mujeres a partir de 40 años; 4945 exámenes. - Mujeres que se presentan en 2 centros médicos de Estados Unidos para mamografía de cribado. | - MD. - Lectura inicial por un radiólogo ciego al resultado de la MC. Pero se discuten luego los casos discordantes y se podía descartar hacer otras pruebas en casos previamente positivos. | - MC. - Lectura inicial por un radiólogo ciego al resultado de la MD. Pero se discuten luego los casos discordantes y se podía descartar hacer otras pruebas en casos previamente positivos. | Diagnóstico de cáncer según la anatomía patológica (biopsia), ya se haya realizado en el momento de una mamografía sospechosa o por la aparición de signos o síntomas durante el año siguiente. | - Tasa de re-llamada - Biopsias - Tasa de detección de cáncer - Cánceres in situ - VPP - Área bajo curva ROC. |
| | | - 4489 mujeres a partir de 40 años; 6736 exámenes. - Edad media: 55,6 años - Mezcla datos de distintas rondas de cribado. - Mujeres que se presentan en 2 centros médicos de Estados Unidos para mamografía de cribado. | - Escala BI-RADS y escala de probabilidad de malignidad de 0-100%. | - Escala BI-RADS y escala de probabilidad de malignidad de 0-100%. | - La MD se compara con MC y con ambas combinadas. | - Tasa de detección de cáncer - Área bajo curva ROC de MD, MC y ambas combinadas |
| LEWIN et al (2002) (34) N= 6736 exámenes en 4489 mujeres Este artículo describe los resultados finales del estudio. | Análisis retrospectivo en el marco del estudio de Lewin . | | | | | |
| GLUECK (2007) (35) N= 4489 Re-análisis de los datos del estudio | | | | | | |

MC (mamografía convencional), MD (mamografía digital), BI-RADS® (Breast Imaging Reporting and Data System), VPP (valor predictivo positivo).

* Las medidas de resultado recogidas en la tabla son las aportadas en los artículos, si bien luego pudieron calcularse medidas adicionales a partir de los datos de los artículos.

TABLA 5. Descripción de las características de los tres estudios realizados en el marco del programa de cribado de Noruega: artículos derivados de estos estudios que aportan datos sin seguimiento de las mujeres.

| AUTOR – AÑO – N | DISEÑO | POBLACIÓN | INTERVENCIÓN (prueba índice) | COMPARACIÓN | GOLD ESTÁNDAR | MEDIDAS DE RESULTADO* |
|---|---|---|--|---|--|--|
| OSLO I Skaane et al (2003) (36) N=3683 | Parte transversal (sin seguimiento) al inicio de un estudio de cohorte prospectivo. | - 3683 mujeres de 50 a 69 años. -Tercera ronda del programa de cribado. | - MD -Doble lectura independiente y posterior reunión de consenso. -Escala de 5 puntos de probabilidad de cáncer. | - MC -Doble lectura independiente y posterior reunión de consenso. -Escala de 5 puntos de probabilidad de cáncer. | Diagnóstico de cáncer de mama según la anatomía patológica (biopsia) | - Tasa de rellamadas (tras consenso) - Biopsias - Tasa detección de cáncer - VPP |
| | Todas las mujeres reciben MC y MD. | | | | | |
| OSLO II Skaane et al (2004) (38) N=25263 | Parte transversal (sin seguimiento) al inicio de un ensayo clínico con aleatorización: 70% MC y 30% MD. | - 25263 mujeres de 45 a 69 años -No especifica la participación en cribados previos ni ronda de cribado en que están las mujeres | -MD, basada en paneles lisos de selenio . -Doble lectura independiente y posterior reunión de consenso. -Escala de 5 puntos de probabilidad de cáncer. | - MC -Doble lectura independiente y posterior reunión de consenso. -Escala de 5 puntos de probabilidad de cáncer. | Diagnóstico de cáncer de mama según la anatomía patológica (biopsia) | - Tasa de rellamadas (tras consenso) - Tasa de detección de cáncer - Cánceres in situ - VPP - Resultados desagregados por grupos de edad |
| | | | | | | |
| Vigeland et al (2007) (40) N= 18239 + 324763 | TRANSVERSAL Un Condado utilizaba MD, mientras los otros 18 utilizaban MC. | - MD: 18239 mujeres de 50-69 años del Condado de Vestfold (primera ronda del programa). - MC: 324763 mujeres de los otros Condados de Noruega (primera ronda del programa de cribado). | | | | - Tasa de rellamadas - Tasa detección de cáncer - Cánceres in situ - VPP |

| TABLA 6. Descripción de las características de los dos estudios realizados en el marco del programa de cribado de Noruega que aportan datos de seguimiento de las mujeres. | | | | | | |
|---|---|--|---|--|---|--|
| AUTOR – AÑO – N | DISEÑO | POBLACIÓN | INTERVENCIÓN (prueba índice) | COMPARACIÓN | GOLD ESTÁNDAR | MEDIDAS DE RESULTADO* |
| OSLO I Skaane et al (2005) (37) N=3683 | COHORTE. PROSPECTIVO. Seguimiento de más de 2 años. Todas las mujeres reciben MC y MD. | - 3683 mujeres - 50-69 años - Tercera ronda del NBCSP. | - MD (soft-copy o monitor) -Doble lectura independiente por radiólogos y posterior reunión de consenso. | - MC -Doble lectura independiente por radiólogos y posterior reunión de consenso. - Escala 5 puntos de probabilidad de cáncer. | Diagnóstico de cáncer de mama según la anatomía patológica (biopsia), independientemente de si ésta se hacía tras la mamografía inicial positiva, por sintomas en el seguimiento o por la mamografía de la siguiente ronda positiva. | Cánceres de intervalo Cánceres detectados en la siguiente ronda de cribado Cánceres in situ Tasa de detección de cáncer |
| OSLO II Skaane et al (2007) (39) N=23929 | ENSAYO CLÍNICO CON ASIGNACIÓN ALEATORIA Y SEGUIMIENTO Grupo de 50-69 años seguidas 2 años. Grupo de 45-49 años seguidas 1,5 años. | - 23929 mujeres - 45-69 años - Contexto del NBCSP (no específica qué ronda o rondas de cribado). | - Doble lectura independiente por radiólogos y posterior reunión de consenso. - Escala 5 puntos de probabilidad de cáncer. | - Escala 5 puntos de probabilidad de cáncer. | Diagnóstico de cáncer de mama según la anatomía patológica (biopsia), independientemente de si ésta se hacía tras la mamografía inicial positiva, por sintomas en el seguimiento o por la mamografía de la siguiente ronda positiva. | Cánceres de intervalo Cánceres detectados en la siguiente ronda de cribado Cánceres in situ Tasa de detección de cáncer |

| TABLA 7. Descripción de las características del estudio DMIST (Digital Mammographic Imaging Screening Trial) | | | | | | |
|--|---|--|--|--|--|---|
| AUTOR – AÑO - N | DISEÑO | POBLACIÓN | INTERVENCIÓN (PRUEBA ÍNDICE) | COMPARACIÓN | GOLD ESTÁNDAR | MEDIDAS DE RESULTADO* |
| DMIST Pisano et al (2005) (42) N=49258 | COHORTE, PROSPECTIVO Seguimiento 11- 15 meses. Todas mujeres reciben MC y MD (se incluyen 5 equipos diferentes de MD) | - 42760 mujeres que se presentan para cribado en 32 centros diferentes de EEUU y Canadá. - Edad sin límite establecido a priori. Edad media: 54,6 años Rango intercuartilico: 47- 61 años | - MD (5 equipos diferentes) - Interpretadas por 2 radiólogos de forma independiente (1 para MC y otro MD). - Se utilizaron 4 escalas: • Probabilidad carcinoma (0-100%), • BI-RADS, • Escala 7 ítems de malignidad, • Escala 5 puntos de rellamada. | - MC - Interpretadas por 2 radiólogos de forma independiente (1 para MC y otro MD). - Se utilizaron 4 escalas: • Probabilidad carcinoma (0-100%), • BI-RADS, • Escala 7 ítems de malignidad, • Escala 5 puntos de rellamada. | - Diagnóstico de cáncer de mama según la anatomía patológica (biopsia), independiente- mente de si ésta se hacía tras la mamografía inicial positiva, por síntomas en el seguimiento o por la mamografía de la siguiente ronda positiva. - La biopsia se interpretó por dos profesionales distintos y los desacuerdos se resolvieron por consenso. | - Área bajo la curva ROC - Sensibilidad - Especificidad - VPP - Datos desagregados por grupos de edad, ronda, densidad mamaria y tipo de anormalidades radiológicas. |

| TABLA 8. Descripción de las características del estudio realizado en el marco del programa de cribado de Florencia | | | | | | |
|---|---|---|---|---|--|--|
| AUTOR--AÑO ULICACIÓN-N | DISEÑO | POBLACIÓN | INTERVENCIÓN (Prueba índice) | COMPARACIÓN | GOLD ESTÁNDAR | MEDIDAS DE RESULTADO |
| Del Turco et al (2007) (43) N=28770 | TRANSVERSAL (Si bien los datos se analizan de forma RETROSPECTIVA) Se comparan 2 grupos apareados por década de edad y por par de radiólogos, cada uno, de dos unidades móviles de un programa de cribado (una con MD y otra con MC). | - 28770 mujeres de 50-69 años - 2 grupos concurrentes del mismo programa de cribado de cáncer mama en dos unidades móviles de Florencia (Italia). Se mezclan mujeres en distintas rondas de cribado. | - MD - Interpretada por 4 radiólogos - Doble lectura no independiente (el segundo radiólogo conoce la opinión del primero). - Escalas de 5 puntos de probabilidad de malignidad y escala BI-RADS. | - MC - Interpretada por 4 radiólogos - Doble lectura no independiente (el segundo radiólogo conoce la opinión del primero). - Escalas de 5 puntos de probabilidad de malignidad y escala BI-RADS. | Diagnóstico de cáncer de mama según la anatomía patológica (biopsia). | - Tasa de re-llamadas - Tasa de detección de cáncer - VPP - Datos desagregados por grupos de edad (50-59 años y 60-69 años) |

4.3 Valoración crítica de los artículos

4.3.1. Calidad según el QUADAS

Como ya se comentó anteriormente, se ha realizado la lectura crítica de 8 de los 11 artículos originales incluidos en ésta revisión; todos salvo los artículos de Lewin 2001, Glueck 2007 y Pisano 2005-a (33, 35, 41) (Tabla 9). Los dos últimos sólo se han usado para ayudar a la cumplimentación del QUADAS de los artículos principales del estudio al que pertenecen.

Todos los artículos incluidos en esta revisión tienen una calidad similar. En general, se sitúan en un rango de 7-9 respuestas afirmativas de las 14 preguntas del QUADAS (19). Sin embargo, es importante precisar que el número máximo de preguntas afirmativas que podía alcanzarse en estos estudios es de 12, ya que las preguntas 5 (¿toda la muestra recibe el *gold* estándar?) y 6 (¿se hace el *gold* estándar independientemente de cual es el resultado de la prueba índice?) sólo pueden responderse negativamente, pues no sería ético hacer biopsias en mujeres con mamografías negativas.

Todos los estudios tuvieron una respuesta afirmativa en cuatro de las preguntas que evalúan la validez interna: 3 (sobre adecuación del *gold* estándar), 7 (el *gold* estándar no contiene a la prueba índice), 8 (detalle con el que se describe la prueba índice) y 10 (la prueba índice se interpreta sin conocer los resultados del *gold* estándar).

Asimismo, la mayoría de los estudios tuvieron respuestas afirmativas en las preguntas que valoran la validez externa (1, 2, 12); todos salvo los estudios de Lewin (34), Pisano (42) y Del Turco (43), que fueron los de menor puntuación. En los estudios de Lewin y de Pisano, el contexto no fue un programa poblacional de cribado, sino mujeres que acudían libremente a hacerse una mamografía (pregunta 1 con respuesta negativa); y en el estudio Del Turco no quedó claro si la interpretación de la mamografía se había hecho en un contexto similar a la práctica clínica habitual en los programas de cribado (pregunta 12). Puede considerarse en general que el espectro de mujeres de los estudios fue representativo de las que recibirían el cribado con MD en la práctica.

Las cuestiones en las que fallaron todos o algunos de los estudios fueron:

- Ninguno de los estudios informó con claridad lo que se hizo con los resultados intermedios o dudosos de la mamografía (pregunta 13). Asimismo, ningún estudio dejó claro si se había interpretado la biopsia sin conocer los resultados de la mamografía (pregunta 11).
- Sólo uno de los estudios describió con suficiente detalle la clasificación de la biopsia (pregunta 9).

- Tres de los estudios no dejaban claro si el tiempo entre la mamografía y la biopsia había sido corto (pregunta 4) y otros tres trabajos no explicaban con claridad el número y motivos de las pérdidas (pregunta 14).

4.3.2. Otros aspectos clave de la validez de los estudios

Además de la valoración de la calidad mediante el QUADAS, hay que tener en cuenta otros aspectos importantes relacionados con los diseños de los estudios que pueden influir en la validez. Los diseños con mayor validez para comparar el rendimiento diagnóstico de la mamografía digital con la convencional serían el ensayo clínico con asignación aleatoria y los estudios de cohorte única (todas las mujeres reciben los dos tipos de mamografía) con seguimiento prospectivo. Estos dos diseños de mayor validez se encuentran en tres de los estudios incluidos en esta revisión: el estudio Oslo II (38,39) es un ensayo clínico y los estudios Oslo I (36,37) y DMIST (41,42) son estudios de cohorte única con seguimiento.

Un diseño inferior en términos de validez interna sería el estudio de dos cohortes seleccionadas de forma separada, una de las cuales recibe la mamografía digital y otra la convencional. Este diseño es el que siguen los estudios de Vigeland (40) y de Del Turco (43). Estos dos estudios son además de tipo transversal, sin datos de seguimiento.

Por último, cabe destacar que el estudio de Lewin et al (33,34) se presenta como una cohorte única con seguimiento; sin embargo, es imposible saber cuantas mujeres fueron realmente seguidas. De hecho, parte de las mujeres del estudio participan en una única ronda de seguimiento mientras que otras participan en dos y tres rondas, siendo contabilizadas varias veces en el estudio (usan como sujeto de estudio las mamografías en vez de las mujeres). En este sentido, la información de este estudio se presenta más bien como si fuera un estudio transversal.

Tabla 9. Evaluación de la calidad de los estudios seleccionados a través de la Herramienta QUADAS (19)

| ITEM QUADAS | LEWIN (34) 2002 | OSLO I (36) 2003 | OSLO I (37) 2005 | OSLO II (38) 2004 | OSLO II (39) 2007 | VIGELAND (40) 2007 | PISANO (42) 2005 | DEL TURCO (43) 2007 |
|--|-----------------|------------------|------------------|-------------------|-------------------|--------------------|------------------|---------------------|
| 1: Espectro de pacientes representativo | No | Sí | Sí | Sí | Sí | Sí | No | Sí |
| 2: Criterios de selección descritos | Sí | Sí | Sí | Sí | Sí | Sí | Sí | Sí |
| 3: GE adecuado | Sí | Sí | Sí | Sí | Sí | Sí | Sí | Sí |
| 4: Tiempo entre PI y GE corto | NC | Sí | Sí | Sí | Sí | Sí | NC | NC |
| 5: Se hace GE a toda la muestra | No | No | No | No | No | No | No | No |
| 6: Se hace GE independiente de resultados de PI | No | No | No | No | No | No | No | No |
| 7: GE independiente (no contiene la PI) ¹ | Sí | Sí | Sí | Sí | Sí | Sí | Sí | Sí |
| 8: PI descrita con detalle | Sí | Sí | Sí | Sí | Sí | Sí | Sí | Sí |
| 9: GE descrito con detalle | Sí | NC | NC | NC | NC | No | NC | NC |
| 10: Interpretación PI sin conocer resultado de GE ² | Sí | Sí | Sí | Sí | Sí | Sí | Sí | Sí |
| 11: Interpretación GE sin conocer resultado de PI | NC | NC | NC | NC | NC | NC | NC | NC |
| 12: Interpretación PI contexto similar a la práctica | Sí | Sí | Sí | Sí | Sí | Sí | Sí | NC |
| 13: Información resultados intermedios/dudosos | NC | NC | NC | NC | NC | NC | No | NC |
| 14: Explicación de las pérdidas | NC | Sí | Sí | NC | NC | Sí | Sí | Sí |
| Resumen | 7 "sí" | 9 "sí" | 9 "sí" | 8 "sí" | 8 "sí" | 9 "sí" | 7 "sí" | 7 "sí" |

GE: Gold estándar (en estos estudios: la anatomía patológica); PI: prueba índice (en estos estudios: la mamografía); NC: No claro

¹ El gold estándar no es una combinación de pruebas que incluya a la prueba índice, sino que es una prueba simple (biopsia). Por eso la pregunta 7 siempre se responde "sí".

² La pregunta 10 se responde siempre afirmativamente, dado que el GE (biopsia) no se ha hecho cuando se está interpretando la PI (mamografía), sino que va detrás en el tiempo.

4.4 Resultados principales de rendimiento diagnóstico

4.4.1. Estudio realizado en dos centros de Estados Unidos por Lewin *et al.*

Los resultados del estudio de Lewin (34) se muestran en la Tabla 10. De este trabajo pueden derivarse datos de cuatro estimadores de rendimiento diagnóstico para comparar la MD con la MC: tasa de detección de cáncer, sensibilidad, área bajo la curva ROC y valor predictivo positivo (VPP). La tasa de detección de cánceres fue de 0,49% con la MC y de 0,40% con la MD, no siendo la diferencia estadísticamente significativa. Asimismo, aunque la sensibilidad y el área bajo la curva ROC también fueron mayores para la MC que para la MD, estas diferencias no alcanzaron significación estadística. Por último, los valores predictivos positivos fueron muy similares con ambas técnicas.

Aunque el artículo de Lewin 2002 no aporta datos sobre las proporciones de cánceres in situ e invasivos, en una publicación anterior con datos preliminares (33) del estudio (35 cánceres detectados en 4945 exámenes), se encontró que no había diferencias en estas proporciones entre MC y MD, siendo del 19% el porcentaje de cánceres in situ en ambos casos.

TABLA 10. Parámetros de rendimiento diagnóstico de la mamografía digital versus la mamografía convencional en el estudio de Lewin

| ESTUDIO N | TÉCNICA | TASA RELAMADAS | BIOPSIAS | Nº CÁNCERES DETECTADOS (TASA DEECCIÓN) | SEN** | VPP | ÁREA ROC |
|---------------------------------------|------------|----------------|-----------|--|-------|------|----------|
| Lewin et al (2002)(34) 6736 exámenes* | MC | 1001 (14,9%) | 143 (2%) | 33 (0,49%) | 78,6% | 3,3% | 0,80 |
| | MD | 793 (11,8%) | 94 (1,4%) | 27 (0,40%) | 64,3% | 3,4% | 0,74 |
| | Total | 1467 | 181 | 42 | - | - | - |
| | Valor de p | p<0,001 | p<0,001 | NS | NS | ND | NS |

MC: mamografía convencional; MD: mamografía digital; Sen: sensibilidad; VPP: valor predictivo positivo; ND: no hay datos; NS: no significativo desde el punto de vista estadístico.

* Cada examen consta de una MC y una MD. Se hicieron 6736 exámenes en 4489 mujeres.

** El artículo no aporta datos de sensibilidad. Ésta se ha calculado a posteriori suponiendo que los 42 cánceres detectados son la totalidad de los cánceres en la población estudiada. Dado que no se tienen en cuenta posibles cánceres de intervalo ni cánceres que podrían no haberse detectado por ser las dos mamografías negativas, este parámetro sobrestima la sensibilidad de ambas técnicas.

Como resultados intermedios, cabe destacar que hubo diferencias estadísticamente significativas entre ambos tipos de mamografía en la tasa de rellamadas y en la frecuencia de biopsias, siendo ambas menores con la MD que con la MC.

Glueck (35) realizó un análisis retrospectivo de los datos del estudio, comparando el rendimiento diagnóstico de hacer las dos mamografías frente a una sola. En este caso se añadieron al análisis los cánceres detectados en el intervalo entre mamografías de cribado, por lo que se hacen los cálculos sobre 49 cánceres (en vez de los 42 del artículo de Lewin 2002 (34)). Usando la MC y MD de manera combinada se incrementa la estimación de la sensibilidad, siendo de 84% con ambas mamografías, frente al 65% de la MC y al 55% de la MD⁸. Las diferencias entre hacer las dos mamografías y hacer sólo una (sea MC o MD) fueron estadísticamente significativas. También se compararon las áreas bajo la curva ROC, encontrando resultados diferentes según el test estadístico utilizado. Así, las diferencias entre hacer las mamografías combinadas y hacer una sólo mamografía no fueron estadísticamente significativas según los test paramétricos, pero si lo fueron según los test no paramétricos, tanto entre el examen combinado y la MC como entre aquel y la MD.

4.4.2. Estudios realizados en el marco del programa nacional de cribado de cáncer de mama de Noruega: Resultados sin considerar el seguimiento

Los tres artículos noruegos presentan datos de análisis transversales (antes de tener en cuenta el seguimiento), que se presentan en la Tabla 11.

En el primer artículo publicado sobre el estudio Oslo I (36), se informa de un total de 31 cánceres detectados, de los cuales 9 (29%) fueron carcinomas in situ. Según los autores, 28 de los 31 cánceres habrían sido detectados con MC y 23 con MD. Con estos datos analizan las diferencias de rendimiento diagnóstico entre los dos tipos de mamografía. Sin embargo, en un artículo posterior, con datos de seguimiento del estudio Oslo I (37), los autores proporcionan una información adicional que hace necesario corregir los datos presentados por Skaane 2003. De hecho, la MC realmente detectó 26 cánceres y la MD 20 (la diferencia de 2 y 3 cánceres respectivamente entre estos datos y los aportados por el artículo se explican a continuación).

⁸ Las estimaciones de sensibilidad de la MC y MD son menores que las estimadas a partir del artículo principal del estudio (Lewin 2002) debido a que el denominador aumenta al incluir los cánceres de intervalo.

Con las correcciones realizadas, los datos presentados en la Tabla 8 sobre el estudio Oslo I no coinciden exactamente con los números del artículo de Skaane 2003⁹. Se observaron tasas de detección de cáncer de mama algo inferiores con la MD (0,54%) que con la MC (0,71%), aunque las diferencias no llegaron a alcanzar significación estadística. Las diferencias a favor de la MC en términos de sensibilidad y valor predictivo positivo fueron de mayor magnitud, pero tampoco fueron estadísticamente significativas. La tasa de rellamadas fue algo superior con la MD (4,6%) que con la MC (3,5%), pero esta diferencia no fue estadísticamente significativa ni se tradujo en un mayor número de biopsias.

En el estudio Oslo II (38), la mayor parte de los resultados de comparación de la MD y MC se dan desagregados en dos grandes grupos de edad (45-49 y 50-69 años). Aunque en este ensayo clínico se aleatorizaron inicialmente 25.263 mujeres, todos los análisis del artículo de Skaane 2004 (38) se hacen sobre 24.908, tras excluir a las mujeres que habían tenido una mamografía distinta a la que le fue asignada por la aleatorización¹⁰.

Uno de los pocos estimadores que se aportan para el total de la población estudiada (conjuntamente para ambos grupos de edad) es la tasa de detección de cáncer de mama, que resultó superior con la MD (0,59%) que con la MC (0,41%), si bien la diferencia no llegó a alcanzar significación

⁹ En el estudio Oslo I hubo dos interpretaciones independientes para cada mamografía (dos radiólogos para cada MC y otros dos para cada MD). Si una de las dos interpretaciones de una mamografía era positiva, la mamografía se discutía en una reunión de consenso. Estas reuniones de consenso fueron distintas e independientes para MC y para MD. Una mamografía se consideraba positiva definitivamente si en el consenso se decidía así, en cuyo caso se pasaba a hacer nuevas pruebas en esa mujer (rellamada). Hubo 2 casos de cáncer con interpretaciones positivas para la MC que fueron descartados en el consenso de MC (por tanto, la MC debe considerarse como negativa, aunque los casos pudieran diagnosticarse gracias a haber sido positivos en el consenso de MD) y 3 casos de cáncer con interpretaciones positivas para la MD que fueron descartados en el consenso de MD (por tanto la MD se debe considerar negativa, aunque los casos pudieran diagnosticarse gracias a haber sido positivos en el consenso de MC). En el artículo de Skaane 2003, los 28 casos de cáncer que se aportan como detectados con MC incluyen a los 2 que fueron negativos tras el consenso de MC. Asimismo, los 23 casos de cáncer que se suponen detectados con MD incluyen a los 3 que fueron negativos tras el consenso de MD. Se han corregido los datos, suponiendo que 26 de los cánceres fueron detectados con MC y 20 con MD.

¹⁰ En el segundo artículo publicado sobre el estudio Oslo II (Skaane 2007(39)), en el que se añaden los resultados tras el seguimiento, la información versa sobre 23.929 mujeres, justificando los autores la exclusión del resto de mujeres por errores iniciales en la inclusión. En este segundo artículo vuelven a rehacer los análisis de rendimiento diagnóstico basal (antes del seguimiento) suponiendo que la población total es de 23.929 mujeres (en vez de 24.908). Sin embargo en el texto de esta revisión se ha decidido aportar los datos de rendimiento basal publicados en Skaane 2004, por no parecer justificada la exclusión adicional de mujeres que hacen los autores en Skaane 2007. Con todo, no hay grandes diferencias en los resultados de ambos análisis.

estadística ($p=0,06$)¹¹. Cuando dicha tasa de detección se estimó en cada subgrupo de edad, se observó como en el de 50-69 años, la MD tenía una mayor tasa de detección de cáncer de mama que la MC (0,83% frente a 0,54%), con una diferencia en los límites de la significación estadística ($p=0,053$). Sin embargo, en el subgrupo de 45-49 años de edad, la tasa de detección de cáncer fue prácticamente igual con ambos tipos de mamografía (0,27% con MD y 0,22% con MC). No se encontraron diferencias entre MD y MC en el valor predictivo positivo en ninguno de los dos grupos de edad. En cuanto al porcentaje de cánceres *in situ* del total de los detectados, cabe señalar que en el grupo de 50-69 años fue muy similar en los detectados por MC (27% de carcinomas *in situ*) y por MD (30%). En las mujeres de 45-49 años, el número total de cánceres detectados es muy pequeño (17 por MC y 8 por MD) para valorar las diferencias entre ambas mamografías en la proporción de cánceres *in situ*¹².

El único parámetro donde se encontraron diferencias estadísticamente significativas entre los dos tipos de mamografía en el estudio Oslo II fue la tasa de rellamadas en el subgrupo de 50-69 años, siendo superior con la MD (3,8%) que con la MC (2,5%). En el grupo de 45-49 años, la tasa de rellamadas fue similar con ambos tipos de mamografía.

Por último, el estudio de Vigeland (40), realizado también en el marco del programa noruego de cribado, comparó el rendimiento de la primera ronda de cribado con MD en Vestfold y con MC en otros 18 condados. Con la MD, la tasa de detección de cáncer fue algo mayor que con la MC; aunque esta diferencia se quedó en el límite de la significación estadística. Sin embargo, al calcular la tasa de detección de forma separada para cánceres *in situ* e invasivos, se observó que no había diferencias entre MD y MC en la tasa de detección de cánceres invasivos, pero la MD detectó más cánceres *in situ* que la MC (tasa de detección de carcinoma ductal *in situ* de 0,21% de la MD frente a 0,11% de la MC; $p<0,001$) (datos no mostrados en la tabla). Esto se traduce también en diferencias en la proporción de carcinomas *in situ* respecto al total de cánceres detectados con cada mamografía: 27% con MD (38 de 140 cánceres) y 16,3% con MC (343 de 2105 cánceres). El valor predictivo positivo de la MD fue también significativamente superior al de la MC.

¹¹ En el re-análisis realizado en Skaane 2007, sobre 23.929 mujeres, la tasa de detección con MD sigue siendo de 0,59% y la de MC baja a 0,38%, alcanzando ahora esta diferencia la significación estadística ($p = 0,03$).

¹² En el re-análisis realizado en Skaane 2007, sobre 23.929 mujeres, hicieron pruebas de significación estadística con toda la población (sin desagregar por subgrupos de edad), encontrando que la tasa de detección de carcinoma *in situ* era similar con MD y MC, pero la tasa de detección de carcinoma invasivo era mayor con MD (0,43% con MD y 0,26% con MC, $p = 0,040$).

En cuanto a la tasa de rellamadas, los autores diferencian entre la que es debida a resultados positivos de la mamografía (mayor parte de las rellamadas) y la que ocurre por problemas técnicos con la imagen. En cuanto a la tasa de rellamadas por mamografías positivas, no se encontraron diferencias entre MD y MC (en torno al 4% en ambos casos). Sin embargo, la tasa de rellamadas por problemas técnicos fue significativamente menor con la MD que con la MC (0,22% y 0,61% respectivamente, $p < 0,001$).

TABLA 11. Parámetros de rendimiento diagnóstico de la mamografía digital versus la mamografía convencional en los estudios noruegos (sin considerar seguimiento)

| Estudio N | Técnica | Tasa rellamadas (tras consenso) | Biopsias | Nº cánceres detectados (Tasa detección) | Sen | VPP |
|--|------------|---------------------------------|-----------|---|--------|-------------|
| OSLO I Skaane et al (2003)(36) 3683 mujeres | MC | 128 (3,5%) | 57 (1,5%) | 26 (0,71%) | 83,9%* | 20,3% |
| | MD | 168 (4,6%) | 52 (1,4%) | 20 (0,54%) | 64,5%* | 11,9% |
| | Total | | | 31 (0,84%) | - | |
| | Valor de p | NS | NS | NS | NS | NS |
| OSLO II Skaane et al (2004)(38) 10619 mujeres de 45-49 años ** | MC | 231 (3,0%) | No datos | 17 (0,22%) | NPC | 7,4% |
| | MD | 112 (3,7%) | No datos | 8 (0,27%) | NPC | 7,1% |
| | Total | -- | | -- | | -- |
| | Valor de p | NS | | NS | | NS |
| OSLO II Skaane et al (2004)(38) 14289 mujeres de 50-69 años ** | MC | 253 (2,5%) | No datos | 56 (0,54%) | NPC | 22,1% |
| | MD | 153 (3,8%) | No datos | 33 (0,83%) | NPC | 21,6% |
| | Total | -- | | -- | | |
| | Valor de p | $p < 0,05$ | | $p = 0,053$ | | NS |
| Vigeland (2007)(40) MC: 324763 MD: 18239 | MC | 13520 (4,2%)*** | No datos | 2105 (0,65%) | NPC | 15,1% |
| | MD | 746 (4,1%)*** | No datos | 140 (0,77%) | NPC | 18,5% |
| | Total | -- | | - | | - |
| | Valor de p | NS*** | | $p = 0,058$ | | $p = 0,015$ |

MC: mamografía convencional; MD: mamografía digital; Sen: sensibilidad; VPP: valor predictivo positivo; NPC: no puede calcularse la sensibilidad; NS: No significativo estadísticamente ($p > 0,05$).

* El artículo no aporta datos de sensibilidad. Ésta se ha calculado a posteriori suponiendo que los 31 cánceres detectados son la totalidad de los cánceres en la población estudiada. Dado que no se tienen en cuenta posibles cánceres de intervalo ni cánceres que podrían no haberse detectado por ser las dos mamografías negativas, este parámetro sobrestima la sensibilidad de ambas técnicas

** En el estudio OSLO II se aleatorizaron las mujeres a MC (17913) y MD (6998). El estudio muestra la mayor parte de resultados diferenciados por grupo de edad, que se presentan en estas dos filas de la tabla.

*** Estos datos se refieren a la tasa de rellamadas por mamografías positivas, pero no incluye las rellamadas por problemas técnicos con la imagen.

4.4.3. Estudios realizados en el marco del programa nacional de cribado de cáncer de mama de Noruega: Resultados considerando el seguimiento

Los artículos de Skaane 2005 (37) y Skaane 2007 (39) relatan los resultados finales de los estudios Oslo I y Oslo II, respectivamente, tras el período de seguimiento (Tabla 12).

En el estudio Oslo I (37), tras dos años de seguimiento, se detectaron 26 cánceres de mama: 10 de intervalo y 16 en la siguiente ronda del cribado. Al añadirlos a los 31 hallados en la ronda inicial de cribado de este estudio, suman un total de 57 cánceres. Los datos de seguimiento, permiten corregir la sensibilidad de la MC y MD (respecto a las estimaciones realizadas con los datos de la ronda inicial del estudio) y calcular la especificidad, el valor predictivo negativo y otros parámetros de rendimiento¹³. Se ha recalculado un indicador que denominamos FN+CN (falsos negativos + nuevos cánceres)¹⁴: la suma serían los 26 cánceres detectados en el seguimiento (intervalo y ronda siguiente) tanto para la MC como para la MD, pues todos ellos se dieron en mujeres que habían tenido mamografías negativas en la ronda inicial. Para calcular la tasa de FN+CN de la MC, a estos 26 cánceres hay que sumar 5 más que fueron detectados en mujeres con MC negativa (gracias a una MD inicial positiva). Con ello, la tasa de FN+CN de la MC es de 0,84% y el valor predictivo negativo de 99,1%. Del mismo modo, para la MD, hay que sumar a los 26 cánceres descubiertos en el seguimiento, 11 que ocurrieron en mujeres con MD inicial negativa (descubiertos gracias a que la MC fue positiva). Con estos datos, la tasa de FN+CN de la MD es del 1,0% y el valor predictivo negativo de 98,9%. Las diferencias entre la

¹³ Para esta revisión, se han calculado los parámetros de rendimiento diagnóstico teniendo en cuenta que una mamografía se considera positiva si tras la reunión de consenso se decide llamar a la mujer para nuevos exámenes. Se insiste en este punto porque en los artículos del estudio Oslo I los autores a veces no tienen en cuenta su propia definición de mamografía positiva (tras consenso) y recalculan los parámetros de rendimiento diagnóstico de forma que suponen como positiva cualquier mamografía en que alguno de los radiólogos la consideró como tal (aunque luego se descartara como negativa tras el consenso).

¹⁴ El indicador FN+CN se crea por la imposibilidad de saber cuántos cánceres del seguimiento (intervalo y siguiente ronda de cribado) eran cánceres nuevos o falsos negativos (cánceres que ya estaban presentes en la ronda de cribado anterior y no se llegaron a diagnosticar por cualquier motivo). Otros parámetros de rendimiento diagnóstico tras el seguimiento (sensibilidad y valor predictivo negativo) se han calculado considerando como falsos negativos a todos los FN+CN. Aunque esto sesga estos indicadores, lo hace por igual para las dos mamografías (MC y MD).

MD y MC en la tasa de FN+CN y el valor predictivo negativo no fueron estadísticamente significativas. Tampoco las diferencias en la sensibilidad (45,6% para MC y 35,1% para MD) alcanzaron significación estadística. Sin embargo, ambos tipos de mamografía tuvieron diferencias estadísticamente significativas en la tasa de falsos positivos (4,0% para la MD y 2,8% para la MC) y en la especificidad (95,9% para la MD y 97,2% para la MD).

En el segundo artículo del estudio Oslo II (39) se aportan los datos de comparación de la MD y MC tras finalizar el seguimiento del ensayo clínico. Aunque se dan algunos datos desagregados para los dos grandes grupos de edad (45-49 años y 50-69 años), los parámetros fundamentales de rendimiento diagnóstico sólo pueden calcularse para el total de la población. Como ya se comentó anteriormente, aunque en este ensayo se aleatorizaron inicialmente 25.263 mujeres, el análisis final con seguimiento incluido se realiza sobre 23.929 (13.912 de 50-69 años, y 10.017 de 45-49 años), justificando los autores la exclusión del resto de mujeres por errores iniciales en la inclusión.

En las 23.929 mujeres se detectaron 105 cánceres en la primera ronda de cribado del estudio, 64 en el grupo con MC y 41 en el de MD¹⁵. A estos cánceres hay que añadir los detectados durante el seguimiento, que en el grupo de 45-49 años fueron los cánceres de intervalo diagnosticados durante 18 meses, mientras que en las mujeres de 50-69 años fueron la suma de cánceres de intervalo diagnosticados durante dos años y los cánceres diagnosticados en la siguiente ronda de cribado¹⁶. Para el total de mujeres, se diagnosticaron 109 cánceres de seguimiento, 52 de intervalo y 57 detectados en la segunda ronda de seguimiento. Estos 109, sumados a los 105 iniciales hacen un total de 214 cánceres.

En las 10.017 mujeres de 45-49 años se diagnosticaron 12 cánceres de intervalo (10 en las 7.082 con MC y 2 en las 2.935 con MD). En las 13.912 mujeres de 50-69 años, se diagnosticaron 40 cánceres de intervalo (30 en las 9.903 con MC y 10 en las 4.009 con MD) y 57 cánceres en la siguiente ronda de cribado (42 con MC y 15 con MD). De los 40 cánceres de intervalo diagnosticados en las mujeres de 50-69 años, todos fueron cánceres invasivos excepto un carcinoma *in situ* diagnosticado en el grupo asignado a la MC¹⁷.

Considerando al total de la población, la tasa de detección de cánceres de intervalo fue superior con la MC (0,23%) que con la MD (0,17%), pero

¹⁵ Nótese la diferencia con los 114 cánceres (73 detectados con MC y 41 con MD) que se informan en el artículo de Skaane 2004 (38), que analiza el rendimiento diagnóstico basal con una población mayor.

¹⁶ En las mujeres de 45-49 años no hubo segunda ronda de cribado dentro del ensayo clínico.

¹⁷ No se aporta información sobre el porcentaje de cánceres invasivos e *in situ* en los cánceres de intervalo del grupo de 45-49 años.

esta diferencia no fue estadísticamente significativa ($p=0,35$). Al realizar el análisis por grupos de edad, pasó algo similar. En el grupo de 50-69 años, la tasa de detección de cánceres de intervalo fue superior para la MC (0,30%) que para la MD (0,25%); y en el grupo de 45-49 años, ésta fue de 0,14% para a MD y 0,07% para la MD; ninguna de estas diferencias fue estadísticamente significativa.

A los efectos de esta revisión, se han desestimado algunos parámetros de rendimiento diagnóstico aportados por los autores, por motivos similares a los ya comentados para el estudio Oslo I¹⁸. Los autores calcularon la sensibilidad de la forma más habitual (excluyendo del denominador los cánceres detectados en la siguiente ronda de cribado¹⁹). Aunque esto es correcto, no permite la comparación con los cálculos realizados para los otros estudios de seguimiento (Oslo I y estudio de Pisano et al), donde los cánceres detectados en la siguiente ronda se han considerado como fallos de la mamografía inicial (vease al inicio de este apartado la explicación sobre el indicador FN+CN). Por estos motivos, a partir de los datos del estudio, se han realizado nuevas estimaciones de los parámetros de rendimiento diagnóstico (Tabla 12). Para ello, se han considerado como mamografías positivas sólo aquellas donde hubo rellamada tras la reunión de consenso y como FN+CN todos los casos de mamografía inicial negativa donde se diagnosticó un cáncer en el seguimiento, ya fuera en el intervalo o con la siguiente mamografía de cribado.

La sensibilidad fue superior con la MD (61,8%) que con la MC (45,2%), diferencia que fue estadísticamente significativa. Las sensibilidades que dan los autores en el artículo (sin incluir en el denominador los cánceres detectados en la siguiente ronda de cribado) tienen valores más altos para ambas mamografías, pero también es superior con la MD (77,4%) que con la MC (61,5%). La especificidad fue un poco más alta en la MC (97,9%) que en la MD (96,3%) y la tasa de falsos positivos fue más alta con la MD (3,62%) que con la MC (2,1%); siendo todas estas diferencias estadísticamente significativas. No se encontraron diferencias entre MC y MD en la tasa de FN+CN y en el valor predictivo de un resultado negativo.

¹⁸ Para esta revisión, se ha considerado que la mamografía inicial fue positiva si tras la reunión de consenso se decidió llamar a la mujer para nuevos exámenes. Al igual que pasaba en los artículos del estudio Oslo I, en el estudio Oslo II los autores a veces no tienen en cuenta su propia definición de mamografía positiva (tras consenso). Así, en Oslo II consideran como verdaderos positivos los casos en que la mamografía de cribado fue calificada como tal por alguno de los radiólogos, pero luego se consideró negativa en la reunión de consenso. Es más, esto sólo lo consideran así si posteriormente se detectó un cáncer de intervalo o en la siguiente ronda.

¹⁹ Consideran como fallos del cribado los cánceres de intervalo y como nuevos cánceres los detectados en la siguiente ronda.

TABLA 12. Parámetros de rendimiento diagnóstico de la mamografía digital versus la mamografía convencional en los estudios noruegos (considerando el seguimiento)

| Estudio N | Técnica | VP | FP | FN+CN* | VN | S* | E | Tasa FN+CN* | Tasa FP | VPN* |
|---|---------|----|-----|--------|-------|-------|--------|-------------|---------|-------|
| OSLO I Skaane et al (2005) (37) 3683 mujeres | MC | 26 | 102 | 31 | 3524 | 45,6% | 97,2% | 0,84% | 2,8% | 99,1% |
| | MD | 20 | 148 | 37 | 3478 | 35,1% | 95,9% | 1,00% | 4,01% | 98,9% |
| | p | - | - | - | - | NS | <0,05 | NS | <0,05 | NS |
| OSLO II Skaane et al (2007) (39) 23929 mujeres** | MC | 66 | 359 | 80 | 16480 | 45,2% | 97,9% | 0,47% | 2,1% | 99,5% |
| | MD | 42 | 252 | 26 | 6624 | 61,8% | 96,5% | 0,37% | 3,62% | 99,6% |
| | p | - | - | - | - | 0,035 | <0,005 | NS | <0,005 | NS |

MC (mamografía convencional), MD (mamografía digital), VP(verdaderos positivos), FP(falsos positivos), FN+CN (falsos negativos + casos nuevos detectados en el intervalo o siguiente ronda), VN (verdaderos negativos), VPN (valor predictivo negativo), S (sensibilidad), E (especificidad). NS: no significativo estadísticamente.

* La Sensibilidad y Valor predictivo negativo se han calculado suponiendo que fueron falsos negativos todos los casos incluidos en el indicador FN+CN.

** En el estudio OSLO II se aleatorizaron las mujeres a MC (16985) y MD (6944).

4.4.4. Estudio DMIST (Digital Mammographic Imaging Screening Trial) realizado en Estados Unidos y Canadá

Los resultados principales del estudio DMIST, realizado en Estados Unidos y Canadá, se publicaron en 2005 (42) (Tablas 13 a 15). Aunque el estudio incluyó inicialmente a 49.528 mujeres, muchas de ellas se excluyeron por motivos diversos, la mayoría porque faltó información sobre el seguimiento. Finalmente se aportan datos de 42.760 mujeres asintomáticas que acudieron para realizarse una mamografía de cribado a uno de los 32 centros sanitarios participantes. Todas las mujeres recibieron una MC y una MD. El seguimiento fue de 15 meses y se hizo una segunda mamografía de cribado al final del mismo.

En ese periodo se diagnosticaron 335 cánceres de mama, 254 (75,8%) en los primeros 365 días (éstos incluyen los descubiertos en la mamografía inicial y en el seguimiento de un año) y 81 (24,2%) entre los 366 y 455 días.

Del total de cánceres detectados, un 69% fueron carcinomas invasivos y un 31% *in situ*, y el estadio T fue menor de T2 en el 89% de los casos. Los hallazgos histológicos y el estadio de los cánceres no fueron diferentes según la positividad de las mamografías iniciales (mujeres con MC y MD positiva, con MC y MD negativas o solo una de las dos positivas).

Los autores presentan diversos tipos de análisis de rendimiento diagnóstico, ya que usan tres sistemas distintos de clasificación de una mamografía como positiva o negativa (una escala de 7 puntos, el BI-RADS® de 5 categorías, y la recomendación del radiólogo de realizar exámenes tras algún hallazgo en la mamografía). El análisis del área bajo la curva ROC se realizó con la escala de 7 puntos. Como se observa en la tabla 10, en la población total no se encontraron diferencias estadísticamente significativas entre MC y MD; sin embargo, la MD resultó tener un mejor rendimiento diagnóstico que la MC en las mujeres menores de 50 años, con mamas heterogéneamente o extremadamente densas, y en las pre o perimenopáusicas. No hubo diferencias entre MC y MD en las mujeres mayores de 50 años ni en las posmenopáusicas (estos últimos datos no se muestran en la siguiente Tabla 13).

TABLA 13. Área bajo la curva ROC para la mamografía digital y convencional basado en la escala de malignidad de 7 puntos de la mamografía

| | Total de mujeres | Mujeres menores de 50 años | Mujeres con mamas de densidad extrema | Mujeres pre y peri menopáusicas |
|-----------------------------------|---------------------------|----------------------------|---------------------------------------|---------------------------------|
| MC | 0,74 | 0,69 | 0,68 | 0,67 |
| MD | 0,78 | 0,84 | 0,78 | 0,82 |
| Diferencia entre MD y MC (IC 95%) | 0,03 (de -0,02 a 0,08) | 0,15 (de 0,05 a 0,25) | 0,11 (de 0,04 a 0,18) | 0,15 (de 0,05 a 0,24) |
| Valor de p | 0,18 | 0,002 | 0,003 | 0,002 |

MC: mamografía convencional; MD: mamografía digital; IC: intervalo de confianza

En las Tablas 14 y 15 se muestran las estimaciones de la sensibilidad, especificidad, y valor predictivo positivo con el uso de la escala BI-RADS® después de 365 días de seguimiento y con la escala de 7 puntos de malignidad después de 455 días de seguimiento. En el total de la población, no se encuentran prácticamente diferencias relevantes en la sensibilidad, especificidad y valor predictivo positivo de la MC y MD, ni en el análisis con BI-RADS® ni en el realizado con la escala de 7 puntos.

TABLA 14 . Exactitud diagnóstica de MC y MD con uso de la escala BI-RADS® * después de 365 días de seguimiento.

| | Todas las mujeres | | | Mujeres menores de 50 años | | | Mujeres pre y peri-menopausicas | | | Mujeres con mamas heterogeneas o densas | | |
|-------------------|-------------------|------|------|----------------------------|------|------|---------------------------------|------|------|---|------|------|
| | S | E | VPP | S | E | VPP | S | E | VPP | S | E | VPP |
| MC | 0,66 | 0,92 | 0,05 | 0,51 | 0,90 | 0,02 | 0,51 | 0,90 | 0,03 | 0,55 | 0,90 | 0,03 |
| MD | 0,70 | 0,92 | 0,05 | 0,78 | 0,90 | 0,03 | 0,72 | 0,90 | 0,04 | 0,70 | 0,91 | 0,04 |
| Valor de p | 0,37 | 0,74 | ND | 0,002 | 0,89 | ND | 0,008 | 0,37 | ND | 0,02 | 0,09 | ND |

MD: mamografía digital, MC: mamografía convencional, S: sensibilidad, E: especificidad, VPP: valor predictivo positivo, IC: intervalo de confianza), NS: no significativo estadísticamente, ND: no datos.

* Para calcular los parámetros de rendimiento diagnóstico, la puntuación de la escala BI-RADS® se dicotomizó en positiva (0, 4 y 5 puntos) y negativa (de 1 a 3 puntos).

TABLA 15 . Exactitud diagnóstica de MC y MD con uso de la escala de malignidad de 7 puntos* después de 455 días de seguimiento.

| | Todas las mujeres | | | Mujeres menores de 50 años | | | Mujeres pre y peri-menopausicas | | | Mujeres con mamas heterogeneas o densas | | |
|-------------------|-------------------|-----------------------|------|----------------------------|------|------|---------------------------------|------|------|---|------|------|
| | S | E | VPP | S | E | VPP | S | E | VPP | S | E | VPP |
| MC | 0,41 | 0,98 | 0,13 | 0,35 | 0,98 | 0,07 | 0,38 | 0,98 | 0,09 | 0,36 | 0,97 | 0,10 |
| MD | 0,41 | 0,98 | 0,12 | 0,49 | 0,97 | 0,08 | 0,47 | 0,97 | 0,10 | 0,38 | 0,97 | 0,10 |
| Valor de p | 0,92 | 0,006 ⁽²⁰⁾ | ND | 0,06 | 0,07 | ND | 0,20 | 0,20 | ND | 0,69 | 0,33 | ND |

MD: mamografía digital, MC: mamografía convencional, S: sensibilidad, E: especificidad, VPP: valor predictivo positivo, IC: intervalo de confianza), ND: no datos.

* Para calcular los parámetros de rendimiento diagnóstico, la puntuación de la escala de malignidad se dicotomizó en positiva (de 4 a 7 puntos) y negativa (de 1 a 3 puntos).

²⁰ El valor de esta p llama la atención, dado que los valores de especificidad son iguales para MC y MD. Es posible que se encuentre significación estadística a una diferencia muy pequeña (no recogida con dos decimales), teniendo en cuenta el gran tamaño muestral. Sin embargo, dada la falta de relevancia desde el punto de vista clínico, se ha decidido considerar que no existen diferencias en la especificidad de ambos tipos de mamografía.

Sin embargo, en las mujeres menores de 50 años, en las pre o perimenopáusicas, y en las que tenían mamas muy densas, la sensibilidad de la MD fue superior a la de la MC en el análisis realizado al año con la escala BI-RADS®, siendo dichas diferencias estadísticamente significativas ($p < 0,05$). Estas diferencias en la sensibilidad entre ambos tipos de mamografía prácticamente se pierden en el análisis realizado con la escala de malignidad de 7 puntos a los 455 días de seguimiento (solo en las mujeres menores de 50 años se queda en los límites de la significación estadística). En ninguno de los análisis de subgrupos realizados a los 365 y 455 días hubo diferencias relevantes en la especificidad y valores predictivos entre MD y MC.

4.4.5. Estudio realizado en Florencia

En este estudio (43) se compararon retrospectivamente dos cohortes de mujeres, una que recibió MD y otra que recibió MC (tabla 16). La mayor parte de los datos se dan desagregados por grupos de edad (50-59 y 60-69 años) y por otro tipo de variables (densidad de la mama, ronda de cribado, etc.).

En este estudio se detectaron 188 cánceres (84 en el grupo de MC y 104 en el grupo de MD). De ellos, 144 fueron invasivos (76,6%) y 44 fueron carcinomas in situ (23,4%). El porcentaje de cánceres en estadios precoces (pTis, pT1mic, pT1a) fue mayor en el grupo de MD (41,3%) que en el de MC (27,3%), estando la diferencia en los límites de la significación estadística ($p = 0,06$).

En la Tabla 12 se puede observar como la tasa de detección de cáncer fue superior con la MD (0,72%) que con la MC (0,58%), pero esta diferencia no fue estadísticamente significativa. Tampoco las diferencias alcanzaron significación estadística, cuando se analizaron en cada subgrupo de edad (50-59 años y 60-69 años), cada subgrupo de densidad mamaria y según si se trataba de la primera o siguientes rondas de cribado (sólo se muestran los datos por subgrupo de edad en la tabla). Sin embargo, al estimarse según tipo de anormalidad detectada en la mama, la tasa de detección resultó significativamente superior con la MD (0,26%) que con la MC (0,12%) en el subgrupo con microcalcificaciones ($p = 0,007$) (datos no mostrados en la tabla).

Respecto al VPP, las diferencias entre MD y MC fueron muy pequeñas y no resultaron estadísticamente significativas (tabla 16). Esto fue así en la población total, para cada subgrupo de edad, para diferentes densidades de la mama y para rondas de cribado incidentes y prevalentes.

En cuanto a la tasa de rellamadas, fue superior con la MD (4,56%) que con la MC (3,96%), siendo dicha diferencia estadísticamente significativa. Cuando se estimó en cada subgrupo de edad, se observó como en las mujeres de 50-59 años, la tasa de rellamada para MD fue significativamente superior que para MC. Sin embargo, en el subgrupo de 60-69 años, las diferencias no fueron estadísticamente significativas.

TABLA 16. Parámetros de rendimiento diagnóstico de la mamografía digital versus la mamografía convencional en el estudio de Del Turco (43)

| Estudio N | Técnica | Tasa rellamadas | N° cánceres detectados (Tasa detección) | VPP |
|-----------------------------|------------|-----------------|---|-------|
| Del Turco (2007) | MC | 3,96% | 0,58% | 14,7% |
| | MD | 4,56% | 0,72% | 15,9% |
| 28770 mujeres (todas) | Valor de p | 0,01 | 0,14 | 0,65 |
| Del Turco (2007) | MC | 4,17% | 0,42% | 10,2% |
| | MD | 5,12% | 0,63% | 12,4% |
| 13510 mujeres de 50-59 años | Valor de p | 0,009 | 0,09 | 0,47 |
| Del Turco (2007) | MC | 3,77% | 0,72% | 19,0% |
| | MD | 4,07% | 0,79% | 19,6% |
| 15260 mujeres de 60-69 años | Valor de p | 0,33 | 0,57 | 0,95 |

4.5. Síntesis de los resultados

Todos los estudios encontrados se centran en la valoración del rendimiento diagnóstico de las mamografías, no aportando información sobre mortalidad u otros resultados en salud.

Para matizar los resultados sobre rendimiento diagnóstico, sobre todo en algunos casos en que ha habido algún tipo de discrepancia entre estudios, se han tenido en cuenta los aspectos de validez y las propias características del contexto (población, método de interpretación de las mamografías, etc.).

Validez de los estudios incluidos en la revisión

Para graduar la importancia de los resultados finales según la calidad de los estudios, éstos se han clasificado en tres grupos según la validez del diseño y

la existencia de seguimiento tras el cribado. Esta clasificación no ha tenido en cuenta el resultado de la valoración con QUADAS, ya que fue similar en los trabajos incluidos y no discrimina apenas entre ellos.

La mayor validez para comparar la MD con la MC la tienen las estimaciones del rendimiento diagnóstico que incluyen información sobre los cánceres detectados durante el seguimiento (cánceres de intervalo y de la ronda siguiente de cribado). En este grupo de mayor validez se incluyen los resultados finales del único ensayo clínico con asignación aleatoria encontrado (Oslo II) (39), y los datos de los dos estudios de cohorte única con seguimiento²¹ (últimos artículos publicados del estudio Oslo I y DMIST) (37,42). En un segundo orden de validez se considerarían las estimaciones del rendimiento diagnóstico que no tienen en cuenta el seguimiento, pero se realizan a partir de diseños que aseguran la comparabilidad de los grupos con MD y MC (ensayos clínicos y estudios de cohorte única). Aquí entrarían las estimaciones del rendimiento diagnóstico realizadas antes del seguimiento en los estudios Oslo I y Oslo II (36,38), así como la aportada en el estudio de cohorte única de Lewin (34). Por último, el tercer grupo de estudios según la validez del diseño, estaría formado por aquellos con dos cohortes de mujeres elegidas de forma separada, una de las cuales recibe la MD y otra la MC. En este grupo se incluyen los estudios de Vigeland y Del Turco, realizados en Noruega y Florencia respectivamente (40,43).

Comparación del rendimiento diagnóstico de MD y MC en los estudios con seguimiento (Oslo II, Oslo I y DMIST) (39,37,42)

En este grupo se incluyen los datos tras el seguimiento de tres estudios: el único ensayo clínico con asignación aleatoria encontrado (estudio Oslo II, con 23929 mujeres), y dos estudios de cohorte única (Oslo I, con 3683 mujeres; y DMIST, con 42760 mujeres).

En el estudio Oslo II, la sensibilidad de la MD fue superior a la de la MC (61,8% versus 45,2%). Sin embargo, en el estudio Oslo I la tendencia fue en sentido contrario (mayor sensibilidad con MC), aunque la diferencia entre ambos tipos de mamografía no fue estadísticamente significativa. Tampoco en el estudio DMIST se encontraron diferencias entre ambas estimaciones de la sensibilidad. Las discrepancias entre los dos estudios Oslo podrían explicarse por el rango de edad de las mujeres incluidas (50-69 años en el estudio Oslo I y 45-69 años en el Oslo II). Sin embargo, esta explicación no valdría para las diferencias entre el estudio Oslo II y el DMIST, ya que ambos incluyen a mujeres menores y mayores de 50 años. Además, el estudio DMIST tiene mayor muestra que el Oslo II, por lo que tendría mayor poder para detectar diferencias.

²¹ Cohorte única: cada una de las mujeres de la cohorte recibe los dos tipos de mamografía.

Un parámetro muy relacionado con la sensibilidad es la tasa de detección de cánceres de intervalo. El único estudio que aportó este dato (Oslo II) no encontró diferencias entre ambas mamografías, ni en la población total ni en ninguno de los dos grupos de edad (mayores y menores de 50 años).

La especificidad de la MD fue menor que la de la MC en los estudios Oslo II y Oslo I, no encontrándose diferencias significativas en el DMIST. Las principales diferencias entre el estudio DMIST y los dos estudios Oslo que pudieran tener alguna relación con estas discrepancias fueron dos. Por un lado, en los estudios Oslo la interpretación de la mamografía y la decisión de seguir haciendo pruebas fue mucho más estricta (dos radiólogos independientes y una reunión de consenso posterior) que en DMIST (lectura por un único radiólogo). Por otro lado, el estudio DMIST no se hizo en el contexto de un programa poblacional, sino partiendo de las mujeres que acudían a diversos centros sanitarios para realizarse una mamografía de cribado.

La tasa de falsos positivos fue mayor con la MD (entre 3,6% y 4,0%) que con la MC (entre 2,1% y 2,8%), tanto en el estudio Oslo II como en el Oslo I. No hubo diferencias entre ambos tipos de mamografía en la tasa FN+CN (falsos negativos más casos de cánceres nuevos detectados en el seguimiento) ni en el valor predictivo negativo en ninguno de los dos estudios Oslo. El estudio DMIST no aportó información sobre estos parámetros.

El estudio DMIST es el único de este grupo que proporciona el área bajo la curva ROC y realiza análisis desagregados por edad y otras variables sobre el rendimiento diagnóstico teniendo en cuenta el seguimiento. Aunque en la población total no encontró diferencias estadísticamente significativas en ninguno de los parámetros analizados (área bajo la curva ROC, valor predictivo positivo, etc.), sí halló diferencias en algunos subgrupos. En las mujeres menores de 50 años, con mamas densas, y pre y peri menopáusicas, la MD presentó una mayor área bajo la curva ROC y una mayor sensibilidad que la MC. Sin embargo, en ninguno de estos subgrupos encontraron diferencias en la especificidad y valor predictivo positivo de ambos tipos de mamografía.

Comparación del rendimiento diagnóstico de MD y MC sin incluir seguimiento en estudios con grupos comparables (Oslo I, Oslo II, Lewin) (34,36,38)

En este grupo se incluyen los datos de tres estudios, sin considerar cánceres detectados en el seguimiento (son datos transversales): el único ensayo clínico con asignación aleatoria encontrado (estudio Oslo II, con 23.929 mujeres), y dos estudios de cohorte única (Oslo I, con 3.683 mujeres; y Lewin, con 4.489 mujeres).

Las tasas de detección de cáncer de mama no presentaron diferencias estadísticamente significativas entre MD y MC en ninguno de los tres estudios, excepto en el subgrupo de mujeres de 50-69 años del estudio Oslo II, donde la tasa de detección fue mayor con MD que con MC, con una diferencia en el límite de la significación estadística ($p=0,053$). Sin embargo, en este mismo estudio no se encontraron diferencias entre MD y MC en las mujeres de 45-49 años. Esta posible diferencia del comportamiento de las dos mamografías según la edad no ha podido corroborarse en el estudio Oslo I (que sólo incluye mujeres de 50-69 años) ni en el de Lewin (que incluye mujeres a partir de los 40 años, pero no aporta datos desagregados por edad). Sin embargo, estos resultados del estudio Oslo II contrastan con los hallados en el seguimiento del estudio DMIST (ya mencionados en el apartado anterior), donde ocurría lo contrario: la MD presentaba mejor rendimiento diagnóstico que la MC en las mujeres menores de 50 años, pero no en las mayores.

Sólo el estudio de Lewin comparó el porcentaje de carcinomas in situ detectados con MD y MC, no encontrando diferencias entre ambas técnicas.

Ninguno de los dos estudios que aportan datos sobre sensibilidad (Lewin y Oslo I) encontraron diferencias entre MC y MD, si bien en ambos casos hubo una tendencia a mayor sensibilidad de la MC (78,6%-83,9%) que de la MD (64,3%-64,5%).

No hubo diferencias estadísticamente significativas en el valor predictivo positivo de MD y MC, hallazgo consistente en los tres estudios, incluidos los dos subgrupos de edad (45-49 años y 50-69 años) analizados por separado en el estudio Oslo II.

El único estudio que aportó información sobre el área bajo la curva ROC fue el de Lewin, encontrando áreas similares para MD y MC.

Comparación del rendimiento diagnóstico de MD y MC en los estudios de dos cohortes elegidas de forma separada (Vigeland y Del Turco) (40,43).

En este grupo se incluyen los datos de dos estudios de cohortes con dos grupos de mujeres seleccionadas de forma separada, ninguno de los cuales consideró cánceres detectados en el seguimiento (son datos transversales): el estudio de Vigeland se realizó en Noruega e incluyó 343.002 mujeres, y el estudio de Del Turco se realizó en Florencia e incluyó 28.770 mujeres.

La tasa de detección de cáncer en ambos estudios era algo mayor para la MD (0,72%-0,77%) que para la MC (0,58%-0,65%), aunque dichas diferencias no alcanzaron significación estadística. En el estudio de Vigeland, el porcentaje de carcinomas in situ detectados fue superior con la MD (27%) que con la MC (16,3%); y en el estudio de Del Turco el porcentaje

de cánceres en estadios precoces fue superior con la MD que con la MC, aunque la diferencia se quedó en los límites de la significación estadística ($p=0,06$).

El valor predictivo positivo fue mayor para la MD (18,5%) que para la MC (15,1%) en el estudio de Vigeland ($p=0,015$), no encontrándose diferencias estadísticamente significativas en el estudio del Turco. Una posible explicación de las diferencias encontradas en el estudio de Vigeland, frente al resto de los estudios (Del Turco, estudios Oslo I y II, y Lewin), es que puedan estar relacionadas con el gran tamaño muestral, con poder suficiente para detectar diferencias muy pequeñas.

Comparación de otros resultados de MD y MC

Los tres estudios noruegos (Oslo I, Oslo II y Vigeland) (36,38 y 40), el estudio de Lewin (34) y el de Del Turco (43) aportan datos sobre la tasa de rellamadas con MD y MC. Los resultados fueron diferentes entre estos estudios. En el estudio de Lewin, dicha tasa fue significativamente menor tras una MD que tras una MC; por el contrario, en el estudio de Del Turco y en el subgrupo de mujeres de 50-69 años del estudio Oslo II fue significativamente mayor tras la MD; y por último, no se encontraron diferencias estadísticamente significativas entre MD y MC en los estudios Oslo I, subgrupo de mujeres de 45-49 años del estudio Oslo II y Vigeland. Estas controversias podrían explicarse en parte por las diferencias ya comentadas entre los estudios: número de mujeres incluidas, diferentes edades, y proceso de decisión para hacer una rellamada (mucho más estricto en los estudios noruegos, donde se requiere doble lectura independiente y reunión de consenso).

El estudio de Vigeland (40), diferenció entre la tasa de rellamadas debida a hallazgos en la mamografía (mayor parte de las rellamadas) y la debida a problemas técnicos. Si bien no encontró diferencias entre MD y MC en la tasa debida a hallazgos en la mamografía, si encontró que la tasa de rellamadas por problemas técnicos fue menor para MD que para MC (0,22% y 0,61%, $p<0,001$).

Por último, el estudio Oslo I (36) y el de Lewin (34) aportaron información sobre el porcentaje de biopsias realizadas, llegando a diferentes conclusiones. De ellos, el estudio Oslo I (con mayor validez por diseño y por valoración con QUADAS) no encontró diferencias entre ambos tipos de mamografía (1,5% y 1,4%) mientras que Lewin encontró mayor tasa de biopsias con MC (2%) que con MD (1,4%), siendo la diferencia estadísticamente significativa ($p<0,001$).

5 Discusión

La evidencia disponible acerca del rendimiento diagnóstico de la mamografía digital frente a la mamografía convencional en el cribado de cáncer de mama procede de 6 estudios, 4 de los cuales tienen un tamaño muestral mayor de 20.000 mujeres. A pesar de ello, la interpretación de la evidencia se ve limitada por la diversidad de diseños y de contextos de los estudios.

La mayor parte de los resultados de estos estudios se ha publicado entre 2005 y 2007, siendo especialmente importantes los artículos que presentan datos después de un tiempo de seguimiento tras el cribado inicial.

Tras esta revisión sistemática sobre el rendimiento diagnóstico comparado de la MD y la MC en el cribado de cáncer de mama, puede decirse que los estudios encontrados tienen características diferentes y que los resultados no son del todo consistentes. Centrándose en los estudios con un diseño de mayor validez (cohortes comparables y con seguimiento) (37,39,41), los resultados obtenidos al comparar el cribado con MD y con MC sugieren que no hay grandes diferencias en el rendimiento diagnóstico de ambas técnicas. En el único ensayo clínico incluido (39), la sensibilidad de la MD supera a la de la MC, pero en el resto de estudios con seguimiento no se encuentran diferencias estadísticamente significativas entre ambos tipos de mamografía. El estudio de Pisano (41), (que actualmente continúa en marcha), es el único trabajo dentro del grupo de estudios de mayor validez que ha analizado diversos parámetros de rendimiento comparado de la MD y MC en diferentes subgrupos poblacionales después de un tiempo de seguimiento. Estos autores encontraron que en las mujeres menores de 50 años, con mamas densas y perimenopaúsicas, la MD tenía mayor área bajo la curva ROC y mayor sensibilidad que la MC. Sin embargo, el artículo que aporta los datos del seguimiento del ensayo clínico Oslo II (39), no encontró diferencias entre MD y MC en las mujeres menores de 50 años en el único parámetro del que da información desagregada por edad: la tasa de detección de cánceres de intervalo.

Otros estudios incluidos en esta revisión, con menor validez de diseño pero con similar puntuación de calidad en la herramienta QUADAS (19), coinciden en que la tasa de detección de cáncer de mama y el valor predictivo positivo no son diferentes entre ambas técnicas mamográficas.

Con todos estos resultados, la evidencia disponible hasta hoy no permite concluir que un tipo de mamografía tenga claramente un mejor rendimiento diagnóstico que la otra en el cribado del cáncer de mama.

5.1. Discusión de las limitaciones y aspectos metodológicos

En una revisión sistemática de la literatura, siempre hay que tener en cuenta que los resultados se basan en los hallazgos de trabajos ya realizados, por lo que está sujeta a las limitaciones de la calidad y número de estudios encontrados. En este caso, cabe destacar que hay un número importante de estudios realizados sobre el tema de interés. Aunque en la valoración de la calidad con la herramienta QUADAS, los estudios revisados presentan una calidad similar, sus diseños son variados, hallándose desde un ensayo clínico con asignación aleatoria, a varios estudios de cohortes con seguimiento, y algunos transversales. Los estudios incluidos se centraron en general en la valoración del rendimiento diagnóstico de ambas técnicas, no aportando en ningún caso información sobre mortalidad u otros resultados en salud.

Los resultados sobre rendimiento diagnóstico encontrados fueron variados y presentaron algunas diferencias entre los estudios. Éstas podrían explicarse en parte por las diferentes características del contexto de cada estudio y de la validez de los diseños.

Por otra parte, los artículos que aparecen tras realizar una búsqueda bibliográfica pueden constituir una muestra sesgada del total de estudios que se han realizado, ya que tienden a publicarse más aquellos trabajos que aportan mejores resultados en relación a la hipótesis inicial (en este caso, por ejemplo, que los resultados de la MD superen a los de la MC). En ésta revisión sólo se han incluido artículos publicados, por lo que no se puede descartar un sesgo de publicación.

Con respecto al análisis de los resultados en una revisión sistemática, lo ideal hubiera sido la realización de una síntesis cuantitativa de los mismos (metanálisis). En este caso, debido a la diversidad de los diseños de los estudios y a la heterogeneidad de los mismos (tanto en términos de las poblaciones como en la forma de lectura de mamografías y protocolo para rellamadas, etc.), no se ha considerado conveniente realizar un metanálisis, por lo que la síntesis de resultados se ha limitado a una descripción cualitativa de los mismos.

Es importante destacar por último, que la valoración crítica de los estudios fue realizada por dos revisoras de manera independiente, obteniéndose una alta concordancia entre las mismas.

5.2 Discusión resultados

5.2.1 Aportación de esta revisión comparada con otras revisiones

Con anterioridad al inicio de este informe, se localizaron tres revisiones sistemáticas relacionadas con el tema de interés (13,17,18). La más reciente de ellas, publicada en 2007 (13), tenía como uno de sus objetivos comparar la evidencia científica disponible sobre los resultados de la MD *versus* la MC en el contexto del cribado de cáncer de mama, e incluía artículos publicados entre 1989 y 2005. Esta revisión se tomó como punto de partida para la revisión actual, incluyéndose los estudios que ya estaban en ella y los publicados posteriormente (entre 2005 y 2007). Sin embargo, dicha revisión no matizó los resultados de los estudios originales incluidos según la calidad de los mismos, y los propios autores advierten de la necesidad de encontrar mejores evidencias respecto al tema.

Una de las aportaciones de la presente revisión sistemática es que se ha incluido una valoración crítica de la evidencia. Así, por una parte se ha realizado una evaluación de la calidad de los artículos mediante la herramienta QUADAS (19) y por otra, se han clasificado los estudios en base a la validez del diseño (según la comparabilidad de los grupos y la existencia o no de seguimiento). Esta evaluación crítica se ha usado para clasificar y matizar los resultados.

Otra de las revisiones previas se había publicado en 2004 (17), y su objetivo se centraba en revisar la precisión de las nuevas tecnologías en el cribado del cáncer de mama. Incluía artículos hasta diciembre de 2002, y la conclusión a la que llegaron los autores fue que aún no existían evidencias suficientes para apoyar el uso de las nuevas tecnologías de imagen, entre ellas la MD.

La tercera y última revisión encontrada, fue publicada en el año 2002 (18), y tenía como objetivo la comparación técnica, clínica y económica de la MD y la MC en el contexto del sistema sanitario canadiense. La conclusión a la que llegaron los autores fue que la MD ofrecía algunos beneficios técnicos y clínicos respecto a la MC, pero dicha evidencia era débil, y la magnitud de las diferencias no compensaba los costes diferenciales de ambas técnicas, por lo que recomendaban el uso de la MC al tratarse de la técnica de menor coste.

En este informe se han incluido todos los estudios relevantes (33,34,36,38,41,42) que estaban ya en las tres revisiones mencionadas (13,17,18), y se han añadido algunos estudios nuevos (35,37,39,40,43). Nuestra aportación principal consiste en la actualización de las revisiones

previas y la matización de los resultados en base a la valoración crítica de la literatura encontrada. Por otra parte, es importante destacar que en varios estudios de los incluidos, han sido recalculados diferentes parámetros, al encontrar algunos errores en los cálculos aportados por los artículos (ya explicado en detalle en el apartado de Resultados) o con la intención de hacer comparables los resultados de los diferentes trabajos. También, partiendo de los datos de los artículos incluidos en la revisión, se han hallado algunos parámetros de rendimiento diagnóstico no calculados por los autores. En ninguna de las revisiones sistemáticas publicadas anteriormente se habían realizado estos cálculos y correcciones.

5.2.2 Interpretación de los resultados

En la revisión llevada a cabo, para matizar los resultados sobre rendimiento diagnóstico, sobre todo en algunos casos en los que hubo algún tipo de discrepancia entre los estudios, se tuvieron en cuenta los aspectos de validez y las propias características del contexto (población, método de interpretación de las mamografías, etc.). Los diseños con mayor validez para comparar el rendimiento diagnóstico de la MD con la MC fueron el ensayo clínico con asignación aleatoria y los estudios de cohorte única con seguimiento prospectivo. El diseño inferior en términos de validez interna fue el estudio de dos cohortes seleccionadas de forma separada, una de las cuales recibió MD y otra MC, y en las que no se aportan datos de seguimiento.

Dos de los tres estudios de mayor validez de diseño no encontraron diferencias estadísticamente significativas en la sensibilidad de la MD respecto a la MC (37,41), y el tercero (único ensayo clínico localizado) (39) encontró una sensibilidad superior de la MD. Los estudios de mediana validez (34,36,38) coinciden en que no hay diferencias en la tasa de detección de cáncer de mama con las diferentes técnicas.

Sin embargo, en el trabajo de Pisano (41) et al, uno de los estudios del grupo de mayor validez, los autores llegaron a la conclusión de que en las mujeres menores de 50 años, con mamas densas y peri-premenopaúsicas, la MD presentaba mayor área bajo la curva ROC y mayor sensibilidad que la MC. Estos resultados ya se habían comentado en la revisión de Carreira et al, publicada (13) en 2007, y se han venido usando como argumento sobre las ventajas de la MD y como base de algunas evaluaciones económicas (44). Es más, algunos investigadores han tomado las conclusiones del estudio de Pisano como punto de partida para recomendar la MD en ciertos grupos de mujeres (13,30,44). Sin embargo, los interrogantes son muchos. Así, si

el rango de edad donde hay mayor acuerdo sobre la idoneidad del cribado de cáncer de mama es el de 50 a 69 años, podríamos preguntarnos si tiene sentido sustituir la MC por la MD, dado que Pisano solo encontró ventajas en las mujeres más jóvenes. Por otro lado, en el caso de programas de cribado que se inician antes de los 50 años (grupo de edad donde hay menor acuerdo sobre la idoneidad del cribado), cabe preguntarse si tendría sentido sustituir la MC por la MD en todas las mujeres o usar simultáneamente dos equipos, uno para cada grupo de edad. Por otra parte, hay que señalar que los resultados de Pisano no coinciden con los del estudio Oslo II. Según los datos transversales del estudio Oslo II (38), en las mujeres menores de 50 años, no habría diferencias en el rendimiento diagnóstico de la MC y la MD. Asimismo, la tasa de cánceres de intervalo en el seguimiento de este mismo estudio (39) fue similar con MD y con MC en el grupo de 45-49 años. Por tanto, no puede concluirse que haya una clara ventaja de la MD en las mujeres jóvenes.

Respecto a la especificidad de la MD, fue menor que la de la MC en la mayoría de los estudios de mayor validez de diseño. Asimismo, la tasa de falsos positivos fue mayor con MD y en la tasa de FN+CN (falsos negativos + casos nuevos detectados en el seguimiento) y el valor predictivo negativo no se encontraron diferencias entre ambos tipos de mamografía. Tan sólo en un estudio transversal se encontró un valor predictivo positivo mayor con la MD, no ocurriendo así en los demás estudios.

En la tasa de rellamadas y en el porcentaje de biopsias, los resultados varían mucho según los estudios, no encontrando diferencias significativas entre ambas técnicas los más estrictos, excepto en el subgrupo de 50-69 años del ensayo clínico (Oslo II), donde la tasa resultó mayor con la MD.

Las discrepancias entre los estudios pueden relacionarse en parte con las diferencias de edad de las mujeres de los diferentes estudios, del contexto (cribado oportunista o en programa poblacional), de la forma de interpretación de las mamografías por los radiólogos (con o sin doble lectura, con o sin consenso, etc.), de los diseños, etc. En cualquier caso, sí puede decirse con la evidencia acumulada hasta ahora, que de haber diferencias en el rendimiento diagnóstico de la MD y la MC, éstas no son de gran magnitud.

5.3 Implicaciones para la práctica clínica y la investigación

Aunque toda nueva tecnología conlleva un cierto entusiasmo por parte de los profesionales, con la investigación disponible hasta ahora, puede decirse que el debate sobre la mamografía digital como prueba de cribado continúa abierto. Aunque la adquisición de equipos para mamografía digital se ha ido extendiendo, no parece haber un uso generalizado en todos los hospitales ni en los programas de cribado poblacionales. Hoy día, en España, sigue predominando la mamografía convencional.

Si se obvia por un momento el tema de los costes y otros criterios que pueden tener relevancia en la toma de decisiones, podría decirse que la evidencia disponible no apoya la sustitución de la MC por la MD en el cribado del cáncer de mama. Es cierto que los estudios publicados en los últimos años son de mejor calidad que los anteriores, incluyen a un mayor número de mujeres y dan resultados de seguimiento; pero aún así, muchos resultados siguen siendo contradictorios, incluido el rendimiento diagnóstico en las mujeres menores de 50 años.

Hay un consenso generalizado en recomendar el cribado de cáncer de mama en las mujeres de 50 a 69 años. En mujeres más jóvenes la controversia es mayor, aunque la última publicación de la U.S. Task Force on Preventive Services amplía la recomendación de cribado al grupo de 45 a 49 años (11), algo que en España ya habían hecho algunos programas. En el caso de seguir ampliándose el cribado a este grupo de edad, podría tener interés la obtención de nuevas evidencias sobre el rendimiento diagnóstico comparado de la MD y MC en mujeres jóvenes. Hasta ahora, muchas de las opiniones a favor de la MD se argumentan en los hallazgos del estudio de Pisano (41), que encontró mejor rendimiento de este tipo de mamografía en las mujeres menores de 50 años, pre y perimenopáusicas y con mamas densas. Sin embargo, como ya se ha comentado, los resultados del estudio Oslo II no coinciden con estos hallazgos (38,39).

Si se tiene en cuenta todo esto, no parece que haya resultados para apoyar fuertemente la recomendación de sustituir la técnica convencional por la digital, al menos en el momento actual y si se toma como criterio fundamental de la decisión el rendimiento diagnóstico.

Sin embargo, es posible que tenga interés la valoración de otros criterios a la hora de formular recomendaciones para la planificación sanitaria. Uno de los principales problemas que surgen aquí es el tema de los costes: ¿cuándo un beneficio extra compensa su coste adicional?. Para responder esta pregunta son fundamentales los estudios de evaluación económica.

Durante la realización de este trabajo, se publicó una evaluación del coste-efectividad de la técnica de MD en el cribado de cáncer de mama (44). Los autores concluyen que la MD como técnica de cribado no es coste-efectiva en todas las mujeres. Parece indicar que sí lo sería en la población de mujeres menores de 50 años, aunque estos resultados se basan fundamentalmente en el estudio de Pisano. Por otra parte, es importante elegir bien el tipo de estudio de evaluación económica: con la evidencia disponible (sin que haya una clara superioridad de una técnica mamográfica sobre otra), lo más adecuado parece un estudio de minimización de costes.

Las diferencias económicas existentes entre ambas tecnologías son conocidas. Los costes de la MD son muy superiores que los de la MC (de 1,5 a 4 veces más) (25,41). Otros estudios publicados sobre el tema han tenido en cuenta que - aparte de los elevados costes de los nuevos equipos de MD- el coste del examen con MD sería de 24,22 a 26,46 euros más que con MC en el supuesto de realizar 5.000 test, y de 9,91 a 12,15 euros más si se realizan 10.000 test (32).

Aunque la MD resulta una técnica más costosa que la MC, es necesario considerar también otros criterios en la toma de decisiones. Algunas ventajas de esta técnica digital respecto a la convencional son claras: permite acortar el tiempo de realización, disminuir la dosis de radiación (15,18), almacenar imágenes en el ordenador, utilizar ayudas informatizadas para la lectura (CAD), realizar teletransmisión y teleconsulta, etc. (15,32). Pero también la técnica tiene algunos inconvenientes. Así, los radiólogos necesitan un cierto volumen de práctica, y presentan una variabilidad llamativa en la interpretación de los resultados tanto en situaciones experimentales como reales (32). Todo esto supone un importante factor limitante en la introducción de las MD en los hospitales. Hay que tener en cuenta que en un contexto de cribado poblacional podría maximizarse el aprovechamiento de las ventajas de la mamografía digital, pero también se maximizaría el impacto de sus inconvenientes.

La investigación sobre las técnicas diagnósticas en el cribado de cáncer de mama sigue produciendo nuevos resultados. La mayoría de los estudios publicados recientemente (44,47-50) se centran en la investigación de las nuevas técnicas digitales para la exploración de la mama, la utilización de la detección ayudada por ordenador, el papel de la resonancia magnética nuclear, etc. La Food and Drug Administration ha aprobado el uso de la resonancia magnética, el ultrasonido y la termografía para el diagnóstico de cáncer de mama, pero aún no para el cribado. (26). La tomosíntesis y las mamografías con contraste también están siendo exploradas como futuras técnicas a desarrollar.

Por otra parte, las técnicas se van mejorando. Así, en los últimos estudios publicados, se comenta ya una disminución de casi el 50 % de los falsos

positivos con las nuevas tecnologías de estereoscopia digital (mamografía 3D) (45).

A modo de resumen, puede concluirse que no hay grandes diferencias en el rendimiento diagnóstico de la MD y MC en el cribado del cáncer de mama, y que no hay evidencia sobre diferencias en términos de resultados de salud. Por tanto, en base a esto, no puede aconsejarse la sustitución de la MC por la MD. Sin embargo, es necesario tener en cuenta otros criterios a la hora de tomar decisiones (evaluación económica, ventajas y limitaciones de ambas técnicas desde el punto de vista de los radiólogos) y seguir avanzando en el conocimiento con nuevas investigaciones realizadas con las técnicas mejoradas y con otras tecnologías de imagen. Por último, es clave realizar investigaciones sobre los aspectos más relevantes y de los que aún no se conoce nada: el impacto comparado de las tecnologías en términos de resultados en salud, especialmente el efecto sobre la mortalidad y la calidad de vida.

6 Conclusiones

6.1. Sobre rendimiento diagnóstico

Rendimiento diagnóstico en los estudios con seguimiento y cohortes comparables:

Los resultados sobre rendimiento diagnóstico comparado de las mamografías digital (MD) y convencional (MC) difieren entre los estudios de mayor validez (con cohortes comparables y seguimiento). Por ello, no puede concluirse que una mamografía tenga claramente mejor rendimiento diagnóstico que la otra.

1. La sensibilidad de la MD fue superior a la de la MC en el único ensayo clínico con asignación aleatoria, no hallándose diferencias estadísticamente significativas en los dos estudios de cohorte única con seguimiento. Asimismo, no se encontraron diferencias en la tasa de detección de cánceres de intervalo con MD y MC.
2. La especificidad de la MD fue menor que la de la MC en dos de los tres estudios con cohortes comparables y seguimiento, no hallándose diferencias en el tercer estudio de estas características, a pesar de ser el de mayor tamaño muestral. Asimismo, la tasa de falsos positivos fue mayor con MD que con MC. Los distintos resultados sobre especificidad podrían relacionarse con la distinta forma de interpretar las mamografías y el contexto de los estudios (programas poblacionales versus cribado oportunista).
3. No se encontraron diferencias entre MD y MC en la tasa de FN+CN (falsos negativos más cánceres nuevos detectados en el seguimiento) ni en el valor predictivo negativo.
4. Sólo un estudio presenta resultados de rendimiento diagnóstico tras el seguimiento desagregados por edad y otras variables. En las mujeres menores de 50 años, con mamas densas y perimenopáusicas, la MD presentó mayor área bajo la curva ROC y mayor sensibilidad que la MC; y no hubo diferencias entre ambos tipos de mamografía en otros parámetros, como la especificidad y el valor predictivo positivo.

Rendimiento diagnóstico en los estudios sin considerar seguimiento:

5. Los tres estudios de cohortes comparables y los dos estudios de cohortes elegidas de forma separada que aportan datos sobre la tasa de detección de cáncer de mama, coinciden en que dicha tasa no es significativamente diferente con el cribado con MD y MC.
6. Sólo dos estudios aportaron datos sobre el porcentaje de carcinomas in situ, presentando resultados discordantes: uno no halló diferencias entre MD y MC y otro encontró un mayor porcentaje con MD.
7. De los cinco estudios que aportaron datos sobre el valor predictivo positivo, los tres con un diseño de mayor validez (grupos comparables) y uno de los estudios con cohortes separadas concluyeron que no había diferencias entre MD y MC. Sólo un estudio con cohortes separadas encontró mayor valor predictivo positivo con la MD.

6.2. Sobre otros resultados

8. No se ha encontrado evidencia científica sobre las diferencias entre las mamografías digital y convencional en términos de mortalidad u otros resultados en salud.
9. Los resultados sobre la tasa de rellamadas tras una mamografía de cribado fueron muy divergentes entre los distintos estudios. Los estudios más estrictos en la lectura de las mamografías no encontraron diferencias entre MD y MC, excepto en el subgrupo de mujeres de 50-69 años del único ensayo clínico, donde la tasa de rellamadas fue mayor con la MD. En el resto de estudios los resultados fueron opuestos: mayor tasa de rellamadas con MD (1 estudio) y con MC (1 estudio).
10. El porcentaje de biopsias se estudió en dos trabajos de cohorte única, siendo diferentes los resultados entre ellos: el estudio de mayor puntuación en la escala de validez no encontró diferencias entre MD y MC, mientras que el de menor validez encontró mayor tasa de biopsias con MC que con MD.

Referencias

1. Asociación Española contra el Cáncer [sitio web en Internet]. Madrid: La asociación; [actualizada Jun 2007; citada 10 Oct 2007]. Disponible en: <http://www.todocancer.com>.
2. Ministerio de Sanidad y Consumo – Instituto de Salud Carlos III. Agencia de Evaluación de Tecnologías Sanitarias (AETS). Cribado del cáncer de mama y cérvix. Uso de la mamografía y de la citología de Papanicolau para la detección precoz del cáncer de mama y de cérvix uterino en España. Madrid: Ministerio de Sanidad y Consumo; 2002. Informe de Evaluación de Tecnologías Sanitarias N° 34.
3. Levi F, Bosetti C, Lucchini F, Negri E, La Vecchia C. Monitoring the decrease in breast cancer mortality in Europe. *Eur J Cancer Prev.* 2005;14:497-502.
4. Ministerio de Sanidad y Consumo - Instituto de Salud Carlos III. Agencia de Evaluación de Tecnologías Sanitarias (AETS). Cribado poblacional de cáncer de mama mediante mamografía. Madrid: Ministerio de Sanidad y Consumo; 1995.
5. Olsen O, Gøtzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet.* 2001;358:1340-2.
6. Gøtzsche PC, Nielsen M. Cribaje (screening) con mamografía para el cáncer de mama; 2006 (Revisión Cochrane traducida). En: La Biblioteca Cochrane Plus, número 2, 2007. Oxford, Update Software Ltd. Disponible en: <http://www.update-software.com>. (Traducida de The Cochrane Library, 2007 Issue 2. Chichester, UK: John Wiley & Sons, Ltd.).
7. US Preventive Services Task Force. Screening for breast cancer [Internet]. Rockville: Agency for Healthcare Research and Quality; 2002 [citada 17 Jul 2007]. Disponible en: <http://www.ahrq.gov/clinic/uspstf/uspstfbrca.htm>.
8. Humphrey L, Helfand M, Chan B, Woolf S. Breast cancer screening: A summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2002;137:347-60.

9. Ringash J, with the Canadian Task Force on Preventive Health Care. Preventive health care, 2001 update: screening mammography among women aged 40-49 years at average risk of breast cancer. *CMAJ*. 2001;164:469-76.
10. Sociedad Española de Medicina de Familia y Comunitaria. Recomendaciones PAPPs. Programa de actividades preventivas y de promoción de la salud. Resumen [monografía en Internet]. Barcelona: Sociedad Española de Medicina de Familia y Comunitaria; 2003 [citada 9 Nov 2007]. Disponible en: http://www.papps.org/recomendaciones/01_recomendaciones.pdf
11. U.S. Preventive Services Task Force. Screening for breast cancer: recommendations and rationale. *Ann Intern Med*. 2002;137:344-6.
12. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. *Ann Intern Med*. 2007;146:502-10.
13. Carreira C, Martínez J, Gómez D, Polanco C, Naranjo P, Del Llano J. Revisión de la evidencia científica sobre la aplicación clínica de la mamografía digital. *Radiología*. 2007;49:145-56.
14. Screening for breast cancer [página principal en internet]. Rockville: National Guideline Clearinghouse (NGC). 1998 [actualizada Mar 2008; citada 10 Jun 2008]. Disponible en: <http://www.guideline.gov>.
15. ECRI Institute. Full-field digital mammography for breast cancer screening [monografía en Internet]. ECRI Institute; 2007 [citado 3 Oct 2007]. Disponible en: <http://target.ecri.org>.
16. Ortega D, Taborga M, Osses D, López A. Mamografía digital. El desafío del presente. *Rev Chile Radiol*. 2004;10:35-7.
17. Irwig L, Houssami N, Van Vliet C. New technologies in screening for breast cancer: a systematic review of their accuracy. *Br J Cancer*. 2004;90:2118-22.
18. Ho C, Hailey D, Warburton R, MacGregor J, Pisano E, Joyce J. Digital mammography versus film-screen mammography: technical, clinical and economic assessments [monografía en Internet]. Ottawa: Canadian Coordinating Office for Health Technology Assessments; 2002 [citado 8 Oct 2007]. Disponible en: <http://www.ccotha.ca>.

19. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Res Methodol.* 2003;3:25.
20. Gitlin JN, Narayan AK, Mitchell CA, Akmal AM, Eisner DJ, Peterson LM, et al. A comparative study of conventional mammography film interpretations with soft copy readings of the same examinations. *J Digital Imaging.* 2007;20:42-52.
21. Berns EA, Hendrick ER, Solari M, Barke L, Reddy D, Wolfman J, et al. Digital and screen-film mammography: comparison of image acquisition and interpretation times. *Am J Roentgenol.* 2006;187:38-41.
22. Dershaw D. Status of mammography after the digital mammography imaging screening trial: digital versus film. *Breast J.* 2006;12:99-102.
23. Skaane P, Kshirsagar A, Stapleton S, Young K, Castellino RA. Effect of computer-aided detection on independent double reading of paired screen-film and full-field digital screening mammograms. *Am J Roentgenol.* 2007;188:377-84.
24. Skaane P, Balleyguier C, Diekmann F, Diekmann S, Pigué JC, Young K, Niklason LT. Breast lesion detection and classification: comparison of screen-film mammography and full-field digital mammography with soft-copy reading. *Radiology.* 2005;237:37-44.
25. Van Ongeval CH, Bosmans H, Steen AV. Current status of digital mammography for screening and diagnosis of breast cancer. *Curr Opin Oncol.* 2006;18:547-54.
26. Hailey D. Digital mammography: an update [monografía en Internet]. Ottawa: Canadian Coordinating Office for Health Technology Assessments; 2006 [citada 12 Oct 2007]. Disponible en: <http://www.ccotha.ca>.
27. Bonardi R, Ambrogetta D, Ciattob S, Gentilec E, Lazzarid B, Mantellinia P, et al. Conventional versus digital mammography in the analysis of screen-detected lesions with low positive predictive value. *Eur J Radiol.* 2005;55:258-63.

28. Jong RA, Jaffe MJ. Digital mammography: 2005. *Can Assoc Radiol J.* 2005;56:319–23.
29. Ge J, Hadjiiski LM, Sahiner B, Wei J, Helvie MA, Zhou C et al. Computer-aided detection system for clustered microcalcifications: comparison of performance on full-field digital mammograms and digitized screen-film mammograms. *Phys Med Biol.* 2007;52:981–1000.
30. Lewin JM. Clinical trials in full-field digital mammography. *Semin Breast Dis.* 2006;9:87-91.
31. Yaffe MF, Bloomquist AK, Mawdsley GE, Pisano ED, Hendrick RE, Fajardo LL, et al. Quality control for digital mammography: Part II. Recommendations from the ACRIN DMIST trial. *Med Phys.* 2006;33:737-52.
32. Ciatto S, Brancato B, Baglioni R, Turci M. A methodology to evaluate differential costs of full field digital as compared to conventional screen film mammography in a clinical setting. *Eur J Radiol.* 2006;57:69–75.
33. Lewin JM, Hendrick RE, D’Orsi CJ, Isaacs PK, Moss LJ, Karellas A et al. Comparison of full-field digital mammography with screen-film mammography for cancer detection: results of 4,94 paired examinations. *Radiology.* 2001;218:873-80.
34. Lewin JM, D’Orsi CJ, Hendrick RE, Moss LJ, Isaacs PK, Karellas A et al. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. *Am J Roentgenol.* 2002;179:671-77.
35. Glueck DH, Lamb MM, Lewin JM, Pisano ED. Two-modality mammography may confer an advantage over either full-field digital mammography or screen-film mammography. *Acad Radiol.* 2007;14:670-76.
36. Skaane P, Young K, Skjennald A. Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading- Oslo I study. *Radiology.* 2003;229:877-84.
37. Skaane P, Skjennald A, Young K, Egge E, Jepsen I, Sager EM et al. Follow-up and final results of the Oslo I Study comparing screen-

film mammography and full-field digital mammography with soft-copy reading. *Acta Radiol.* 2005;46:679-89.

38. Skaane P, Skjennald A. Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program-- the Oslo II Study. *Radiology.* 2004;232:197-204.
39. Skaane P, Hofvind S, Skjennald A. 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. *Radiology.* 2007;244:708-17.
40. Vigeland E, Klaasen H, Klingen TA, Hofvind S, Skaane P. Full-field digital mammography compared to screen film mammography in the prevalent round of a population-based screening programme: the Vestfold County Study. *Eur Radiol.* 2008;18:183-91.
41. 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. *New Engl J Med.* 2005;353:1773-83.
42. Pisano ED, Gatsonis CA, Yaffe MJ, Hendrick RE, Tosteson AN, Fryback DG et al. American College of Radiology Imaging Network Digital Mammographic Imaging Screening Trial: objectives and methodology. *Radiology.* 2005;236:404-12.
43. Del Turco MR, Mantellini P, Ciatto S, Bonardi R, Martinelli F, Lazzari B et al. Full-field digital versus screen-film mammography: comparative accuracy in concurrent screening cohorts. *Am J Roentgenol.* 2007;189:860-6.
44. Tosteson AN, Stout NK, Fryback DG, Acharyya S, Herman BA, Ana LG et al. Cost-effectiveness of digital mammography breast cancer screening. *Ann Intern Med.* 2008;148:1-10.
45. Getty D, D'Orsi C, Pickett R, Newell M, Gundry K, Roberson S, et al. Improved accuracy of lesion detection in breast cancer screening with stereoscopic digital mammography. *RSNA Meeting 2007: Connecting Radiology.* Chicago: November 25-30 2007. Available from: http://rsna2007.rsna.org/rsna2007/v2007/conference/event_display.cfm?id=66601&em_id=5002357

Anexo 1: análisis económico

1.- Introducción

El cáncer de mama es el tipo de cáncer más frecuente entre las mujeres. Nuestro país sigue siendo uno de los que registra una menor incidencia, diagnosticándose cada año 16.000 nuevos casos. Aunque durante los últimos diez años la mortalidad por cáncer de mama se ha reducido a un ritmo anual del 1,4%, es importante constatar que esta enfermedad causa en España la muerte de aproximadamente 6000 mujeres al año (1). En Andalucía el cáncer de mama es el tipo de cáncer que presenta mayor mortalidad entre las mujeres (2). Los datos de supervivencia en nuestro país también son también algo superiores al de la media europea, según refleja el estudio EUROCARE (European Cancer Registrie-based study on Survival and Care of cancer patients), en la actualidad, la supervivencia del cáncer de mama a los cinco años del diagnóstico es de un 78%, frente al 70% que representaba en el periodo anterior (3).

Este cambio de tendencia en la mortalidad por cáncer de mama pone de manifiesto que los adelantos en el diagnóstico y el tratamiento desempeñan un papel fundamental para el control de esta enfermedad. La detección precoz mejora la supervivencia entre las mujeres que padecen este tipo de cáncer, y en este sentido, los programas de cribado poblacional mediante mamografía suponen una estrategia de diagnóstico útil y eficaz que contribuye a mejorar las expectativas para estas pacientes (4).

Como ya se ha mencionado con anterioridad en este informe, la mamografía es la prueba de diagnóstico utilizada para el cribado poblacional del cáncer de mama. Además de la tomografía axial computerizada (TAC), las dos tecnologías disponibles actualmente para detectar este tipo de tumor son la MC y la MD. Existen dos tipos de mamógrafos digitales, los mamógrafos digitales directos (MDD) y los mamógrafos digitales indirectos (CR), una vez producida la imagen no hay diferencia en el tipo de imagen ni en los costes de explotación.

Durante años, para el cribado del cáncer de mama, se procedía a realizar un examen de la mama mediante rayos x convencionales. Sin embargo, el campo del diagnóstico por imagen ha experimentado un importante avance tecnológico, y las placas o radiografías convencionales están siendo sustituidas por mamógrafos digitales. La mamografía digital es una mamografía, que en su procedimiento, utiliza exclusivamente soportes digitales (monitores, cámaras, lectores e impresoras) para la obtención y registro de las imágenes.

Una vez que se realiza la exposición, la imagen aparece en el monitor, el operador por tanto no tiene que realizar ningún trabajo adicional, por eso a estos sistemas de los denomina habitualmente “*shoot and see systems*”. Posteriormente, la imagen se puede enviar electrónicamente a un sistema de manejo de imágenes PACS (*Picture Archiving and Communications Systems*) o almacenarse localmente.

En el Sistema Público de Andalucía (SSPA), los mamógrafos utilizados para el cribado del cáncer de mama son digitales, bien porque al expirar la vida útil de los mamógrafos convencionales éstos se han sustituido por mamógrafos digitales directos, o bien porque los mamógrafos convencionales se han convertido en digitales mediante la incorporación de un lector CR.

A la luz de las conclusiones de la revisión sistemática llevada a cabo por AETSA, esto es, que tanto la mamografía convencional como la digital tienen el mismo rendimiento diagnóstico, desde el punto de vista del análisis económico, el único tipo de evaluación posible sería una minimización de costes.

Sin embargo, y al margen de las conclusiones de la revisión sistemática, la realidad es que la mamografía digital empieza a sustituir a la mamografía convencional. En el SSPA, según la información aportada por el servicio de Electromedicina del área hospitalaria de Valme, son varios los hospitales que cuentan con MDD en sus servicios. El hospital Universitario Virgen del Rocío cuenta con dos MDD, el primero adquirido en el año 1995 y el segundo en 1997. Los hospitales Juan Ramón Jiménez, Reina Sofía, San Cecilio y Virgen de la Victoria (CPE San José) también cuentan con esta tecnología, al igual que los centros concertados de San Juan de Dios en Sevilla y el hospital Costa del Sol en Marbella. Durante el año 2008 estaban pendientes de adjudicación los hospitales de Valme, Virgen de la Victoria, Universitario Virgen Macarena, Infanta Margarita, La Rosaleda y el Centro de Salud del Cachorro en Sevilla.

Es probable que algunos servicios estén esperando a que se desarrollen otros ensayos clínicos que proporcionen nuevos datos de evidencia a favor de la tecnología digital. Sin embargo, todo parece indicar que si la MD no está sustituyendo más rápidamente a la MC es debido a su elevado coste de adquisición y mantenimiento. En este sentido, y teniendo en cuenta lo expuesto anteriormente, revisaremos la literatura económica que compara en términos de eficiencia ambas tecnologías, y a partir de las conclusiones que se deriven de la evidencia, y teniendo en cuenta el actual grado de implementación de la mamografía digital en nuestro sistema sanitario, y al mismo tiempo, la disponibilidad de datos, estudiaremos la posibilidad de realizar un análisis de impacto presupuestario que proporcione una visión sobre la sustitución progresiva de la tecnología convencional que resulte coste efectiva desde el punto de vista sanitario.

2. Objetivos

Explorar los aspectos económicos relacionados con la mamografía digital frente a la analógica.

3. Método

Se realizó una revisión de la literatura y se consultaron las siguientes bases de datos: MEDLINE a través de Ovid, EMBASE, las bases mantenidas por el Center of Reviews and Dissemination (CRD), Cochrane Library, ECRI Institute, CEA Registry, y Euronheed, con la siguiente estrategia de búsqueda estrategias de búsqueda para EMBASE y MEDLINE:

EMBASE 1980 to 2008 Week 23. Fecha de ejecución: 12 junio de2008
Searches Results.

- 1 (mammograph\$ or mammogram\$).ti. 6620
- 2 (digital or comput\$).ti. 69584
- 3 screening.ti,ab. 167228
- 4 and/1-3 189
- 5 Digital mammography/ 163
- 6 exp Breast Tumor/ 149074
- 7 exp Mass Screening/ 59483
- 8 exp screening/ 181040
- 9 exp cancer screening/ 27145
- 10 screening.ti,ab. 167228
- 11 or/6-10 400903
- 12 5 and 11 104
- 13 4 or 12 270
- 14 Health Economics/ 10025
- 15 exp Economic Evaluation/ 96820
- 16 exp "Health Care Cost"/ 99001
- 17 exp PHARMACOECONOMICS/ 52354
- 18 or/1-4 238244
- 19 (econom\$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic\$).ti,ab. 218199
- 20 (expenditure\$ not energy).ti,ab. 9241

21 (value adj2 money).ti,ab. 420
 22 budget\$.ti,ab. 8403
 23 or/19-22 226483
 24 18 or 23 449501
 25 letter.pt.415922
 26 editorial.pt. 211329
 27 note.pt. 231661
 28 or/25-27 858912
 29 24 not 28 432350
 30 (metabolic adj cost).ti,ab. 368
 31 ((energy or oxygen) adj cost).ti,ab. 1654
 32 ((energy or oxygen) adj expenditure).ti,ab. 9510
 33 or/30-32 11050
 34 29 not 33 429849
 35 exp animal/ 18242
 36 exp animal-experiment/ 1243097
 37 nonhuman/ 3070934
 38 (rat or rats or mouse or mice or hamster or hamsters or animal or animals or dog or dogs or cat or cats or bovine or sheep).ti,ab,sh. 1987115
 39 or/35-38 3484921
 40 exp human/ 6202771
 41 exp human-experiment/ 247538
 42 or/40-41 6203636
 43 39 not (39 and 42) 2872978
 44 34 not 43 377908
 45 13 and 44 224

 Ovid MEDLINE(R)
 # Searches Results
 1 (mammograph\$ or mammogram\$).ti. 9646
 2 digital.ti. 16840
 3 screening.ti,ab. 210990
 4 and/1-3 119
 5 Mammography/ 19333
 6 Radiographic Image Enhancement/ 11828
 7 Radiographic Image Interpretation, Computer-Assisted/ 4292
 8 Image Processing, Computer-Assisted/ 59178
 9 6 or 7 or 8 71957
 10 5 and 9 1708
 11 Breast Neoplasms/pc, di, ra [Prevention & Control, Diagnosis, Radiography] 33754
 12 exp Mass Screening/ 89766

13 or/11-12 116744
 14 10 and 13 1018
 15 4 or 14 1061
 16 Economics/ 25670
 17 exp “Costs and Cost Analysis”/ 138127
 18 “Value of Life”/ 5038
 19 Economics, Medical/ 6990
 20 exp Economics, Hospital/ 15513
 21 Economics, Nursing/ 3838
 22 Economics, Pharmaceutical/ 1926
 23 or/16-22 182519
 24 (econom\$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic\$).ti,ab. 289631
 25 (expenditure\$ not energy).ti,ab. 12079
 26 (value adj1 money).ti,ab. 14
 27 budget\$.ti,ab. 12304
 28 or/24-27 302577
 29 23 or 28 398175
 30 letter.pt.651117
 31 editorial.pt. 234649
 32 historical-article.pt. 251506
 33 or/30-32 1126912
 34 29 not 33 376833
 35 Animals/ 4276967
 36 Humans/ 10452718
 37 35 not (35 and 36) 3221200
 38 34 not 37 358069
 39 15 and 38 60

La búsqueda inicial para identificar estudios económicos relevantes desde el punto de vista de la evaluación de tecnologías sanitarias se llevo a cabo en junio del 2008.

Dado que este informe es complementario a la revisión sistemática sobre “Rendimiento diagnóstico de la mamografía digital en el cribado del cáncer de mama”, en el que se estudia la eficacia y seguridad de dicha técnica comparada con el cribado convencional, la pregunta de investigación para la revisión sistemática de estudios económicos que ahora nos ocupa la hemos planteado en términos similares, es decir, comparar la eficiencia de la mamografía digital frente a la mamografía convencional para el cribado del cáncer de mama en mujeres entre 40 y 70 años.

4. Resultados y Discusión

Se localizaron un total de 267 artículos, y de éstos, tras la lectura por título y *abstract*, 25 fueron seleccionados para lectura a texto completo. Se excluyeron aquellos artículos que no fuesen estudios de cribado o que no aportasen información económica. Tras la lectura a texto completo se eliminaron 15 artículos (13 por no aportar información económica y 2 por tratar aspectos técnicos de los modelos económicos estudiados). Finalmente fueron 10 los artículos incluidos en la revisión. Para evaluar la calidad de los estudios que suponían una evaluación económica completa se utilizó la lista de comprobación de Drummond (5).

Encontramos una revisión sistemática publicada por la CCOHTA en el año 2002 y actualizada en el año 2006, comparando los aspectos técnicos, clínicos y económicos de la MD en relación a la MC, de buena calidad metodológica, que incluye los principales estudios económicos y de costes que comparan la MD con la MC publicados hasta la fecha. Dicha actualización recoge la evidencia de los cuatro estudios más relevantes que comparan las dos tecnologías mencionadas llevados a cabo desde la publicación del informe original. Se trata de los mismos estudios que se analizan en el informe de efectividad de AETSA: Oslo I (Skaane et al., 2003), Oslo II (Skaane & Skjennald, 2004), y dos de EEUU, Lewin et al. (2002) y DMIST (Pisano et al., 2005).

Durante el año 2008 se publican tres informes relevantes desde el punto de vista económico. En marzo el National Health Service (NHS) elabora un informe de evaluación de tecnologías sanitarias exponiendo la mejor estrategia clínica para implantar la MD en el programa de cribado escocés y que resulte al mismo tiempo coste efectiva. Se publica también un análisis coste efectividad a partir de los datos recogidos en el ensayo clínico DIMIST (Digital Mammographic Imaging Screening Trial). Y por último, durante el mes de julio, y por tanto, posteriormente a la fecha de nuestra búsqueda, de nuevo el NHS, a través de la Purchasing and Supply Agency, del Centre for Evidence-based Purchasing, publica un nuevo estudio de coste efectividad comparando las dos modalidades de la MD (directa e indirecta) con la MC, que incluye además una revisión de la literatura económica.

De los 10 artículos seleccionados para nuestra revisión, 6 están incluidos y revisados en el último informe del NHS, y otros 3 en el informe de evaluación de la CCOHTA. Sólo encontramos un estudio que resulta exclusivo de nuestra búsqueda, el publicado en el contexto español por Martínez Cantarero, J.F en la Revista Española de Economía de la Salud (ReES). Este hecho supone la principal limitación de nuestro estudio, y en este sentido, la contribución de este trabajo se limita a reunir en un mismo

documento la evidencia que existe hasta la fecha en relación a los aspectos económicos de la MD en relación a la MC.

En el año 2000, Hiatt et al (6) estudiaron, para el contexto de Estados Unidos, la diferencia en costes entre la MD y la MC, comparando los costes fijos y variables de ambas tecnologías. Concluyen que aunque los costes fijos de la MD son mayores que los de la MC, ya que el precio de los equipos y las estaciones de trabajo es mayor, a partir de cierto volumen de mamografías, concretamente a partir de 12.000 pruebas, se produce un cierto ahorro por mamografía realizada si el cribado se realiza mediante la tecnología digital. Teniendo en cuenta el volumen de actividad anual de su contexto, este punto de equilibrio tardaría en alcanzarse aproximadamente 3 años. Como ventajas de la MD destacan que evita el coste de las películas y los productos químicos asociados, así como que permite cribar a un mayor número de mujeres y reducir el tiempo de exploración. No se consideran los costes de personal.

Hasta el año 2004 destacar dos publicaciones. La primera de Ho “et al.” para el contexto canadiense (7), publicada en el 2002 y actualizada en el 2006 (8); y la segunda por Leegod “et al.” para el contexto británico (9). Ambos estudios consideraban la misma efectividad para ambas tecnologías.

Tras realizar una modelización de costes para el contexto canadiense, Ho et al., concluyen que los costes de la MD directa son superiores tanto a los de la MD indirecta como a los de la MC. A igual rendimiento diagnóstico recomendaban la tecnología más barata, en este caso la MC. La CCOHTA actualiza este informe en el año 2006, reconociendo la mayor precisión de la MD para el diagnóstico del cáncer de mama en mujeres con mamas radiológicamente densas, menores de 50 años, y peri o premenopáusicas. En cuanto a sus costes, los de adquisición siguen siendo entre dos y cuatro veces más caros que los de la MC, los de capital y de explotación un 38% más caros. Los costes de la MD indirecta resultaron equivalentes a los de la MC. Los autores sugieren que esta diferencia en costes probablemente irá disminuyendo a medida que los sistemas digitales y sus componentes se abaraten.

En el contexto británico, Leegod R and Gray AM, llevaron a cabo un análisis de costes comparando los costes directos de cribar a 10.000 mujeres mediante ambas tecnologías. Los autores concluyeron que el cribado mediante MD resultaba un 38,6% más caro, sin embargo, al aumentar el número de mamografías realizadas y reducir como consecuencia el número total de equipos necesarios en un 20%, esta diferencia disminuye hasta un 12.8%. Es decir, respaldan que aunque los costes de la MD son más elevados, en la medida en la que se prescindan de la impresión de imágenes digitales y se reduzca el número de equipos necesarios, el coste de ambas tecnologías tendería a igualarse. La principal limitación de este estudio es

que los autores no tenían información exacta sobre costes exactos de los mamógrafos digitales ni de los contratos de mantenimiento.

En el año 2005, Ciatto et al (10) estimaron el impacto que suponía la tecnología digital en relación al coste total de mamografías en Italia. Se valoraron dos escenarios, suponiendo un volumen total de 5.000 y 10.000 mamografías al año, considerando además tres posibilidades para la MD según el número de imágenes impresas entregadas a las mujeres. Basan los datos de efectividad en el estudio Oslo II. La principal limitación de este estudio de cara a nuestro análisis, es que no se trata de un programa de cribado, sino de mujeres asintomáticas que por prevención acuden voluntariamente a realizarse un estudio mamográfico. La MD resulta más cara para ambos escenarios, debido sobre todo a los altos costes de compra y mantenimiento, sin embargo, a medida que aumenta el volumen de pruebas realizadas la diferencia disminuye. Los autores resaltan que los costes de personal y estructura son los que tienen un mayor impacto sobre el coste total de la tecnología analógica, mientras que el coste de adquisición de los equipos hace que la tecnología digital se encarezca, y en este sentido, mientras que los primeros tienden a aumentar, el coste de la tecnología tiende a abarataarse a medida que se difunde su implementación. Destacar el estudio de cargas de trabajo que realizan. Según sus mediciones la MC supone una mayor carga para el personal encargado de archivar y descargar imágenes, y la MD para los radiólogos, indicando que en ciertos casos necesitan el doble de tiempo para la lectura de las imágenes digitales. Concluyen que aunque la MD resulta más cara que la MC, las ventajas que ésta ofrece, relacionadas sobretodo con la tecnología de la información, podrían compensar la diferencia en costes, sugiriendo que una reducción de los costes de compra y mantenimiento de la tecnología digital podría incluso invertir la situación a favor de la MD.

En el contexto español hay que destacar dos publicaciones, Roldan et al (11), en el año 1992, estimaron que los costes de adquisición de la MD eran 2.6 veces más caros que los de la MC, su coste de mantenimiento 2.2 veces superior. Sin embargo estimaron ahorros para la tecnología digital para los costes de personal (-2,7%), fungibles (-7,2%) y archivo películas (-24,3%). También estimaron una reducción importante del número de rellamadas mediante la MD (del 6,4% al 0,4%). En el año 2006 Martínez Cantarero et al. (12) publican en la ReES un análisis comparativo de los distintos tipos de costes y ahorros asociados a cada tecnología, que también incorpora una revisión de la literatura económica publicada hasta la fecha. Estos autores destacan que para aprovechar el potencial de la mamografía digital es necesario alcanzar un determinado volumen de actividad, y al mismo tiempo, incorporar cambios en los servicios de radiología que permitan adaptar la tecnología digital a los servicios sanitarios.

Durante el año 2008 el NHS del Reino Unido publica dos estudios en relación a este tema. Durante el mes de marzo, el NHS Quality Improvement Scotland elabora un informe de evaluación de tecnologías sanitarias (13), exponiendo la estrategia más efectiva para llevar a cabo la progresiva sustitución de la tecnología analógica por la digital, tanto desde el punto de vista clínico como económico, en el programa de cribado escocés. Este informe incide en que a pesar de no existir diferencias en el rendimiento diagnóstico entre la MC y la MD, lo cierto es que ésta aporta ciertos beneficios adicionales, relacionados principalmente con la reducción de las dosis de radiación recibidas y la tasa de rellamadas. El informe destaca que aun no existe evidencia suficiente para comparar la MD directa con la CR, aunque afirma que la indirecta es más barata, y que la directa, al reducir el tiempo necesario para la exploración, permite cribar a un mayor número de mujeres. El informe ofrece una serie de recomendaciones de cara a la implementación de la tecnología digital que permiten garantizar el uso óptimo de los recursos. Desde el punto de vista de la tecnología, supone una guía útil para los servicios de aprovisionamiento hospitalarios responsables de las decisiones de compras, ya que detalla de forma exhaustiva los requisitos, desde el punto de vista técnico, que deberían reunir los equipos digitales, tanto en las unidades fijas como en las móviles. Los autores recomiendan que antes de implantar la MD se adapten los servicios con las infraestructuras necesarias para gestionar la información, mencionando explícitamente los sistemas PACS y BSIS (Breast Screening Information System). Sugieren que aunque la MD sea más cara que la MC, tanto por su coste de adquisición como de mantenimiento, la MD también supone importantes ahorros, al evitar el consumo de fungibles por ejemplo, y que la evidencia disponible indica que el coste por mujer cribada puede llegar a ser similar para ambas tecnologías siempre que el programa de cribado garantice un volumen de productividad suficiente. Resaltar que el informe aborda desde la perspectiva de las pacientes la diferencia, entre ambas tecnologías, en términos de calidad de vida y resultados. Aunque estiman que desde el punto de vista de la mujer cribada el procedimiento resulte similar, éstas se beneficiarían de menores dosis de radiación y menores tasas de rellamada. Realizan un impacto presupuestario de lo que costaría una sustitución progresiva de la MC por la MD para el sistema. Teniendo en cuenta unos precios de adquisición de los equipos de aproximadamente 252.500 euros para la DDR, de 76.850 euros para la CR y de 54.850 euros para la MC; y además suponiendo un aumento del 30% en el número de mamografías realizadas al año por máquina, sustituir la MC por MD directa costaría 8,8 millones de euros, y en caso de sustituirla por indirecta 4.9 millones de euros. El coste total para un programa con DDR durante 10 años sería de 31,17 millones de euros, comparado con los 23,27 millones euros que costaría el

programa de CR. Estos valores aparecen en el informe original en GBP y se han transformado a euros según tipo oficial de cambio. (14)

El Centre for Evidence Based Purchasing publica en julio del 2008 un análisis de coste efectividad comparando la MD, tanto directa como indirecta, con la MC (15). El objetivo de este estudio es determinar en qué medida cada una de estas tecnologías resulta coste efectiva para el cribado y diagnóstico del cáncer de mama en relación a los distintos flujos de trabajo que implican el uso de las mismas, y como varía el coste efectividad en función de los distintos volúmenes de actividad. El análisis indica que la MD es coste efectiva cuando la comparamos con la MC, siendo en términos generales la MD directa más coste efectiva que la indirecta, aunque esto sólo para volúmenes de trabajo suficientemente altos y siempre que se consiga materializar el ahorro de personal como consecuencia de los menores tiempos de procesamiento requeridos por la tecnología digital

La CR resultó ser la tecnología más coste efectiva para los centros de cribado, pero por un margen muy pequeño en relación a la MD. En cuanto a la distribución de los costes, estos son sustancialmente diferentes para ambas tecnologías. Los costes de personal suponen aproximadamente un 65-75% de los costes totales para el caso de la MD, y del 55-60% para la MC. Los costes de capital suponen un 12-15% sobre el total de los costes en el caso de la MD directa, del 5-9% para la CR y de un 3-8% en el caso de la MC. El mayor ahorro viene determinado por el consumo de fungibles, mientras que para la MC suponían hasta un 38% sobre el total de los costes, en el caso de la MD éstos son como máximo del 15%.

También durante el año 2008 se publicó un análisis coste efectividad a partir de los datos de efectividad y económicos recogidos en el ensayo DMIST (16). La evaluación concluye que en relación a la MC, el cribado del cáncer de mama mediante MD no es coste efectivo. Si se tienen en cuenta los subgrupos de población en los que la mamografía digital presenta mejores valores de sensibilidad, es decir, en mujeres con mamas radiológicamente densas, menores de 50 años, y peri o premenopáusicas, el cribado poblacional por grupos de edad parece ser la estrategia más coste efectiva, mientras que el cribado orientado por la densidad mamaria resultaría una opción demasiado cara en relación al beneficio extra que proporciona.

5. Conclusiones

El uso de la tecnología digital aplicada al diagnóstico por imagen ha evolucionado muy rápidamente, y aunque actualmente la MC sigue siendo la prueba de diagnóstico indicada para el cribado y diagnóstico del cáncer de mama, lo cierto es que se está avanzando de forma progresiva hacia la completa implantación de la MD, tanto para actividades de cribado como de diagnóstico, y muchas de las unidades que utilizan MC están siendo reconvertidas para adaptarse a la nueva tecnología digital.

La mayoría de los estudios revisados coinciden en que la MD resulta más cara que la MC, al menos en sus fases iniciales de implementación, debido sobre todo a los precios de adquisición de los equipos y su mantenimiento. A pesar de que ambas tecnologías ofrecen un rendimiento diagnóstico similar, sin embargo, ciertas ventajas de la tecnología digital, no sólo clínicas, sino también organizacionales y económicas, justificarían el que a pesar de su precio, ésta empiece a sustituir, de forma progresiva, a la MC. Si se tienen en cuenta ciertos factores, como el aumento del volumen de productividad, el ahorro por la reducción del consumo de fungibles y el progresivo abaratamiento de los equipos digitales, la situación podría invertirse a favor de la MD.

Los análisis de coste efectividad indican que la MD sólo resulta coste-efectiva para volúmenes de trabajo suficientemente altos y siempre que se consigan materializar los ahorros que esta tecnología facilita. El cribado poblacional por grupos de edad parece ser la estrategia más coste-efectiva para los programas de cribado. En cualquier caso habría que considerar el contexto de cada organización, ya que las diferencias en la práctica clínica y el hacer de los profesionales afecta de forma definitiva a los resultados del análisis de coste-efectividad.

No existe evidencia que determine si la MD directa es más o menos recomendable que la indirecta, en este sentido, la decisión de optar por una u otra debe tomarse en base a la discusión promovida desde cada uno de los servicios o programas.

6. Recomendaciones

A la luz de las conclusiones de esta revisión, si se pretende una sustitución técnica de la mamografía convencional por la digital, es necesario tener en cuenta una serie de recomendaciones que garanticen un proceso de implementación de la tecnología digital que permita el uso óptimo de los

recursos, facilitando la integración organizativa y la toma de decisiones en los distintos niveles de intervención:

1. Necesidad de una guía que proporcione información útil para gestores sanitarios interesados en la adquisición de equipos de mamografía digital, y que al mismo tiempo, oriente el proceso de implementación de la tecnología digital y facilite el proceso de toma de decisiones, especificando los requisitos técnicos, organizacionales y profesionales que permitan obtener los máximos beneficios de la tecnología y recuperar el coste de la inversión.
2. El proceso de implantación de la mamografía digital debería ser integrado dentro de los sistemas de información PACS, y en este sentido se debería trabajar para lograr que las mamografías tomadas desde el programa de cribado se incorporasen al sistema PACS de referencia. Avanzar en el diagnóstico a partir de las estaciones de trabajo reduce la necesidad de impresión de las imágenes en películas, permite minimizar los costes de explotación de la mamografía digital y optimizar el uso de la tecnología.
3. Sería además aconsejable que los PACS, a su vez, se integrasen en sistemas de archivo, administración y tratamiento estadístico de datos del propio programa de cribado, facilitando de esta manera la evaluación de la calidad del programa y la utilización de la tecnología.
4. Es conveniente que el programa de cribado garantice un volumen de actividad que permita alcanzar niveles de rendimiento suficientemente altos. La evidencia sugiere que los costes de ambas tecnologías pueden llegar a igualarse si se aumenta el número de horas del programa y por tanto, el número de mujeres cribadas. Cada servicio o programa de cribado debería hacer un esfuerzo por estudiar los cambios organizativos que serían necesarios para rentabilizar la tecnología y conocer el número de radiografías que marcan su umbral de rentabilidad.
5. La implantación de la mamografía digital, que afecta de forma directa a las tareas que de forma habitual realizan los distintos profesionales implicados en el proceso de cribado, hace recomendable facilitar el proceso de formación de todos y cada uno de ellos en relación al uso de la nueva tecnología.
6. Son necesarios nuevos estudios económicos específicos, que permitan estimar costes reales de nuestro entorno y que faciliten el análisis coste-efectividad de la tecnología digital.

7. Referencias

- (1) López-Abente G, Pollán Santamaría M, aragonés Sanz N, Pérez Gómez B, Hernández Barrera V, Berta Suárez Rodríguez, et al. La situación del cáncer en España: Madrid: Ministerio de Sanidad y Consumo; 2005.
- (2) Consejería de Salud. Dirección General de Salud Pública y Participación. Vigilancia del cáncer en Andalucía. Evaluación de resultados en Salud Pública. 2006.
- (3) EUROCARE en: <http://www.eurocare.it/>
- (4) Castells X, Sala M, Ascunce N, Salas D, Zubizarreta R, Casamitjana M. Descripción del cribado del cáncer en España. Ministerio Sanidad y Consumo, editor. AATRM Num. 2006/01. 2007. Informes, Estudios e Investigación.
- (5) Drummond M et al. Methods for the economic evaluation of health care programmes. 2nd ed. Oxford. Oxford University Press. 1997
- (6) Hiatt MD, Carr JJ, Manning RL. The length of time necessary to break even after converting to digital mammography. Journal of Telemedicine & Telecare 2000; 6(4):222-224
- (7) Ho C, Hailey D, Warburton R, MacGregor JH, Pisano ED, Joyce j. Digital Mammography versus Film-Screen Mammography: Technical, Clinical and Economic Assessments. 2002. Canadian Coordinating Office for Health Technology Assessment. Ottawa. Canada
- (8) Hailey, D. Digital mammography: un update (Issues in emerging health technologies issue 91). Ottawa: Canadian Agency for Drugs and Technologies in Health; 2006.
- (9) Legood R, Gray A. A cost comparasion of full field digital mammography (FFDM) with film-screen mammography in breast cancer screening. NHSBSP Equipement Report 0403, editor.
- (10) Ciatto S, Brancato B, Baglioni R, Turci M. A methodology to evaluate differential costs of full field digital as compared to conventional screen film mammography in a clinical setting. European Journal of Radiology 2006; 57(1):69-75

- (11) Roldan J. Tecnología digital y convencional en radiología. Estudio comparativo de dosis y costes. *Radiología* 1996; 38 (3):191-197
- (12) Martintez J, Polaco C, Naranjo P, Del Llano J, Carreira C. Valoración económica de la mamografía digital vs convencional. *Revista Española de Economía de la Salud* 2006
- (13) NHS Quality Improvement Scotland. Determining the most clinically and cost-effective way of implementing digital mammography services for breast screening in NHSScotland. 2008. Health Technology Assessment Report.
- (14) European Central Bank. European Central Bank (EUROSYSTEM). ECB 2008. disponible en <http://www.ecb.int/stats/exchange>.
- (15) NHS Centre for Evidence-based Purchasing 08015. Cost-effectiveness of full field digital mammography (FFDM) and computed radiography (CR) versus film/screen imaging for mammography. 2008.
- (16) Tosteson AN, Stout NK, Fryback DG, Acharyya S, Herman BA, Hannah LG et al. Cost-effectiveness of digital mammography breast cancer screening. *Annals of Internal Medicine* 2008;148(1):1-10.

Anexo 2: QUADAS

| Items | Questions | Answer | | | |
|-------|--|--------|---|---|----|
| | | Y | N | U | NA |
| 1 | Was the spectrum of patients representative of the patients who will receive the test in practice? | | | | |
| 2 | Were selection criteria clearly described? | | | | |
| 3 | Is the reference standard likely to correctly classify the target condition? | | | | |
| 4 | Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two test? | | | | |
| 5 | Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis? | | | | |
| 6 | Did the patients receive the same reference standard regardless of the index test result? | | | | |
| 7 | Was the reference standard independent of the index test? | | | | |
| 8 | Was the execution of the index described in sufficient detail to permit replication of the test? | | | | |
| 9 | Was the execution of the reference standard described in sufficient detail to permit its replication? | | | | |
| 10 | Were the index test results interpreted without knowledge of the results of the reference standard? | | | | |
| 11 | Were the reference standard results interpreted without knowledge of the results of the index test? | | | | |
| 12 | Were the same clinical data available when test results were interpreted as would be available when the test is used in practice? | | | | |
| 13 | Were uninterpretable/intermediate test results reported? | | | | |
| 14 | Were withdrawals from the study explained? | | | | |

Y: yes ; N= no; U= unclear; NA= the item does not apply

Fuente: Whiting P et al. The development of QUADAS. BMC Medical Res Methodol. 2003;3:25.

ISBN 978-84-96990-40-1



9 788496 990401

Precio 10€