

Eficacia del diagnóstico asistido por ordenador aplicado al cribado de cáncer de mama

Revisión sistemática

Informe corto

Effectiveness of computer-aided diagnosis applied to breast cancer screening. Systematic review. *Short report. Executive summary*

INFORMES DE EVALUACIÓN DE TECNOLOGÍAS SANITARIAS
AETSA

INFORMES, ESTUDIOS E INVESTIGACIÓN



MINISTERIO
DE SANIDAD, SERVICIOS SOCIALES
E IGUALDAD



RED ESPAÑOLA DE AGENCIAS DE EVALUACIÓN
de Tecnologías y Prácticas en el Sistema Nacional de Salud



JUNTA DE ANDALUCÍA
CONSEJERÍA DE IGUALDAD, SALUD Y POLÍTICAS SOCIALES

Eficacia del diagnóstico asistido por ordenador aplicado al cribado de cáncer de mama

Revisión sistemática

Informe corto

Effectiveness of computer-aided diagnosis applied to breast cancer screening. Systematic review. *Short report. Executive summary.*

INFORMES DE EVALUACIÓN DE TECNOLOGÍAS SANITARIAS
AETSA

Fernández Prada, María

Eficacia del diagnóstico asistido por ordenador aplicado al cribado de cáncer de mama. María Fernández Prada, Aurora Llanos Méndez y Rebeca Isabel Gómez — Sevilla: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía, 2013.

79 p; 24 cm. (Colección: Informes, estudios e investigación. Ministerio de Sanidad, Servicios Sociales e Igualdad. Serie: Informes de Evaluación de Tecnologías Sanitarias)
ISBN: 978-84-15600-48-0

1. Cribado / métodos 2. Diagnóstico asistido por computador 3. Neoplasias de la mama / prevención y control I. Llanos Méndez, Aurora II. Isabel Gómez, Rebeca. III. Andalucía. Agencia de Evaluación de Tecnologías Sanitarias IV. España. Ministerio de Sanidad, Servicios Sociales e Igualdad V. España. Ministerio de Economía y Competitividad.

Autora: María Fernández Prada, Aurora Llanos Méndez y Rebeca Isabel Gómez

Este documento se ha realizado al amparo del convenio de colaboración suscrito por el Instituto de Salud Carlos III, organismo autónomo del Ministerio de Economía y Competitividad, y la Fundación Progreso y Salud de la Consejería de Igualdad, Salud y Políticas Sociales de la Junta de Andalucía, en el marco del desarrollo de actividades de la Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del SNS, financiadas por el Ministerio de Sanidad, Servicios Sociales e Igualdad

**Edita: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía
Consejería de Igualdad, Salud y Políticas Sociales
JUNTA DE ANDALUCÍA
Avda. de la Innovación, s/n. Edificio ARENA 1. Planta baja.
41020 Sevilla
España – Spain**

ISBN: 978-84-15600-48-0

NIPO: 680-14-040-6

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

Eficacia del diagnóstico asistido por ordenador aplicado al cribado de cáncer de mama

Revisión sistemática
Informe corto

Effectiveness of computer-aided diagnosis applied to breast cancer screening. Systematic review.
Executive summary. Short report

INFORMES DE EVALUACIÓN DE TECNOLOGÍAS SANITARIAS
AETSA



MINISTERIO
DE SANIDAD, SERVICIOS SOCIALES
E IGUALDAD



RED ESPAÑOLA DE AGENCIAS DE EVALUACIÓN
de Tecnologías y Prácticas de Salud



JUNTA DE ANDALUCÍA
CONSEJERÍA DE IGUALDAD, SALUD Y POLÍTICAS SOCIALES

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.

Contribución de los autores

María Fernández Prada. Especialista en Medicina Preventiva y Salud Pública. Planificación y diseño de la investigación, documentación, obtención de datos, análisis y presentación de resultados y elaboración del manuscrito.

Aurora Llanos-Méndez: Doctora en medicina y cirugía, especialista en Medicina Preventiva y Salud Pública. Agencia de Evaluación de Tecnologías Sanitarias (AETSA). Coordinadora del proyecto, planificación y diseño de la investigación, documentación, obtención de datos, análisis y presentación de resultados y elaboración del manuscrito.

Rebeca Isabel Gómez. Licenciada en Documentación y Biblioteconomía. Servicio de Documentación e Información de AETSA. Desarrollo del proyecto: búsqueda bibliográfica y documentación. Elaboración del informe: aprobación del informe final.

Este manuscrito ha sido leído y aprobado por todos los autores.

Índice

Resumen ejecutivo	9
Executive Summary	11
Antecedentes	13
Cáncer de mama	13
Programa de detección precoz de cáncer de mama.....	14
Descripción de la tecnología.....	15
Estado de desarrollo de la tecnología y difusión (aspectos legales)....	18
Tecnologías alternativas	18
Justificación y objetivos.....	21
Metodología	23
Tipo de estudio	23
Búsqueda.....	23
Criterios de selección de los artículos recuperados	24
Criterios de inclusión.....	24
Criterios de exclusión	25
Extracción de los datos.....	25
Resultados.....	27
Resultado de la búsqueda	27
Descripción y calidad de los artículos.....	27
Revisiones sistemáticas.....	27
Informe de evaluación de tecnologías	32
Artículos originales.....	35
Principales resultados.....	43
Validez diagnóstica	43
Aspectos económicos.....	47
Coste por unidad	47
Estudios de evaluación económica	47
Discusión.....	51
Consecuencias e implicaciones prácticas	52
Limitaciones	53
Conclusiones.....	55
Referencias.....	57
Anexos.....	63
Anexo 1. Estrategias de búsqueda	63
Anexo 2. Diagrama de flujo de la gestión documental	67

Índice de tablas y figuras

Tabla 1. Descripción de las revisiones sistemáticas	31
Tabla 2. Descripción del estudio 1 del Informe de Evaluación de Tecnologías.....	35
Tabla 3. Descripción de los artículos originales	41
Tabla 4. Resultados de los estudios incluidos	45
Tabla 5. Listado de precios, precio promedio cotizado, desviación estándar y rango de precios cotizados.....	47

Resumen ejecutivo

Antecedentes

El programa de cribado de cáncer de mama se puso en marcha en España en 1990 y la mamografía es la prueba de elección. La lectura e interpretación de estas imágenes difiere según las Comunidades Autónomas ya que en algunas se hace mediante lectura simple y en otras se aplica una doble lectura. El diagnóstico asistido por ordenador (DAO) aplicado al cribado de cáncer de mama como herramienta de apoyo al radiólogo está siendo explorado como futura metodología como sustitución de uno de los radiólogos.

Objetivos

Conocer la eficacia en términos de validez y precisión diagnóstica del DAO como herramienta de apoyo en el cribado de cáncer de mama en comparación con la lectura convencional.

Fuentes de información

Se consultaron las bases de datos referenciales MedLine, EMBASE y Web of Science. También se consultaron fuentes de información a texto completo: Red Internacional de Agencias de Evaluación de Tecnologías (INAHTA) a través de la base de datos del *Center for Reviews and Dissemination* (CRD), The Cochrane Library, y en la plataforma de Agencias y Unidades de Evaluación de Tecnologías Sanitarias españolas (AUnETS). También se contempló *The Emergency Care Research Institute* (ECRI).

Criterios de selección

Se seleccionaron revisiones sistemáticas y estudios primarios de pruebas diagnósticas que incluyeran mujeres pertenecientes a un programa de cribado en el que se hubiera aplicado el DAO como prueba de estudio. Sólo se consideraron aquellos trabajos en los que se recogieran resultados sobre validez diagnóstica y precisión. El límite temporal (desde noviembre de 2011 hasta abril de 2013) se estableció teniendo en cuenta la fecha de cierre de búsqueda de la revisión sistemática seleccionada como de mayor calidad y que fue tomada de referencia para su actualización (Azevedo *et al* 2012²⁹).

Extracción y recogida de los datos

La extracción de la información procedente de los trabajos seleccionados fue realizada por un solo revisor y la evaluación de la calidad por pares. El análisis crítico y el nivel de evidencia se establecieron en base a lo descrito

por el *Scottish Intercollegiate Guidelines Network* (SIGN). Se realizó una síntesis cualitativa de los resultados.

Resultados

Se recuperaron 963 referencias tras eliminar duplicados, de las que se seleccionaron 4 revisiones sistemáticas, 1 informe de tecnologías sanitarias y 4 estudios originales. Todos los autores mostraron la validez diagnóstica del DAO en términos de sensibilidad. No fue así con la especificidad. La precisión no fue contemplada en ningún trabajo. Los trabajos fueron divididos en función del tipo de intervención (DAO aislado o DAO más lectura simple). De las cuatro revisiones sistemáticas recuperadas, sólo dos fueron de alta calidad, mostrando sensibilidades del DAO más la lectura simple entre 86 % y 87 %, con especificidades entre 88 % y 96 %. Los estudios originales recuperados con el mismo tipo de intervención mostraron un amplio rango en la sensibilidad, con valores entre 79 % y 97 %. Noble *et al.* 2009³¹ obtuvieron un aumento en la tasa de rellamada del 4,1 % (IC 95 %: 2,7-6,3), con una proporción de falsos positivos de 80 de cada 100.000 mujeres cribadas al emplear el DAO como herramienta de apoyo a la lectura mamográfica simple.

En cuanto al DAO utilizado de forma aislada, la sensibilidad osciló entre el 48,7 y el 100, mientras la especificidad fue del 95 %.

Limitaciones

Estos resultados deben ser valorados con cautela puesto que la calidad de los trabajos fue entre moderada y baja, a excepción de las dos revisiones sistemáticas que registraron una alta calidad a pesar de que los originales que las compusieron fueron de calidad moderada y baja.

Conclusiones

Las revisiones sistemáticas de mejor calidad (nivel de evidencia 1++) obtuvieron resultados de validez diagnóstica para el DAO más lectura simple entre 86 % y 87 % para la sensibilidad y 93 % para la especificidad. DAO aislado mostró una sensibilidad entre 94 % y 100 % y especificidad entre 42 % y 95 % en la mayoría de los trabajos. La evaluación económica resultó ser incierta. Ningún trabajo contempló la eficacia de la prueba en términos de precisión.

Executive Summary

Background

The screening program for breast cancer was launched in Spain in 1990 and mammography is the test of choice. The reading and interpretation of these images differs by regions where it can be done by single or double reading. The computer-aided diagnosis (CAD) applied to breast cancer screening as a tool to support the radiologist is being explored as a future one radiologist replacement methodology.

Objectives

Knowing the efficiency in terms of validity and diagnostic accuracy of CAD as a support tool in screening for breast cancer compared to conventional reading.

Sources of information

Were consulted reference databases Medline, EMBASE and Web of Science. It was also consulted some full text information sources: International Network of Agencies for Health Technology Assessment (INAHTA), Centre for Reviews and Dissemination (CRD), The Cochrane Library, and the platform of Agencies and Units of Spanish Health Technology Assessment (AUnETS). It has been also contemplated The Emergency Care Research Institute (ECRI).

Selection criteria

Systematic reviews and primary studies of diagnostic tests that included women from a breast screening program with CAD were selected. We only considered those papers which had results on diagnostic validity and accuracy. The deadline (from November 2011 to April 2013) was established taking into account the higher quality systematic review (Azevedo *et al.* 2012²⁹).

Extraction and data collection

The extraction of information was performed by a single reviewer. Quality assessment was done by peers. The critical analysis and the level of evidence were established on the basis described by the Scottish Intercollegiate Guidelines Network (SIGN). It was performed a qualitative synthesis of the results.

Results

963 references were retrieved after eliminating all the duplicates. Finally, there were selected four systematic reviews, one health technology report and four original studies. All authors showed the diagnostic validity of CAD in terms of sensitivity. Not so with specificity. The accuracy was not referred in any paper. All papers were divided according to the type of intervention (the use of CAD isolated or the use of CAD with simple reading). Only two of four systematic reviews were of high quality, showing sensitivities CAD with simple reading between 86 % and 87 %, and specificities between 88 % and 96 %. The original studies retrieved with the same type of intervention showed a wide range in sensitivity, with values between 79 % and 97 %. Noble *et al.* 2009³¹ obtained an increase in recall rate of 4.1 % (95 % CI 2.7 to 6.3), with a false positive rate of 80 per 100,000 women screened by using the CAD as a tool to support mammographic reading simple.

Regarding the isolated CAD, sensitivity ranged was from 48,7 % to 100 %, while the specificity was 95 %.

Limitations

These results should be interpreted with caution because the quality of the work was moderate to low, with the exception of two systematic reviews that reported a high quality even though the original papers which composed them were of moderate and low quality.

Conclusions

Systematic reviews of best quality (evidence level 1++) obtained a CAD diagnostic accuracy between 86% and 87% for sensitivity and 93% for specificity. Isolated CAD showed a sensitivity of 94% and 100%, and specificity between 42% and 95%. Economic evaluation was uncertain. None studies looked at the effectiveness of the test in terms of accuracy.

Antecedentes

Cáncer de mama

Globalmente, el cáncer de mama es el cáncer más frecuente entre las mujeres, representando el 23% de todos los cánceres femeninos que son diagnosticados cada año. En 2008, la incidencia mundial fue de 1,29 millones de casos nuevos¹. En la Unión Europea, se ha estimado una incidencia media de 93,4 en el año 2006². En España, la tasa de incidencia de este cáncer ajustada por edad en 2008 fue de 61 casos por 100.000 mujeres³. Cada año se diagnostican 16.000 casos nuevos^{4,5} y existe una heterogénea distribución geográfica por provincias. Así pues, en el período entre 1991 y 1997 se observó cómo Gerona, Navarra, Tarragona y Canarias registraron las tasas más altas, mientras que para Granada, Cuenca y Asturias fueron las más bajas⁶.

Además de ser un importante problema en cuanto a magnitud de la enfermedad, el cáncer de mama tiene una gran trascendencia social ya que, actualmente, es una de las principales causas de mortalidad femenina. En los países desarrollados se encuentra como primera causa de muerte por cáncer entre las mujeres seguido del cáncer de pulmón y colorrectal, aunque sabemos que tiene una tendencia decreciente según lo publicado por Cabanes y Pollán en un estudio poblacional realizado entre 1980 y 2007⁷. En este sentido, en España en 2008 este cáncer produjo 12,9 muertes por cada 100.000 mujeres³.

Sin embargo, a día de hoy el cáncer de mama representa una de las enfermedades oncológicas con mayor supervivencia. Según los últimos datos disponibles de EURO CARE en el que se incluyen a las mujeres diagnosticadas en el período de 2000 a 2002, la supervivencia global a los 5 años en Europa fue del 79 %, mientras que en EEUU se observó un 90 %. Teniendo en cuenta el territorio europeo, los Países Nórdicos, Suiza, Italia, Holanda, España y Austria fueron los que registraron una mayor supervivencia, al contrario que los países del Este, Irlanda y Reino Unido. Concretamente en España, la cifra ascendió al 83 % y fue superior a la media europea⁸.

Por último, según datos proporcionados por la Organización Mundial de la Salud (OMS), se estima que para el año 2020 habrá, en el mundo, 4.482 cánceres de mama y 1.194 muertes más que en 2008³.

Programa de detección precoz de cáncer de mama

El programa de cribado de cáncer de mama fue puesto en marcha en España por primera vez en 1990 en la Comunidad Foral de Navarra, y a lo largo de dicha década se implantó en el resto de Comunidades Autónomas. La Ciudad Autónoma de Ceuta es la que tiene la trayectoria más corta ya que comenzó en el año 2001.

La población diana del programa es variable en el territorio español. Así pues, mientras que todas las Comunidades incluyen mujeres asintomáticas entre 50 y 64 años, Castilla-La Mancha, Castilla y León, Ceuta, Comunidad Valenciana, La Rioja y Navarra abarcan también mujeres entre 45 y 49 años. Posteriormente, en base a la Recomendación del Consejo Europeo de 2 de diciembre de 2003, muchos programas incorporaron a las mujeres entre 65 y 69 años⁹.

La lectura mamográfica simple o doble, acompañada de biopsia en caso de ser necesario, se considera la prueba de elección (*gold standard*) para el cribado de cáncer de mama¹⁰. Se realizan dos proyecciones en cada mama, tanto en el cribado inicial como en los sucesivos. La lectura de la mamografía se realiza por uno o dos radiólogos, en función de la Comunidad Autónoma. Además, siguiendo las directrices europeas, es recomendable que el radiólogo tenga experiencia previa, acreditada por la realización de un mínimo de 5.000 lecturas anuales. Sin embargo, esto no se exige en todas las Comunidades (este es el caso de Andalucía, Asturias, Aragón, Canarias, Castilla La Mancha, Cataluña, Ceuta y La Rioja)^{9,11,12}.

El último informe sobre los indicadores de proceso y resultados de los programas españoles de cribado de cáncer de mama en 2009, destacó que la participación media fue de 69,1 %, habiendo importantes diferencias entre Comunidades (48-91,6 %), con cifras de adherencia superiores al 90 %¹².

En el caso de Andalucía, el programa comenzó en 1995, alcanzando la cobertura total en el año 2006. Esta Comunidad Autónoma presta el servicio a casi el 16 % de las mujeres del territorio español (689.000 andaluzas de 4.313.603 mujeres en España) y dispone de 45 unidades de exploración (37 fijas y 8 móviles; 35 propias del programa y 10 propias de la Asociación Española Contra el Cáncer). Igualmente, es la que más profesionales sanitarios tiene a su disposición ya que 170 radiólogos lectores y 57 técnicos especialistas en radiología forman parte del programa⁹.

Descripción de la tecnología

El diagnóstico asistido por ordenador (DAO) es un procedimiento que utiliza un *software* basado en algoritmos creados para identificar regiones sospechosas de malignidad que señalará, posteriormente, con una marca (asterisco, triángulo, figura ovalada, etc). No se conocen detalles acerca de los parámetros utilizados en el desarrollo algorítmico. El sistema puede ser aplicado sobre cualquier imagen digitalizada (resonancia magnética nuclear, tomografía computerizada, ultrasonidos, radiografía convencional, etc.), aunque fundamentalmente ha sido desarrollado para imágenes de mama, pulmón y colon. Su aplicación sobre cerebro, hígado y columna vertebral se encuentra en desarrollo, al igual que su uso en oftalmología y dermatología.

La interpretación de la mamografía digital con el uso del DAO requiere varias etapas. En primer lugar, el profesional sanitario hace su lectura habitual e interpretación de la mamografía. En segundo lugar, activa el DAO para que éste procese la imagen e identifique las regiones sospechosas, si las hay. Por último, el profesional reevalúa la mamografía comparando su interpretación personal con el resultado del DAO¹³. De esta forma, el objetivo de este sistema no es sustituir al radiólogo en la lectura de la mamografía, sino ayudarlo en la identificación de lesiones potencialmente malignas. De esta manera, podría mejorar la detección de posibles alteraciones patológicas que hubieran sido pasadas por alto con el ojo humano.

Existen diferentes tipos de sistemas de DAO disponibles en el mercado y otros se están desarrollando, tanto por casas comerciales, como por grupos de investigación. Los más destacados son^{10,14}:

- **ImageChecker** (R2 technologies Inc., Sunnyvale, CA, USA): este fue el primer sistema desarrollado en 1998 por R2 *Technology* y centrado en la mamografía digital. En ese año fue también aprobado por la FDA. En 2006 Hologic, Inc. adquirió esta tecnología.
- **Second Look** (ICad Systems, Canada): aprobado por la FDA en 2006 y aplicado a la mamografía digital.
- **Mammex TR** (Scanis, Inc., Foster City, CA, USA): desarrollado de forma conjunta con *Center for Medical Image Analysis of TRW Healthcare Solutions* (Pleasanton, CA, USA; business unit of TRW, Inc., Cleveland, OH, USA).
- **Kodak Mammography Computer-Aided Detection System** (*Kodak Health Group now Carestream Health, Inc. Rochester, NY, USA*).
- **ERLM Fuji®**.

- **IssiCAD** (*Intelligent system, Inc.*)
- **Image Clear** (*Titan Systems, Inc.*)

El funcionamiento del DAO se basa en cinco fases¹⁵:

1. **Digitalización de la imagen:** es indispensable que la imagen médica, en este caso la mamografía, se encuentre en formato digital para ser valorada mediante este sistema.
2. **Preprocesamiento:** en esta fase se aumentan las diferencias entre el tejido normal y el sano. Así pues, las estrategias más utilizadas para preprocesar una imagen médica digitalizada son:
 - a. **La modificación global del histograma:** entendido como la representación gráfica de la distribución de los distintos tonos de una imagen: esto se consigue mediante distintas aproximaciones, como son la ecualización del histograma (transformación de una imagen a un histograma con distribución uniforme), la ecualización del histograma multipico (partición del histograma según los picos que contenga) y la expansión del histograma (contiene el rango mayor de valores posible).
 - b. **El procesamiento local:** aplicación de técnicas en ventanas locales de la imagen.
 - c. **El procesamiento multiescala:** se basa en el uso de *wavelets* (pequeñas ondas cuya energía está concentrada en tiempo) que modifican los coeficientes para realzar las anomalías y, posteriormente, devuelven la imagen a su dominio original.
3. **Segmentación:** detecta y localiza las regiones sospechosas de malignidad (regiones de interés o ROIs) y las aísla del tejido normal. Este proceso determinará la sensibilidad del sistema, esto es, su capacidad para detectar correctamente el tejido patológico. Puede realizarse mediante tres grupos de técnicas:
 - a. **Técnicas clásicas:** la más sencilla es la umbralización (segmentación de la imagen según los valores de intensidad de los píxeles) global aunque no tiene una gran precisión para la detección de ROIs y, por tanto, se utiliza como entrada de una técnica posterior, como puede ser la umbralización local o adaptativa. La *region growing* es una de las más populares para la segmentación de algunos tipos de tumor que presentan un nivel de brillo superior al del tejido circundante y comparte características con la *region clustering*. Por último, la que se ha estudiado más ampliamente es la basada en el emparejamiento de patrones o *template matching* y trata de segmentar las

anomalías buscando patrones que se asemejen a prototipos obtenidos previamente de una muestra de entrenamiento.

- b. **Substracción de imágenes:** esto se emplea cuando el paciente es sometido a seguimiento ya que mediante el alineamiento elástico de las imágenes médicas que se desean comparar, se identifican las ROIs.
 - c. **Técnicas multiescala:** se basa en la aplicación de diferentes escalas para determinar si existe variación en el tamaño del tumor en dos momentos distintos.
4. **Extracción de características y selección:** una vez segmentada la imagen puede ser necesario extraer las características de las distintas ROIs para determinar con mayor precisión el nivel de sospecha. Esta etapa es clave para la detección de los tumores ya que el rendimiento del DAO depende más de la extracción y selección de las características de los ROIs que del proceso de segmentación. Estas características pueden dividirse en:
- d. **Características de intensidad:** es la aproximación más simple y se basa en el uso de los niveles de gris de cada ROI para caracterizar la imagen.
 - e. **Características geométricas o morfológicas:** se basan en la forma del ROI y se calculan a partir de su área y borde.
 - f. **Características de texturas:** pueden obtenerse a partir de matrices de co-ocurrencia (método estadístico que estudia la relación espacial entre grupos de datos).
5. **Clasificación:** cuando las características han sido extraídas y seleccionadas deben ser clasificadas para determinar el grado de sospecha. Existen diferentes tipos de clasificadores, sin embargo lo habitual es combinarlos entre sí para aumentar el rendimiento global del proceso. Esta combinación puede ser en paralelo, en cascada o de forma jerárquica y el resultado se observa mediante marcas superpuestas sobre la imagen. Así pues, unos sistemas utilizan un asterisco o una marca ovalada en el caso de la clasificación de masas mientras que el triángulo o el rectángulo correspondería a microcalcificaciones¹⁴.

Estado de desarrollo de la tecnología y difusión (aspectos legales)

El DAO fue introducido en 1960 por Lusted, sin embargo, hasta 1998 no obtuvo en EEUU la aprobación de la *Food and Drug Administration* (FDA) para su uso en la mamografía digital¹⁶. Tres años después, el 10 % de los centros de mamografía de EEUU había aprobado el uso del DAO en ese ámbito y, actualmente, Medicare y otras compañías de seguros lo incluyen entre sus servicios. Más aún, la mayoría de las mamografías que se realizan en EEUU a día de hoy son interpretadas con la ayuda del DAO¹⁷. La mayoría de los sistemas han obtenido también la Conformidad Europea (CE)¹⁰.

España no tiene incorporado el sistema DAO de manera establecida dentro del cribado de cáncer de mama. En 2008, según datos proporcionados por el Gobierno de Castilla-La Mancha y la compañía de tecnologías INDRA, el Servicio de Salud de Castilla-La Mancha fue el primero en disponer de esta tecnología sanitaria para el estudio de pulmón, hígado y mama¹⁸. Se desconoce su uso en otras Comunidades Autónomas o en el ámbito privado.

Tecnologías alternativas

En España, el sistema de lectura de las mamografías dentro del programa de cribado es de dos tipos: simple o doble.

- Lectura simple de mamografía: entendemos por lectura simple aquella en la que la mamografía es interpretada por un solo radiólogo. En este sentido, 6 Comunidades Autónomas han optado por el sistema de lectura simple (Aragón, Cantabria, Ceuta, Navarra, País Vasco y Murcia).
- Lectura doble de mamografía: la imagen es valorada, independientemente, por dos radiólogos. Este sistema es el más extendido en el territorio español y el juicio diagnóstico puede establecerse con consenso o sin consenso:
 - Con consenso: cuando no existe concordancia entre las interpretaciones de ambos profesionales, estos se reúnen y deciden el diagnóstico conjuntamente. Esto es aplicado en Baleares, Canarias, Comunidad Valenciana, Extremadura, La Rioja, Madrid y Cataluña.
 - Sin consenso: en este caso no se realiza una interpretación conjunta sino que, automáticamente, se le asigna a la imagen el

resultado más desfavorable y que pudiera tener peores consecuencias para la salud de la paciente⁹. Andalucía y Galicia son las que utilizan esta alternativa.

Es necesario aclarar que en tres Comunidades Autónomas conviven los dos sistemas de lectura, de tal forma que se realiza una lectura simple en la mayoría de las mamografías, pero una proporción de ellas se hace con lectura doble. Este es el caso de Asturias, Castilla-La Mancha y Castilla y León⁹.

Justificación y objetivos

Este informe ha sido realizado por la Agencia de Evaluación de Tecnologías Sanitarias de Andalucía bajo petición de la Dirección General de Cartera Básica de Servicios del Sistema Nacional de Salud y Farmacia perteneciente al Ministerio de Sanidad, Servicios Sociales e Igualdad.

El cribado del cáncer de mama se inició en España en 1990. Tras la experiencia de la última década, existen pruebas razonables para poder recomendar el cribado de cáncer de mama en mujeres de 50-69 años.

La prueba de cribado y diagnóstico utilizado hasta ahora ha sido la mamografía analógica. No obstante, en los últimos veinte años se ha incorporado la mamografía digital como técnica de mejora. Ésta ofrece ventajas operativas como una mayor rapidez y menor dosis de radiación¹⁹, además de la oportunidad para la incorporación de nuevas aplicaciones informáticas tales como la aplicación del Diagnóstico Asistido por Ordenador (DAO). Este sistema no pretende sustituir al profesional sanitario en la interpretación de las imágenes médicas, sino que se ha planteado como herramienta de apoyo para la identificación de las estructuras²⁰ que quizás, debido a la escasa calidad de la imagen, la ocultación de objetos por el paso de 3D a 2D, el reducido tamaño de las lesiones o el cansancio y falta de concentración, el ojo humano no pueda reconocer¹⁵.

Actualmente, España no tiene incorporado el DAO dentro del programa de cribado de cáncer de mama, sin embargo, esta tecnología podría ser asumida en un futuro ya que a medio o largo plazo podría suponer una disminución de los costes sanitarios. Por tanto, y teniendo en cuenta lo anterior, queda justificada la revisión de la evidencia disponible sobre el DAO aplicado al cribado de cáncer de mama.

Con este trabajo se pretende dar respuesta a la siguiente pregunta de investigación:

¿Es el Diagnóstico Asistido por Ordenador (DAO) eficaz en términos de validez y precisión diagnóstica como herramienta de apoyo en el cribado de cáncer de mama, en comparación con la lectura convencional?

Los objetivos principales de este estudio son conocer la eficacia en términos de validez y precisión diagnóstica del DAO como herramienta de apoyo en el cribado de cáncer de mama en comparación con la lectura convencional.

Metodología

Tipo de estudio

Revisión sistemática de la literatura, siguiendo las recomendaciones recogidas por la declaración PRISMA²¹. Para la valoración del riesgo de sesgo en el establecimiento de la calidad de los artículos originales se siguieron los criterios recomendados por Colaboración Cochrane²². La síntesis de los resultados se realizó de forma cualitativa. El método seguido para establecer el nivel de evidencia fue el descrito por el *Scottish Intercollegiate Guidelines Network (SIGN)*²³ para las revisiones sistemáticas y el *National Institute for Health and Care Excellence (NICE)*²⁴ para los estudios de pruebas diagnósticas.

Búsqueda

Se siguió un procedimiento de selección escalonado de estudios de acuerdo con la metodología. Inicialmente, se trataron de localizar y recuperar revisiones sistemáticas de la literatura, con o sin meta análisis, e informes de evaluación de tecnologías sanitarias. Si no se localizaron estudios con esos diseños, se tomaron en cuenta los estudios primarios de pruebas diagnósticas, siempre y cuando hayan sido realizados con fecha posterior a la fecha de cierre de la última revisión sistemática seleccionada. Las siguientes bases de datos referenciales fueron consultadas [hasta abril de 2013]: Medline, Embase y *Web of Science*. También se consultaron fuentes de información a texto completo: Red Internacional de Agencias de Evaluación de Tecnologías (INAHTA) a través de la base de datos del *Center for Reviews and Dissemination (CRD)*ⁱ, *The Cochrane Library*, y en la plataforma de Agencias y Unidades de Evaluación de Tecnologías Sanitarias españolas (AUnETS)ⁱⁱ. Asimismo se consultó la *Food and Drug Administration (FDA)*ⁱⁱⁱ, la Conformidad Europea (CE)^{iv} y en el registro de ensayos clínicos norteamericano *ClinicalTrials.gov*^v.

ⁱ <http://www.crd.york.ac.uk/crdweb/SearchPage.asp>

ⁱⁱ <http://aunets.isciii.es/web/guest/home>

ⁱⁱⁱ <http://www.fda.gov>

^{iv} <http://www.marcado-ce.com>

^v <http://www.clinicaltrials.gov>

Se realizó una revisión manual en los sitios WEB de agencias no incluidas en INAHTA y de instituciones nacionales e internacionales como el Ministerio de Sanidad, Servicios Sociales e Igualdad^{vi}, la Organización Mundial de la Salud (OMS)^{vii}, los *Centers for Disease Control and Prevention* (CDC)^{viii}, *The Emergency Care Research Institute* (ECRI)^{ix}, *National Institute for Health and Clinical Excellence* (NICE)^x y la Asociación Española Contra el Cáncer^{xi}, así como una revisión secundaria a partir de las referencias bibliográficas de los artículos recuperados.

La estrategia de búsqueda se muestra en el Anexo 1.

Se realizó un análisis crítico por pares utilizando, para las revisiones sistemáticas las escalas AMSTAR²⁵ y SIGN 50²³, para los estudios primarios de pruebas diagnósticas las recomendaciones QUADAS²⁶ y QUADAS-2²⁷ y, para la valoración económica la *Critical Appraisal Skills Programme* (CASP) adaptadas por CASP España (CASPe)²⁸

Criterios de selección de los artículos recuperados

Criterios de inclusión

- **Población:** mujeres que se encuentran incluidas en el programa de cribado de cáncer de mama.
- **Intervención:** aplicación del DAO en la lectura de las mamografías digitalizadas.
- **Comparación:** sistemas convencionales de lectura de las mamografías (lectura simple/doble).
- **Resultados:** eficacia en términos de validez y precisión diagnóstica.
- **Diseño:** revisiones sistemáticas y estudios primarios de pruebas diagnósticas.

^{vi} <http://www.msssi.gob.es>

^{vii} <http://www.who.int/es/>

^{viii} <http://www.cdc.gov>

^{ix} <http://www.ecri.org/Pages/default.aspx>

^x <http://www.nice.org.uk>

^{xi} <http://www.aecc.es>

Criterios de exclusión

- Estudios no originales: revisiones narrativas, cartas al director, editoriales, notas.
- *Abstracts* de congresos.
- Estudios preclínicos realizados sobre animales, *ex vivo* o *in vitro*.
- Estudios primarios incluidos en las revisiones sistemáticas seleccionadas.

Extracción de los datos

Toda la información fue extraída de los artículos seleccionados por un solo revisor. Las variables recogidas incluyeron información general como el autor, el país, el año de publicación, los objetivos, las características de los pacientes, así como de la intervención y el seguimiento. Variables específicas incluyeron indicadores de eficacia, medidos en términos de validez (sensibilidad, especificidad, valor predictivo positivo y negativo, valor global de la prueba, y razón de verosimilitud positiva y negativa) y fiabilidad (concordancia inter e intraobservador).

Resultados

Resultado de la búsqueda

Como resultado de la búsqueda en las bases de datos referenciales Medline, Embase y Web of Science así como otras fuentes de datos, se recuperaron 999 referencias bibliográficas. Después de eliminar las 36 referencias duplicadas, se procedió a la selección de las citas por título y resumen, según los criterios de inclusión y exclusión expuestos anteriormente. Se recuperaron 20 artículos a texto completo, 10 de los cuales se descartaron en el posterior proceso de discriminación, quedando para su análisis 10 artículos originales de las cuales 4 fueron revisiones sistemáticas²⁹⁻³², un informe de evaluación de tecnologías³³ y 4 estudios originales³⁴⁻³⁷ (Anexo 2).

Descripción y calidad de los artículos

Revisiones sistemáticas

Se han seleccionado para el análisis un total de cuatro revisiones, de las cuales las dos más recientes se publicaron en 2012 y la más antigua en 2004. Tres de ellas tuvieron origen en países europeos (Reino Unido, Suecia)²⁹⁻³¹ mientras que el restante perteneció a autores de Australia³².

De las cuatro revisiones sistemáticas, tres²⁹⁻³¹ de ellas estudiaron el uso combinado del DAO más la lectura mamográfica simple comparado con la biopsia, mientras que la última³² analizó el DAO de forma aislada.

El número de estudios incluidos en las revisiones osciló entre cuatro y siete. Noble *et al.* 2009³¹ fueron los únicos que realizaron un metaanálisis complementario a la revisión sistemática. Este trabajo fue publicado en 2010 por el Instituto ECRI añadiendo, este último, datos sobre costes del *software*¹⁰. Por otro lado, la revisión de Azevedo *et al.* 2012²⁹ en la que se seleccionaron cuatro estudios, contuvo dos trabajos no incluidos en las revisiones anteriores.

La revisión de Eadie *et al.* 2012³⁰ incluyó otros tipos de tumores además del mamario, como fueron el de pulmón y piel. Así pues, estos autores analizaron un total de 48 estudios, de los cuales 8 fueron relacionados con la aplicación del DAO sobre la mamografía.

Descripción de la población

Los estudios que formaron parte del análisis en las revisiones sistemáticas seleccionadas y de los que se disponen datos tuvieron muestras superiores a 5.000 mujeres. El rango osciló entre 5.016 en el estudio de Ko *et al.* 2006 y 429.345 en el estudio de Fenton *et al.* 2007.

Toda la población se compuso de mujeres asintomáticas que se encontraban dentro de un programa de cribado poblacional real o simulado, éste último en el caso de algunos estudios incluidos en la revisión de Eadie *et al.* 2012³⁰. Además, Irwig *et al.* 2004³² incluyeron mujeres con predisposición genética al cáncer de mama por antecedentes familiares y aquellas en las que la mamografía pudiera ser una prueba menos eficaz por ser jóvenes o por tener tejido mamario radiológicamente denso.

Las edades de las participantes fueron variables ya que el diseño de los programas de cribado cambió según el país. En general, dos de las cuatro revisiones incluyeron mujeres mayores de 40 años y se desconocen las edades correspondientes a las otras dos revisiones.

Descripción de la intervención

Prueba a estudio

Tres de las revisiones sistemáticas estudiaron el DAO como herramienta de apoyo a la lectura simple²⁹⁻³¹, mientras que una lo utilizó de forma aislada³².

Las revisiones no ofrecieron información detallada sobre el tipo de *software* utilizado en cada estudio.

Prueba de referencia

La prueba de referencia fue la anatomía patológica en tres de las cuatro revisiones seleccionadas²⁹⁻³¹. No se aportaron datos referentes al procedimiento de realización ni a la clasificación histopatológica seguida. En la revisión sistemática de Eadie *et al.* 2012³⁰, se incluyeron, además, estudios en los que el diagnóstico radiológico inicial fue la prueba de referencia, sin especificar en qué consistió dicho diagnóstico ni cuáles habían sido esos trabajos.

La revisión de Irwig *et al.* 2004³² no ofreció información sobre la prueba de referencia utilizada.

Descripción de las medidas de resultado

Eficacia

- Sensibilidad y especificidad: todas las revisiones seleccionadas midieron la eficacia del DAO en términos de sensibilidad. Las tres más recientes analizaron también la especificidad²⁹⁻³¹.

Aunque sólo uno de los autores realizó un metaanálisis, Eadie *et al.* 2012³⁰ calcularon la media ponderada de estos parámetros.

- **Odds ratio diagnóstica y curva de ROC:** estas medidas de poder discriminativo y diagnóstico fueron calculadas por Eadie *et al.* 2012³⁰. Esta última fue construida siguiendo el método de Moses-Shapiro-Littenerg.

Los valores predictivos y los cocientes de probabilidad no formaron parte de los resultados y tampoco pudieron ser calculados a partir de los datos de los trabajos.

Otros resultados

- **Tasa de detección de cáncer:** definida como el cociente entre las mujeres con diagnóstico final de cáncer y el total de mujeres examinadas ($VP/total$)²⁹.
- **Incremento de la tasa de detección de cáncer:** definida como la diferencia entre la tasa de detección de cáncer medida mediante DAO como herramienta de apoyo a la lectura simple, y la tasa de detección de cáncer medida mediante la lectura simple³¹.
- **Tasa de rellamada:** definida como el cociente entre las mujeres con prueba a estudio positiva y el total de mujeres examinadas ($VP+FP/total$)^{29,32}.
- **Incremento de rellamada:** definida como la diferencia entre la tasa de rellamada cuando se aplica el DAO como herramienta de apoyo a la lectura simple, y la tasa de rellamada cuando sólo se emplea la lectura simple³¹
- **Proporción de mujeres que fueron rellamadas y resultaron sanas:** definida como el cociente entre las mujeres que fueron inicialmente positivas siendo sanas y todas las mujeres que fueron inicialmente positivas ($FP/VP+FP$)³¹.

Descripción de la calidad

En general, de las cuatro revisiones sistemáticas seleccionadas sólo dos fueron valoradas con alta calidad según SIGN y AMSTAR mientras que las restantes fueron de baja calidad (Anexos 3 y 4).

- **Estudios cuya intervención fue DAO + lectura simple**

Azevedo *et al.* 2012²⁹ y Noble *et al.* 2009³¹ fueron revisiones sistemáticas clasificadas como de alta calidad (nivel de evidencia 1++). La búsqueda bibliográfica incluyó las bases de datos PubMed, Embase y *The Cochrane Library*, entre otras. La revisión se realizó por pares. Se

contempló el análisis de la calidad de los estudios originales a través de las herramientas QUADAS y GRADE en el trabajo de Azevedo *et al.* 2012²⁹, y una escala basada en la modificación de los criterios QUADAS (*ECRI Institute Internal Validity Scale for Diagnostic Studies*), en el de Noble *et al.* 2009³¹. Los criterios de inclusión de los artículos estuvieron bien definidos y eso hizo que los trabajos seleccionados fueran homogéneos. La principal limitación fue que los autores no proporcionaron un listado con los artículos incluidos/excluidos para análisis, sin embargo, la probabilidad de aparición de sesgos que pudieran modificar los resultados del estudio fue baja.

En cuanto al metaanálisis de Noble *et al.* 2009³¹, estos autores aplicaron el modelo de efectos aleatorios por el método de *DerSimonian and Laird* para la combinación estadística de 3 estudios para el cálculo de la sensibilidad y especificidad, y de 5 estudios para el resto de resultados. Como parámetro de inconsistencia entre los resultados usaron I^2 , donde un $I^2 > 50\%$ implicaba gran heterogeneidad de los estudios. Sin embargo, los autores especificaron que no realizaron una exploración de la misma a través de la meta-regresión debido al escaso número de trabajos que contuvieron las medidas de resultados buscadas. Tampoco lo hicieron mediante el análisis de sensibilidad ni el de subgrupos, aún cuando los estudios analizados para el cálculo de la sensibilidad y especificidad mostraron heterogeneidad. No presentaron el gráfico *Forest Plot* ni tuvieron en cuenta el sesgo de publicación.

La revisión de Eadie *et al.* 2012³⁰ fue clasificada como de moderada calidad (nivel de evidencia 1+). A pesar de que el trabajo incluyó un número elevado de bases de datos y que mostró la estrategia de búsqueda así como el diagrama de flujo y los criterios de inclusión de los artículos, no se especificó si la revisión fue realizada por pares y no proporcionaron tablas con la descripción de los estudios analizados. Tampoco mostró sus resultados según el tipo de patología (cáncer de mama, pulmón o piel) ni un listado de los artículos incluidos/excluidos. Por el contrario, tuvo en cuenta la calidad científica de los artículos incluidos (escala QUADAS).

- **Estudios cuya intervención fue DAO aislado**

La revisión de Irwig *et al.* 2004³² fue de baja calidad (nivel de evidencia 1-). La probabilidad de sesgo de selección fue alta, ya que emplearon exclusivamente Medline para la búsqueda, no mostraron la estrategia ni el diagrama de flujo y no realizaron una revisión por pares. Además, no se realizó la valoración de la calidad científica de los estudios originales, lo que podría suponer una interpretación sesgada de los resultados. Tampoco mostraron si hubo conflicto de intereses.

Tabla 1. Descripción de las revisiones sistemáticas

	Azevedo et al 2012²⁹	Eadie et al 2012³⁰	Noble et al 2009³¹	Irwig et al 2004³²
Características generales				
País de realización	Suecia	Reino Unido	Reino Unido	Australia
N	4	8	7	4
Temporalidad	3 prospectivos 1 prospectivo + retrospectivo	ND	4 prospectivos 3 retrospectivos	ND
Población de los estudios incluidos para análisis				
N	Gilbert (N=28204) Gromet (N=118808 y 112413) Georgian-Smith (N=6381) Khoo (N=6111)	ND	Gromet (N=118808) Fenton (N=429345) Georgian-Smith (N=6381) Ko (N=5016) Morton (N=21349) Birdwell (N=8682) Freer (N=12860)	ND
Calidad de los estudios incluidos	QUADAS Y GRADE Moderada 1 estudio Baja 3 estudios	QUADAS Alta 8 estudios	ECRI Moderada 7	ND
Edad (años)	40-74	ND	> 50	ND
Otras características	Sin interés	Sin interés	Sin interés	-Predisposición genética -Mamas densas
Intervención				
Prueba de estudio	DAO + lectura simple	DAO + lectura simple	DAO + lectura simple	DAO aislado
Prueba referencia	Biopsia	-Biopsia -Diagnóstico radiológico inicial	Biopsia	ND
Medidas de resultado				
	-Sensibilidad* -Especificidad* -Tasa de detección de cáncer -Tasa de rellamada -Coste-efectividad	-Sensibilidad † -Especificidad † -ORD y AUC	-Sensibilidad ‡ -Especificidad ‡ -Incremento de la tasa de detección de cáncer con respecto lectura simple -Incremento de la tasa de rellamada y biopsia -Proporción de mujeres rellamadas y biopsiadas que resultaron sanas	-Sensibilidad* -Tasa de rellamada
AUC: área bajo la curva ROC; DAO: diagnóstico asistido por ordenador; ND: datos no disponibles; ORD: odds ratio diagnóstica.				
* síntesis cualitativa; † media ponderada; ‡ metanálisis.				

Informe de evaluación de tecnologías

En el año 2005, la *Health Technology Assessment* publicó un informe de evaluación dirigido por Taylor *et al.*³³ sobre el impacto del DAO en la sensibilidad y especificidad de la mamografía de cribado. Este informe se compuso de dos estudios primarios independientes realizados en el Reino Unido entre los años 2001-2002 y 2003-2004, respectivamente. El primero de ellos tuvo como objetivo principal conocer la diferencia en la sensibilidad y especificidad obtenida con o sin el uso del DAO. Además, analizaron la influencia de los diferentes tipos de lectores (médico radiólogo, técnicos en radiología y médico especialista en mama) sobre estos parámetros. Este estudio tuvo un diseño caso-control (fase II de D. Sackett).

El segundo trabajo no cumplió con nuestros criterios de selección al no mostrar datos de validez diagnóstica desglosada según el tipo de lector. Su objetivo fue determinar la mejora de la sensibilidad y especificidad para la detección imágenes sospechosas sin tener en cuenta si el lector fue especialista o no en radiología.

Descripción de la población

Las mujeres que formaron parte del estudio original seleccionado se encontraban dentro del programa de cribado y tuvieron edades comprendidas entre los 50 y los 64 años. El número de mujeres que participaron en el estudio fue de 180, de las cuales 60 fueron cánceres confirmados. Las mamografías fueron extraídas del *South-West London Breast Screening Service*. La técnica de diagnóstico habitual en este centro fue la doble lectura con consenso. Se excluyeron aquellas participantes que tuvieron inicialmente una imagen con apariencia sospechosa, pero resultaran negativas tras la biopsia.

Descripción de la intervención

Prueba a estudio

El *software* empleado fue el R2 ImageChecker M1000. La lectura por parte de profesionales tuvo en cuenta tres tipos de profesionales: médicos radiólogos (n=30), técnicos en radiología (n=15) y médicos especialistas en mama (n=5). A pesar de que todos participaron en un riguroso programa de entrenamiento y que trabajaban de manera habitual en los programas de cribado de cáncer de mama leyendo al menos 5.000 mamografías anuales, para la presente revisión sistemática sólo se tuvo en cuenta la intervención que considera el grupo de médicos radiólogos.

Por tanto, el estudio evaluó la sensibilidad del DAO como herramienta de apoyo a la lectura simple por parte de un médico radiólogo.

Prueba de referencia

La confirmación o no de la enfermedad tuvo lugar mediante la anatomía patológica (biopsia). No se aportaron datos sobre la técnica de realización, el tipo de biopsia (aguja fina/gruesa) o la clasificación histopatológica del tumor.

Descripción de las medidas de resultado

El estudio midió la validez diagnóstica (sensibilidad y especificidad) del DAO junto con el médico radiólogo, en comparación con prueba de referencia.

El trabajo no contempló la eficacia de la prueba en términos de precisión.

Descripción de la calidad

Este trabajo fue clasificado como de baja calidad según la escala QUADAS y QUADAS-2 (nivel de evidencia 2-). El estudio tuvo un alto riesgo de sesgo en los aspectos relacionados con la “selección de los pacientes” y en el “flujo y cronograma”. Para la “prueba de estudio” fue dudoso y para la “prueba de referencia” bajo. En cuanto a la aplicabilidad, fue considerada baja para todos los apartados.

Se desarrollan a continuación los aspectos que presentaron problemas metodológicos relacionados con el riesgo de sesgo del estudio siguiendo el esquema de la herramienta QUADAS-2:

- **Relacionados con la selección de los pacientes:**

El diseño de este estudio fue de tipo caso-control y, por tanto, el espectro de los pacientes no se consideró representativo de aquellos que se someterán a la prueba a estudio en la práctica clínica habitual. Además, el estudio no ofreció datos sobre los factores de riesgo de las mujeres ni comorbilidades. En este sentido, el riesgo de sesgo fue alto.

- **Relacionados con la prueba diagnóstica en estudio:**

Los autores no dejaron claro si la prueba a estudio se había realizado una vez conocidos los resultados de la biopsia.

No obstante, sí expusieron con claridad cuál fue el procedimiento a seguir para la aplicación del DAO. En primer lugar, todos los lectores fueron entrenados para el uso del *software*. Además, se insistió en que los investigadores realizaran su sistemática de lectura habitual de la mamografía. Por último, se pusieron en común los resultados de la lectura mamográfica y los del DAO.

Así pues, el riesgo de sesgo para este apartado fue dudoso.

- **Relacionados con la prueba de referencia:**

La prueba no fue descrita con claridad por parte de los autores. A pesar de mencionar que se trataba de la biopsia, no se detalló la técnica de realización, el tipo de biopsia (aguja fina/gruesa) ni la clasificación histopatológica del tumor. La biopsia se realizó previamente a la lectura del DAO.

La biopsia es el *gold standard* en el diagnóstico del cáncer de mama. Sin embargo, el seguimiento, utilizado como prueba de referencia en algunas mujeres, no es adecuado para clasificar correctamente a las enfermas.

El riesgo de sesgo en este aspecto fue bajo.

- **Flujo y cronograma:**

No todas las pacientes fueron sometidas a la prueba de referencia. Solamente lo hicieron aquellas que tuvieron un resultado mamográfico positivo.

Esto podría estar justificado por el conflicto ético que supondría realizar una biopsia a mujeres con resultados negativos en las pruebas radiológicas. No obstante, implicaría un sesgo de verificación diferencial, situación que lleva a clasificar erróneamente a algunos falsos negativos como verdaderos negativos, con lo que se sobrestimarían tanto la sensibilidad como la especificidad de la prueba.

De esta manera, el riesgo de sesgo fue alto.

Aplicabilidad

Los resultados de este estudio no podrían considerarse extrapolables a la población general de cribado de cáncer de mama ya que se trata de un estudio con un diseño caso-control. Aparte de eso, deberían haberse tenido en cuenta los antecedentes personales y familiares de las mujeres, sus factores de riesgo o comorbilidades.

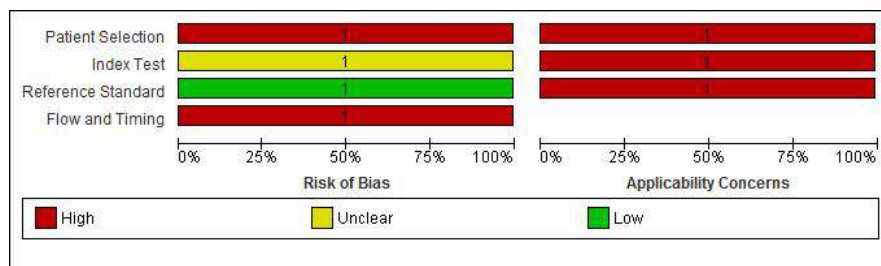


Tabla 2. Descripción del estudio 1 del Informe de Evaluación de Tecnologías

	Taylor <i>et al.</i> 2005 ³ Estudio 1
Características generales	
País de realización	Reino Unido
Tipo de estudio	Caso-control
Población del estudio	
N mamografías (mujeres)	180 (180)
Edad mujeres	50-64
Criterios inclusión/exclusión	Se excluyeron los casos con apariencia sospechosa y aquellos que fueron negativos en la biopsia
Intervención	
Prueba de estudio	DAO + lectura simple
Prueba de referencia	Biopsia
Software utilizado	R2 ImageChecker M1000
Medidas de resultado	
	- Sensibilidad - Especificidad

DAO: diagnóstico asistido por ordenador

Artículos originales

Se han seleccionado cuatro artículos originales posteriores a la fecha de cierre de búsqueda (noviembre de 2011) de la revisión sistemática más reciente y de mayor calidad recuperada (Azevedo *et al.* 2012²⁹).

En cuanto al lugar de desarrollo de la investigación, sólo un trabajo fue realizado en EEUU³⁵ mientras que el resto lo hicieron en Europa (Holanda, Noruega y España)^{34,36,37}.

Tres de ellos tuvieron un diseño retrospectivo de tipo caso-control (fase II de D. Sackett)³⁴⁻³⁶, mientras que uno evaluó una cohorte prospectiva (fase III de D. Sackett)³⁷. El objetivo principal de los trabajos era conocer la eficacia del DAO aplicado al programa de cribado de cáncer de mama, uno de ellos centrado exclusivamente en la identificación de masas³⁴, y el resto en todo tipo de lesiones³⁵⁻³⁷.

Descripción de la población

El número de participantes incluidos en el estudio fue variable y, en algunos casos, no coincidió con el número de mamografías evaluadas debido a que se realizó más de una proyección (craneo-caudal y mediolateral oblicua) por mujer. Mientras que Hupse *et al.* 2013³⁴ evaluaron 200 mamografías, el resto de los autores superaron las 6.000.

Incluso Skaane *et al.* 2012³⁶ y Sánchez *et al.* 2011³⁷ emplearon más de 20.000 porque analizaron a toda la población cribada.

Todas las mamografías pertenecieron a mujeres que se encontraban dentro del programa poblacional de cribado de cáncer de mama y, por tanto, no manifestaban sintomatología en el momento de la evaluación. Las edades fueron variables por países. Zheng *et al.* 2012³⁵ no detallaron cuál había sido la edad de las participantes. Con los datos disponibles, podemos decir que el mayor rango de edad fue el de Holanda (mujeres entre 50 y 75 años) del trabajo de Hupse *et al.* 2013³⁴. Ningún trabajo incluyó otras características de las participantes tales como antecedentes familiares/personales u otras comorbilidades.

Los criterios de exclusión de los artículos fueron mencionados, exclusivamente, por Sánchez *et al.* 2011³⁷ que especificaron que aquellas mamografías en las que no fuera aplicable el uso del DAO o aquellas mujeres en las que el cáncer sólo hubiera sido visible por ecografía, se excluirían de la población de estudio.

Hupse *et al.* 2013³⁴ y Zheng *et al.* 2012³⁵ seleccionaron una muestra dentro de la población general del programa de cribado. Los autores no mencionaron si se realizó un cálculo del tamaño muestral o el tipo de muestreo. En el primer caso, se tuvieron en cuenta 200 mujeres de las cuales 80 habían sido diagnosticadas de cáncer de mama, mientras que 120 se encontraban libres de tumor. Se desconocen los criterios de inclusión para cada grupo. Por otro lado, Zheng *et al.* 2012³⁵ dividieron su muestra de estudio en los siguientes cinco grupos: (1) 477 con cáncer confirmado, (2) 48 con cáncer intermedio, entendiéndose que el diagnóstico se realizó en la segunda mamografía, (3) 19 mujeres que fueron rellamadas por presentar imágenes de alto riesgo de malignidad, (4) 222 mujeres que fueron rellamadas por presentar imágenes de bajo riesgo de malignidad y (5) 354 mujeres sin cáncer de mama. Igual que Hupse *et al.* 2013³⁴, los autores no explicaron si hubo algún criterio para la selección de dichas pacientes y su inclusión en cada grupo.

Descripción de la intervención

Prueba a estudio

Zheng *et al.* 2012³⁵ y Hupse *et al.* 2013³⁴ realizaron más de una proyección mamaria. El uso del DAO fue de forma aislada en tres de los cuatro trabajos, para posteriormente compararlo con la sistemática habitual (lectura simple/doble). Sánchez *et al.* 2011³⁷ lo usaron de apoyo al radiólogo.

Todos los profesionales fueron sometidos a un programa de entrenamiento antes del uso de DAO. En general los estudios no ofrecieron detalles sobre el contenido del periodo de formación.

El número de radiólogos incluidos en los estudios osciló entre 6 y 12 con experiencia entre 2 y 24 años en la lectura de mamografías. Un trabajo incluyó, además, médicos especialistas en formación³⁴. En ningún estudio hubo técnicos en radiología ni médicos especialistas en mama como lectores.

Todos los autores, a excepción del Skaane *et al.* 2012³⁶ detallaron el *software* empleado para la realización del estudio, que fueron: Hologic Selenia (Hologic Inc) FFDM system, ImageChecker V8.0 and V8.3 (Hologic/R2, Santa Clara, CA, USA) system y CAD system Second Look version 1.1 (CADx System Inc., Quebec, Canada). No hubo coincidencias en este aspecto.

Prueba de referencia

Zheng *et al.* 2012³⁵ fueron los únicos autores que no definieron la prueba de referencia. Para todos los demás, ésta fue la histopatología a través de la biopsia.

Ningún autor ofreció datos sobre la técnica de realización, procesamiento ni clasificación histológica del tejido.

Descripción de las medidas de resultado

Eficacia

- Sensibilidad y especificidad: estudiada en todos los trabajos, excepto en uno en el que no se calculó la especificidad³⁶. En dos trabajos^{36,37} se pudo desglosar la sensibilidad según el tipo de lesión (microcalcificaciones, distorsiones de la arquitectura, masas). Además, el trabajo de Zheng *et al.* 2012³⁵ ofreció resultados de sensibilidad según el momento de diagnóstico de cáncer de mama (actual o intervalo).
- Valores predictivos positivos y negativos: sólo en el trabajo de Sánchez *et al.* 2011³⁷.
- Área bajo la curva: Hupse *et al.* 2013³⁴ ofrecieron los resultados desglosados según el tipo de lector (radiólogo especialista o especialista en formación) en forma de curvas de ROC.

Ningún trabajo no contempló la eficacia de la prueba en términos de precisión.

Descripción de la calidad

Según las escalas QUADAS y QUADAS-2, de los 4 trabajos incluidos para análisis uno de ellos tuvo una moderada calidad (nivel de evidencia 2+)³⁷ mientras que el resto fueron de baja calidad (nivel de evidencia 2-)³⁴⁻³⁶.

Se exponen a continuación los epígrafes que presentaron problemas metodológicos relacionados con la validez interna del estudio siguiendo el esquema de la herramienta QUADAS-2:

Relacionados con la selección de los pacientes:

- Riesgo de sesgo: El espectro de los pacientes pudo considerarse representativo de las mujeres cuyas mamografías serían evaluadas mediante el DAO en otras circunstancias solamente para el trabajo de Sánchez *et al.* 2011³⁷. Este trabajo evaluó una cohorte prospectiva incluyendo más de 20.000 mujeres (fase III de D. Sackett), mientras que el resto fueron retrospectivos de tipo caso-control (fase II de D. Sackett). No obstante, Skaane *et al.* 2012³⁶, aún siendo un estudio caso-control incluyó un amplio tamaño muestral, con 23.000 mujeres evaluadas. En este sentido, el estudio de Sánchez *et al.* 2011³⁷ se consideró como bajo riesgo de sesgo, el de Skaane *et al.* 2012³⁶ tuvo un dudoso riesgo de sesgo y, los dos últimos registraron un alto riesgo de sesgo.
- Aplicabilidad: los trabajos de Hupse *et al.* 2013³⁴, y Zheng *et al.* 2012³⁵ no son extrapolables a la población general de cribado debido su diseño inicial. Skaane *et al.* 2012³⁶ realizaron, igualmente, un estudio caso-control, sin embargo tuvieron en cuenta toda la población de cribado y no una muestra. Por último, los resultados de Sánchez *et al.* 2011³⁷, sí que podrían considerarse extrapolables a la población general de cribado de cáncer de mama ya que incluyeron toda la cohorte de mujeres. No obstante, ninguno de los trabajos ofrecieron datos sobre los factores de riesgo de las mujeres, antecedentes personales o familiares ni comorbilidades.

Relacionados con la prueba diagnóstica de estudio:

- Riesgo de sesgo: el DAO fue descrito correctamente por todos los autores. No obstante, ninguno dejó claro si la prueba de estudio se había realizado una vez conocidos los resultados de la prueba de referencia. Es más, el trabajo de Hupse *et al.* 2013³⁴ pudo haber ofrecido resultados sesgados relacionados con el observador ya que los lectores fueron informados previamente de

que las imágenes que evaluarían no contendrían lesiones de tipo microcalcificaciones y también la proporción de mamografías con imágenes sospechosas de malignidad. Tampoco se especificó previamente si emplearon un punto de corte o umbral para la clasificación de las lesiones en patológicas o no patológicas. Por tanto, el trabajo de Hupse *et al.* 2013³⁴ registró un alto riesgo de sesgo, mientras que el resto tuvieron un riesgo de sesgo dudoso.

- Aplicabilidad: todos los autores describieron la prueba a estudio, los aspectos técnicos y el procedimiento en la interpretación. Sin embargo, la variación de la interpretación de las imágenes en el trabajo de Hupse *et al.* 2013 pudo mermar su aplicabilidad.

Relacionados con la prueba de referencia:

- Riesgo de sesgo: la biopsia es el *gold standard* en el diagnóstico del cáncer de mama. Sin embargo, el seguimiento, utilizado como prueba de referencia en algunas mujeres, no fue adecuado para clasificar correctamente a las enfermas. Hupse *et al.* 2013³⁴ utilizaron la biopsia como prueba de referencia, mientras que Skaane *et al.* 2012³⁶ y Sánchez *et al.* 2011³⁷ utilizaron la mamografía, y Zheng *et al.* 2012³⁵ se desconoce.

Por otro lado, tanto los resultados de la biopsia como los de la mamografía fueron interpretados independientemente de los resultados del DAO en todos los trabajos.

Así pues, para Zheng *et al.* 2012³⁵, Skaane *et al.* 2012³⁶ y Sánchez *et al.* 2011³⁷ el riesgo de sesgo fue alto, y para Hupse *et al.* 2013³⁴ fue bajo.

- Aplicabilidad: la biopsia no fue descrita con claridad. No se describió la técnica de realización, el tipo de biopsia (aguja fina/gruesa) ni la clasificación histopatológica del tumor. Tampoco lo fue la mamografía donde no se detalló cuál fue la clasificación radiológica llevada a cabo para el diagnóstico de la enfermedad. Por tanto, la aplicabilidad para todos los trabajos fue dudosa en este aspecto.

Relacionados con el flujo y el cronograma:

- Riesgo de sesgo: no todas las pacientes fueron sometidas a la prueba de referencia. Solamente lo hicieron aquellas que tuvieron un resultado mamográfico positivo. Esto podría estar justificado por el conflicto ético que supondría realizar una biopsia a mujeres con resultados negativos en las pruebas radiológicas. No obstante, implicaría un sesgo de verificación

diferencial, situación que lleva a clasificar erróneamente a algunos falsos negativos como verdaderos negativos, con lo que se sobrestimarían tanto la sensibilidad como la especificidad de la prueba. Por tanto, el riesgo de sesgo fue considerado alto para todos los trabajos en este apartado.

	<u>Risk of Bias</u>				<u>Applicability Concerns</u>		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Hupse et. al 2013	●	●	+	●	●	●	?
Sánchez Gómez et. al 2011	+	?	●	●	+	+	?
Skaane et. al 2012	?	?	●	●	?	+	?
Zheng et. al 2012	●	?	●	●	●	+	?

● High	● Unclear	● Low
--------	-----------	-------

Tabla 3. Descripción de los artículos originales			
	Hupse et al 2013 ³⁴	Zheng et al 2012 ³⁵	Skaane et al 2012 ³⁶
Características generales			
País de realización	Holanda	EEUU	Noruega
Temporalidad	retrospectivo	retrospectivo	prospectivo
Población de los estudios			
N mamografías (mujeres)	200 (se desconoce)	6.478 (1.120)	23.923 (23.923)
Edad mujeres	50-75	ND	50-69
Criterios inclusión/exclusión	<ul style="list-style-type: none"> - No está claro - Los autores mencionan que incluyen 200 casos, de los cuales 80 son cánceres y 120 no. -No especifican los criterios de inclusión/exclusión 	<ul style="list-style-type: none"> - No está claro - Se incluyeron las siguientes mujeres: <ul style="list-style-type: none"> - 477 con cáncer confirmado (diagnosticado en la segunda mamografía) - 19 rellamadas con alto riesgo - 222 rellamadas con bajo riesgo - 354 negativas 	<ul style="list-style-type: none"> - Mujeres del programa de cribado poblacional entre noviembre del año 2000 y diciembre de 2001 - No menciona criterios de exclusión
			21.855 (21.855)
			45-65
			- Se excluyeron las que no era aplicable el uso del DAO y los cánceres sólo visibles por ecografía

Tabla 3. Descripción de los artículos originales (Continuación)				
	Hupse et al 2013 ³⁴	Zheng et al 2012 ³⁵	Skaane et al 2012 ³⁶	Sánchez et al 2012 ³⁷
Intervención				
Proyecciones	2	2	1	1
Profesionales lectores	9 radiólogo 3 MIR	ND	8 radiólogo	6 radiólogo
Experiencia	1 y 24 años	ND	4 y 10 años	2 y 10 años
Entrenamiento	ND	ND	ND	De 3 a 9 meses
Prueba de estudio	DAO aislado vs. lectura simple	DAO aislado vs. lectura habitual	DAO aislado vs. lectura doble con consenso	DAO + lectura simple vs. lectura simple
Software utilizado	ND	Hologic Selenia (Hologic Inc) FFDM system	ImageChecker V8.0 and V8.3 (Hologic/R2, Santa Clara, CA, USA) system	CAD system Second Look version 1.1 (CADX System Inc., Quebec, Canada)
Prueba referencia	Biopsia	ND	Biopsia	Biopsia
Medidas de resultado	- Sensibilidad - Especificidad	- Sensibilidad general, cáncer actual y cáncer intervalo - Especificidad general, cáncer actual y cáncer intervalo	- Sensibilidad general y según el tipo de lesión	- Sensibilidad general y según el tipo de lesión - Especificidad general - Valor predictivo positivo - Valor predictivo negativo
DAO: diagnóstico asistido por ordenador; MIR: médico interno residente; ND: datos no disponibles; vs: versus.				

Principales resultados

Validez diagnóstica

1. Revisiones sistemáticas

- **Estudios cuya intervención fue DAO + lectura simple**
 - **Sensibilidad:** 86 % (IC95%: 84,2-87,6) en el metaanálisis (estudio con nivel de evidencia 1++). Entre 87,2 y 90,4 % en la revisión de Azevedo *et al.* 2012²⁹, siendo el valor más bajos el obtenido en el trabajo original de mayor calidad. En la revisión de Eadie *et al.* 2012³⁰ la sensibilidad fue 90,79 ($\pm 1,01$)
 - **Especificidad:** 88,2 % (IC95 %: 88,1-88,3) en el metaanálisis. Entre 91,5 % y 96,9 % en la revisión de Azevedo *et al.* 2012²⁹ siendo el porcentaje más bajo el obtenido en el grupo de estudios clasificado como de peor calidad. La media ponderada fue 92,70 ($\pm 7,48$)³⁰.
 - **Curva ROC:** Eadie *et al.* 2012³⁰ incluyeron la curva ROC donde no hubo diferencias estadísticamente significativas entre la lectura simple (AUC: 0,86 y DE: 0,07) y la lectura simple más DAO (AUC: 0,95 y DE: 0,03).
- **Estudios cuya intervención fue DAO aislado**
 - **Sensibilidad:** entre 95 y 100 %

2. Informe de evaluación de tecnologías y estudios primarios

- **Estudios cuya intervención fue DAO + lectura simple**^{33,37}
 - **Sensibilidad:** 97,1 % para los estudios primarios y 79 % (IC95 %: 76%-81 %) en el informe de evaluación de tecnologías.
 - **Sensibilidad según el tipo de lesión**³⁷:
 - Nódulos: 77,4 %
 - Microcalcificaciones: 91,3 %
 - Distorsiones de arquitectura: 70 %
 - Lesiones mixtas: 100 %
 - **Especificidad:** 92,8 %
 - **Valores predictivos:** positivo 6,1 % y negativo 99,5 %.

- **Estudios cuya intervención fue DAO aislado³⁴⁻³⁶:**
 - **Sensibilidad:** entre 48,7 %, en el estudio en el que los lectores tuvieron información previa sobre el contenido de las mamografías a evaluar (exclusión de microcalcificaciones y conocimiento de proporción de imágenes patológicas)³⁴, y 94 %.
 - **Sensibilidad según el tipo de lesión³⁶:**
 - Microcalcificaciones: 100 %
 - Masas: 90,4 %
 - **Sensibilidad según el momento del diagnóstico³⁵:**
 - Cánceres actuales: 75,9 %
 - Cánceres de intervalo: 72,9 %
 - **Especificidad:** 41,9%-95 %, siendo el trabajo de Hupse *et al.* 2013³⁴ el que obtuvo los valores más elevados en este parámetro.

Otros resultados

- Tasa de detección de cáncer: entre 2/100.000 y 7,02/100.000²⁹
- Incremento de la tasa de detección de cáncer: 50/100.000 (IC 95 % 30-80)³¹
- Tasa de rellamada: 3,9 % y 10,6 %²⁹.
- Incremento de rellamada: 4,1 % (2,7-6,3 %)³¹.
- Proporción de mujeres que fueron rellamadas y resultaron sanas: 80/100.000³¹.

Tabla 4. Resultados de los estudios incluidos

Autor y año	Nivel evidencia	Intervención	Sensibilidad	Especificidad
Revisiones sistemáticas				
<i>Azevedo et al. 2012</i> ²⁹	1++	DAO + lectura simple	Alta calidad: 87,2* Baja calidad: 90,4-91,5*	96,9* ND
<i>Eadie et al. 2012</i> ³⁰	1+	DAO + lectura simple	90,79 (±1,01) †	92,70 (±7,48) †
<i>Noble et al. 2009</i> ³¹	1++	DAO + lectura simple	86 (84,2-87,6) ‡	88,2 (86,1-88,3) ‡
<i>Irwig et al. 2004</i> ³²	1-	DAO aislado	95-100*	ND
Informe de evaluación de tecnologías				
<i>Taylor et al. 2005</i> ³³	2-	DAO + lectura simple	79 (76-81)	ND
Estudios originales				
<i>Hupse et al. 2013</i> ³⁴	2-	DAO aislado	48,7 62	95 80
<i>Zheng et al. 2012</i> ³⁵	2-	DAO aislado	Global 75,6 Sólo casos actuales 75,9 Sólo casos intervalo 72,9	Global 41,9 Sólo casos actuales ND Sólo casos intervalo ND
<i>Skaane et al. 2012</i> ³⁶	2-	DAO aislado	94	ND
<i>Sánchez et al. 2011</i> ³⁷	2+	DAO + lectura simple	97,1	92,8

DAO: diagnóstico asistido por ordenador; ND: datos no disponibles.

*: expresado como síntesis cualitativa.

†: expresado como media ponderada y desviación típica.

‡: expresado como metanálisis.

Aspectos económicos

Coste por unidad

Según el informe de evaluación de tecnologías publicado por Taylor *et al.* 2005³³, el coste del ImageChecker, sistema con vida útil de 7 años, para el *National Health Service British Screening Programme* en 2001 fue 108.000 £ más el mantenimiento anual estimado en 10.000 £ al año.

En 2010, el Instituto *ECRI* publicó un informe de evaluación de tecnologías que contenía el listado de precios, el precio promedio cotizado, la desviación estándar y el rango de precios cotizados de algunos de los sistemas DAO disponibles en el mercado (Tabla 5).

Tabla 5. Listado de precios, precio promedio cotizado, desviación estándar y rango de precios cotizados

Sistema DAO	Lista de precios	Precio promedio cotizado	Desviación estándar	Rango de precios cotizados
ImageChecker M1000 (R2 Technology)	78.000 \$	243.997 \$	286.842 \$	Not reported
ImageChecker D (MF100H) (R2 Technology)	1.071.450 \$	687.265 \$	210.646 \$	548.750 \$ a 1.000.850 \$
ImageChecker Dmax (R2 Technology)	265.000 \$	174.325 \$	31.855 \$	151.800 \$ a 196.950 \$
ImageChecker LS (R2 Technology)	265.000 \$	67.600 \$	20.419 \$	42.000 \$ a 108.000 \$
Second Look 200 (iCAD)	69.960 \$	—	—	—
Second Look 200 (iCAD)	139.950 \$	—	—	—
Second Look 200 (iCAD)	149.950 \$	—	—	—

Fuente: ImageChecker precios del ECRI Institute Sourcebase; Second Look precios del SourceOne Radiology Distribution Products and Pricing Report 2010.

Estudios de evaluación económica

De los trabajos recuperados para su análisis, solamente el informe de evaluación de tecnologías contempló la evaluación económica del DAO33.

El objetivo de este trabajo fue conocer el coste-efectividad del DAO más la lectura simple frente a la sistemática de lectura habitual, desde la perspectiva del Sistema Sanitario. Los costes utilizados fueron obtenidos de cinco centros de cribado de cáncer de mama y se presentaron de

acuerdo con los precios estimados en libras de los años 2001/02. Se incluyeron los siguientes costes:

- Gastos de inversión: 108.000 £.
- Gastos de mantenimiento: estimado en 10.000 £ al año.
- Entrenamiento y tiempo de lectura de los profesionales: 45,11 £ por hora en el caso del radiólogo y 13,71 £ por hora para el técnico en radiología).
- Digitalización de las mamografías: 5,90 £ por hora.
- Costes de invitación de las mujeres a la participación en el programa, el coste de los films y el material fungible: 25,49 £ por unidad.

A todos ellos se aplicó una tasa de descuento del 3,5 %.

Descripción de la metodología

El análisis se presentó de dos maneras:

- **Coste por cáncer detectado a los 12 meses.** Se obtuvo mediante un árbol de decisión con cuatro resultados: verdaderos positivos, falsos negativos, falsos positivos y verdaderos negativos. Este análisis se dividió en la presencia de cáncer o no cáncer. Para cada uno de los resultados se estimaron los costes correspondientes. Los datos de sensibilidad y especificidad fueron obtenidos de los estudios incluidos en el informe de evaluación de tecnologías. La variable prevalencia fue incluida en el modelo para determinar el número de cánceres detectados. El coste por cáncer detectado (resultado coste-efectividad) se estimó del cociente entre las diferencias de los costes y las diferencias de los cánceres adicionales detectados.
- **Coste por años de vida ajustados por calidad (AVAC) a los 10 años.** Se empleó el modelo de Markov de 10 años de duración y ciclos anuales. Este modelo incluyó 5 estados de salud (cáncer diagnosticado, recurrencia local, recurrencia regional, recurrencia distal y muerte). En este caso, se aplicó una tasa de descuento del 3,5 % en costes y del 1,5 % en AVAC. El ratio de coste-efectividad incremental se calculó como el cociente entre la diferencia de costes y la diferencia de AVAC. Para confirmar la robustez del modelo e identificar los parámetros con mayor influencia en los resultados finales se realizaron dos tipos de análisis de sensibilidad: análisis univariante y análisis probabilístico (usando la simulación de Monte Carlo). También

fue calculada la curva de aplicabilidad coste-efectividad para indicar la probabilidad con la incorporación del DAO resultaría coste-efectivo teniendo en cuenta un umbral de 30.000 £ que se aplica en el Reino Unido.

Descripción de la calidad

Teniendo en cuenta los parámetros que contempla la herramienta de evaluación CASP (Anexo 6) para estudios de tipo económico este trabajo ha sido de alta calidad ya que la pregunta de investigación y el objetivo estuvieron bien definidos, realizaron una correcta descripción de las alternativas y sus consecuencias, la metodología fue adecuada, hubo un desglose de los costes y se aplicaron las tasas de descuento correspondientes. Los resultados de la evaluación estuvieron bien definidos y el análisis de sensibilidad fue realizado correctamente.

Descripción de los resultados

- **Coste por cáncer detectado a los 12 meses.** El coste por cáncer detectado a los 12 meses mediante el uso del DAO por cada 1.000 mujeres cribadas fue de 23.272 £, teniendo en cuenta que los costes de maquinaria, entrenamiento, lectura y material fungible adicionales fueron 4.538 £ y los cánceres adicionales detectados fueron 0,195. En el caso del estudio 1, estos fueron 10.476 £ y 0,13, respectivamente.
- **Coste por años de vida ajustados por calidad (AVAC) a los 10 años.** En este caso sólo pudo contemplarse el resultado del estudio 2 ya que en el 1 los datos no estuvieron disponibles. Así pues, este coste-efectividad sería de 26.069 £, para unos costes de 4.171 £ y unos AVAC de 0,16. El análisis de sensibilidad fue realizado para 5.000 simulaciones y la curva de aceptabilidad indicó que, para un umbral de 30.000 £, hubo un 50% de probabilidad de que el DAO fuera una tecnología coste-efectiva.

Discusión

Los programas de detección precoz en el cáncer de mama mediante la mamografía tienen un efecto potencial en la reducción de la mortalidad por esa causa³⁸. En la línea de las nuevas modalidades para su cribado, se han planteado distintos estudios que comparan la eficacia del DAO como herramienta de apoyo al radiólogo frente a la sistemática de lectura habitual. En este contexto, la aplicación del DAO al cribado de cáncer de mama, como modificación del procedimiento de lectura mamográfica, está siendo explorada como futura metodología.

La eficacia de la mamografía aplicada al programa de cribado de cáncer de mama, en términos de sensibilidad, se ha estimado entre el 85 % y el 95 %. La especificidad ronda el 90 %³⁹.

De las cuatro revisiones sistemáticas recuperadas, sólo dos fueron de alta calidad, mostrando sensibilidades del DAO más la lectura simple entre 86 % y 87 %, con especificidades entre 88 y 96 %. Los 2 estudios originales recuperados con el mismo tipo de intervención mostraron un amplio rango en la sensibilidad, con valores de 79 % y 97 %, siendo el valor más alto el obtenido por el trabajo de calidad moderada, con un 93 % de especificidad. De este modo, los **valores de sensibilidad y especificidad no difirieron de manera importante con los de la mamografía**. Aún así, Noble *et al.* 2009³¹ obtuvieron un aumento en la tasa de rellamada del 4,1 % (IC 95 %: 2,7-6,3), con una proporción de falsos positivos de 80 de cada 100.000 mujeres cribadas al emplear el DAO como herramienta de apoyo a la lectura mamográfica simple.

En cuanto al DAO utilizado de forma aislada, todos los trabajos recuperados tuvieron una baja calidad, bien fuera por tratarse de diseños de tipo caso-control en el caso de los originales³⁴⁻³⁶, o por no contener suficiente información de la búsqueda, el tipo de revisión, la prueba de estudio y la de referencia, en la revisión sistemática³². La sensibilidad del DAO osciló entre el 48,7 y el 100, siendo la menor para el estudio de Hupse *et al.* 2013³⁴ donde los profesionales conocían tanto la proporción de imágenes malignas como el hecho de que se habían eliminado todas las imágenes compatibles con microcalcificaciones. Este trabajo registró, además, una especificidad del 95%, que fue la más alta de todos los trabajos de esta categoría. Teniendo en cuenta lo anterior, **la sensibilidad y especificidad para la mamografía fueron superiores a lo obtenido por estos autores para el DAO**.

Algunos trabajos se han mostrado más favorables en la detección de cierto tipo de lesiones. Sánchez *et al.* 2011³⁷ y Skaane *et al.* 2012³⁶ fueron los

únicos que mostraron la eficacia del DAO desglosada según el tipo de lesión, encontrándose una sensibilidad para las microcalcificaciones del 91,3 % en el primer caso, y del 100 % en el segundo. En este sentido, otros autores^{13,40}, han obtenido los mismos resultados, destacado que el DAO posee un elevado rendimiento diagnóstico para las microcalcificaciones. En el caso de las masas, la sensibilidad fue de 90,4 % para Skaane *et al.* 2012³⁶, lo que converge con lo obtenido por otros autores⁴⁰⁻⁴². Sánchez *et al.* 2011³⁷ incluyeron los nódulos, las distorsiones de la arquitectura mamaria y las lesiones mixtas, cuya sensibilidad fue 77,4 %, 70 % y 100 %, respectivamente.

No obstante, estos resultados hay que valorarlos con cautela debido, fundamentalmente, a que la calidad de los trabajos originales fue entre moderada y baja, incluso los que formaron parte de las dos revisiones sistemáticas de mayor calidad. De esta manera, estos resultados, a pesar de no diferir de lo encontrado en la literatura para la mamografía, podrían estar sobreestimados lo que implicaría que, realmente, la eficacia de la prueba a estudio sería menor que la referida a la mamografía.

Consecuencias e implicaciones prácticas

La realización de cualquier prueba de cribado conlleva una serie de riesgos intrínsecos. En primer lugar, la radiación y las complicaciones derivadas de la realización de la biopsia, en el caso del cáncer de mama, debe ser contemplado. En segundo, puede producirse el efecto de “etiquetado de enfermo”, que puede ir acompañado o no del sobrediagnóstico ya que el cribado podría detectar lesiones cancerosas que quizás nunca se hubieran manifestado clínicamente y mucho menos hubieran ocasionado la muerte. Por último, el incremento en la detección de falsos positivos tendría también consecuencias a nivel psicosocial. Estas mujeres sufrirían altos niveles de estrés y ansiedad durante la espera del diagnóstico definitivo, serían sometidas a pruebas complementarias de segundo orden con sus respectivos riesgos y, además, esta situación les provocaría disrupciones en el medio familiar y laboral³⁹.

En cuanto a las consecuencias económicas, la puesta en marcha del DAO en el programa de cribado supondría la contratación de un solo radiólogo en los centros donde la lectura mamográfica doble sea la técnica habitual. *A priori*, parece lógico pensar que esta decisión ocasionaría un ahorro relacionado con los costes del profesional sanitario. Sin embargo, debería considerarse que la aplicación del DAO conllevaría, entre otras cosas, un incremento en los falsos positivos y la tasa de rellamada que sería directamente proporcional a los costes en recursos humanos y materiales (personal cualificado para la realización de la biopsia, materiales para su realización y procesamiento de la muestra, personal para el análisis e

interpretación de los resultados, etc.). De este modo, según la evaluación económica realizada por Taylor et al. 2005³³ y para un umbral de 30,000 £ habría un 50 % de probabilidades de que el DAO fuera una tecnología coste-efectiva. A pesar de la alta calidad del documento, sus resultados generan dudas sobre los beneficios de su incorporación en el programa de cribado y deben tomarse con precaución al mismo tiempo que ser considerados según el contexto donde se desee aplicar.

Limitaciones

Debe hacerse explícito que este trabajo no se encuentra exento de limitaciones.

En primer lugar, hay que considerar que los resultados de una revisión sistemática se basan en los estudios recuperados tras la realización de una búsqueda bibliográfica. La inclusión sólo de estudios publicados pudo favorecer el sesgo de localización de los trabajos ya que las revistas no indexadas en alguna de las principales bases de datos no estuvieron disponibles. No obstante, para minimizar este sesgo no se establecieron limitaciones en cuanto al idioma de publicación y la selección de los estudios se realizó en base a criterios predefinidos.

En segundo lugar, debe tenerse en cuenta que la revisión por pares no pudo ser realizada en todas las fases del estudio. Ésta se llevó a cabo en la lectura crítica de los trabajos seleccionados.

Por último, es importante aclarar que esta revisión sistemática no cuenta con un metanálisis posterior. Esto es debido a que los artículos originales no proporcionaron los datos suficientes para completar la tabla 2x2, por lo que no fue posible realizar el análisis estadístico.

Conclusiones

- Las revisiones sistemáticas de mejor calidad (nivel de evidencia 1++) obtuvieron resultados de validez diagnóstica para el DAO más lectura simple entre 86 % y 87 % para la sensibilidad y 93 % para la especificidad. No obstante, los estudios originales que las componen fueron de calidad moderada o baja, lo que pudo haber sobrestimado los resultados. Los 2 estudios originales publicados con posterioridad mostraron un amplio rango en la sensibilidad, entre 79 % y 97 % (nivel de evidencia 2- y 2+, respectivamente).
- DAO aislado mostró una sensibilidad entre 94 % y 100 % y especificidad entre 42 % y 95 % en la mayoría de los trabajos evaluados excepto en uno, donde la sensibilidad fue inferior al 50 %. Este resultado pudo explicarse al excluir las microcalcificaciones del análisis.
- La validez diagnóstica según el tipo de lesión fue convergente con lo encontrado en la literatura. Para las microcalcificaciones, la sensibilidad fue entre 91,3 % y 100 %, mientras que para las masas fue 90,4 %.
- La tasa de rellamada fue 4,1 % (IC 95 %: 2,7-6,3) en el metaanálisis, con una proporción de falsos positivos de 80 de cada 100.000 mujeres cribadas.
- La evaluación económica resultó ser incierta. El coste por cáncer detectado mediante el uso del DAO por cada 1.000 mujeres cribadas fue de 23.272 £ y curva de aceptabilidad indicó que, para un umbral de 30.000 £, hubo un 50% de probabilidad de que el DAO fuera una tecnología coste-efectiva.
- Ningún trabajo de los rescatados contempló la precisión de la prueba diagnóstica.

Referencias

1. Boyle P, Levin B. World cancer report 2008 [Internet]. Lyon: International Agency for Research on Cancer; 2008 [citado 16/07/2013]. URL: <http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/index.php>. (Archivado por WebCite® en <http://www.webcitation.org/5w0V3eALT>)
2. Ferlay J, Autier P, Boniol M, Heanue M, Colombet M, Boyle P. Estimates of the cancer incidence and mortality in Europe in 2006. *Ann Oncol.* 2007;18:581-92.
3. International Agency for Research on Cancer. Globocan 2008 [Internet]. Lyon: World Health Organization; 2008 [citado 16/07/2013]. URL: <http://globocan.iarc.fr/factsheet.asp>. (Archivado por WebCite® en <http://www.webcitation.org/6IJ3ctMvY>)
4. Boyle P, Ferlay J. Cancer incidence and mortality in Europe, 2004. *Ann Oncol.* 2005;16:481-8.
5. López-Abente G, Pollán M, Aragonés N, Pérez-Gómez B, Hernández V, Lope V, et al. La situación del cáncer en España. Madrid: Ministerio de Sanidad y Consumo; 2005.
6. Pollán M, García Medizábal MJ, Pérez-Gómez B, Aragonés N, Lope V, Pastor R, et al. Situación epidemiológica del cáncer de mama en España. *Psicooncología.* 2007;4:231-48.
7. Cabanes A, Vidal E, Aragonés N, Pérez-Gómez B, Pollán M, Lope V, et al. Cancer mortality trends in Spain: 1980-2007. *Ann Oncol.* 2010;21:14-20.
8. Verdecchia A, Francisci S, Brenner H, Gatta G, Micheli A, Mangone L, et al. Recent cancer survival in Europe: a 2000-02 period analysis of EURO CARE-4 data. *Lancet Oncol.* 2007;8:784-96.
9. Castells X, Sala M, Ascunce N, Salas D, Zubizarreta R, Casamitjana M. Descripción del cribado del cáncer en España. Proyecto DESCRIC. Madrid: Plan de Calidad para el Sistema Nacional de Salud, Ministerio de Sanidad y Consumo. Barcelona: Agència d'Avaluació de Tecnologia i Recerca Mèdiques de Catalunya; 2007. Informes, estudios e investigación; AATRM 2006/01.

10. ECRI Institute. Computer-aided detection mammography for breast cancer screening. Plymouth: ECRI Institute; 2010.
11. European Commission. En: Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L, editores. European guidelines for quality assurance in breast cancer screening and diagnosis. 4ª ed. Luxemburgo: Office for official publications of the European Communities; 2006.
12. Resultados de los programas de cribado de cáncer de mama 2009 [Internet]. [s.l.]: Red de programas de cribado de cáncer; 2013 [citado 16/07/2013]. URL: <http://www.programascancerdemama.org>. (Archivado por WebCite® en <http://www.webcitation.org/6IJ3jiD4X>)
13. Dromain C, Boyer B, Ferré R, Canale S, Delalogue S, Balleyguier C. Computed-aided diagnosis (CAD) in the detection of breast cancer. *Eur J Radiol.* 2013;82:417-23.
14. Boyer B, Balleyguier C, Granat O, Pharaboz C. CAD in questions/answers: Review of the literature. *Eur J Radiol.* 2009;69:24-33.
15. Llobet Azpitarte R. Aportaciones al diagnóstico de cáncer asistido por ordenador [tesis doctoral]. Valencia: Universidad Politécnica de Valencia; 2006.
16. Suárez Cuenca JJ. Desarrollo de un sistema de diagnóstico asistido por computador para detección de nódulos pulmonares en tomografía computarizada multicorte [tesis doctoral]. Santiago de Compostela: Universidad de Santiago de Compostela; 2009.
17. Rao VM, Levin DC, Parker L, Cavanaugh B, Frangos AJ, Sunshine JH. How widely is computer-aided detection used in screening and diagnostic mammography? *J Am Coll Radiol.* 2010;7:802-5.
18. Castilla-la Mancha contará con un Centro de Excelencia en Imagen Médica Digital. Servicio de Salud de Castilla-La Mancha [Internet]. 19/07/2009 [citado 16/07/2013]. URL: <http://sescam.jccm.es/web1/home.do?main=/ciudadanos/noticias/20090719CentroExcelenciaImagenMedica.html>. (Archivado por WebCite® en <http://www.webcitation.org/6IA5uTq52>)
19. Carreira C, Martínez J, Gómez D, Polanco C, Naranjo P, del Llano D. Revisión de la evidencia científica sobre la aplicación clínica de la mamografía digital. *Radiologia.* 2007;49:145-56.

20. Sarriugarte G, Rueda JR, Martínez-LaFuente O, Nicolás P, Otegi A, Rosales FJ, et al. La mamografía digital en el cribado del cáncer de mama; evaluación de su eficacia y de su introducción en el Programa de Detección Precoz del Cáncer de Mama de la Comunidad Autónoma del País Vasco. Investigación Comisionada. Vitoria-Gasteiz: Gobierno Vasco, Departamento de Sanidad y Consumo; 2011. Informe nº: Osteba D-11-04.
21. Urrutia G, Bonfill X. Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metanálisis. *Med Clin (Barc)*. 2010;135:507-11.
22. Higgins JPT, Green S, editors. Assessing risk of bias in included studies. En: *Cochrane handbook for systematic reviews of interventions Version 5.1.0* [Internet]. The Cochrane Collaboration; 2011. [citado 16/07/2013]. URL: <http://www.cochrane-handbook.org>. (Archivado por WebCite® en <http://www.webcitation.org/6IJ3pRLiK>)
23. Scottish Intercollegiate Guidelines Network (SIGN). SIGN 50: A guideline developer's handbook. Edinburgo: SIGN; 2011 [citado 08/03/2013]. URL: <http://www.sign.as.uk>. (Archivado por WebCite® en <http://www.webcitation.org/6IJ3t2bjS>)
24. National Institute for Health and Clinical Excellence. The guidelines manual. Londres: National Institute for Health and Clinical Excellence; 2007 [citado 16/07/2013]. URL: <http://www.nice.org.uk/niceMedia/pdf/GuidelinesManualAllChapters.pdf>. (Archivado por WebCite® en <http://www.webcitation.org/6IJ3xeRIj>)
25. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10.
26. Withing P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
27. González MP, Velarde C. Listas guía de comprobación de estudios sobre pruebas diagnósticas incluidos en las revisiones sistemáticas: declaración QUADAS. *Evid Pediatr*. 2012;8:20.

28. Guías CASPe de Lectura Crítica de la Literatura [Internet]. Alicante: CASPe España; 2005 [citado 16/07/2013]. URL: <http://www.redcaspe.org/que-hacemos/herramientas/>. (Archivado por WebCite® en <http://www.webcitation.org/6IJ4ZDrNg>)
29. Azevedo E, Zackrisson S, Mejàre I, Arnlind MH. Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. *BMC Med Imaging*. 2012;12:22.
30. Eadie LH, Taylor P, Gibson AP. A systematic review of computer-assisted diagnosis in diagnostic cancer imaging. *Eur J Radiol*. 2012;81:70-6.
31. Noble M, Bruening W, Uhl S, Schoelles K. Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis. *Arch Gynecol Obstet*. 2009;279:881-90.
32. 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.
33. Taylor P, Champness J, Given-Wilson R, Johnston K, Potts H. Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography. *Health Technol Assess*. 2005;9:1-58.
34. Hupse R, Samulski M, Lobbes M, den Heeten A, Imhof-Tas MW, Beijerinck D, et al. Standalone computer-aided detection compared to radiologists' performance for the detection of mammographic masses. *Eur Radiol*. 2013;23:93-100.
35. Zheng B, Sumkin JH, Zuley ML, Lederman D, Wang X, Gur D. Computer-aided detection of breast masses depicted on full-field digital mammograms: a performance assessment. *Br J Radiol*. 2012;85:153-61.
36. Skaane P, Kshirsagar A, Hofvind S, Jahr G, Castellino RA. Mammography screening using independent double reading with consensus: is there a potential benefit for computer-aided detection? *Acta Radiol*. 2012;53:241.
37. Sanchez S, Torres M, Vega A, Sainz M, Baroja A, Ruiz M, et al. Impact of a CAD system in a screen-film mammography screening program: a prospective study. *Eur J Radiol*. 2011;80:317-21.

38. World Health Organization. Breast cancer: prevention and control [Internet]. Ginebra: World Health Organization; 2011 [citado 16/07/2013]. URL: <http://www.who.int/cancer/detection/breastcancer/en/index3.html>. (Archivado por WebCite® en <http://www.webcitation.org/6ICbkPucc>)
39. Márquez-Calderón S. Los efectos negativos de las intervenciones preventivas basadas en el individuo. *Gac Sanit.* 2008;22:205-15.
40. James JJ, Gilbert FJ, Wallis MG, Gillan MGC, Astley SM, Boggis CRM, et al. Mammographic features of breast cancers at single reading with computer-aided detection and at double reading in a large multicenter prospective trial of computer-aided detection: CADET II. *Radiology.* 2010;256:379-86.
41. Taplin SH, Rutter CM, Lehman CD. Testing the effect of computer-assisted detection on interpretive performance in screening mammography. *AJR Am J Roentgenol.* 2006;187:1475-82.
42. Brem RF, Hoffmeister JW, Zisman G, DeSimio MP, Rogers SK. A computer-aided detection system for the evaluation of breast cancer by mammographic appearance and lesion size. *AJR Am J Roentgenol.* 2005;184:893-6.

Anexos

Anexo 1. Estrategias de búsqueda

OID MEDLINE(R) <1946 TO APRIL WEEK 2 2013> ESTRATEGIA DE BÚSQUEDA:

1. ("18832239" or "17888770" or "18353630" or "16990670").ui.
2. Mass screening/
3. diagnosis/ or exp early diagnosis/
4. (diagnos* or screening*).ti,ab. (1663339)
5. Mammography/
6. (mammogr* or (RADIOGRAPH* adj3 (mamm* or breast))).ti,ab.
7. 2 or 3 or 4
8. 5 or 6
9. 7 and 8
10. exp Diagnosis, computer-assisted/is, mt, sn, ut
[Instrumentation, Methods, Statistics & Numerical Data,
Utilization]
11. exp diagnostic imaging/is, mt, sn, ut
12. 1 ((computer or computer?assisted or computer?aided) adj3
(analysis or diagnos* or detect* or radiograph*).ti,ab.
13. ((imag* or scan*) and diagnos*).ti,ab.
14. 10 or 11 or 12 or 13
15. 9 and 14
16. reading/
17. (read* and (simple or double)).ti,ab.
18. 16 or 17
19. 15 and 18
20. exp "reproducibility of results"/
21. exp "sensitivity and specificity"/
22. "outcome assessment (health care)"/
23. (sensitivit* or specificit* or (predictive adj2 value*) or
likelihood or ((false or true) adj2 (positiv* or
negativ*))).ti,ab.
24. ((confiance or detect*) adj3 (rate? or ratio?)).ti,ab.
25. 20 or 21 or 22 or 23 or 24
26. 19 and 25

**EMBASE (R) <1946 TO APRIL WEEK 2 2013>
ESTRATEGIA DE BÚSQUEDA:**

27. "cancer screening/exp OR "early diagnosis/exp
28. diagnosis*:ab,ti OR screening*:ab,ti
29. #1 OR #2
30. "mammography"/de
31. mammogr*:ab,ti OR (radiograph* NEAR/3 (mamm* OR breast)):ab,ti
32. #4 OR #5
33. #3 AND #6
34. "diagnosis"/de OR "computer assisted radiography"/exp OR "diagnostic imaging"/exp
35. computer:ab,ti OR computer\$:ab,ti AND assisted:ab,ti OR computer\$:ab,ti AND aided:ab,ti
36. radiograph*:ab,ti OR diagnose*:ab,ti OR detect*:ab,ti OR analysis:ab,ti
37. #9 AND #10
38. imag*:ab,ti AND diagnos*:ab,ti
39. l#8 OR #11 OR #12
40. #7 AND #13
41. "reading"/exp
42. read*:ab,ti AND (simple:ab,ti OR double:ab,ti)
43. #15 OR #16
44. #14 AND #17
45. "reproducibility"/exp OR "sensitivity and specificity"/exp OR "outcome assessment"/exp
46. sensitivit*:ab,ti OR specificit*:ab,ti OR (predictive NEAR/2 value*):ab,ti OR likelihood:ab,ti OR ((false OR true) NEAR/2 (positiv* OR negative*)):ab,ti
47. ((confiance OR detect*) NEAR/3 (rate\$ OR ratio\$)):ab,ti
48. #19 OR #20 OR #21
49. #18 AND #22

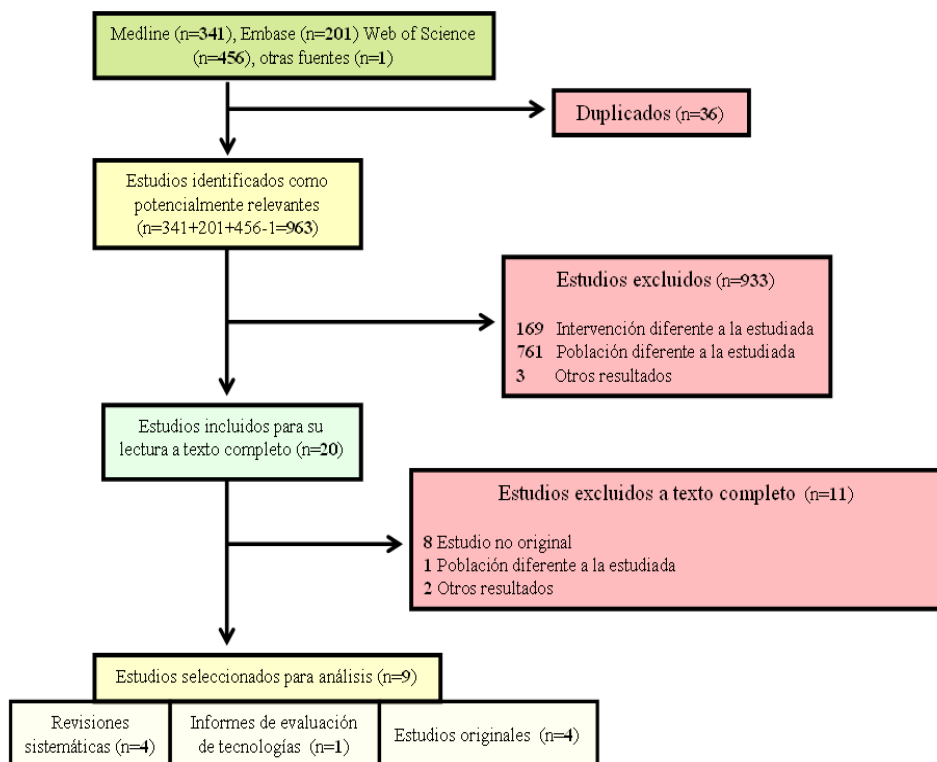
WEB OF SCIENCE (R) <1946 TO APRIL WEEK 2 2013> ESTRATEGIA DE BÚSQUEDA

1. TI=(diagnos* or screening*) or TS=(diagnos* or screening*)
2. TI=(mammogr* or (RADIOGRAPH* near/3 (mamm* or breast))) or TS=(mammogr* or (RADIOGRAPH* near/3 (mamm* or breast)))
3. #2 OR #1
4. TI=((computer or computer\$aided) near/3 (analysis or diagnos* or detect* or radiograph*)) or TS=((computer or computer\$aided) near/3 (analysis or diagnos* or detect* or radiograph*))
5. TI=((imag* or scan*) and diagnos*) or TS=((imag* or scan*) and diagnos*)
6. #5 OR #4
7. TI=(sensitivit* or specificit* or (predictive near/2 value*) or likelihood or ((false or true) near/2 (positiv* or negativ*))) or TS=(sensitivit* or specificit* or (predictive near/2 value*) or likelihood or ((false or true) near/2 (positiv* or negativ*)))
8. TI=((confiance or detect*) near/3 (rate\$ or ratio\$)) or TS=((confiance or detect*) near/3 (rate\$ or ratio\$))
9. #8 OR #7
10. #9 AND #6 AND #3
11. #10 AND Document Types=(Review)
12. #10 AND Document Types=(Review)
13. #10 AND Document Types=(Review)
14. #10 AND Document Types=(Review). Refined by: Web of Science Categories=(RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING OR COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR ONCOLOGY OR COMPUTER SCIENCE CYBERNETICS OR COMPUTER SCIENCE INFORMATION SYSTEMS OR MEDICINE RESEARCH EXPERIMENTAL OR MEDICAL INFORMATICS OR PHYSICS APPLIED OR INSTRUMENTS INSTRUMENTATION)
15. #10 AND Document Types=(Review). Refined by: Web of Science Categories=(RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING OR COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR ONCOLOGY OR COMPUTER SCIENCE CYBERNETICS OR COMPUTER SCIENCE INFORMATION SYSTEMS OR MEDICINE RESEARCH EXPERIMENTAL OR MEDICAL INFORMATICS OR PHYSICS APPLIED OR INSTRUMENTS INSTRUMENTATION) AND Languages=(ENGLISH)
16. #10
17. #10 Refined by: Web of Science Categories=(RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING OR ONCOLOGY OR PHYSICS APPLIED OR OBSTETRICS GYNECOLOGY OR MEDICAL INFORMATICS OR COMPUTER SCIENCE INFORMATION SYSTEMS OR ENGINEERING BIOMEDICAL OR COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR MEDICINE RESEARCH EXPERIMENTAL OR INSTRUMENTS INSTRUMENTATION)
18. #10 Refined by: Web of Science Categories=(RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING OR ONCOLOGY OR PHYSICS APPLIED OR

OBSTETRICS GYNECOLOGY OR MEDICAL INFORMATICS OR COMPUTER SCIENCE INFORMATION SYSTEMS OR ENGINEERING BIOMEDICAL OR COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR MEDICINE RESEARCH EXPERIMENTAL OR INSTRUMENTS INSTRUMENTATION) AND Document Types=(ARTICLE)

19. #10 Refined by: Web of Science Categories=(RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING OR ONCOLOGY OR PHYSICS APPLIED OR OBSTETRICS GYNECOLOGY OR MEDICAL INFORMATICS OR COMPUTER SCIENCE INFORMATION SYSTEMS OR ENGINEERING BIOMEDICAL OR COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR MEDICINE RESEARCH EXPERIMENTAL OR INSTRUMENTS INSTRUMENTATION) AND Document Types=(ARTICLE) AND Publication Years=(2013)
20. #10 Refined by: Web of Science Categories=(RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING OR ONCOLOGY OR PHYSICS APPLIED OR OBSTETRICS GYNECOLOGY OR MEDICAL INFORMATICS OR COMPUTER SCIENCE INFORMATION SYSTEMS OR ENGINEERING BIOMEDICAL OR COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR MEDICINE RESEARCH EXPERIMENTAL OR INSTRUMENTS INSTRUMENTATION) AND Document Types=(ARTICLE) AND Publication Years=(2013) AND Document Types=(ARTICLE)
21. #10 Refined by: Web of Science Categories=(RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING OR ONCOLOGY OR PHYSICS APPLIED OR OBSTETRICS GYNECOLOGY OR MEDICAL INFORMATICS OR COMPUTER SCIENCE INFORMATION SYSTEMS OR ENGINEERING BIOMEDICAL OR COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR MEDICINE RESEARCH EXPERIMENTAL OR INSTRUMENTS INSTRUMENTATION) AND Document Types=(ARTICLE) AND Publication Years=(2013) AND Document Types=(ARTICLE) AND Languages=(ENGLISH OR SPANISH)

Anexo 2. Diagrama de flujo de la gestión documental



Anexo 3. Cuestionario de calidad de las revisiones sistemáticas (SIGN)

Evaluación de la calidad de revisiones sistemáticas y metaanálisis según la escala propuesta por SIGN				
	Azevedo 2012 ²⁹	Eadie, 2012 ³⁰	Noble, 2009 ³¹	Irwig, 2004 ³²
Sección 1: Validez interna				
En qué medida se cumple este criterio:				
Criterio de evaluación	-Se cumple adecuadamente -Se cumple parcialmente -No se cumple adecuadamente	-No realizado -No recogido -No aplicable		
El estudio está dirigido a una pregunta adecuada y claramente formulada	Adecuadamente	Adecuadamente	Adecuadamente	Adecuadamente
Se incluye una descripción de la metodología utilizada	Adecuadamente - PubMed desde 1950 hasta 2011 - EMBASE desde 1950 hasta 2011 - The Cochrane Library desde 1950 hasta 2011 - No muestra la estrategia de búsqueda - Revisión por pares - Plantean la búsqueda de acuerdo con la pregunta PICO	Parcialmente - Medline - EMBASE - IEEExplore - SCOPUS - Web of Science - Science Direct - CINAHL/EBSCO - (todo lo anterior desde enero 1990 hasta agosto 2010) - Muestra la estrategia de búsqueda - Se desconoce si la revisión ha sido por pares	Adecuadamente - Refiere haber buscado en 17 bases de datos - Destaca Medline - Destaca EMBASE - Destaca The Cochrane Library - Destaca PsycINFO - No especifica la fecha de búsqueda - Muestra la estrategia de búsqueda Embase/Medline - Revisión por pares	No cumple adecuadamente - Medline desde 1966 hasta dic 2002 - No muestra la estrategia de búsqueda - Se desconoce si se ha realizado revisión por pares

Evaluación de la calidad de revisiones sistemáticas y metaanálisis según la escala propuesta por SIGN (Continuación)					
Azevedo 2012 ²⁹		Eadie, 2012 ³⁰		Noble, 2009 ³¹	Irwig, 2004 ³²
Sección 1: Validez interna					
La búsqueda fue suficientemente rigurosa para identificar los estudios relevantes	Adecuadamente	Adecuadamente	Parcialmente	No se cumple adecuadamente	
	- Incluye la literatura gris y manuales - Muestra diagrama de flujo	- Incluye gran número de bases de datos. - Muestra diagrama de flujo - Incluye los artículos en inglés menos los que el texto completo no estuviera disponible (n=9)	- Se desconocen el resto de bases de datos que utiliza - Incluye la literatura gris - No muestra diagrama de flujo - Sólo incluye trabajos en inglés	- No utiliza otra bases de datos relevantes - No menciona incl/exc de literatura gris - No muestra diagrama de flujo - Se desconoce criterios de incl/excl de idiomas. - No información sobre sesgo de publicación	
La calidad de los estudios originales fue evaluada y tenida en cuenta	Adecuadamente	Adecuadamente	Adecuadamente	No realizado	
	- Aplicaron QUADAS - Aplicaron GRADE	- Aplicaron QUADAS	- Aplicaron ECR (QUADAS modificado)		

Evaluación de la calidad de revisiones sistemáticas y metaanálisis según la escala propuesta por SIGN (Continuación)				
	Azevedo 2012 ²⁹	Eadie, 2012 ³⁰	Noble, 2009 ³¹	Irwig, 2004 ³²
Sección 1: Validez interna				
Hay suficiente similitud entre los estudios seleccionados para realizar una combinación	<p>Adecuadamente</p> <ul style="list-style-type: none"> - Los trabajos incluidos para análisis cumplen los criterios de inclusión propuestos por los autores: <ul style="list-style-type: none"> * mujeres dentro de un programa de cribado * uso del DAO +lectura simple * comparación con lectura doble * medición de sensibilidad, especificidad, tasa de detección de cáncer, tasas de rellamada 	<p>No aplicable</p> <ul style="list-style-type: none"> - Los autores no muestran el listado de los artículos incluidos para análisis separados por patología (mama, pulmón o piel). 	<p>Adecuadamente</p> <ul style="list-style-type: none"> - Los trabajos incluidos para análisis cumplen los criterios de inclusión propuestos por los autores: <ul style="list-style-type: none"> * mujeres dentro del programa de cribado * trabajos en inglés * en aquellos trabajos en los que se contemplan mujeres en cribado y fuera de cribado se tendrá en cuenta el trabajo si los resultados están estratificados por esa condición. * estudios prospectivos y retrospectivos * estudios aleatorizados * medición de sensibilidad, especificidad, tasa de detección de cáncer, tasas de rellamada 	<p>Adecuadamente</p> <ul style="list-style-type: none"> - Los trabajos incluidos para análisis cumplen los criterios de inclusión propuestos por los autores: <ul style="list-style-type: none"> * Mujeres asintomáticas o con factores de riesgo de cáncer de mama por antecedentes familiares o mujeres con mamas densas o aquellas donde la mamografía no se considere una técnica diagnóstica eficaz.

Evaluación de la calidad de revisiones sistemáticas y metaanálisis según la escala propuesta por SIGN (Continuación)				
	Azevedo 2012 ²⁹	Eadie, 2012 ³⁰	Noble, 2009 ³¹	Irwig, 2004 ³²
Sección 2: Evaluación global del estudio				
¿En qué medida el estudio minimiza los sesgos? (+ +, +, -)	++ Todos o la mayoría de los criterios se han cumplido. Aquellos no cumplidos es poco probable que modifique las conclusiones.	- No consta el tipo de revisión (1 ó 2 investigadores) No se proporcionó listado de estudios diferenciados por tipo de tumor incluido	+ + La descripción de la metodología de búsqueda es insuficiente. No incluye diagrama de flujo. Solo incluye trabajos en inglés.	- No consta el tipo de revisión (1 ó 2 investigadores) El número de bases de datos consultadas es insuficiente. No incluye literatura gris No fue valorada y documentada la calidad científica de los estudios incluidos No muestra conflicto de intereses
Si la respuesta es + o -, ¿en qué dirección el sesgo puede haber desviado los resultados del estudio?	No procede	Sesgo de selección Sesgo de publicación	No procede	Sesgo de publicación

Evaluación de la calidad de revisiones sistemáticas y metaanálisis según la escala propuesta por SIGN (Continuación)				
	Azevedo 2012 ²⁹	Eadie, 2012 ³⁰	Noble, 2009 ³¹	Irwig, 2004 ³²
Sección 3: Descripción de los estudios				
¿Qué tipo de estudios están incluidos en la revisión?	3 prospectivos 1 prospectivo + retrospectivo	ND	4 prospectivos 3 retrospectivos	ND
¿De qué modo la revisión ayuda a responder la pregunta de investigación? (resumir la principal conclusión y cómo está relacionada con la pregunta de investigación)	La revisión no arrojó suficiente evidencia para determinar si la eficacia del DAO +lectura simple fue, al menos, equivalente a la obtenida en la práctica habitual, es decir, la doble lectura.	La revisión sirvió para determinar que el uso del DAO aplicado en la mamografía podría obtener algún beneficio para los radiólogos. Sin embargo, la aplicación del DAO en patología de pulmón y piel no evidenció efectos beneficiosos.	La revisión sirvió para conocer que los resultados derivados de la evaluación del DAO en el cáncer de mama pueden diferir según la experiencia y habilidad del profesional sanitario así como las poblaciones de estudio.	La revisión concluyó que no hubo suficiente apoyo para el uso de las tecnologías evaluadas en poblaciones de cribado y que deberían someterse a más estudios de evaluación.
ND: datos no disponibles.				

Anexo 4. Cuestionario de calidad de las revisiones sistemáticas (AMSTAR)

Evaluación de la calidad de las revisiones sistemáticas y metaanálisis según la escala AMSTAR				
	Azevedo 2012 ²⁹	Eadie, 2012 ³⁰	Noble, 2009 ³¹	Irwig, 2004 ³²
1. ¿Se brindó un diseño a priori? La pregunta de investigación y los criterios inclusión deben ser definidos antes de realizar la revisión.	Sí	Sí	Sí	Sí
2. ¿Hubo selección de estudios y extracción de datos por duplicado? Debe haber al menos dos extractores de datos independientes y un procedimiento de consenso establecido para los desacuerdos.	Sí	No consta	Sí	No consta
3. ¿Se llevó a cabo una búsqueda exhaustiva de la literatura? Debe buscarse en al menos dos fuentes electrónicas. El reporte debe incluir los años y las bases de datos utilizadas (Ejemplo, Central, EMBASE y MEDLINE). Deben mencionarse las palabras clave y/o los términos MESH y si fuera factible, debería brindarse la estrategia de búsqueda. Todas las búsquedas deben ser complementadas consultando current contents*, revisiones, libros de texto, registros especializados o a expertos en el campo de estudio en particular, y examinando las referencias en los estudios encontrados.	Sí (PubMed, EMBASE)	(PubMed, EMBASE, ieexplore, SCOPUS, Web of Science, CINAHL/EBSCO, Science Direct)	Sí (Medline, EMBASE, Cochrane Library)	No (sólo Medline)

Evaluación de la calidad de las revisiones sistemáticas y metaanálisis según la escala AMSTAR (Continuación)				
	Azevedo 2012 ²⁹	Eadie, 2012 ³⁰	Noble, 2009 ³¹	Irwig, 2004 ³²
<p>4. ¿Se usó el estado de publicación (ej. literatura gris) como un criterio de inclusión? Los autores deben afirmar que buscaron reportes sin considerar su tipo de publicación. Los autores deben afirmar si excluyeron o no cualquier reporte (de la revisión sistemática), basados en su estado de publicación, idioma, etc.</p>	Dudoso (no consta, como se desconoce la prueba de referencia se desconoce si es aplicable)	Sí (la prueba de estudio se basa en el análisis de la prueba de referencia)	Sí (lo iban haciendo al mismo tiempo)	Sí (la prueba de estudio se basa en el análisis de la prueba de referencia)
<p>5. ¿Se proporcionó una lista de estudios (incluidos y excluidos)? Debe suministrarse una lista de los estudios incluidos y excluidos.</p>	Sí	No consta	Sí	No consta
<p>6. ¿Fueron suministradas las características de los estudios incluidos? Deben suministrarse los datos de los estudios originales, de forma agregada como en una tabla, sobre los participantes, las intervenciones y los resultados. Deben ser reportados los rangos de las características en todos los estudios analizados por ejemplo, edad, raza, sexo, datos socioeconómicos relevantes, estado de enfermedad, duración, gravedad, u otras enfermedades.</p>	Sí	No	Sí	Sí
<p>7. ¿Fue valorada y documentada la calidad científica de los estudios incluidos? Deben suministrarse los métodos de la valoración a priori (Ej, para estudios de efectividad, si el autor decidió incluir solamente ensayos aleatorizados, doble-ciego y controlados con placebo, o la asignación oculta como criterio de inclusión); para otros tipos de estudios, serán relevantes ítems alternativos.</p>	Sí (QUADAS y GRADE)	Sí (QUADAS)	Sí (ECRI → QUADAS modificado)	No
<p>8. ¿Fue usada apropiadamente la calidad científica de los estudios incluidos para formular las conclusiones? Los resultados de rigor metodológico y la calidad científica deben ser considerados en el análisis y las conclusiones de la revisión, y declarado explícitamente para formular las recomendaciones.</p>	Sí	No	No	No

Evaluación de la calidad de las revisiones sistemáticas y metaanálisis según la escala AMSTAR (Continuación)				
	Azevedo 2012 ²⁹	Eadie, 2012 ³⁰	Noble, 2009 ³¹	Irwig, 2004 ³²
9. ¿Fueron apropiados los métodos utilizados para combinar los resultados de los estudios? Para los resultados combinados, debe realizarse una prueba para asegurar que los estudios eran combinables, que valore su homogeneidad (Ej, la prueba de Chi2 para la homogeneidad, ¹²). Si existiese heterogeneidad debe usarse un modelo de efectos aleatorios y/o considerar si es clínicamente apropiado la combinación (Ej, ¿Es sensato combinar?).	No aplicable	No aplicable	No aplicable	No aplicable
10. ¿Fue valorada la probabilidad de sesgo de publicación? Una valoración del sesgo de publicación debe incluir una combinación de las ayudas gráficas (Ej, gráfico de embudo u otras pruebas disponibles) y/o pruebas estadísticas (Ej, prueba de regresión de Egger).	No	Sí	No	No
11. ¿Fue incluido el conflicto de intereses? Las fuentes potenciales de financiamiento deben ser claramente agradecidas tanto de la revisión sistemática como de los estudios incluidos.	Sí	Sí	Sí	No

Anexo 5. Cuestionario de calidad de los estudios originales e informe de tecnologías sanitarias (QUADAS)

Evaluación de la calidad de los estudios originales e informe de evaluación de tecnologías según la escala QUADAS					
	Zheng B, 201235	Skaane P, 201236	Sánchez S, 201137	Hupse R, 201334	Taylor P, 200533 (STUDY 1*)
1. ¿El espectro de pacientes fue representativo de los pacientes que se someterán a la prueba a estudio en la práctica clínica habitual?	No (477)	Dudoso (23923)	Sí (21855)	No (200)	Dudoso (180)
2. ¿Los criterios de selección estuvieron claramente descritos?	Sí	Sí	Sí (se excluyeron las mamografías sin DAO y los cánceres sólo visibles por ECO)	No (dicen que cogen 200 casos de los cuales 80 cáncer y 120 no dicen criterios inc/exc)	Sí (casos con apariencia sospechosa y biopsia negativa fueron excluidos)
3. ¿La prueba de referencia es la correcta para detectar la condición estudiada?	Dudoso (la patología es confirmada por la prueba habitual, pero no mencionan cuál es)	Sí (mamografía)	Sí (mamografía)	Sí (mamografía + biopsia)	Sí (biopsia)

Evaluación de la calidad de los estudios originales e informe de evaluación de tecnologías según la escala QUADAS					
	Zheng B, 201235	Skaane P, 201236	Sánchez S, 201137	Hupse R, 201334	Taylor P, 200533 (STUDY 1*)
4. ¿El período de tiempo transcurrido entre la prueba de referencia y la prueba a estudio es suficientemente corto para asegurar razonablemente que la condición estudiada no cambia entre la realización de las dos pruebas?	Dudoso (no consta, como se desconoce la prueba de referencia se desconoce si es aplicable)	Sí (la prueba de estudio se basa en el análisis de la prueba de referencia)	Sí (lo iban haciendo al mismo tiempo)	Sí (la prueba de estudio se basa en el análisis de la prueba de referencia)	Sí (1 semana)
5. ¿Todos los individuos del estudio se sometieron a la prueba de referencia para la confirmación del diagnóstico?	No	No	No	No	No
6. ¿Los pacientes recibieron la misma prueba de referencia a pesar de los resultados de la prueba a estudio?	No	No	No	No	No
7. ¿La prueba de referencia fue independiente de la prueba a estudio (la prueba a estudio no formó parte de la prueba de referencia)?	Dudoso	Sí	Sí	Sí	Sí
8. ¿La ejecución de la prueba a estudio se describió con suficiente detalle como para permitir su replicación?	Sí	Sí	Sí	Sí	Sí
9. ¿La ejecución de la prueba de referencia se describió con suficiente detalle como para permitir su replicación?	No (no sabemos cuál es la prueba de referencia)	Dudoso (no ofrecen detalles sobre la biopsia)	Sí	Dudoso (no ofrecen detalles sobre la biopsia)	Dudoso (no ofrecen detalles sobre la biopsia)
10. ¿Los resultados de la prueba a estudio se interpretaron sin conocimiento de los resultados obtenidos en la prueba de referencia?	Dudoso	Sí	Sí	Sí	Sí

Evaluación de la calidad de los estudios originales e informe de evaluación de tecnologías según la escala QUADAS					
	Zheng B, 201235	Skaane P, 201236	Sánchez S, 201137	Hupse R, 201334	Taylor P, 200533 (STUDY 1*)
11. ¿Los resultados de la prueba de referencia se interpretaron sin conocimiento de los resultados obtenidos en la prueba de estudio?	Sí	Sí	Sí	Sí	Dudoso
12. ¿Cuando se interpretaron los resultados de la prueba a estudio, estuvieron disponibles los mismos datos clínicos de los que se dispondría en la práctica clínica habitual?	Dudoso	Sí	Sí	Sí	Sí
13. ¿Se recogieron los resultados no interpretables/intermedios de la prueba a estudio?	Dudoso	Dudoso	Dudoso	Dudoso	Dudoso
14. ¿Se explicaron las pérdidas del estudio?	No se registraron pérdidas	No se registraron pérdidas	No se registraron pérdidas	No se registraron pérdidas	No se registraron pérdidas
*: Informe de Evaluación de Tecnologías					

Anexo 6. Cuestionario de calidad del estudio de evaluación económica del Informe de Evaluación de Tecnologías de Taylor P, 2005³³.

Evaluación de la calidad del estudio de evaluación económica de Taylor P, 2005³³	
A/ ¿Es válida esta evaluación?	
1. ¿Está bien definida la pregunta u objetivo de la evaluación?	Sí
2. ¿Existe una descripción suficiente de todas las alternativas posibles y sus consecuencias?	Sí
3. ¿Existen pruebas de la efectividad, de la intervención o del programa evaluado?	Sí
4. ¿Los efectos de la intervención (o intervenciones) se identifican, se miden y se valoran o consideraran adecuadamente?	Sí
5. ¿Los costes en que se incurre por la intervención (intervenciones) se identifican, se miden y se valoran adecuadamente?	Sí
6. ¿Se aplican tasas de descuento a los costes de la intervención/es? ¿y a los efectos?	Sí
B/ ¿Cuáles son los resultados?	
7. ¿Cuáles son los resultados de la evaluación?	-cáncer detectado a los 12 meses -años de vida ajustados por calidad a los 10 años
8. ¿Se realizó un análisis adecuado de sensibilidad?	Sí
9. ¿Sería el programa igualmente efectivo en tu medio?	No sé
10. ¿Serían los costes trasladables a tu medio?	No sé
11. ¿Vale la pena aplicarlos a tu medio?	No sé

