



ORIGINAL

Auditoría de la mamografía de cribado del cáncer de mama y variabilidad de la práctica radiológica



M.D. Moreno-Ramos* y E. Ruíz-García

Servicio de Radiodiagnóstico, Hospital San Juan de Dios del Aljarafe, Bormujos, Sevilla, España

Recibido el 10 de julio de 2015; aceptado el 6 de septiembre de 2015

Disponible en Internet el 17 de noviembre de 2015

PALABRAS CLAVE

Mamografía;
Cribado;
Auditoría de calidad;
Variabilidad en la
práctica médica

Resumen

Objetivo: Auditoría de las mamografías de cribado de cáncer de mama realizadas en un hospital general, valorando la presencia de la variabilidad en la práctica médica (VPM) en el proceso diagnóstico.

Material y métodos: Revisión de informes de las mamografías de cribado realizadas entre el 1 de mayo de 2010 y el 30 de abril de 2011, con seguimiento clínico durante los 2 años y comparación con los estándares publicados.

Resultados: De 3.878 mujeres exploradas, 368 (9,48%) fueron citadas de nuevo para completar el estudio (97 [16,1%] en cribado inicial y 271 [8,2%] en revisiones). Se indicaron 43 biopsias (1,1%), de las que 24 fueron diagnósticas de cáncer. El valor predictivo positivo (VPP) en estudios de cribado (VPP1) fue de 6,52%, el VPP2 (biopsia recomendada) de 55%, la sensibilidad de 100%, la especificidad de 91% y la tasa de detección de cáncer de 6,1/1.000. No hubo falsos negativos. Veinte tumores fueron invasivos (83,3%). No se apreció infiltración de ganglios axilares en 15. En 6 casos el tamaño del tumor fue menor o igual a 10 mm, y en 17 fue menor de 15 mm. Hubo un porcentaje más elevado de nuevas citaciones en 2 radiólogos (12% y 17,2% frente a 7,3%) ($p < 0,001$). En 217 casos (58,96%) ($p < 0,001$) solo un radiólogo indicó nuevas citaciones. De este grupo el 73% fue dado de alta en la primera visita, frente al 47,68% en el grupo no discrepante ($p < 0,001$). Cuatro de los cánceres se detectaron entre estas 217 pacientes.

Conclusiones: Los resultados observados se ajustan a los valores de referencia. El dato discordante es la tasa de nuevas citaciones, tanto en cribado inicial como en revisión, con importante variabilidad en función del radiólogo lector.

© 2015 SECA. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

* Autor para correspondencia.

Correo electrónico: lola67@wanadoo.es (M.D. Moreno-Ramos).

KEYWORDS

Mammography;
Screening;
Quality audit;
Variability in medical
practice

An audit of breast cancer screening mammograms and the variability of radiological practice**Abstract**

Objective: To audit the breast cancer screening mammograms performed in a general hospital and to assess the variation in medical practice in the diagnostic process.

Material and methods: A review was carried out on the screening mammograms performed between 1 May 2010 and 30 April 2011, with clinical follow up for two years, and a comparison with the published standards.

Results: Of the 3,878 women examined, 368 (9.48%) were called back to complete the study (97 [16.1%] in the initial screening and 271 [8.2%] in revisions). Forty three biopsies (1.1%) were indicated, of which 24 were diagnosed with cancer. The positive predictive value (PPV) in screening studies (PPV1) was 6.52%. For the recommended biopsy (PPV2) it was 55%, with a sensitivity of 100%, a specificity of 91% and a cancer detection rate of 6.1/1,000. There were no false negatives. Twenty tumours were invasive; with no axillary lymph node infiltration was observed 15 of them. In 6 cases, the size of the tumour was less than or equal to 10 mm, and in 17 it was less than 15 mm. There were a higher percentage of new appointments by two radiologists (12% and 17.2% versus 7.3%) ($P < .001$). In 217 cases (58.96%; $P < .001$) only one radiologist indicated new appointments. Of this group, 73% were discharged in the first visit, compared to 47.6% in the non-discrepant group ($P < .001$). Four of the cancers were detected in these 217 patients.

Conclusions: The observed results are adjusted to the reference values. The discordant data are the new appointments rate, both in the initial screening and in the review, with a significant variation depending on the radiology reader.

© 2015 SECA. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

El cáncer de mama es la neoplasia más frecuente en la mujer, y la primera causa de muerte por cáncer en el sexo femenino¹. Para este año, 2015, se estima que constituirá del 24% al 28% del total de cánceres diagnosticados en la mujer^{2,3}.

Dado que se ha relacionado con factores de riesgo difícilmente modificables (edad, genéticos, hormonales...), los esfuerzos para disminuir sus efectos se centran en el diagnóstico precoz mediante mamografías periódicas, para lograr mejorar el pronóstico y aumentar la supervivencia con tratamientos menos agresivos, y conseguir así una mayor calidad de vida. De ahí el diseño de programas de detección precoz de cáncer de mama (PPDPCM). La disminución de la mortalidad por cáncer de mama conseguida por estos es el argumento fundamental para justificarlos, y claramente supera los posibles perjuicios (sobrediagnóstico, falsos positivos, etc.)⁴. En nuestra comunidad se inició en 1995⁵. Diferentes estudios ponen en evidencia una amplia variabilidad interobservador, tanto a la hora de interpretar las mamografías como de indicar recomendaciones de seguimiento⁶⁻⁸. Esto puede dar lugar tanto a falsos positivos, con rellamadas innecesarias que generan ansiedad en la mujer y aumento de costes, como a falsos negativos, afectando a la calidad de los programas. Esta variabilidad se ha atribuido a diferentes factores, unos relacionados con la mujer y otros con la experiencia y formación del radiólogo, así como con características técnicas de la mamografía⁹.

En este estudio nos proponemos realizar una auditoría de calidad del programa de detección precoz de cáncer de mama en nuestro centro, y comparar los resultados con los estándares de referencia publicados. Por otro lado, analizar la variabilidad interindividual, o variabilidad en la práctica médica (VPM), y su posible influencia en los resultados obtenidos.

Material y métodos

Estudio descriptivo y retrospectivo de las mamografías de cribado realizadas durante el periodo del 1 de mayo de 2010 al 30 de abril de 2011, en el Hospital de San Juan de Dios de Aljarafe, tanto en mujeres que acudían por primera vez (primer cribado) como en aquellas con controles sucesivos bianuales (cribado sucesivo). A todas ellas se les realizó un seguimiento clínico a los 2 años posteriores a la realización de la técnica, con revisión de la historia clínica y consulta al Registro de cáncer de Andalucía, con el fin de detectar la aparición de cáncer en este periodo. Se excluyeron las mujeres que durante el periodo de estudio se realizaron una mamografía de control a los 6 meses, indicada con anterioridad a la fecha de inicio del mismo.

Se realizaron en nuestro centro 3.878 mamografías de cribado a mujeres procedentes del programa de detección precoz de cáncer de mama en Andalucía⁵, dirigido a mujeres con edades comprendidas entre los 50 y los 69 años. La primera exploración se denominó primer cribado, y posteriormente se realizaron controles bianuales (cribado sucesivo). En cada ocasión se llevó a cabo una

Tabla 1 Categorización de los hallazgos mamográficos

Categoría Bi-rads	Descripción
0	Se necesitan estudios adicionales
1	Normal
2	Benigna
3	Probablemente benigna
4	Probablemente maligna
5	Maligna
6	Confirmación AP de malignidad

Fuente: American College of Radiology¹⁰.

mamografía con doble proyección que fue estudiada por 2 radiólogos distintos (doble lectura) de modo independiente (no consensado)⁵. Los hallazgos se categorizan en función de la nomenclatura descrita por el *American College of Radiology* y que se denomina *breast imaging-reporting and data system* (Bi-rads) y que se describe en la [tabla 1](#). A todas las mujeres con mamografías con categoría 0, 3, 4 o 5 se les deben realizar estudios complementarios (ecografía, proyección localizada, magnificación, biopsia). A este proceso se llama recitación o nueva citación. En nuestro centro esta consulta de recitación se lleva a cabo en su mayor parte por un grupo de 3 radiólogos, que son también los que realizan las exploraciones intervencionistas. En la definición recogida por el Bi-rads¹⁰ de examen de cribado positivo incluye aquellos con categorías 0, 4 y 5 y negativo los catalogados como 1, 2 y 3, aunque indica que esta última no se incluye en los exámenes de cribado que originan una recitación, sino solo en los controles subsiguientes. En el formulario a cumplimentar en nuestro programa, diseñado por la Consejería de Salud de Andalucía, no se indica la categoría 0 y sí la 3. Por lo tanto, y siguiendo las recomendaciones de la guía, todos los estudios en los que se indicó categoría 3 por la necesidad de completar estudio serían en realidad categoría 0, y así se han contabilizado. Denominamos grupo «no discrepante» a aquellas mamografías que fueron derivadas a la consulta de nueva citación por los 2 radiólogos lectores y grupo «discrepante» a aquellas en las que solo fue un radiólogo quien recomendó la re-citación.

Los indicadores valorados se recogen en la [tabla 2](#) y son los que se recomiendan por el Bi-rads 2003 (edición española de 2006)¹⁰ para una auditoría básica de calidad, que están basados en los publicados por la *Agency for health policy and researchs (Public Health Service US)*¹¹. Se ha ampliado el estudio, añadiendo otros indicadores referidos al tamaño y grado de invasión del tumor, tomados de la *European guidelines for quality assurance in breast cancer screening and diagnosis* (4.^a ed.)¹². Los resultados se compararon con los estándares publicados en la Guía europea¹² y los de la Agencia Americana¹¹.

Las exploraciones se realizaron en un equipo digital Senograph 2000 (*General Electric Medical Systems*, EE. UU.). En la realización de los informes de las mamografías participaron todos los radiólogos del Servicio de Radiología del Hospital San Juan de Dios de Aljarafe, un total de 9 profesionales, sin dedicación exclusiva a la patología mamaria que contaban entre 6 y 12 años de ejercicio profesional, con una media de 7,8 años. El número de mamografías informadas por cada radiólogo fue similar.

Tabla 2 Indicadores valorados y su comparación con los valores de referencia

Indicadores	Valores de referencia	Resultados obtenidos
Sensibilidad	> 80%	100%
Especificidad	> 75	91%
VPP1 ^a	5-10%	6,2%
VPP2 ^b	25-40%	55%
% biopsias	< 1%	1,16%
Tasa detección cáncer	6-10/1.000	6,18/1.000
Porcentaje de tumores invasivos	< 90%	83,33%
Tumor invasivo y ganglio negativo	Cribado inicial > 70%	Cribado inicial 50%
	Cribado subsiguiente > 75%	Cribado subsiguiente 81,25%
Tumor invasivo ≤ 10 mm	≥ 25%	31,25%
Tumor invasivo < 15 mm	50%	70%
% recitaciones	Cribado inicial 7%	Cribado inicial 16,11%
	Cribado subsiguiente 5%	Cribado subsiguiente 8,27%

^a VPP1: porcentaje de todos los exámenes de cribado (categorías 0, 4 y 5) que tienen un resultado AP de cáncer en el término de un año.

^b VPP2: porcentaje de todos los estudios de cribado recomendados para biopsia (categorías 4 y 5) que en el término de un año tuvieron diagnóstico AP de cáncer.

Valores de referencia basados en Basset et al.¹¹ y Perry et al.¹².

El análisis estadístico se realizó con el programa estadístico *Statistical Package for Social Sciences* (SPSS) V19.0. Para todas las variables cualitativas se calcularon los totales y las frecuencias (en porcentajes). Se realizó el test de Chi-cuadrado o el test de Fisher (si hubo más de un 20% de casillas con frecuencia esperada menor del 5% o al menos una casilla fue inferior al 1%). Para estudiar la concordancia entre los profesionales se calculó el índice Kappa.

Resultados

Del total de mujeres incluidas en el estudio 3.276 correspondían a cribados sucesivos y 602 eran mujeres que acudían por primera vez al programa (primer cribado). Trescientas sesenta y ocho (9,4%) fueron catalogadas como categoría Bi-rads 0, 4 o 5 y citadas de nuevo para completar estudio, 97 (16,1%) en cribado inicial y 271 (8,2%) en revisiones sucesivas. No hubo falsos negativos, ya que ninguna de las mujeres catalogadas como Bi-rads 1 o 2, tanto en lectura inicial como en la consulta de recitación, desarrolló cáncer en los 2 años siguientes.

Se indicaron 43 biopsias derivadas de los estudios complementarios en las recitaciones (1,1%), 39 en primera visita y 4 en la segunda. Este porcentaje es ligeramente superior

al valor estándar (<1%). Veinticuatro fueron diagnósticas de cáncer; 2 de los carcinomas se diagnosticaron por indicación de biopsia en la segunda visita de revisión. La relación resultado benigno/maligno en biopsia fue de 1:1,2.

Se obtuvo un valor predictivo positivo en estudios de cribado (VPP1) del 6,5%, y un VPP2 (biopsia recomendada) del 55%. La sensibilidad fue del 100% y la especificidad del 91%. En cuanto a la tasa de detección de cáncer fue del 6,1/1.000.

De los 24 tumores diagnosticados 20 fueron invasivos (83,3%), 4 en cribado inicial y 16 en el subsiguiente. De estos 20 en 15 pacientes no se apreció infiltración de ganglios axilares en el estudio anatomopatológico tras el acto quirúrgico. Dos correspondieron a mujeres en cribado inicial (50%) y 13 a cribado subsiguiente (81,2%). En 6 casos el tamaño del tumor fue menor o igual a 10 mm, correspondiendo a 1 de cribado inicial (25%) y 5 de subsiguiente (31,2%). En 17 ocasiones el tumor fue menor de 15 mm (70%).

Todos estos valores (tabla 2), excepto el porcentaje de recitación y el porcentaje de tumores invasivos sin infiltración metastásica ganglionar en cribado inicial, cumplieron los estándares. Dado la escasa desviación del porcentaje de biopsias indicadas se consideró este ítem dentro de la normalidad.

Para valorar la influencia de la VPM en el porcentaje de recitaciones, se comparó el número de las mismas por cada radiólogo, existiendo una proporción muy elevada en los radiólogos 5 y 7. Al realizarse doble lectura, una misma mujer pudo ser citada nuevamente por un solo radiólogo o por los 2, por lo que el número de recitaciones fue más elevado que el número de mujeres citadas para completar el estudio. Ajustando los datos al número absoluto de recitaciones (n=368), la media de estos radiólogos fue del 12% y del 17,2%, mientras que la del resto fue del 7,3% (p < 0,001) (fig. 1).

En 217 casos (58,96%) solo un radiólogo indicó la recitación (p < 0,001), denominando a este grupo «discrepante». Destacó el radiólogo número 5, con un porcentaje significativamente mayor a los demás, con el dato coincidente que también fue uno de los tenía un mayor porcentaje de recitación (p < 0,001) (fig. 1). Tras realizar los estudios complementarios, en el grupo «discrepante», 160 mujeres (73%) fueron dadas de alta en la primera visita frente a 72 mujeres (47,6%) en el grupo «no discrepantes» (p < 0,001). En cuanto a la categoría al alta, en el grupo «discrepante» a 135 mujeres (62%) se asignó la categoría Bi-rads 1 (normal), mientras que en el grupo «no discrepante» esta categoría se adjudicó a 52 mujeres (34,4%) (p < 0,001). Sin embargo, hay que resaltar que 4 de los cánceres se detectaron en el grupo «discrepante».

La correlación entre los radiólogos para la categoría Bi-rads 5 fue del 100%

Los resultados se resumen en la tabla 2.

Discusión

Como corroboran nuestros resultados, nos ajustamos de modo aceptable a los valores de referencia, excepto en el número de recitaciones, sobre todo en las mamografías de primer cribado, con una importante variabilidad en función del radiólogo lector. Si tomamos como referencia el estándar americano¹¹ nos situamos en el límite superior. Esta guía

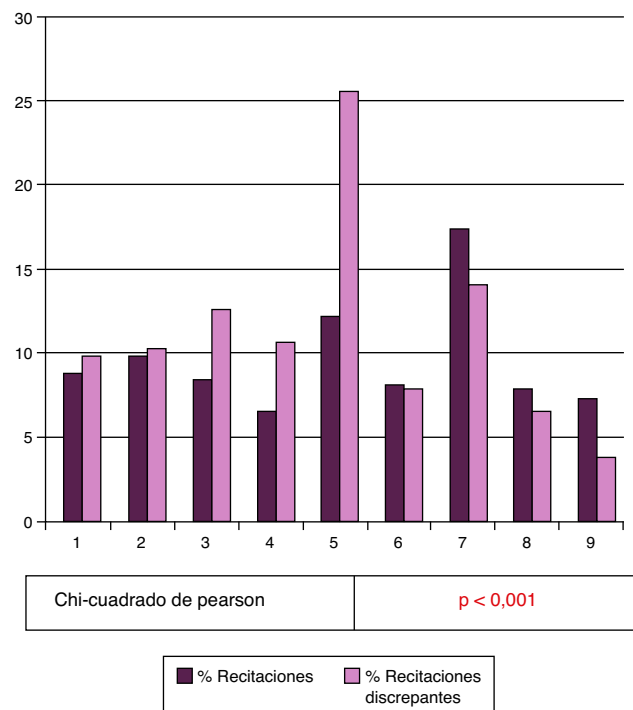


Figura 1 Porcentaje de recitaciones realizadas y discrepantes por cada radiólogo lector. Obsérvese el mayor porcentaje de recitaciones de los radiólogos 5 y 7 (p < 0,001) y el elevado número de recitaciones discrepantes del radiólogo 5 (p < 0,001).

considera que el valor deseable debe ser inferior al 10%, sin diferenciar entre primer cribado y cribado sucesivo. Aunque está descrita una mayor tasa de recitación en el primer cribado¹³, solo hemos encontrado un resultado similar en la literatura¹⁴, correspondiendo a un programa de cribado norteamericano. El dato obtenido es aún más remarcable si tenemos en cuenta que la tasa de cáncer de mama en esta franja de edad en nuestra área es significativamente menor que en el resto de la provincia, 140 casos/100.000 hab frente a una media de 231 casos/100.000 hab¹⁵.

Diferentes estudios han demostrado la variabilidad existente entre los radiólogos tanto en la interpretación de las mamografías como en las recomendaciones de seguimiento^{6,8,16-18}, con una mayor correlación en el caso de la categoría Bi-rads 5^{19,20}, como en nuestro caso. El aumento de las recitaciones que ocasiona puede afectar a la calidad del programa de cribado²¹.

En nuestro caso influye de modo muy negativo, ya que el porcentaje de recitaciones, sobre todo de primer cribado, es superior al estándar deseado, ocasionando problemas a la hora de gestionar las citas y aumentando el gasto⁹. Con estos resultados sería esperable un aumento del número de procedimientos intervencionistas. Sin embargo, nuestro porcentaje de biopsias está prácticamente dentro de la normalidad con un VPP2 por encima del estándar, lo que se traduce en que se indican un número adecuado de biopsias con alta especificidad. Esto, pensamos, va en consonancia con el hecho de que la consulta de recitación es llevada en su mayor parte por el mismo equipo de 3 radiólogos, que al poder hacer un seguimiento de los casos han podido beneficiarse de un *feedback* positivo. Esta hipótesis de mejora

del rendimiento en la detección de enfermedad mediante un proceso de retroalimentación por la autogestión de las recitaciones ha sido puesta de manifiesto también por otros autores²²⁻²⁴. Otro aspecto negativo del aumento de las recitaciones es la generación de ansiedad en las mujeres, en un porcentaje elevado de los casos innecesaria^{9,13}, que puede hacer disminuir la adhesión y la satisfacción con el programa. Si bien 4 de los cánceres fueron diagnosticados en el grupo «discrepante», coincidimos con Apesteguía et al.⁴ en que debería establecerse el mejor punto de encuentro para conseguir una buena sensibilidad sin comprometer excesivamente la especificidad; es decir, el número de falsos positivos, entendidos como recitaciones, debería ser tan bajo como sea posible sin reducir el porcentaje de detección de cáncer.

En general, existe consenso en determinar que a mayor especialización y años de la experiencia del radiólogo, en definitiva formación, es mayor es la especificidad^{7,22,25}, sin variar la sensibilidad²². Existe controversia en cuanto a la influencia del volumen de mamografías leídas, encontrando estudios tanto en uno como otro sentido en mamografía de cribado^{22,23,26} y diagnóstica²⁷.

Nuestro centro es un hospital comarcal. No hay radiólogos con dedicación exclusiva a la patología mamaria. La media de edad es de 39,5 años (34-48), y en su mayoría realizaron la formación MIR en el mismo centro hospitalario. No hay diferencias significativas en el número de mamografías estudiadas por cada uno de ellos. Por lo tanto, constituimos un grupo muy uniforme en el que la experiencia es similar. Debido al alto número de recitaciones discrepantes (58,9%, $p < 0,001$), y a las diferencias encontradas entre los diferentes radiólogos (12 y 17,2% frente a 7,3%; $p < 0,001$) pensamos que en nuestro centro la actitud depende en gran medida de características y actitudes muy personales de cada radiólogo lector. De hecho, la mayoría de las recitaciones en el grupo «discrepante» se catalogan de Bi-rads 1 en la primera y única visita ($p < 0,001$), y no hubo diagnóstico de cáncer en estas mujeres en los 2 años posteriores. La hipótesis que nos planteamos es si nos encontramos ante actitudes derivadas de la medicina defensiva. Esta tiene por único objeto buscar la protección ante una eventual reclamación/denuncia. Se observa en todos los campos de la medicina, integrándose a veces de modo inconsciente, y lleva a seguir procedimientos alejados de la evidencia científica y de las recomendaciones publicadas. Este punto entronca con el concepto general de VPM. Esta es una circunstancia inherente a la práctica médica. Podría considerarse un concepto positivo desde la perspectiva de la elección de una adecuada alternativa tras valorar en cada caso los riesgos y características individuales de la enfermedad y del paciente. Sin embargo, aunque la medicina debería basarse en práctica analítica y toma de decisión con conocimientos científicos demostrados, a veces se mueve en el terreno de la incertidumbre, constituyendo un acto interpretativo basado en experiencias propias. La hipótesis causal más extendida sugiere que la incertidumbre (entendida como la ausencia de evidencias científicas y/o consenso profesional sobre qué debe hacerse ante una situación concreta) está en la base de las VPM, aunque en otros casos, más que de incertidumbre estaríamos hablando de ignorancia (los médicos no conocen las evidencias o, aun conociéndolas, hacen caso omiso de ellas)²⁸.

En cuanto al otro dato discordante encontrado, la disminución del porcentaje de tumores invasivos sin infiltración metastásica ganglionar en estudios de cribado inicial, pensamos que el bajo número de casos de nuestra serie ($n = 4$) distorsiona los resultados.

Destacar la alta sensibilidad obtenida (100%), ya que no hubo falsos negativos, dato importante tratándose de una enfermedad grave y de gran impacto social.

Un PPDCM útil y bien concebido debe monitorizarse permanentemente desde el punto de vista epidemiológico⁴. En nuestra opinión, esto nos ha permitido identificar problemas y áreas de mejora, como paso previo a plantear soluciones. Así se mejorará la gestión de este proceso, disminuyendo el gasto y evitando la generación de ansiedad en las pacientes ante derivaciones innecesarias.

Finalmente, concluimos que nos ajustamos a los valores de referencia, excepto en la tasa de recitaciones, tanto en primer cribado como en sucesivos, con importante variabilidad en función del radiólogo lector. Teniendo en cuenta nuestros resultados, no hay que infravalorar la influencia de la VPM y el peso que la medicina defensiva pudiera tener en ella. El consenso general existente es que a mayor formación y experiencia de los radiólogos lectores, menor debe ser el porcentaje de recitaciones. Por esto apoyamos la existencia de unidades específicas de mama, para optimizar así la curva de aprendizaje. Si esto no fuera posible, sobre todo en hospitales pequeños, mejorar la formación de los radiólogos con cursos, y sobre todo con sesiones de revisión de las recitaciones con resultado final del estudio, fundamentales para un *feedback* positivo.

Conflicto de intereses

Los autores declaran que no existe conflicto de intereses.

Agradecimientos

A la Dra. M.^a Victoria Ruiz Romero, por su ayuda y colaboración en el tratamiento estadístico de los resultados.

A la Dra. Isabel Olea Comas por facilitar la recogida de datos.

Bibliografía

1. Instituto de Estadística y Cartografía de Andalucía. Consejería de Economía, Innovación, Ciencia y Empleo [consultado 8 Ene 2015]. Disponible en: <http://www.juntadeandalucia.es/institutodeestadisticaycartografia>.
2. Sociedad Española de Oncología Médica [consultado 8 Ene 2015]. Disponible en: <http://www.seom.org/>.
3. International Agency for research on cancer. World Health Organization. Globocan [consultado 8 Ene 2015]. Disponible en: http://globocan.iarc.fr/Pages/burden_sel.aspx.
4. Apesteguía Ciriza L, Pina Insausti LJ. Cribado poblacional de cáncer de mama. Certezas, controversias y perspectivas de futuro. *Radiología*. 2014;56:479-84.
5. Proceso asistencial integrado cáncer de mama. Detección precoz de cáncer de mama. 2.^a ed. Sevilla: Consejería de Salud; 2005.
6. Elmore JG, Wells CK, Howard DH, Feinstein AR. Variability in radiologists interpretation of mammograms. *N Engl J Med*. 1994;331:1493-9.

7. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists. Findings from a national sample. *Arch Intern Med.* 1996;156:209–13.
8. Carney PA, Elmore JG, Abraham LA, Gerrity MS, Hendrick RE, Taplin SH, et al. Radiologist uncertainty and the interpretation of screening. *Med Decis Making.* 2004;24:255–64.
9. Mellado Rodríguez M, Osa Labrador AM. Cribado de cáncer de mama. Estado actual. *Radiología.* 2013;55:305–14.
10. Reston VA. Breast Imaging reporting and Data System (BI-RADS®). 4th ed. 2003. Ed española. Madrid: American College of Radiology/Sociedad Española de Radiología Médica; 2006.
11. Basset LW, Hendrick RE, Bassford TL, Butler PF, Carter D, DeBor M. Clinical practice guideline n.º 13. AHCPR publication No. 95-0632. Quality determinants of mammography. Rockville, Md: Agency for health policy and researchs, Public Health Service, US Department of health and human services; 1994.
12. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. 4th ed. Luxembourg: Office for official Publications of the European communities; 2006.
13. Castells X, Molins E, Macià F. Cumulative false positive recall rate and association with participant related factors in a population based breast cancer screening programme. *J Epidemiol Community Health.* 2006;60:316–21.
14. Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false positive recall or biopsy recommendation after 10 years of screening mammography. *Ann Intern Med.* 2011;155:4181–92.
15. Registro del Cáncer de Andalucía. Secretaría General de Salud Pública y Participación de la Consejería de Salud de la Junta de Andalucía. Acceso a través del registro de cáncer de Granada en 30 enero de 2015. Disponible en: <http://www.cancergrada.org/es/>.
16. Skaane P, Engendal K, Skjennald A. Interobserver variation in the interpretation of breast imaging. *Acta Radiol.* 1997;38:497–502.
17. Baker JA, Kornguth PJ, Floyd CE. Breast imaging reporting and data system standardized mammography lexicon: Observer variability in lesion description. *AJR Am J Roentgenol.* 1996;166:773–8.
18. Ciccone G, Vineis P, Frigerio A, Segnan N. Interobserver and intra-observer variability of mammogram interpretation: A field study. *Eur J Cancer.* 1992;28A:1054–8.
19. Mendoza H-N, Cisneros L, Martin-Ramos J, Arango AJ. Bi-rads 3. ¿Realmente son hallazgos benignos? Variabilidad interobservador. *Ann Radiol Mex.* 2009;8:173–6.
20. Berg WA, Campassi C, Langerberg P, Sexton MJ. Breast imaging reporting and data system inter and intraobserver variability in feature analysis and final assessment. *AJR Am J Roentgenol.* 2000;174:1769–77.
21. McCann J, Stockton D, Godward S. Impact of false positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer Res.* 2002;4:R11.
22. Molins E, Macià F, Ferrer F, Maristany MT, Castells X. Association between radiologists' experience and accuracy in interpreting screening mammograms. *BMC Health Services Research.* 2008;8:91.
23. Beam CA, Conant EF, Sickles EA. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. *J Natl Cancer Inst.* 2003;95:282–90.
24. Buist DS, Anderson ML, Smith RA, Carney PA, Miglioretti DL, Monsees BS, et al. Effect of radiologists' diagnostic work-up volume on interpretive performance. *Radiology.* 2014;273:351–64.
25. Elmore JG, Wells CK, Howard DH. Does diagnostic accuracy in mammography depend on radiologist's experience. *J Womens Health.* 1998;7:443–9.
26. Théberge I, Chang SL, Vandal N, Daigle JM, Guertin MH, Peltier E, et al. Radiologist interpretive volume and breast cancer screening accuracy in a Canadian organized screening program. *J Natl Cancer Inst.* 2014;106: djt461.
27. Haneuse S, Buist DS, Miglioretti DL, Anderson ML, Carney PA, Onega T, et al. Mammographic interpretive volume and diagnostic mammogram interpretation performance in community practice. *Radiology.* 2012;262:69–79.
28. Peiró S, Bernal-Delgado E. Variaciones en la práctica médica: apoyando la hipótesis nula en tiempos revueltos. *Rev Esp Salud Pública.* 2012;86:213–7.