

Análisis de lecturas «falso negativo» en un programa de screening de mama

Félix Martín* • Luis Andreo** • Marta Redondo*** • Isabel Perona***

Programa de Prevención de Cáncer de Mama de la Comunidad Valenciana. Servicios de Radiodiagnóstico: *Hospital Lluís Alcanyís de Xàtiva, **Hospital Verge dels Lliris de Alcoi, ***Hospital General de Castelló.

Objetivo: Revisar los carcinomas incidentes (CI), analizando las lecturas que en una vuelta anterior habían sido consideradas negativas. Describir el tipo de imágenes radiológicas más frecuentemente desapercibidas. Comparar tamaño y estadios tumorales entre falsos negativos (FN) y verdaderos negativos (VN). Valorar la influencia de la densidad mamaria, la edad y otros factores en la aparición de FN.

Material y método: Seleccionamos 267 CI, y sometemos las mamografías previas a una lectura triple ciego. Ciento diecisiete de estos casos fueron mezclados para su lectura con mamografías normales, elegidas al azar, en relación 1/3.

Resultados: Se detectaron un total de 77 casos (28,8%): 25 por los tres lectores, 21 por dos y 31 por uno. Los 190 estudios restantes (71,16%) fueron considerados VN. El grado de detección fue mayor en los 150 casos leídos directamente (32%) que en los 117 casos mezclados con otros normales (24,7%). La imagen radiológica que más frecuentemente pasó inadvertida fue la asimetría, aunque varió según la unidad de origen. Entre los factores que indujeron al error de lectura destacamos diez casos que estudiados con proyecciones adicionales fueron considerados normales, y 11 casos tras dos años de seguimiento sin cambios. La densidad mamaria y la edad influyen pero no nos parecen determinantes. Respecto al estadio, tamaño y afectación ganglionar, no encontramos diferencias estadísticamente significativas entre los FN y VN, si bien la proporción de *in situ* es algo menor en el grupo de los FN.

Conclusiones: El estudio de FN es un parámetro necesario de control de calidad y autoevaluación de las lecturas. La incidencia de FN va a estar en función de las características del programa. Los objetivos de estadio y tamaño tumoral siguen cumpliéndose a pesar del diagnóstico más tardío en el grupo de los FN.

Palabras clave: Mamografía. Rendimiento diagnóstico. Incidencia. Cáncer de mama. Screening de mama.

Martín F, Andreo L, Redondo M, et al. Análisis de lecturas «falso negativo» en un programa de screening de mama. Radiología 2002;44(1):1-7.

Correspondencia:

FÉLIX MARTÍN DÍEZ. Servicio de Radiodiagnóstico. Hospital Lluís Alcanyís de Xàtiva. Ctra. Xàtiva-Silla, Km 2. 46800 Xàtiva.

Recibido: 10-VII-2001.

Aceptado: 3-I-2002.

Analysis of «false negative» readings in a breast screening program

Objective: To review a series of cases of incidental carcinomas (IC), analyzing the readings that had been considered negative in previous studies. To describe the radiological images that most often go undetected. To compare the tumor sizes and stages associated with false negatives and true negatives. To assess the influence of breast density, age and other factors on the occurrence of false negatives.

Material and methods: We selected 267 IC and subjected the previous mammographies to blind readings by three radiologists. For their reading, 117 of these cases were mixed in with randomly chosen normal mammographies, in a ratio of 1 to 3.

Results: A total of 77 cases (28.8%) of false negatives were detected: 25 by three radiologists, 21 by two and 31 by one. The remaining 190 studies (71.16%) were considered to be true negatives. The rate of detection was greater among the 150 cases read directly (32%) than in the 117 cases mixed with normal mammographies (24.7%). The radiological image that most frequently went undetected was asymmetry, although it varied depending on the unit used to film the studies. Among the factors leading to error in the readings, we point out the existence of ten cases studied with additional projections that were considered normal and 11 cases in which the lesions remained unchanged after two years of follow-up. Breast density and age had some influence but did not appear to be determinant. We found no statistically significant differences between false and true negatives in terms of stage, size and lymph node involvement, although the proportion of *in situ* lesions is somewhat lower among false negatives.

Conclusions: The study of false negatives is a necessary parameter for quality control and reading self-evaluation. The incidence of false negatives depends on the characteristics of the screening program. The objectives of tumor stage and size continue to be fulfilled, despite a later diagnosis in the false negative group.

Key words: Mammography. Diagnostic yield. Incidence. Breast cancer. Breast screening.

Los programas de prevención del cáncer de mama tienen una implantación cada vez mayor en nuestra sociedad.

Cada programa tiene sus características propias: población diana, sistema de lectura, doble proyección o no, intervalos de estudio, etc. y se marca unos objetivos a cumplir que garanticen su rentabilidad.

Los programas utilizamos un sistema de lectura habitualmente basado en el «Breast Imaging reporting and data system» (BI-RADS) publicado por el American College of Radiology (ACR)¹ y coincidimos básicamente en las categorías y en la actitud que van a generar.

Coincidimos por tanto en que las lecturas normales o benignas van a suponer el seguimiento normal, en el intervalo previamente definido, que suele ser a los dos años. El resto de categorías siempre van a determinar una detección y por tanto una nueva actitud o prueba diagnóstica. Y es en este punto, crucial para el diagnóstico precoz del cáncer de mama, donde mayor unificación de criterios debe existir, y tan sólo el error de interpretación, motivado o no por un defecto previo de técnica, va a determinar una lectura «falso negativo» que podría condicionar una detección más tardía, en estadios más avanzados.

Hasta ahora, se han realizado pocos estudios sobre la incidencia de FN en las lecturas de los programas de cribado. El objetivo de nuestro trabajo es, una vez terminada la tercera vuelta, analizar las mamografías previas de los CI, que por tanto habían sido consideradas en una vuelta anterior como lecturas normales o benignas, con el fin de determinar si ya existían hallazgos que podían haber sido detectados más precozmente (FN).

Intentaremos objetivar si se trata de un error de lectura o bien de signos mínimos, de difícil detección. Analizaremos la descripción de los hallazgos detectados y no considerados en su momento. Valoraremos la influencia de parámetros como la edad y densidad mamaria en la incidencia del FN e intentaremos conocer otros motivos que pudieron inducir, en algunos casos, a error de lectura.

Finalmente, compararemos las características biológicas de estos carcinomas cuando fueron detectados más tarde, con los casos en que no se detecta ningún hallazgo previo, es decir VN.

MATERIAL Y MÉTODO

El Programa de Prevención del Cáncer de Mama de la Comunidad Valenciana comienza en mayo de 1992 con la apertura, en una primera fase, de cinco unidades.

Desde entonces, se ha ido ampliando el número de unidades y actualmente, con 23 en funcionamiento, se encuentra cubierto todo el territorio de nuestra comunidad.

El programa está dirigido a mujeres entre 45 y 65 años, citadas por carta personalizada, según los datos de los padrones municipales, con intervalos de dos años.

Cada unidad funciona con un sistema de doble lectura independiente y consenso entre los dos lectores, con doble proyección en la primera vuelta y única, oblicua medio-lateral, en las sucesivas.

El sistema de lectura es muy similar al BIRADS¹ con alguna pequeña modificación. En nuestro programa se consideran siete descripciones de posibles hallazgos: 1 Nódulo/masa, 2 Desestructuración/estrella, 3 Calcificaciones, 4 Asimetría, 5 Alteración piel/pezón, 6 Combinadas/Múltiples, y 7 Otras, que van a dar lugar a seis posibles categorías: 1 No realizada, 2.1 Normal grasa, 2.2 Normal densa, 3 Benigna, 4 Probablemente Benigna, 5 Probablemente Maligna, y 6 Maligna.²

Los objetivos generales de nuestro programa se han visto cumplidos a lo largo de estos años, pero nos faltaba poder evidenciar un último parámetro de control de calidad como es el estudio de los FN en nuestras lecturas.

Nos proponemos hacer una valoración de cuatro de nuestras unidades tras, como mínimo, cinco años en funcionamiento y

TABLA 1
DATOS GENERALES DE ESTAS CUATRO UNIDADES

	Resultados		
	Objetivos	1.ª Vuelta	Sucesivas
Tasa Particip.	> 70%	78,33%	82,25%
% Proyec. Adic.	< 10%	7,86%	5,74%
% Citas Precoz	< 5%	5,13%	4,02%
% Ecos	< 5%	3,48%	2,25%
Tasa detección (1.ª V)	4-5%	4,6%	3,7%
VPP	> 50%		65,5%
Ca <i>in situ</i>	> 10%		16,5%
Sin Afect. Gangl.	> 70%		76,8%
Tamaño < 2 cm	> 50%		76%

que no hayan sufrido cambios importantes en sus primeros y segundos lectores. En la tabla 1 exponemos los objetivos del programa³ y los resultados de estas cuatro unidades, diferenciando en alguno de ellos los que han sido obtenidos en la primera vuelta y en la segunda o sucesivas.

El estudio se inició el uno de mayo de 1999, habiendo sido estudiadas hasta esta fecha 64.110 mujeres. Se realizaron un total de 193.531 mamografías, detectándose carcinoma en 734 mujeres, siendo 432 carcinomas prevalentes, es decir detectados en mujeres que acudían por primera vez al programa (tasa de detección del 4,6% en primera vuelta), y 302 CI, es decir, detectados en segunda vuelta o sucesivas (tasa de detección del 3,71%).

De los CI, seleccionamos 267 que tenían completos todos los datos que pretendemos analizar.

El estudio previo de cada unidad participante, considerado normal o benigno, es remitido para que sea informado según nuestro protocolo a los segundos lectores de las otras tres unidades que, por supuesto, no conocen los casos. Cuando la exploración objeto de estudio es de segunda vuelta o posterior, adjuntamos la mamografía anterior para que se tenga la posibilidad de comparar, si se considera necesario. Si hubiera alguna proyección adicional, también se adjunta, reproduciendo exactamente la situación en que fueron leídas las mamografías en su momento. Naturalmente, cada lector desconoce el informe de los otros, no existe la posibilidad de exploración clínica, ni acceso a la historia, tratándose de una lectura a triple ciego mamográfica pura.

De los 267 casos, 172 eran mamografías de primera vuelta con doble proyección, y 95 eran de segunda o tercera vuelta con proyección única oblicua medio-lateral pero disponiendo de estudios previos para comparar.

Como pensamos que podía existir cierto componente subjetivo en los lectores a la hora de buscar imágenes sospechosas, 150 de los casos (96 de primera vuelta y 54 de segunda) fueron leídos de forma directa, sabiendo los lectores que en todos ellos se iba a desarrollar un carcinoma y los otros 117 casos (76 de primera vuelta y 41 de segunda) fueron leídos mezclados aleatoriamente con 351 mamografías que serán normales (relación 1/3), elegidas al azar, del archivo de cada unidad, guardando la misma proporción de mamografías de primera y segunda vuelta. La finalidad es doble: aproximarnos más a una situación habitual de lectura, y poder cuantificar, con la valoración que se hace de estos casos normales, el grado de condicionamiento de cada lector.

TABLA 2
N.º DE LECTORES QUE DETECTAN ALGÚN HALLAZGO PARA ESTUDIAR

	<i>L. Directa</i> <i>n</i> = 150	<i>L. Aleatoria</i> <i>n</i> = 117	<i>1.ª Vuelta</i> <i>n</i> = 172	<i>2.ª y suces.</i> <i>n</i> = 95	<i>TOTAL</i> <i>N</i> = 267
Por 3	16 (10,6)	9 (7,7)	20 (11,6)	5 (5,2)	25 (9,36)
Por 2	12 (8)	9 (7,7)	16 (9,3)	5 (5,2)	21 (7,86)
Por 1	20 (13,3)	11 (9,4)	22 (12,7)	9 (9,4)	31 (11,61)
No detectado	102 (68)	88 (75,2)	114 (66,2)	76 (80)	190 (71,16)

En su informe, los radiólogos reflejan: descripción del hallazgo, localización en la mamografía, categoría en que lo incluirían y actitud que tomarían en primer lugar, con independencia de que a raíz de ella se recurriera o no a nuevas pruebas complementarias.

En el análisis de los resultados, consideraremos como FN los casos que fueron correctamente detectados por alguno de los lectores. Los clasificaremos a su vez como «error de lectura» si fueron detectados por los tres, y «hallazgos mínimos» o de «difícil detección» si fueron detectados por uno o dos lectores. Los casos que no fueron detectados son considerados VN.

Estos VN fueron revisados retrospectivamente en cada unidad de origen, comparando con las mamografías en que se diagnosticó el carcinoma. El objetivo es determinar en cuantos casos, sólo conociendo la localización y características del carcinoma es posible detectar hallazgos radiológicos previos, si por tanto el estudio de los FN podría realizarse con este sistema más sencillo, si estos hallazgos coinciden con los encontrados en los FN, y valorar si la densidad mamaria es un factor influyente o no.

Simplificando los patrones mamográficos establecidos originalmente por Wolfe⁴ y posteriormente empleados por el BI-RADS, en nuestro programa consideramos como «grasa» a la mama de fácil lectura y «densa» a aquella cuya estructura puede dificultar la detección de imágenes sospechosas.

Para analizar la posible influencia de la densidad mamaria en la detección de alteraciones, examinamos en cada uno de los 267 CI objeto de estudio, si previamente habían sido incluidos en la categoría normal grasa o normal densa, y los compararemos con la densidad mamaria considerada en los 467 carcinomas restantes diagnosticados en el programa y con el total de 64.110 mujeres que han sido estudiadas en estas unidades.

Como la valoración de grasa o densa es poco objetivo, hemos comparado también, por tramos de edad, el grupo de los CI (en el momento de las mamografías objeto de estudio) con el resto de los carcinomas.

Intentamos determinar las causas que pudieron inducir a un error de lectura revisando todos los casos que podrían haber sido detectados: defectos de técnica, proyecciones adicionales completamente normales, seguimiento de más de dos años sin cambios, y multifocalidad o multicentricidad.

Finalmente, comparamos el estadio, tamaño tumoral y afectación ganglionar entre FN y VN. El análisis de las frecuencias relativas de estas características biológicas se llevó a cabo mediante una prueba de Ji cuadrado al ser mayor de cinco el número de casos esperados en todas las comparaciones realizadas. Se tomó como nivel de significación una $p < 0,05$.

Al comparar el tamaño tumoral, reflejaremos también los menores de 1,5 cm porque según los estudios de Tabar and cols.,

hasta este tamaño, la supervivencia a los 12 años es de alrededor del 90%, con independencia de que exista o no afectación ganglionar.⁵

RESULTADOS

Se valoran un total de 267 casos. De los 150 casos leídos directamente se detectó correctamente algún hallazgo en 48 ocasiones (32%), y de los 117 casos mezclados aleatoriamente con otros normales se detectó algún hallazgo en 29 ocasiones (24,7%). Total 77 casos (tabla 2).

Entre las 172 mamografías correspondientes a la primera vuelta, se valoró correctamente algún hallazgo en 58 ocasiones (33,7%), y de las 95 mamografías correspondientes a la segunda vuelta o sucesivas, se observó algún hallazgo en 19 ocasiones (20%).

Veintiuno de los casos correspondieron a carcinomas de intervalo que acudieron a la unidad, siendo correctamente valorado algún hallazgo por alguno de los lectores en cuatro (19%).

Del total de 267 casos, 25 fueron detectados por los tres lectores (9,36%), por lo que los consideraremos como error de lectura. 21 casos (7,86%) fueron detectados por dos lectores, y 31 (11,61%) por un solo lector, en total 52 casos (19,47%) que fueron detectados por uno o dos lectores y que consideramos como hallazgos mínimos o de difícil detección. Los 190 estudios restantes (71,16%) serán considerados como VN (tabla 2).

La descripción de los hallazgos detectados en cada unidad A, B, C y D, son los reflejados en la tabla 3. En ella podemos observar que globalmente, la descripción más frecuentemente infravalorada fue la asimetría. Pero en cambio, en la unidad A, el hallazgo más frecuentemente subestimado fueron las microcalcificaciones, y en la unidad B las asimetrías y desestructuraciones.

De los 77 hallazgos correctamente detectados, 56 (72,7%) fueron incluidos en la categoría *probablemente benigna*, y 21

TABLA 3
DESCRIPCIÓN DE LOS HALLAZGOS POR UNIDADES

<i>Unidad</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>TOTAL</i>	<i>%</i>
Nódulo	3	2	2	3	10	13
Desestructuración	0	5	0	0	5	6,5
Microcalcificaciones	10	5	1	4	20	26
Asimetría	8	12	5	9	34	44,1
Alteración piel	0	0	0	0	0	0
Combinadas	4	2	0	2	8	10,4
					77	100

TABLA 4

ANÁLISIS DE LECTORES

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
% Casos detectados	17,3	15,6	18,6	23,5
% Actuación en normales	14,9	11,4	15,1	24,3

(27,2%) como *probablemente malignas*, teniendo en cuenta que consideramos siempre la mayor de ellas tan sólo con que uno de los lectores la clasifique como tal.

En la tabla 4 valoramos el grado de actuación de cada lector (*a*, *b*, *c*, y *d*). En la primera línea reflejamos el porcentaje de casos correctamente detectados por cada uno de ellos. En la segunda línea, la valoración que hacen de los casos normales que se extrajeron del programa y que fueron mezclados con los casos reales objeto del estudio.

Comparando los 190 casos no detectados por ningún lector (VN), con las mamografías del carcinoma ya diagnosticado, nos encontramos con que en 58 de ellos (30,5%) se observaba ya alguna imagen que daría lugar al posterior carcinoma. A este grupo de carcinomas incidentes los consideraremos como potencialmente «detectables» (tabla 5). Llama nuestra atención que en este caso son más frecuentes los nódulos que las calcificaciones respecto a los hallazgos radiológicos encontrados en los FN. Los 15 nódulos eran todos ellos menores de cinco milímetros, las calcificaciones eran en su mayoría una o dos microcalcificaciones, y la combinada era de asimetría, de muy difícil identificación, con calcificación.

Los 132 casos restantes de nuestro estudio consideramos que eran absolutamente «indetectables».

En la tabla 6 comparamos la densidad mamaria considerada en cada uno de los tres grupos en que hemos dividido estos 267 CI objeto de estudio, con la de los 467 carcinomas restantes diagnosticados en el programa y con el total de 64.110 mujeres que han sido estudiadas en estas unidades (tabla 6).

Podemos observar que en los CI, tanto los FN (primera columna) como los VN (dos columnas siguientes), tenían la mitad de las mamas consideradas grasas y la otra mitad densas, prácticamente sin diferencias entre los tres grupos. En cambio, el resto de los carcinomas detectados en el programa, no incluidos en el estudio, asentaron sobre mamas grasas en un 67,4% y sobre mamas densas en un 32,5% de los casos, resultados casi idénticos al analizar la densidad mamaria considerada en el total de las mujeres que acudieron al programa.

Comparando por tramos de edad estos carcinomas observamos resultados similares: apenas hay diferencias entre los tres

TABLA 5

ANÁLISIS RETROSPPECTIVO DE LOS VN. HALLAZGOS «DETECTABLES»

Descripción	<i>n</i> = 58	%
Nódulo	15	25,8
Desestructuración	4	6,9
Microcalcificaciones	8	13,7
Asimetría	30	51,7
Alteraciones piel	0	0
Combinadas	1	1,8

grupos de CI (FN y VN), pero sí de estos con el resto de los carcinomas que en un 47,5% de los casos se detectan en mujeres con más de 60 años (tabla 7).

Hemos analizado entre los 77 casos detectados y los 58 que no fueron detectados por ningún lector, pero que sí presentaban ya algún hallazgo previo, las causas que creemos pudieron inducir a un error de lectura (tabla 8).

Los cuatro casos considerados como defecto de técnica fueron: dos lesiones no incluidas en la mamografía, una pequeña asimetría en pliegue inframamario no desplegado, y unas microcalcificaciones en una mama densa con técnica demasiado blanda.

Hubo cinco casos de primera vuelta con asimetrías en una proyección pero totalmente normales en la otra, que fueron considerados normales. Y lo mismo ocurrió en diez casos en que tras realizarse proyecciones adicionales, bien la cráneo-caudal (dos casos) porque eran de segunda vuelta, o bien una localizada con compresión (ocho casos), se consideró que la mamografía era *normal* o *benigna* y en cambio allí se desarrolló un carcinoma.

Cinco casos de asimetría, cuatro de microcalcificaciones y dos lesiones estrelladas con muy bajo grado de sospecha, fueron sometidas a seguimiento cada seis meses durante al menos dos años, considerándose finalmente como benignas. Se pasaron a cita normal a los dos años, pero en estudios posteriores se desarrolló un carcinoma sobre esas imágenes aparentemente estables.

En la tabla 9 hacemos un estudio de las características biológicas de todos los CI en el momento de su diagnóstico y tratamiento, comparando los 77 casos que tuvieron hallazgos «detectados» por algún lector (FN), con los 190 VN.

Podemos observar en el grupo de los FN una menor proporción de carcinomas *in situ* respecto a los VN (10,3% frente a 17,8%), diferencia que si bien clínicamente parece relevante, no llega a ser estadísticamente significativa ($p = 0,13$) al no ser una muestra muy numerosa.

TABLA 6
ANÁLISIS COMPARATIVO POR DENSIDAD MAMARIA

	<i>Carcinomas incidentes</i>			<i>Resto</i> <i>Carcinomas</i> <i>n</i> = 467	<i>Mujeres</i> <i>Programa</i> <i>n</i> = 64.110
	<i>Detectados</i> <i>n</i> = 77	<i>Detectables</i> <i>n</i> = 58	<i>Indetectables</i> <i>n</i> = 132		
Grasa	50,6	48,2	48,4	67,4	66,6
Densa	49,3	51,7	51,1	32,5	33,3

TABLA 7
ANÁLISIS COMPARATIVO POR EDAD

	Cáncer incidentes			Resto Cáncer n = 467
	Detectados n = 77	Detectables n = 58	Indetectables n = 132	
Edad				
45-49	23,3	24,1	26,5	17,3
50-54	23,3	22,4	20,4	15,6
55-59	22,0	27,5	19,6	19,4
Más de 60	31,1	25,8	33,3	47,5

Pero también podemos observar que sumadas la proporción de *in situ* y estadio I se obtiene resultados muy similares en ambos grupos (66,2% y 68,4%), con un $p = 0,73$. Es decir que de alguna forma el menor número de *in situ* se compensa con un mayor porcentaje de estadios I.

Algo similar observamos que ocurre con el tamaño tumoral, donde el menor porcentaje de *in situ* en los FN es también compensado por un mayor porcentaje de tumores inferiores a 1,5 cm, de muy buen pronóstico, con independencia de que haya o no afectación ganglionar.⁵

La proporción de tumores mayores de 2 cm es mayor en el grupo de los FN, pero tampoco es estadísticamente significativa ($p = 0,3$).

Tampoco hay diferencias significativas respecto a la afectación ganglionar ($p = 0,72$).

De los 267 cánceres objeto de estudio, 21 eran cánceres de intervalo que se diagnosticaron dentro del programa. De ellos, cuatro (19%) fueron detectados, uno por tres lectores, y tres por un lector. Dos eran de primera vuelta, y dos de segunda. En el momento del diagnóstico el tamaño era de 1,4, 1,5, 2 y 3 cm; dos en un estadio I, uno en IIA, y uno en IIB; dos tenían afectación ganglionar y otros dos no; los cuatro fueron sometidos a mastectomía radical modificada, pero quizás lo más llamativo fue que dos de ellos tenían un grado de diferenciación III y dos tenían un grado II.

DISCUSIÓN

El estudio de los FN es un parámetro más de control de calidad.⁶ La tasa de FN estará en función de las características de los programas: población a estudiar, edad, intervalo entre estudios, si es screening puro o a demanda, si se realiza doble lectura, una o dos proyecciones, exploración clínica sistemáticamente, a veces, o en ningún caso, y sobre todo de los objetivos previamente fijados por el programa.^{3,7}

TABLA 8
CAUSAS DE «NO DETECCIÓN»

Defecto de técnica	4
Sólo se detectaba en una proyección	5
Al hacer complementarias, confunden	10
Seguimiento de dos años, sin cambios	11
Aparece sobre lesión benigna	1
Se detecta un foco y había más	4

La fijación de objetivos en un programa de screening se realiza teniendo en cuenta todos los factores coste-beneficio, incluyendo los efectos adversos que podrían producir un grado de actuación excesivo. Los programas que persiguen una mayor precocidad en el diagnóstico van a producir más efectos adversos. Es necesario encontrar un equilibrio entre ambos, buscando una disminución de la mortalidad con los menores efectos adversos posibles.^{8,9,10}

Los objetivos del programa van a condicionar por tanto la agresividad de las lecturas, y el grado de actuación sobre los diferentes niveles de sospecha, lo que se verá reflejado en los resultados: tamaño, estadios y afectación ganglionar,^{11,12} pero también en el número de exploraciones complementarias y biopsias quirúrgicas realizadas.

En nuestro medio, y probablemente condicionado por el menor número de reclamaciones médicas, hay un porcentaje mayor de biopsias quirúrgicas positivas (VP), y por tanto, un mayor número de FN respecto a otros estudios de este tipo realizados en EE.UU., a pesar de que, en nuestro programa, se realiza sistemáticamente doble lectura lo que según los estudios de Beam et al incrementa los verdaderos positivos, pero también los FN, aunque siempre en función de la experiencia de los dos lectores.¹³

Otro factor, no menos importante a tener en cuenta en el planteamiento de objetivos es el coste económico del programa. En nuestro caso, la gratuidad absoluta de todo el proceso, tanto diagnóstico como terapéutico, probablemente haría difícil sostener un programa con un valor predictivo positivo para las biopsias quirúrgicas inferior al 50%.

TABLA 9
ANÁLISIS COMPARATIVO DE LAS CARACTERÍSTICAS BIOLÓGICAS ENTRE LOS FN Y VN

		FN n = 77	VN n = 190
Estadio	<i>in situ</i>	8 (10,3)	34 (17,8)
	I	43 (55,8)	96 (50,5)
	IIA	10 (12,9)	37 (19,4)
	IIB	12 (15,6)	19 (10,0)
	III (o más)	4 (5,1)	4 (2,1)
	<i>in situ</i>	8 (10,3)	34 (17,8)
Tamaño	< 1,5 cm	45 (58,3)	103 (54,2)
	1,5-2 cm	6 (7,8)	20 (10,5)
	> 2 cm	18 (23,3)	33 (17,3)
	Afectación ganglionar	19 (24,6)	43 (22,6)

El estudio de los FN es más complicado de realizar que cualquier otro parámetro de control de calidad, y lo primero que nos debemos plantear es ¿cómo hacerlo?

Sobre la metodología:

Puesto que una gran parte de las mamografías presentan alguna imagen sobre la que fijaríamos nuestra atención si nos dijeron que allí hay un carcinoma, creemos que lo más razonable es hacer un estudio ciego pero no por uno, sino por varios lectores, que deben ser, por supuesto, expertos en lectura mamográfica. Deben ser varios lectores porque cada uno tiene un nivel psico-visual diferente, incluso él mismo lo tiene diferente en función del momento en que efectúe la lectura como hemos observado en nuestro estudio.^{7,14} Por esta razón, en alguno de los trabajos revisados, no se consideran FN los hallazgos que son detectados por un solo lector.¹⁵

En nuestro trabajo, uno de los lectores tuvo un grado de actuación muy elevado (24,3%) en las mamografías normales, condicionando obviamente una mayor detección en los casos reales (23,5%). La realización de doble o incluso tercera lectura en los programas de screening, naturalmente incrementa la detección, pero también los falsos positivos.

Entre los lectores hemos constatado pequeñas diferencias en lo que supone la primera actitud en el seguimiento de estudio. Unos comenzarían más frecuentemente por hacer proyecciones adicionales, otros por ecografía, y otros con citas precoces. Hemos de insistir en que se trata de una primera actitud, de la que muy probablemente pueden derivarse otras actuaciones complementarias: ecografía, cita precoz, punción, biopsia quirúrgica, etc.

En cualquier caso, lo que nos parece más interesante de que sean varios lectores, es que se puede cuantificar el grado de error en la «no detección» de un hallazgo previo, en función de cuantos de ellos lo detectan.

Consideramos que deberían incluirse entre estos casos objeto de estudio mamografías que se ha comprobado serán normales, para disminuir el nivel de condicionamiento psico-visual y lograr valoraciones más objetivas. En nuestro caso, observamos que se produce una mayor detección cuando el lector sabe que está leyendo solamente casos que serán carcinomas. Analizados los casos normales, elegidos al azar, observamos un grado de actuación de los lectores del 16,8% mientras que normalmente, en las lecturas de sus unidades es del 9,2%.

Sobre los resultados:

En nuestro estudio, son detectados correctamente un total de 77 hallazgos sospechosos por alguno de los lectores (FN = 28,8%) cuando se desconoce la localización y características del carcinoma. Retrospectivamente es posible detectar hallazgos previos en 58 de los 190 casos restantes cuando son leídos conociendo estos datos (30,5%). Por tanto, en el mejor de los casos, se podría detectar algún hallazgo previo en 135 casos (50,56%). Estas conclusiones coinciden con las proporcionadas en estudios similares, que consideran que siempre hay aproximadamente un 50% de los casos totalmente negativos.^{15,16}

Observamos que en las mamografías analizadas de primera vuelta se detectaron más FN que en las sucesivas (33,7% frente al 20%). Consideramos que el hecho de que los lectores, en su unidad, dispongan de estudios previos para comparar condiciona menos errores de lectura, porque aunque es cierto que el lector podía tener menos experiencia en lectura de screening en la pri-

mera vuelta (no en informe mamográfico convencional), también es cierto que disponía de doble proyección y el porcentaje de solicitud de nuevos estudios era mayor en esa primera vuelta como podemos observar en la tabla 1.

Aunque sí existe cierta tendencia a ser más frecuente el FN en la mama densa, consecuencia lógica de que los CI globalmente aparecen todos ellos sobre mamas más densas, en nuestro trabajo no parece tan determinante como en otros la densidad mamaria, aunque naturalmente no podemos objetivar lo que un estudio y otro considera como mama grasa o densa.^{12,17}

Otros autores evidencian que las mujeres con FN son entre siete y ocho años más jóvenes.¹⁸ Nosotros constatamos que en nuestro programa hay unas tasas de detección más bajas de 45 a 59 años y en cambio más altas a partir de los 60. Estas diferencias se hacen más ostensibles cuando comparamos los carcinomas incidentes, entre ellos los FN, con el resto de los carcinomas detectados en el programa, si bien es cierto, que la incidencia de cáncer de mama también se incrementa con la edad. Creemos que el estudio más minucioso en los casos de mamas más densas o mujeres más jóvenes evitará FN, tendiendo a igualar la diferencia de detección que las diferentes investigaciones demuestran.

En algunos trabajos se considera FN solamente a los casos en que aparece un carcinoma durante el intervalo.¹² Creemos que todo CI diagnosticado en un programa de screening debe considerarse potencialmente susceptible de haber sido detectado en la vuelta anterior, con independencia del momento en que haya sido diagnosticado y de los intervalos con que el programa efectúe las revisiones.

Las características de los hallazgos detectados son difíciles de comparar con otros programas que han utilizado el sistema de lectura BIRADS ya que existen matices que nos diferencian. Quizás lo más significativo es que nuestro programa no diferencia entre nódulo bien o mal definido, solapándose este último con nuestra definición de la asimetría. Teniendo en cuenta esta salvedad, nuestros resultados pueden ser similares a otros estudios.^{11,15,16,19}

Como hemos observado en la tabla 3 la descripción radiológica de los FN ha variado en las diferentes unidades. Estos datos son de gran valor para corregir en un futuro las actuaciones individuales de cada unidad.

Entre las causas de no detección, no hemos podido distinguir las que fueron por considerar una imagen ya presente como benigna, puesto que a veces, una o dos calcificaciones aisladas, un pequeño nódulo o asimetría poco denso y de apenas dos o tres milímetros, los podemos haber considerado en la lectura previa indistintamente como lectura normal o benigna.

Comparando estadio, tamaño tumoral y afectación ganglionar entre el grupo de FN y VN no hemos encontrado diferencias significativas, coincidiendo con otros trabajos realizados en este sentido.¹⁸ En nuestro caso, hemos encontrado un menor porcentaje de *in situ* en el grupo de los FN, pero que no llega a ser estadísticamente significativo, y clínicamente parece que se compensa con una mayor proporción de tumores menores de 1,5 cm y de estadios I en este grupo de FN.

Es importante resaltar que los objetivos del programa siguen cumpliéndose a pesar del diagnóstico más tardío en el grupo de los FN.

Actuaciones mas agresivas, con mayor numero de biopsias

quirúrgicas van a incrementar el numero de detecciones y por tanto una disminución de los FN, pero por el contrario, se producirá un notable incremento de los falsos positivos con mayor morbilidad y encarecimiento del programa.

Creemos que debemos dirigir nuestros esfuerzos a optimar todo lo posible los procedimientos prequirúrgicos, que mejoren la detección sin necesidad de incrementar la tasa de biopsias quirúrgicas.

CONCLUSIONES

El control de calidad de un programa de screening de mama debe incluir no sólo los parámetros más convencionales y fáciles de obtener, entre ellos los verdaderos y falsos positivos, sino que además son necesarios estudios de los FN que, a título personal, nos ayudarán a conocer y reflexionar sobre aquellos hallazgos que nos pasan más frecuentemente desapercibidos, y cuales fueron los motivos.

Creemos que el estudio debe ser ciego, realizado por varios expertos en lectura de screening, al menos tres, contrastando sus detecciones para poder objetivar si se trata de un error de lectura, de hallazgos mínimos o de verdaderos negativos. Creemos también que para ser más objetivos, deberían mezclarse con casos normales elegidos al azar que además nos permitirán valorar el grado de condicionamiento de los lectores.

La incidencia de FN va a estar en función de las características del programa y de los objetivos fijados. Son más frecuentes en mujeres más jóvenes y con mamas más densas, pero no nos parecen tan determinantes.

Podremos valorar las características biológicas de estos carcinomas cuando son diagnosticados, y compararlos con los verdaderos negativos y los generales del programa. En nuestro caso, hemos podido comprobar que se siguen cumpliendo los objetivos.

BIBLIOGRAFÍA

1. American College of Radiology (ACR). Breast Imaginng reporting and data system (BIRADS™). Second Edition. Reston [VA]: American College of Radiology; 1995.
2. Consellería de Sanitat i Consum. Programa de Prevención de cáncer de mama de la Comunidad Valenciana (Monografies Sanitaries serie E, n.º 25) Valencia 1998.
3. Kirkpatrick A, Törnberg S, Thijssen MAO. European Guidelines for quality assurance in mammography screening. Bruselas: Comision of the European Communities, 1993.
4. Wolfe JN. A study of breast parenchima by mammography in the normal woman and those with benign and malignant disease of the breast. Radiology 1976;89:201-5.
5. Tabar L, Fagerberg G, Day N. Breast cancer treatment and natural history: new insights from results of screening. Lancet 1992; 339:412-4.
6. Murphy WA, Destouet JM, Monses BS. Professional quality assurance for mammography screening programs. Radiology 1999; 175:319-20.
7. Kopans DB. Mammography screening for breast cancer. Cancer 1993;72:1809-12.
8. Ong G, Austoker J, Brett J. Breast screening: adverse psychological consequences one month after placing women on early recall because of a diagnostic uncertainty. A multicentre study. J Med Screen 1997;4(3):158-68.
9. Elmore JG, Barton MB, Moceri VM. Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med 1998 Apr 16;338(16):1089-96.
10. Mattsson A, Leitz W, Rutqvist LE. Radiation risk and mammographic screening of women from 40 to 49 years of age: effect on breast cancer rates and years of life. Br J Cancer 2000 Jan;82(1):220-6.
11. Martin J, Moskowitz M, Milbrath JR. Breast cancers missed by mammography. AJR 1979;132:737-9.
12. Sickles EA, Ominsky SH, Solitto RA. Medical audit of a rapid- throughput mammography screening practice: methodology and results of 27,114 examinations. Radiology 1990;175:323-7.
13. Beam CA, Sullivan DC, Layde PM. Effect of human variability on independent double reading in screening mammography. Acad Radiol 1996;3:891-7.
14. Duncan AA, Wallis MG. Classifying interval cancers. Clinical Radiology 1995;50:774-7.
15. Duncan KA, Needham G, Gilbert FJ. Incident round cancers: What lessons can we learn? Clinical Radiology 1998;53:29-32.
16. Daly CA, Apthorp L, Field S. Second round cancers: How many were visible on the first round of the National Breast Screening Programme, Three years earlier? Clinical Radiology 1998;53:25-8.
17. Baines CJ, Vidmar M, McKeown-Eyssen G. Impact of menstrual phase on false-negative mammograms in the Canadian National Breast Screening study. Cancer 1997;80:720-4.
18. Chew SB, Hughes M, Kennedy C. Mammographically negative breast cancer at the Strathfield Breast Center. Aust NZ J Surg 1996; 66:134-7.
19. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancer missed at