

Validez y reproducibilidad de la interpretación de la mamografía por radiólogos mexicanos, mediante el sistema BI-RADS

Gabriela Torres-Mejía,* Yolanda Villaseñor-Navarro,** Elsa Yunes-Díaz,*
Angélica Ángeles-Llerenas,* Olga Georgina Martínez-Montañez,**^{***} Eduardo Lazcano-Ponce*

* Centro de Investigación en Salud Poblacional, Instituto Nacional de Salud Pública.

** Instituto Nacional de Cancerología. *** Centro Nacional de Equidad de Género y Salud Reproductiva de la Secretaría de Salud.

Validity and reliability of mammographic interpretation by Mexican radiologists, using the BI-RADS system

ABSTRACT

Objective. Mammographic interpretation using the BI-RADS system determines if a patient will continue in the regular screening program or needs additional studies for breast cancer diagnosis, hence the importance of its evaluation. Sensitivity, specificity, diagnostic accuracy of mammographic interpretation and inter- and intra-observer variation, were assessed in a random sample of 29 radiologists from the Mexican Ministry of Health. **Material and methods.** Eighty diagnostic and screening mammographic studies were selected from the National Cancer Institute in Mexico City. Under the same conditions, two interpretations using the BI-RADS system were performed by each of the participating radiologists. **Results.** Sensitivity in the first and second measurement was 72 and 74% respectively, while the specificity was 80% for the first and 82% for second. Multivariate analysis showed a positive association between the last year number of interpreted mammograms, the number of courses taken and diagnostic accuracy ($\beta = 0.00002$, 95% CI 0.000004, 0.00003, $p = 0.02$, $\beta = 0.005$, CI 95% -0.0004, 0.01, $p = 0.07$, respectively). Inter-observer agreement was low in both measurements ($\kappa = 0.24$, $\kappa = 0.25$, respectively). However, after combining BI-RADS categories in 2, agreement increased to 0.46 for both measurements. The intra-observer agreement was good ($\kappa = 0.67$). **Conclusions.** The number of mammographic studies interpreted the previous year is a good measure of the experience of the radiologist and is reflected in the diagnostic accuracy of mammographic interpretation. Courses on mammographic interpretation also have a positive impact on performance.

Key words. BI-RADS. Reliability. Breast cancer. Sensitivity. Specificity.

RESUMEN

Objetivo. La interpretación mamográfica mediante el sistema BI-RADS determina que una paciente continúe en el programa de tamizaje regular o sea canalizada para realizarle estudios adicionales diagnósticos para cáncer de mama, de ahí la importancia de su evaluación. Se determinó la sensibilidad, especificidad y exactitud diagnóstica de la interpretación mamográfica así como la reproducibilidad inter- e intra-observador, en una muestra aleatoria de 29 radiólogos de la Secretaría de Salud de México. **Materiales y métodos.** Se seleccionaron 80 estudios mamográficos diagnósticos y de tamizaje del Instituto Nacional de Cancerología de la ciudad de México. Bajo las mismas condiciones, los radiólogos participantes realizaron dos interpretaciones mediante el sistema BI-RADS. **Resultados.** La sensibilidad en la primera y segunda medición fue de 72 y 74%, respectivamente, mientras que la especificidad fue de 80% en la primera y de 82% en la segunda. El análisis multivariado mostró una asociación positiva entre el número de mamogramas interpretados el año previo, el número de cursos realizados y el índice de exactitud diagnóstica ($\beta = 0.00002$; IC 95% 0.000004, 0.00003; $p = 0.02$; $\beta = 0.005$; IC 95% -0.0004, 0.01; $p = 0.07$, respectivamente). La reproducibilidad inter-observador fue baja en ambas mediciones ($\kappa = 0.24$; $\kappa = 0.25$, respectivamente). Sin embargo, después de agrupar las categorías del sistema BI-RADS en dos, la concordancia se incrementó a 0.46 para ambas mediciones. La concordancia intra-observador fue buena ($\kappa = 0.67$). **Conclusiones.** El número de mamogramas leídas el año previo es una buena medida de la experiencia del radiólogo y se refleja en la exactitud diagnóstica de la interpretación mamográfica. Los cursos de actualización también tienen un impacto positivo en la interpretación.

Palabras clave. BI-RADS. Reproducibilidad. Cáncer de mama. Sensibilidad. Especificidad.

INTRODUCCIÓN

El objetivo final de un programa de detección temprana de cáncer de mama (CM) mediante la mamografía,^{1,2} consiste en reducir las tasas de mortalidad por esta patología.³⁻⁵ Ciertos estudios han mostrado que el impacto de un programa de tamizaje organizado sobre la reducción de la mortalidad por CM va de 25 a 30% en mujeres de 50 a 69 años y de 16% en mujeres de 40 a 49 años.⁶⁻⁹ Recientemente, la *U.S. Preventive Services Task Force* (USPSTF) recomendó que la decisión de realizarse la mamografía en mujeres de 40 a 49 años debe tomarse conjuntamente con el médico, tomando en cuenta los factores de riesgo y preferencias de la paciente.¹⁰ Sin embargo, aún existe controversia.^{11,12} En países en vías de desarrollo, el impacto de este programa en la mortalidad por CM ha sido bajo debido, entre otros factores, a:

- Baja cobertura de mujeres con alto riesgo.¹³
- Deficiente calidad en el proceso de atención médica.
- Falta de homologación de criterios diagnósticos debido a la carencia de sistemas de control de calidad y de capacitación continua de su personal.¹⁴

En México, se sabe que hay gran disparidad en cuanto a la adherencia de los sistemas de mamografía con las normas establecidas por instancias internacionales como el Colegio Americano de Radiología y la Comunidad Europea. Dentro de las incongruencias más importantes se resaltan: El procesamiento de la placa, acondicionamiento del cuarto oscuro y el tipo de negatoscopios.¹⁵

En México, la interpretación de la mamografía diagnóstica y de tamizaje se realiza mediante el sistema BI-RADS (del inglés: *Breast Imaging Reporting and Data System*), desarrollada por el Colegio Americano de Radiología, y tiene como propósito que el médico radiólogo cuente con una herramienta para analizar los hallazgos, estandarizar y homologar la terminología del reporte mamográfico y favorecer el monitoreo mediante recomendaciones específicas.¹⁶

En nuestro país, se ha planteado la necesidad de mejorar el programa de tamizaje de CM debido a que, actualmente, entre las mujeres de 25 años o más, ocupa el primer lugar como causa de muerte por neoplasia maligna y ha desplazado de esta posición, al cáncer cérvico uterino.^{17,18} La mortalidad por CM puede reducirse mediante la detección temprana y el tratamiento oportuno.¹⁹ Las recomendaciones derivadas de la interpretación mamográfica

desempeñan un papel importante en la progresión de la lesión y, por lo tanto, sobre la mortalidad por esta neoplasia.

Es necesario implementar acciones continuas para evaluar a las unidades de primero, segundo y tercer nivel de atención que cuentan con mastógrafo, con el propósito de mejorar la calidad del servicio y su eficiencia. Dentro de las características que deben ser evaluadas destacan el control de calidad, que incluye: Infraestructura/equipo/material, posición, técnica de la imagen e interpretación por el radiólogo.²⁰ Todas estas acciones propician la disminución de los resultados falsos positivos y negativos.^{21,22}

OBJETIVO

Evaluar la certeza diagnóstica y los factores asociados, así como la reproducibilidad de la interpretación mamográfica (i.e. mediante el uso del sistema BI-RADS).

MATERIAL Y MÉTODOS

El estudio consistió en evaluar la validez y reproducibilidad de la interpretación mamográfica mediante el sistema BI-RADS hecha por 30 médicos radiólogos de la Secretaría de Salud de 30 unidades con mastógrafo, de 20 estados de la República Mexicana en 2006.

La unidad de observación fueron los médicos radiólogos. El estudio fue aprobado por la Comisión de Ética del Instituto Nacional de Salud Pública; los resultados individuales son confidenciales.

Para la selección de los radiólogos, se estratificó a los estados de la República Mexicana de acuerdo con el índice de marginación de la CONAPO, que clasifica en cinco a las áreas geográficas del país en función de sus carencias y rezagos estructurales.²³ Del total de unidades hospitalarias de primero, segundo y tercer nivel de atención de la Secretaría de Salud que contaban con mastógrafo en enero 2006 (n = 98), se seleccionaron aleatoriamente 30 unidades con el objeto de contar con una muestra representativa del total de unidades.

Una vez hecha la selección aleatoria de las unidades hospitalarias, mediante consentimiento informado se invitó a participar en el estudio a los médicos radiólogos responsables de la interpretación de las mamografías de cada una de las unidades seleccionadas. La tasa de respuesta fue de 79%, de un total de 38 radiólogos, aceptaron participar 30. El motivo de no aceptación de los radiólogos fue debido a compromisos laborales o a problemas de salud.

El universo se integró con 30 radiólogos de los centros de lectura de mamografías distribuidos de la siguiente manera: De la zona 1, tres integrantes: Chihuahua, Guerrero y Oaxaca; de la zona 2, ocho integrantes: Tabasco, Yucatán, Michoacán, Zacatecas; de la zona 3, seis integrantes: Sinaloa, Morelos, Querétaro; de la zona 4, siete integrantes: Tamaulipas, Estado de México, Colima y Jalisco; de la zona 5, seis integrantes: Distrito Federal, Baja California, Coahuila, Nuevo León. Todos participan en el Programa Nacional de Prevención y Control del Cáncer de Mama. De los 30 radiólogos que aceptaron participar, 29 realizaron ambas mediciones, por lo que únicamente se reportan los resultados de 29 radiólogos.

La selección de los mamogramas se realizó de la siguiente manera:

- Del total de mamografías obtenidas de 2004 a 2005 se seleccionaron 80 estudios mamográficos con y sin diagnóstico de cáncer de mama de un centro de referencia hospitalario de tercer nivel (Instituto Nacional de Cancerología, ciudad de México).
- Cada estudio mamográfico contó con cuatro placas, una medio-lateral y una cráneo-caudal para cada mama.
- Las mamografías fueron clasificadas mediante el sistema BI-RADS por un médico radiólogo calificado (estándar de referencia).

Los estudios mamográficos con sospecha de cáncer de mama, BI-RADS 4 y 5, contaban con un estudio histopatológico confirmatorio. Los estudios con BI-RADS 1, 2 y 3, correspondieron a estudios que no presentaron lesión sugestiva de cáncer de mama en dos años consecutivos. Antes de seleccionar la muestra, los estudios se estratificaron por edad y por el tipo de interpretación BI-RADS. Los mamogramas con BI-RADS 4 o 5, de acuerdo con el estudio histopatológico, correspondieron a mujeres con cáncer, o bien, a mujeres con patología benigna, mientras que los mamogramas con BI-RADS 1, 2 y 3 fueron clasificados como normales.

Se incluyeron mamografías de 80 mujeres de 40 a 79 años de edad que fueron obtenidas mediante un mastógrafo digital y que no requerían estudios adicionales (i. e. placas, ultrasonido). Las placas fueron tomadas siguiendo los estándares de calidad que dicta la NOM-229-SSA1-2002, Salud ambiental;²⁴ es decir, bajo los criterios normativos, responsabilidades sanitarias, generales, de garantía de calidad y de funcionamiento para equipos de mamografía que dicta la

norma. Dado que es más difícil identificar una tumoración en mamogramas de mujeres premenopáusicas que en el de mujeres posmenopáusicas –debido a que las primeras tienen mayor densidad que las segundas– fue que se incluyeron 40 estudios de mujeres de 40 a 49 años, 20 de mujeres de 50 a 59 años y 20 de mujeres de 60 a 69 años.²⁵

Se excluyeron las placas de mujeres que contaban con alguna de las siguientes características:

- Historia de cáncer de mama.
- Mamas muy grandes que requirieron de más de una placa o de placas especiales.
- Cirugía cosmética previa.

El tamaño de muestra fue de 80 estudios; se estimó un coeficiente de reproducibilidad de 0.80, nivel alfa de 0.05, intervalo de confianza de ± 0.10 . El cálculo del tamaño de muestra se realizó mediante el método propuesto por Streiner y Norman (1995) que es válido para coeficientes de correlación intraclass y κ .²⁶

Los estudios mamográficos se identificaron con un folio consecutivo, los nombres fueron ocultados para mantener la confidencialidad de las pacientes. La interpretación realizada por el médico radiólogo, estándar de referencia, mediante el sistema BI-RADS no fue revelado, ni al radiólogo participante, ni al investigador responsable de la distribución de las placas. La distribución de las placas seleccionadas de acuerdo con el sistema BI-RADS, se hizo de la siguiente forma:

- BI-RADS 1 (15%).
- BI-RADS 2 (15%).
- BI-RADS 3 (10%).
- BI-RADS 4 (10%).
- BI-RADS 5 (50%).

A los médicos radiólogos se les invitó a participar vía telefónica, se les preguntó si deseaban participar en el estudio y se les explicó que el objetivo era evaluar la variabilidad de la interpretación mamográfica. Una vez reunidos los médicos radiólogos para interpretar los estudios, se les solicitó que firmaran una carta de consentimiento. Asimismo, se les solicitó que contestaran un cuestionario para obtener la siguiente información: Total de mamografías que interpretaron en el último año, años interpretando mamografías, horas al día dedicadas a interpretar estos estudios y antecedente de cursos en interpretación mamográfica.

Para la interpretación, los médicos únicamente contaron con la edad de la paciente y se les solicitó

que clasificaran cada uno de los estudios mamográficos mediante el sistema BI-RADS, utilizando únicamente la clasificación del 1 al 5 y una última categoría en donde se les solicitó incluir los estudios que a su juicio no debían ser interpretados porque la calidad del estudio fue considerada insuficiente. A cada uno de los médicos radiólogos se le proporcionó una tarjeta con la clasificación de BI-RADS para su consulta y un formato de llenado, donde anotaron la clasificación de BI-RADS. El estudio se realizó en las instalaciones de Radiología del Instituto Nacional de Cancerología, programándose dos fechas para realizar las mediciones. Hubo una diferencia de hasta ocho semanas entre la primera y la segunda medición.

Análisis estadístico

Para evaluar la certeza diagnóstica, se estimó la sensibilidad y especificidad, para lo cual se agruparon los estudios en dos categorías de acuerdo con el sistema BI-RADS (recomendación de seguimiento rutinario o a corto plazo: BI-RADS 1-3 vs. recomendación de biopsia: BI-RADS 4-5), tomando como referencia la interpretación del experto del Instituto Nacional de Cancerología. La sensibilidad se consideró como la proporción de mujeres con diagnóstico histopatológico de CM a quienes el radiólogo indicó biopsia (Sistema BI-RADS 4-5). La especificidad se consideró como la proporción de mujeres sin CM a quienes el radiólogo recomendó seguimiento de rutina o a corto plazo (BI-RADS 1-3) o cuando consideró que el estudio había sido realizado con una técnica deficiente. Se calculó también el porcentaje de falsos negativos que correspondió al complemento de la sensibilidad y el porcentaje de falsos positivos que correspondió al complemento de la especificidad.

Con el objeto de evaluar cuáles factores se asociaron a una interpretación adecuada, se calculó el índice de exactitud diagnóstica mediante las curvas ROC (del inglés *Receiver operating characteristic*), que proveen un índice que no requiere de un punto de corte único y es, por lo tanto, una medida en donde la prevalencia de la enfermedad brindada al experimento por el radiólogo, no tiene ningún efecto. El valor va de 0 a 1, donde el 1 corresponde a una sensibilidad y especificidad de 100%. Mediante una regresión lineal múltiple, se identificaron aquellas características que se asociaron con una mejor interpretación mamográfica.

Para el análisis de reproducibilidad inter-observador, primero se analizó la concordancia entre las

lecturas realizadas por los radiólogos tomando en cuenta las interpretaciones del 1 al 5 mediante el sistema BI-RADS. El análisis se realizó después de condensar los resultados de acuerdo con las recomendaciones que surgen de la clasificación BI-RADS:

- Seguimiento rutinario y a corto plazo (BI-RADS 1, 2 y 3).
- Biopsia (BI-RADS 4 y 5).

Se determinó el coeficiente de concordancia diagnóstica a través de la estimación del coeficiente κ ,^{27,28} y la evaluación de la concordancia diagnóstica global entre los observadores. Para el análisis de la reproducibilidad intra-observador, se compararon la primera y segunda lectura de cada uno de los radiólogos y se siguió el mismo procedimiento. Para el análisis de los resultados se empleó el paquete Stata v8.²⁹ De acuerdo con el valor de κ , los resultados de concordancia fueron considerados como:

- Sin concordancia, < 0.20.
- Baja, 0.20-0.39.
- Moderada, 0.40-0.59.
- Moderadamente alta, 0.60-0.79.
- Alta, > 0.79.²⁰

RESULTADOS

Descripción de la población de estudio

Se obtuvieron resultados de 29 radiólogos. Del total, 62.1% fueron hombres y 37.9% mujeres, la mediana de edad fue de 43 años con un rango intercuartil de 38 a 51 años. La distribución de los médicos radiólogos de acuerdo con el estado de procedencia según el índice de marginación de la CONAPO se puede observar en el cuadro 1. En cuanto al nivel de atención, 10.3% de los médicos radiólogos laboraban en unidades de primer nivel de atención, 58.6% en unidades de segundo y 31% en unidades de tercer nivel. Su experiencia, mediana de años desde su graduación, mamogramas leídos el año previo, número de meses interpretando mamogramas y número de horas diarias interpretando mamogramas se indican en el cuadro 1. En cuanto al número total de cursos de interpretación mamográfica tomados hasta el día del estudio, la mediana fue de 3 (rango intercuartil de 2 a 4) (Cuadro 1).

Cuadro 1. Características de los médicos radiólogos mexicanos de acuerdo con el estado de procedencia clasificado según el índice de marginación de la CONAPO.*

Marginación de acuerdo con la CONAPO	Núm. de radiólogos	Edad†	Hombres (%)	Núm. de mamogramas leídos el año previo†	Núm. de meses interpretando†	Núm. de cursos de interpretación mamográfica†	Sensibilidad†	Especificidad†	Falsos positivos†	Falsos negativos†
1	3	63 (37-67)	66.7	600 (0-700)	72 (40-108)	2 (0-4)	68.8 (54.2-75)	84.4 (81.3-90.6)	15.6 (9.4-18.8)	31.3 (25-45.8)
2	8	43 (36.5-47)	50	1 725 (1000-3000)	31.5 (18-92.5)	2 (1-4)	71.9 (62.5-84.4)	87.5 (70.3-90.6)	12.5 (9.4-29.7)	28.2 (15.7-37.5)
3	6	40 (37-42)	83.3	1 000 (500-1000)	46 (40-87)	2 (1-5)	59.4 (54.2-64.6)	90.6 (87.5-90.6)	9.4 (9.4-12.5)	40.7 (35.4-45.8)
4	6	51.5 (45-53)	83.3	1 595 (454-2 800)	45 (36-82)	4 (3-5)	87.5 (68.8-91.7)	76.6 (68.8-87.5)	23.5 (12.5-31.3)	12.5 (8.3-31.3)
5	6	42 (34-50)	50	1 900 (1200-2500)	34 (27-36)	4 (3-4)	80.3 (79.2-83.3)	79.7 (78.1-90.6)	17.2 (9.4-21.9)	19.8 (16.7-20.8)
Total	29	43 (38-51)	62.1	1 120 (700-2400)	40 (27-82)	3 (2-4)	72.9 (60.4-83.3)	84.4 (78.1-90.6)	15.6 (9.4-21.9)	27.1 (16.7-39.6)

* CONAPO: Los estados se clasifican del 1 al 5 iniciando con los de mayor marginación: Zona 1 (Tres integrantes: Chihuahua, Guerrero y Oaxaca); Zona 2 (Ocho integrantes: Tabasco, Yucatán, Michoacán, Zacatecas); Zona 3 (Seis integrantes: Sinaloa, Morelos, Querétaro); Zona 4 (Siete integrantes: Tamaulipas, Estado de México, Colima y Jalisco) y Zona 5 (Seis integrantes: Distrito Federal, Baja California, Coahuila, Nuevo León). †: Mediana y rango intercuartil.

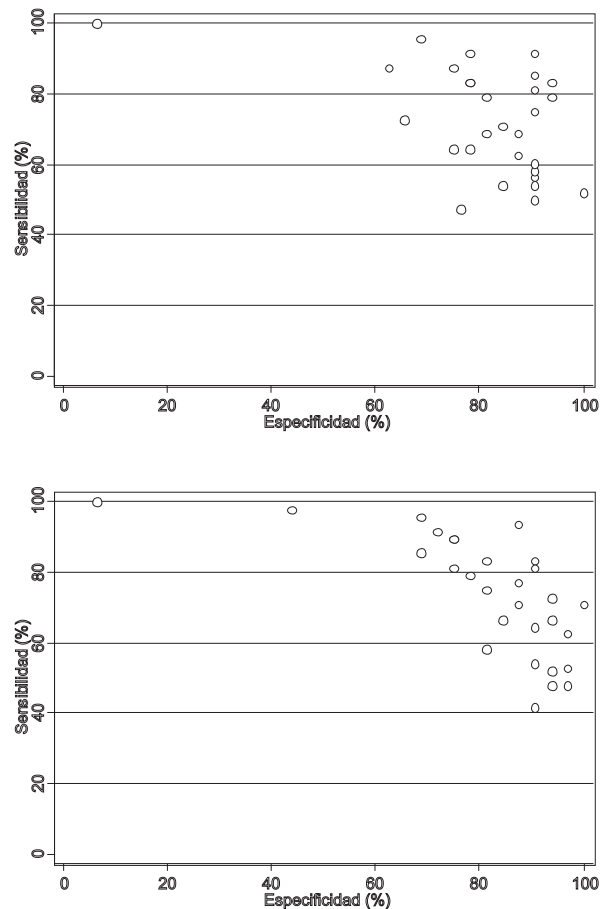


Figura 1. Sensibilidad y especificidad del uso del sistema BI-RADS para el diagnóstico de cáncer por médicos radiólogos de la Secretaría de Salud, primera medición y segunda medición (2006).

Sensibilidad, especificidad, falsos positivos y falsos negativos

En la primera medición, el promedio de la sensibilidad (verdaderos positivos) fue de 72% y la especificidad (verdaderos negativos) de 81%, mientras que 28% correspondió a falsos negativos y 19% a falsos positivos. Las medianas correspondientes a estos valores y los rangos intercuantiles se reportan en el cuadro 1. En términos generales, la proporción de falsos positivos fue mayor cuando el número de cursos sobre interpretación mamográfica se incrementó. Por el contrario, el porcentaje de falsos negativos fue mayor cuando el número de cursos reportado fue menor (Cuadro 1).

Aunque se observó una tendencia a la disminución de la sensibilidad conforme se incrementó la especificidad, algunos radiólogos mostraron combinaciones de altas sensibilidades y especificidades, mientras

Cuadro 2. Relación entre sensibilidad, especificidad, falsos positivos, falsos negativos e índice de exactitud diagnóstica derivada de la interpretación mamográfica realizada por 29 radiólogos mexicanos (2006).

Índice de exactitud diagnóstica	Sensibilidad (Media)	DE	Especificidad (Media)	DE	Falsos positivos	DE	Falsos negativos	DE
< 0.806*	66.7	15.2	74.2	29.2	25.8	29.2	33.3	15.2
> 0.893†	82.4	11.3	88.4	8.8	11.6	8.9	17.6	11.3

*: Percentil 25 del índice de exactitud diagnóstica. †: Percentil 75 del índice de exactitud diagnóstica.

Cuadro 3. Reproducibilidad entre radiólogos mexicanos* interpretando 80 estudios mamográficos mediante el Sistema BI-RADS† para describirlos de acuerdo con la sospecha de malignidad (2006).

BI-RADS†	κ	Primera medición		κ	Segunda medición	
		Z	p		Z	p
1	0.31	57.5	< 0.001	0.32	59.9	< 0.001
2	0.25	46.5	< 0.001	0.26	48.6	< 0.001
3	0.09	16.2	< 0.001	0.07	12.3	< 0.001
4	0.13	25.0	< 0.001	0.12	22.9	< 0.001
5	0.40	74.5	< 0.001	0.43	79.6	< 0.001
Mala calidad‡						
9	0.06	11.8	< 0.001	0.04	7.0	< 0.001
Combinado§	0.24	84.3	< 0.001	0.25	87.4	< 0.001

*: Corresponde a 29 radiólogos participantes y un radiólogo estándar de referencia. †: BI-RADS (*Breast Imaging Reporting and Data System*: 1 = Negativo, 2 = Benigno, 3 = Probablemente benigno, 4 = Anomalía sospechosa, 5 = Altamente sugestiva de malignidad). ‡: Estudios con calidad técnica deficiente. §: Incluye el análisis de los valores 1-5 y 9.

Cuadro 4. Reproducibilidad entre radiólogos mexicanos* interpretando 80 estudios mamográficos mediante el Sistema BI-RADS,† para describirlos de acuerdo con la recomendación (2006).

BI-RADS†	κ	Primera medición		κ	Segunda medición	
		Z	p		Z	p
1-3	0.46	85.8	< 0.001	0.46	85.4	< 0.001
4-5	0.46	86.7	< 0.001	0.47	87.2	< 0.001
Mala calidad‡						
9	0.46	86.9	< 0.001	0.45	87.4	< 0.001
Combinado§	0.46	85.8	< 0.001	0.46	85.4	< 0.001

*: Corresponde a 29 radiólogos participantes y un radiólogo estándar de referencia. †: BI-RADS (*Breast Imaging Reporting and Data System*: 1-3 = Seguimiento, 4-5 = Biopsia). ‡: Estudios con calidad técnica deficiente. §: Incluye análisis de los valores 1-3 = Seguimiento, 4-5 = Biopsia, y 9.

que otros mostraron sensibilidades y especificidades bajas. Un comportamiento similar se observó en la segunda medición (Figura 1).

Índice de exactitud diagnóstica

El valor del índice de exactitud diagnóstica va de 0 a 1 correspondiendo el 1, a una sensibilidad y espe-

cificidad de 100%. En el presente trabajo, la media de este índice fue de 0.85 (DE = 0.063; rango intercuartil de 0.81 a 0.89) para la primera medición y de 0.86 (DE = 0.063; rango intercuartil de 0.84 a 0.89) para la segunda medición. En el cuadro 2, se observa que cuando el índice de exactitud diagnóstica está por arriba del percentil 75 (> 0.89) la sensibilidad y la especificidad tienen un valor mayor y los falsos

Cuadro 5. Modelo multivariado para evaluar los factores asociados a la exactitud diagnóstica,* mediante el sistema BI-RADS,[†] en 29 radiólogos mexicanos (2006).

	β^{\dagger}	IC 95% [†]	p
Núm. de mamogramas interpretados el año previo	0.00002	0.000004 , 0.00003	0.02
Núm. meses interpretando mamogramas	-0.001	-0.001 , -0.0001	0.03
Núm. de cursos	0.005	-0.0004 , 0.010	0.07
Intercepto	0.834	0.787 , 0.881	
R ^{2†}	0.43		

*: Índice de exactitud diagnóstica mediante curvas ROC; modelo ajustado por edad y sexo. [†]: BI-RADS (*Breast Imaging Reporting and Data System*: 1-3 = Seguimiento y mamografías con calidad deficiente, 4-5 = Biopsia). [‡]: Fuerza de la relación lineal entre las variables y el índice de exactitud diagnóstica.

positivos y negativos un valor menor que cuando el índice se encuentra por debajo del percentil 25 (< 0.81).

Factores asociados al índice de exactitud diagnóstica

Las características que se asociaron con un incremento en la exactitud diagnóstica fueron: Número de estudios mamográficos leídos el año previo y el número de cursos sobre interpretación mamográfica tomados durante su práctica profesional ($\beta = 0.00002$; IC 95% 0.000004, 0.00003; $p = 0.02$; $\beta = 0.005$; IC 95% -0.0004, 0.01; $p = 0.07$, respectivamente). Conforme se incrementó el tiempo en meses de interpretación mamográfica, el índice disminuyó ($\beta = -0.001$; IC 95% -0.001, -0.0001; $p = 0.02$); esta variable se midió para evaluar el efecto de la experiencia sobre la exactitud diagnóstica. El sexo y la edad no se asociaron de forma estadísticamente significativa, pero se ajustó en el modelo (Cuadro 4).

Para interpretar mejor los resultados, se pueden cambiar las unidades de β para que expresen el cambio del índice en desviaciones estándar y utilizando como incremento de la variable independiente el rango intercuartil. Por ejemplo, si se incrementa el número de mamogramas interpretados el año previo, de 700 a 2,400, es decir en 1,700, se incrementa el índice 0.034 ($0.00002 \times 1,700 = 0.034$), aproximadamente media desviación estándar del índice de exactitud diagnóstica, en donde 0.063 es la desviación estándar del índice ($0.034/0.063 = 0.54$ desviaciones estándar).

Reproducibilidad inter- e intra-observador

El coeficiente de concordancia inter-observador fue bajo en ambas mediciones ($\kappa = 0.24$ para la primera medición; $\kappa = 0.25$ para la segunda) (Cuadro

3). El coeficiente de reproducibilidad más alto se obtuvo para los valores de BI-RADS 1 y 5, y el más bajo para el valor 3 y para aquellos en los que el estudio se reportó de mala calidad. Al agrupar las categorías de BI-RADS en dos (BI-RADS 1-3; y BI-RADS 4-5 más los reportados como estudios de mala calidad), se observó un incremento en la concordancia de 0.24 a 0.46 en la primera medición y de 0.25 a 0.46 en la segunda medición (Cuadro 4). Por otro lado, la concordancia intra-observador, es decir la observada entre la primera y la segunda medición realizada por cada uno de los médicos radiólogos fue buena ($\kappa = 0.67$).

DISCUSIÓN

En relación con la interpretación mamográfica, los resultados muestran una sensibilidad de 72%, especificidad 81%, falsos negativos 28% y falsos positivos 19%. El índice de exactitud diagnóstica fue de 0.85 (DE = 0.063) y la variable que se le asoció más fue el número de mamogramas leídos el año previo ($p = 0.02$). Por último, la reproducibilidad inter-observador agrupando BI-RADS 1-3 y 4-5 fue de 0.46 y la intra-observador de 0.67.

Aunque en promedio, la sensibilidad y especificidad en el estudio de México, fueron en general altas (72 y 81%, respectivamente), se observó un rango de 48 a 100% para la sensibilidad y de 6 a 100% para la especificidad. Un estudio realizado en Estados Unidos³⁰ en donde 108 radiólogos de centros acreditados por el Colegio Americano de Radiología analizaron 79 mamografías de tamizaje –de las cuales, 45 correspondían a cáncer– reportó una sensibilidad con un rango de 47 a 100%, similar al del presente estudio y una especificidad con un rango de 36 a 99%, menos amplio que el del estudio mexicano. Aunque en promedio la sensibilidad y especificidad reportadas en el presente estudio fueron altas, esto pudo deberse –además de las habilidades profesionales de los

radiólogos– a la calidad de las imágenes evaluadas que se obtuvieron bajo los criterios normativos de la NOM-229-SSA1-2002, Salud ambiental.²⁴ Los autores del estudio realizado en Estados Unidos refieren que dentro de los criterios de certificación debe tomarse en cuenta no únicamente la calidad técnica del equipo radiológico y el de las imágenes, sino también la exactitud diagnóstica de la interpretación,³⁰ que es lo que se analizó en el presente trabajo.

La Organización Mundial de la Salud (OMS) reconoce que los tamizajes de cáncer de mama pueden tener un efecto negativo en la población, aun cuando el rigor metodológico sea respetado en las pruebas.³¹ Se han reportado diversos daños atribuibles al tamizaje mediante mastografía: Exposición a radiación, dolor durante el procedimiento, ansiedad, falsos positivos, falsos negativos, biopsias no necesarias y el sobrediagnóstico.^{32,33} En el presente estudio, la media de falsos positivos fue de 19% (DE = 17%) en la primera medición y 18% (DE = 19) en la segunda medición. Diversos estudios han reportado que más de 50,000 mujeres por año han recibido un diagnóstico mamográfico falso positivo,³⁴ y se han reportado casos de suicidio a causa del resultado de cáncer de mama.³⁵

En cuanto a los falsos negativos, se han reportado valores que van de 4 a 34%, dependiendo del tipo de población y de la definición de falso negativo utilizada.³⁶ En el presente estudio, la media de falsos negativos en la primera y segunda medición fue de 27% (DE = 14) y 26% (DE = 17), respectivamente. A los tumores que son diagnosticados después de una mastografía clasificada como negativa y antes de la siguiente mastografía de tamizaje se denominan cánceres de intervalo.^{37,38} Un incremento en los falsos negativos puede deberse a características histopatológicas del tumor así como a la densidad mamográfica. Se ha reportado que la posibilidad de presentar un cáncer de intervalo es seis veces mayor en las mujeres con densidad mamográfica muy alta en comparación con densidades bajas (OR = 6.14; IC95% 1.95-19.4).³⁷ Las características mamográficas varían de acuerdo con el tejido glandular, la grasa, el estroma y el epitelio,³⁹ de ahí las diferencias en cuanto a exactitud diagnóstica en mujeres pre- y posmenopáusicas.⁴⁰ Los problemas actuales relacionados con la interpretación de la mamografía incluyen la incapacidad para reconocer imágenes anormales.⁴¹ De ahí, que la capacitación y la certificación son indispensables para mejorar la interpretación. A la fecha, son pocos los países que han instituido programas de acreditación en un esfuerzo por minimizar este tipo de error.^{42,51,52}

En relación con los factores que se asocian con la exactitud diagnóstica, dos estudios que evaluaron

la asociación entre el número de estudios mamográficos interpretados y la exactitud diagnóstica, encontraron que dicho indicador aumenta conforme se incrementa el número de mamogramas interpretados.^{43,44} Por ejemplo, el estudio de Kan, *et al.* (2000), realizado a 35 radiólogos con al menos tres años de experiencia en interpretación mamográfica, reportó que un mínimo de 2,500 interpretaciones al año se asoció con menos tasas de interpretación anormal y con mejores tasas de detección de cáncer cuando se compararon con un número menor de interpretaciones anuales.⁴³

Los resultados muestran que, en promedio, los radiólogos participantes interpretan entre 700 y 2,400 mamogramas anuales (rango intercuartil), debido, probablemente, a que sus actividades no se limitan únicamente a la interpretación mamográfica. Sickles, *et al.* (2002) han mostrado que radiólogos especialistas en la interpretación mamográfica detectan cánceres en estadios más tempranos que los radiólogos generales.⁴⁴

Sin embargo, otras investigaciones han reportado que más allá del número de interpretaciones, es importante considerar la calidad de las mastografías, los equipos y la experiencia del médico que incluye la retroalimentación que recibe el radiólogo.^{45,46}

La sensibilidad y especificidad reportadas en el presente estudio dependen, además de las habilidades profesionales de los radiólogos, de la calidad de las imágenes evaluadas, obtenidas bajo los criterios normativos de la NOM-229-SSA1-2002, Salud ambiental.²⁴ Debido a que en este trabajo se mantuvo constante la calidad de las mastografías, no se evaluó este aspecto como indicador de una buena interpretación y tampoco se preguntó si contaban o no con una retroalimentación sistemática sobre el diagnóstico definitivo.

Un estudio evaluó la experiencia de 21 radiólogos que rutinariamente interpretaron mastografías y siete que no lo hacían bajo rutina, respecto a la exactitud diagnóstica, medida a través de la sensibilidad y especificidad, reportó una mejor especificidad en los primeros. Sin embargo, al interior de este grupo encontró que el número de estudios interpretados no se asoció con una mayor especificidad. Las diferencias entre aquellos resultados con los presentes, pueden explicarse por contar con una mayor variabilidad en cuanto a la densidad (mamogramas de mujeres de 40 a 69 años de edad en este estudio vs. mamogramas de mujeres de 50 a 64 años en aquél), al uso del índice de exactitud diagnóstica como variable dependiente en este estudio y al uso de la sensibilidad y especificidad en aquél.⁴⁷

En el presente trabajo el número de cursos sobre interpretación mamográfica se asoció con un mayor índice de exactitud diagnóstica; sin embargo, la asociación no fue estadísticamente significativa ($p = 0.07$) (Cuadro 5). Al respecto, de acuerdo con las Guías Europeas para el tamizaje de cáncer de mama,⁴⁸ el entrenamiento específico en interpretación mamográfica es básico y se recomiendan al menos 60 hrs.

La literatura internacional reporta que el entrenamiento incide sobre la calidad de la lectura mamográfica, particularmente sobre la tasa de detección de lesiones malignas.⁴⁸⁻⁵⁰ Peer, *et al.* (2005) reportaron una ausencia de programas institucionales de entrenamiento y en su estudio encontraron que los radiólogos y técnicos radiólogos reportaron que 5 a 15 hrs de entrenamiento son apropiadas para un entrenamiento inicial; sin embargo, consideran que un entrenamiento continuo de 5-10 hrs cada tres años es apropiado. Esto no solamente aplicó a la radiología en general sino también a la mamografía digital.⁵¹ En Bélgica, se requiere acreditar a los radiólogos (dedicados a realizar mastografía de tamizaje y diagnóstica) con 8 hrs de entrenamiento (4 hrs prácticas y 4 hrs teóricas sobre mamografía digital) y se recomienda una recertificación cada cinco años para garantizar estándares de calidad.⁵²

En el caso de los radiólogos que participan en tamizaje en Reino Unido, realizan una lectura anual hasta de 5,000 casos por año. En contraste, en Estados Unidos, los radiólogos tienden a realizar un menor número de lecturas; sin embargo, su exactitud tiende a estar más relacionada con otros factores como años de experiencia interpretando mamogramas.⁵³ En el estudio de los radiólogos mexicanos, la edad fue un factor que no se asoció estadísticamente con la exactitud diagnóstica. A este respecto, en el estudio de Molins, *et al.* (2008) aquéllos de 45 años y más mostraron un incremento en la sensibilidad en 33% (OR 1.33; 95% IC 1.12 a 1.59) en comparación con los menores de 45 años después de ajustar por el número de estudios mamográficos interpretados, retroalimentación, porcentaje del tiempo dedicado a la radiología mamaria, años de práctica interpretando mamogramas y consulta con otros radiólogos al interpretar mamogramas.⁴⁷

En el presente estudio el número de meses interpretando mamogramas se consideró un indicador de la experiencia. Sin embargo, se observó una relación inversa de esta variable con la exactitud diagnóstica. Los resultados indican que el tiempo que lleva un radiólogo interpretando mamogramas no es un buen indicador de la experiencia, ya que no se observó

una correlación estadísticamente significativa con el número de horas al día que dedican a la interpretación de mamogramas ($r = -0.07$, $p = 0.73$) y se observó una correlación positiva con el número de estudios interpretados el año previo pero no estadísticamente significativa ($r = 0.29$, $p = 0.15$).

La repetitividad de las interpretaciones entre los 29 radiólogos, fue baja cuando se tomaron las cinco categorías de BI-RADS (concordancia inter-observador: $\kappa = 0.24$ para la primera medición; $\kappa = 0.25$ para la segunda medición). Sin embargo, al agrupar los resultados en BI-RADS 1-3 y 4-5, el coeficiente se incrementó a 0.46, tanto en la primera como en la segunda medición. Este incremento se ha observado en otros estudios.^{20,54,55} La agrupación de los estudios en dos grupos (BI-RADS 1, 2, 3 vs. BI-RADS 4 y 5) se ha utilizado consistentemente para evaluar reproducibilidad. A partir de la introducción del sistema BI-RADS, la variabilidad del observador ha sido revisada por algunos investigadores.^{15,55-57} Estos estudios indican que a pesar de la existencia de un sistema estandarizado, la variabilidad en el reporte de la mastografía persiste.¹⁶

CONCLUSIONES

Los resultados del presente trabajo sugieren que el número de mamogramas leídos el año previo es una buena medida de la experiencia del radiólogo y se ve reflejada en la exactitud diagnóstica de la interpretación mamográfica. Asimismo, los cursos de actualización también tienen un impacto positivo en la interpretación. En México, se requiere contar con un sistema de control de calidad, no únicamente de las instalaciones, equipo (función, mantenimiento)²⁴ y de la calidad de las mamografías,⁴² sino de la interpretación mamográfica.

REFERENCIAS

1. Moreira C, Svoboda K, Poulos A, Taylor R, Page A, Rickard M. Comparison of validity and reliability of two image classification systems for the assessment of mammogram quality. *J Med Screen* 2005; 12: 38-42.
2. Bigenwald RZ, Warner E, Gunasekara A, Hill KA, Causer PA, Messner SJ, et al. Is mammography adequate for screening women with inherited BRCA mutations and low breast density? *Cancer Epidemiol Biomarkers Prev* 2008; 17: 706-11.
3. Yankaskas BC, Klabunde CN, Ancelle-Park R, Renner G, Wang H, Fracheboud J, et al. International Breast Cancer Screening Network. International comparison of performance measures for screening mammography: can it be done? *J Med Screen* 2004; 11: 187-93.
4. Barlow WE, Chi C, Carney PA, Taplin SH, D'Orsi C, Cutter G, et al. Accuracy of screening mammography interpretation by characteristics of radiologists. *J Natl Cancer Inst* 2004; 96: 1840-50.

5. Uematsu T. Screening and diagnosis of breast cancer in augmented women. *Breast Cancer* 2008; 15: 159-64.
6. Blanks RG, Moss SM, McGahan CE, Quinn MJ, Babb PJ. Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990-8: comparison of observed with predicted mortality. *BMJ* 2000; 321: 665-9.
7. Duffy SW, Tabar L, Chen HH, Holmqvist M, Yen MF, Abdallah S, et al. The impact of organized mammographic service screening on breast carcinoma mortality in seven Swedish counties. *Cancer* 2002; 95: 458-69.
8. Kerlikowske K. Efficacy of screening mammography among women aged 40 to 49 years and 50 to 69 years: comparison of relative and absolute benefit. *J Natl Cancer Inst Monogr* 1997; (22): 79-86.
9. Sarkeala T, Anttila A, Saarenmaa I, Hakama M. Validity of process indicators of screening for breast cancer to predict mortality reduction. *J Med Screen* 2005; 12: 33-7.
10. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2009; 151: 716-26.
11. Kopans DB. The recent US preventive services task force guidelines are not supported by the scientific evidence and should be rescinded. *J Am Coll Radiol* 2010; 7: 260-4.
12. Quanstrum KH, Hayward RA. Lessons from the mammography wars. *N Engl J Med* 2010; 363: 1076-9.
13. Mishra SI, Bastani R, Crespi CM, Chang LC, Luce PH, Baquet CR. Results of a Randomized Trial to Increase Mammogram Usage among Samoan Women. *Cancer Epidemiol Biomarkers Prev* 2007; 16: 2594-604.
14. Shen Y, Parmigiani G. A model-based comparison of breast cancer screening strategies: Mammograms and clinical breast examinations. *Cancer Epidemiol Biomarkers Prev* 2005; 14: 529-32.
15. Brandan ME, Ruiz-Trejo C, Verdejo-Silva M, Guevara M, Lozano-Zalce H, Madero-Preciado L, et al. Evaluation of equipment performance, patient dose, imaging quality, and diagnostic coincidence in five Mexico City mammography services. *Arch Med Res* 2004; 35(1): 24-30.
16. Obenauer S, Hermann KP, Grabbe E. Applications and literature review of the BI-RADS classification. *Eur Radiol* 2005; 15: 1027-36.
17. Lozano R, Gómez-Dantés H, Arreola-Ornelas H, Méndez O, Knaut FM. Tendencias en la mortalidad por cáncer de mama en México, 1979-2006. Observatorio de la Salud. Documento de trabajo. Competitividad y Salud, Fundación Mexicana para la Salud 2008.
18. Programa Nacional de Salud 2007-2012. Secretaría de Salud. México. Disponible en: http://portal.salud.gob.mx/sites/salud/descargas/pdf/pns_version_completa.pdf [Consultado 28 de abril 2010].
19. Bassett LW, Cardenosa G, D'Orsi CJ, Dempsey PJ, Dershaw DD, Destouet JM, et al. Risk of risk-based mammography screening, ages 40-49. American College of Radiology Task Force on Breast Cancer. *J Clin Oncol* 1999; 17: 735-8.
20. McKay C, Hart CL, Erbacher G. Objectivity and accuracy of mammogram interpretation using the BI-RADS final assessment categories in 40- to 49-year-old women. *J Am Osteopath Assoc* 2000; 100: 615-20.
21. Wang J, Shih TT, Hsu JC, Li YW. The evaluation of false negative mammography from malignant and benign breast lesions. *Clin Imaging* 2000; 24: 96-103.
22. Chida K, Zuguchi M, Sai M, Saito H, Yamada T, Ishibashi T, et al. Optimization of tube potential-filter combinations for film-screen mammography: a contrast detail phantom study. *Clin Imaging* 2005; 29: 246-50.
23. Disponible en: <http://www.conapo.gob.mx/>
24. Comisión Federal para la Protección contra Riesgos Sanitarios. Normas Oficiales Mexicanas [NOM-229-SSA1-2002, monografía en internet]. México COFEPRIS. Disponible en: <http://www.cofepris.gob.mx/work/sites/cfp/resources/LocalContent/596/5/229ssa1.pdf> [Consultado 28 de abril 2010].
25. Aitken Z, Walker K, Stegeman BH, Wark PA, Moss SM, McCormack VA, Silva Idos S. Mammographic density and markers of socioeconomic status: a cross-sectional study. *BMC Cancer* 2010; 10: 35.
26. Streiner DL, Norman GR. Health Measurement Scales: A practical guide to their development and use. New York: Oxford University Press; 1995, p. 79-96.
27. Feinstein AR. A bibliography of publications on observer variability. *J Chron Dis* 1985; 38: 619-32.
28. Raab SS. Diagnostic accuracy in cytopathology. *Diag Cytopathol* 1994; 10: 68-75.
29. Stata 8. Statistics/data analysis. Stata Corporation, 702 University Drive East. College Station. Tx 778740, U.S.A.
30. 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.
31. Brodersen J, Thorsen H, Cockburn J. The adequacy of measurement of short and long-term consequences of false-positive screening mammography. *J Med Screening* 2004; 11: 39-44.
32. Mandelblatt JS, Cronin KA, Bailey S, Berry DA, De Koning HJ, Draisma G, et al. Breast Cancer Working Group of the Cancer Intervention and Surveillance Modeling Network. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Intern Med* 2009; 151: 738-47.
33. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan B, Nygren P, Humphrey L. Screening for Breast Cancer: Systematic Evidence Review Update for the US Preventive Services Task Force. *Ann Intern Med* 2009; 151: 727-37.
34. Department of Health. NHS Cancer Screening Programs. Breast screening programme 2000-2001.
35. Weil JG, Hawker JI. Positive findings of mammography may lead to suicide. *BMJ* 1997; 314: 754-5.
36. Wang J, Shih TT, Hsu JC, Li YW. The evaluation of false negative mammography from malignant and benign breast lesions. *J Clinical Imaging* 2000; 24: 96-103.
37. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst* 2000; 92: 1081-7.
38. Roger P, Delfour C, Ragu N, Serre I, Baldet P, Taourel P. Carcinoma with false negative mammogram: when and how? *J Radiol* 2004; 85: 2063-7.
39. Johns PC, Yaffe MJ. X-ray characterization of normal and neoplastic breast tissues. *Phys Med Biol* 1987; 32: 675-95.
40. Sala E, Warren R, McCann J, Duffy S, Day N, Luben R. Mammographic parenchymal patterns and mode of detection: implications for the breast screening programme. *J Med Screen* 1998; 5: 207-12.
41. Smith-Bindman R, Chu P, Miglioretti DL, Quale C, Rosenberg RD, Cutter G, et al. Physician predictors of mammographic accuracy. *J Natl Cancer Inst* 2005; 97: 358-67.
42. Norma Oficial Mexicana (NOM-041SSA2-2002), para la prevención, diagnóstico, tratamiento, control y vigilancia epidemiológica del cáncer de mama. Secretaría de Salud. Normas Oficiales Mexicanas [NOM-041-SSA2-2002, monografía en internet]. México, SS. Disponible en: <http://bibliotecas.salud.gob.mx/gsd/collect/nomssa/index/assoc/HASHbe3f.dir/doc.pdf> [Consultado 28 de abril 2010].

43. Kan L, Olivotto I, Burhenne LW, Sickles E, Coldman A. Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. *Radiol* 2000; 215: 563-7.
44. Sickles E, Wolverton D, Dee K. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiol* 2002; 224: 861-9.
45. Beam C, 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.
46. Beam C, Conant E, Sickles E. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. *J Natl Cancer Inst* 2003; 95: 282-90.
47. Molins E, Maciá F, Ferrer F, Maristany MT, Castells X. Association between radiologists' experience and accuracy in interpreting screening mammograms. *BMC Health Serv Res* 2008; 8: 91.
48. Van Engen R, Young K, Bosmans H, Thijssen M. The European protocol for the quality control of the physical and technical aspects of mammography screening. Part B: Digital mammography. In: The Fourth Edition of the European Guidelines for Breast Cancer Screening. European Commission. Available from: <http://www.euref.org> [Accessed 19 August 2009].
49. Skaane P, Young K, Skjennald A. Population based mammography screening: comparison of screen film and full-field digital mammography with soft-copy reading-The Oslo I study. *Radiol* 2003; 229: 877-84.
50. Skaane P, Skjennald A. Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program-The Oslo II Study. *Radiol* 2004; 232: 197-204.
51. Peer S, Faulkner K, Torbica P, Peer R, Busch HP, Vetter S, et al. Relevant training issues for introduction of digital radiology: results of a survey. *Radiat Prot Dosim* 2005; 117: 154-61.
52. Van Ongeval C, Van Steen A, Bosmans H. Teaching syllabus for radiological aspects of breast cancer screening with digital mammography. *Radiat Prot Dosimetry* 2008; 129: 191-4.
53. Moss SM, Blanks RG, Bennett RL. Is radiologists' volume of mammography reading related to accuracy? A critical review of the literature. *Clin Radiol* 2005; 60: 623-6.
54. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. *AJR* 2000; 174: 1769-77.
55. Baker JA, Kornguth PJ, Floyd CE Jr. Breast imaging reporting and data system standardized mammography lexicon: observer variability in lesion description. *Am J Roentgenol* 1996; 166: 773-8.
56. Orel SG, Kay N, Reynolds C, Sullivan DC. BI-RADS categorization as a predictor of malignancy. *Radiol* 1999; 211: 845-50.
57. Kerlikowske K, Grady D, Barclay J, Frankel SD, Ominsky SH, Sickles EA, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology breast imaging reporting and data system. *J Natl Cancer Inst* 1998; 90: 1801-9.

Reimpresos

Dra. Gabriela Torres-Mejía

Instituto Nacional de Salud Pública
 Centro de Investigación en Salud Poblacional,
 Av. Universidad 655,
 Col. Sta. Ma. Ahuacatlán,
 62508, Cuernavaca, Mor.
 Tel.: 52 777 311-2343 Fax 52 777 311-1148
 Correo electrónico: gtorres@correo.insp.mx

Recibido el 16 de diciembre de 2009.

Aceptado el 5 de octubre de 2010.