

Trabajo Original

Curva de aprendizaje para el diagnóstico ecográfico específico de masas anexiales

Learning curve for the specific ultrasound diagnosis of adnexal masses

Guillermo Avilés L. ¹, Humberto Vaccaro C. ², Linder Díaz C. ³, Fernando Amor L. ², Alberto Iturra A. ⁴, Mauricio León A. ⁴, Víctor Toledo S. ⁵, Raimundo Avilés D. ^{1,6}.

¹ Clínica Las Condes. Santiago, Chile.

² Ultrasonic Panorámico. Santiago, Chile.

³ Centro de Ultrasonografía AGB, Clínica Sanatorio Alemán. Concepción, Chile.

⁴ Clínica INDISA. Santiago, Chile.

⁵ Servicio de Obstetricia y Ginecología, Hospital de la Fuerza Aérea de Chile. Santiago, Chile.

⁶ Departamento de Obstetricia y Ginecología Universidad Finis Terrae, Hospital Dr. Luis Valentín Ferrada. El Carmen, Maipú, Chile.

Autor para correspondencia:

Linder Díaz Colmenarez, ginecologo.obstetra@gmail.com.  <https://orcid.org/0000-0001-9420-0406>

RESUMEN

Objetivo: Determinar el tiempo que requiere una curva de aprendizaje para diagnóstico ecográfico específico histopatológico en masas anexiales basándonos en cálculos estadísticos no influidos por la prevalencia según diferentes grados de experiencia. **Métodos:** Estudio observacional, descriptivo, transversal. Se estudiaron imágenes de 108 masas anexiales. La prueba estándar de oro fue el reporte histopatológico definitivo. Se comparó el rendimiento diagnóstico de 4 examinadores con la siguiente experiencia en diagnóstico ecográfico de patología anexial: A > 20 años, B ≤ 20 hasta > 10 años, C ≤ 10 hasta > 5 años y D ≤ 5 años, analizando solo imágenes y sin datos clínicos de las pacientes, para emitir un diagnóstico específico a libre escritura. **Resultados:** Prevalencia de masas malignas 17,2 % (15/87). Nivel de confianza en los examinadores se consideró según falta de respuesta diagnóstica: alto (<6 %) con experiencia de más de 10 años y moderado a bajo ≤ 10 años. Examinadores con más de 5 años siempre mostraron likelihood ratio positivo mayor a 10, exactitud diagnóstica mayor a 90 % y Odds ratio diagnóstica mayor a 46, no así para examinador con menor tiempo de experiencia, quién presentó resultados con mala utilidad clínica. El cambio de probabilidad de acierto específico pre-test a post-test mejoró consistentemente con los años de experiencia. **Conclusión:** Se necesitarían más de 10 años de experiencia con especial dedicación a ecografía ginecológica avanzada para un rendimiento diagnóstico específico deseado junto con alta confianza en ecografía de masas anexiales.

Palabras claves: cáncer, ovario, masas anexiales, ecografía, curva de aprendizaje, experiencia, caracterización subjetiva, patrón de reconocimiento, diagnóstico específico.

ABSTRACT

Objective: To determine the time required for a learning curve of histopathological specific ultrasound diagnosis in adnexal masses based on statistical calculations not influenced by prevalence according to different degrees of experience. **Methods:** Observational, descriptive, cross-sectional study. Images of 108 adnexal masses were studied. The gold standard test was the definitive histopathological report. The diagnostic performance of 4 examiners with the following experience in ultrasound diagnosis of adnexal pathology: A > 20 years, B ≤ 20 to > 10 years, C ≤ 10 to > 5 years and D ≤ 5 years was compared, analyzing only images and blinded of clinical data of the patients, to issue a specific diagnosis with free writing. **Results:** Prevalence of malignant masses 17.2% (15/87). The level of confidence in the examiners was considered according to the lack of diagnostic response: high (<6%) with experience of more than 10 years and moderate to low ≤ 10 years. The examiners with more than 5 years always showed likelihood ratio positive greater than 10, diagnostic accuracy greater than 90% and diagnostic Odds ratio greater than 46, not so for the examiner with less experience time who presented results with little clinical utility. The change in specific probability from pre-test to post-test improved consistently with years of experience. **Conclusion:** More than 10 years of experience with special dedication to advanced gynecological ultrasound are probably needed for a desired specific diagnostic performance coupled with high confidence in adnexal mass ultrasound.

Keywords: cancer, ovary, adnexal masses, ultrasound, learning curve, experience, subjective characterization, pattern recognition, specific diagnosis.

CUADRO ABREVIATURAS

Abreviatura	Español	Inglés
TBPM	Tumor de bajo potencial de malignidad	Tumor of low malignancy potencial
S	Sensibilidad	Sensibility
E	Especificidad	Specificity
LR+	Likelihood ratio positivo	Likelihood ratio positive
LR-	Likelihood ratio negativo	Likelihood ratio negative
ED	Exactitud diagnóstica	Diagnostic accuracy
DOR	Odds ratio diagnóstica	Diagnostic Odds ratio
IC	Intervalo de confianza	Confidence interval
GIRADS	Sistema de informes y datos de imágenes ginecológicas	Gynecological Imaging Reporting and Data System

INTRODUCCIÓN

Los avances tecnológicos en los equipos de ultrasonido diagnóstico - en especial la incorporación de transductores endocavitarios- y el mayor conocimiento y experiencia que han adquirido los operadores, han permitido una mejoría notable en la precisión del diagnóstico ginecológico de las masas anexiales¹⁻³.

Dos enfoques se han desarrollado paralelamente en este sentido. Por una parte, la elaboración de modelos de puntaje, con la intención de estandarizar el diagnóstico y hacerlo accesible a un gran número

de operadores con independencia de su experiencia en este campo^{4,5}. Por otra parte, la caracterización subjetiva de las masas, lo que necesariamente amerita una sólida experiencia acumulada tras muchos años de observación de la patología anexial en tiempo real⁶⁻⁹, pudiéndose acortar la temporalidad con entrenamiento intensivo específico con volúmenes 3D y/o clips de videos de masas acumuladas, clasificadas y retroalimentadas por un experto y cotejados con su respectivo diagnóstico anatomopatológico¹⁰⁻¹⁴ en lo que podemos definir

como “eyes-on” que estimula la memoria por trabajo visual ¹⁵.

La ecografía en cualquiera de sus campos clínicos es operador dependiente, así como prácticamente sucede en todas las acciones médicas ¹⁵. En la elaboración de un diagnóstico preciso inciden varios factores: la calidad del equipo utilizado, las características físicas de la paciente en el caso de la ultrasonografía ginecológica, la habilidad innata del operador, el entrenamiento y/o experiencia que ha acumulado a lo largo del tiempo y el criterio clínico ^{1-8,10-17}. El orden de la enumeración anterior no refleja necesariamente la importancia relativa de cada uno de esos puntos.

Es bien conocido que el método con mejor rendimiento discriminatorio de masas anexiales malignas versus benignas, superando a la TAC y RMN para sensibilidad, especificidad y exactitud diagnóstica desde hace más de 2 décadas y manteniéndose hasta la actualidad, es la caracterización subjetiva ecográfica por operador con experiencia ^{1,3-5,16-21}. También se sabe que los valores predictivos positivos y negativos son altamente influidos por la prevalencia de la enfermedad, no así para cálculos estadísticos en los que en la fórmula el dividendo y divisor incluyan sensibilidad y especificidad o se calcule en sentido vertical (columnas) y no en sentido horizontal (filas) de tablas 2x2, siendo likelihood ratio (razón de verosimilitud) el principal parámetro en este sentido ^{9,22-29}. Las series publicadas mayoritariamente reportan prevalencias de malignidad que varían entre 10 % y 40 % ^{1,6,8,10,16,17,19,29,30} por ser centros o unidades de derivación, no correspondiendo a la prevalencia real del día a día considerando toda la población. En estos estudios los valores predictivos son más útiles y representativos para unidades de centros terciarios que manejan con más frecuencia masas de alto riesgo, pero no así en el diagnóstico ultrasonográfico de estas masas en población general no seleccionada sin discriminar por edad o estado menopáusico, al considerarse funcionales, estructurales benignas y malignas, sean estas uniloculares, multiloculares con o sin componente sólido y sólidas puras, donde la prevalencia de malignidad al incluirlas todas difícilmente supere el 1 % ^{9,31-33}.

El trabajo expuesto a continuación, pretende señalar que se requieren años de experiencia para poder efectuar un diagnóstico ecográfico anexial específico histopatológico confiable, acorde a las exigencias actuales en esta patología en población no seleccionada. Como objetivo secundario en base a la ausencia de respuesta se pretendió analizar la confianza según la experiencia de los examinadores al intentar acertar por ecografía el diagnóstico histopatológico.

MÉTODOS

Estudio observacional, descriptivo y transversal. La muestra aportada por dos de los autores (G.A.L. y R.A.D.) desde Clínica Las Condes, en pacientes ≥ 18 años y no embarazadas, con una serie de 108 masas anexiales unilaterales con diagnóstico confirmado por histopatología durante el año 2014, desde historias médicas y registros de imágenes de la institución. Todas las masas finalmente incluidas fueron extirpadas quirúrgicamente máximo 30 días luego de la detección ecográfica, con su respectivo diagnóstico definitivo. Criterios de exclusión inicial: masas intervenidas de urgencia como gestaciones ectópicas anexiales, torsiones en las masas, abscesos tubo-ováricos y masas catalogadas como funcionales por su resolución espontánea en el seguimiento o en reporte anatomopatológico. Se recibió autorización escrita desde la dirección académica y comité de ética de Clínica Las Condes para realizar la investigación, conservándose confidencialidad para datos de las pacientes. Un tercer autor (H.V.C.), operador experimentado evaluó las imágenes ciego al diagnóstico definitivo, considerando la serie de alta dificultad diagnóstica por un gran número de masas sin patrón ecográfico de reconocimiento tumoral típico, eliminando 13 de las masas porque no reunían las condiciones mínimas para hacer un diagnóstico en base solo a las imágenes suministradas.

Un total de 4 examinadores que no participaron en etapas previas evaluaron por separado 4 imágenes y volúmenes para cada masa en 2D, Doppler y 3D adquiridas con sistema de ultrasonido Voluson E8 (GE Healthcare, Milwaukee, Wisconsin, USA) con transductores multifrecuencia endocavitario 5–9 MHz y/o convexo abdominal de 2–6 MHz cuando fue

necesario la exploración transabdominal. Sin recibir información clínica adicional de la paciente salvo el hecho que todas las masas correspondían a patología estructural no funcional, se les solicitó primero diferenciar entre tumor maligno (primario, borderline o metastásico) o benigno en base caracterización subjetiva y a los criterios de la clasificación GIRADS y en estas últimas la mayor precisión diagnóstica histopatológica posible escrita libremente sin aportarse lista de opciones diagnósticas a los examinadores. No se consideró como objetivo de estudio ni se le solicitó a los examinadores que especificara el tipo de tumor maligno (Borderline, primario o metastásico). Ninguna de las imágenes y volúmenes 3D en masas malignas mostraban ascitis.

Experiencia en ecografía ginecológica: Examinador A: más de 20 años y con especial dedicación a ecografía ginecológica avanzada. Examinador B: entre > 10 años y ≤ 20 años y con entrenamiento por tutor experto en ecografía ginecológica avanzada. Examinador C: entre > 5 años y ≤ 10 años y con entrenamiento por tutor experto en ecografía ginecológica avanzada. Examinador D: ≤ 5 años, ginecólogo general, sin entrenamiento en ecografía ginecológica avanzada.

El diagnóstico anatómico-patológico fue considerado como la “prueba de oro” de referencia. Para el análisis estadístico fueron considerados como malignos los tumores con bajo potencial de malignidad (TBPM).

El análisis estadístico fue efectuado por dos de los autores (H.V.C. y L.D.C.), cegados a datos de identificación y experiencia de cada examinador. Se calcularon números absolutos, porcentajes, medias, desviación estándar en las variables según programa SPSS versión 24,0 para Windows® (IBM, Chicago, IL, USA). También se utilizó el programa Excel 2019 versión 365 (Microsoft Office, Redmond, Washington, USA). Se determinó desde tablas 2x2: sensibilidad (S), especificidad (E), likelihood ratio positivo (LR+) y negativo (LR-), exactitud diagnóstica (ED) y Odds ratio diagnóstica (DOR). Los límites superior e inferior de IC95 % para S, E y ED se calcularon mediante método de Wilson³⁴; IC95 % de LR+ y LR- se calcularon según método descrito por Simel y cols³⁵; finalmente IC95 % para DOR calculado como describe Armitage y col³⁶.

Se utilizó la prueba de Chi cuadrado (X^2) y test exacto de Fisher cuando ambas variables fueron cualitativas, siendo este último método el que aportó los valores de p, ya que todas las tablas 2x2 presentaron por lo menos 1 casilla con valor menor de 5. Se consideró valores significativos estadísticamente cuando $p \leq 0,05$.

Likelihood ratio fue considerado como utilidad clínica^{23,26-28}:

- Altamente relevante, cuando LR+ > 10 y LR- < 0.1.
- Buena, cuando LR+ > 5 hasta 10 y LR- 0.1 hasta < 0.2.
- Regular, cuando LR+ > 2 hasta 5 y LR- 0.2 hasta < 0.5.
- Mala, cuando LR+ 0 hasta 2 y LR- 0.5 hasta 1.

Los resultados se expresaron en tablas y figuras. Se calcularon y graficaron nomogramas de Fagan con sus respectivas proyecciones desde probabilidad pre-test hasta probabilidad post-test según valor de likelihood ratio, siempre que pudo obtenerse en cada examinador para diferenciar malignos de benignos y también en detección específica para masas benignas.

Se solicitó a los examinadores emitir diagnóstico solo cuando el nivel de seguridad lo consideraran alto, y por esto se consideró alto nivel de confianza para emitir diagnóstico ecográfico específico si el examinador no respondió en menos del 10%, moderado si no respondió entre 10 % y 25 %, bajo cuando la falta de respuesta ocurrió en más del 25%.

Se utilizó guía STARD para elaboración del manuscrito³⁷.

RESULTADOS

De la muestra inicial de 108, se analizaron 87 masas como se detallan en figura 1. La prevalencia de malignos fue de 17,2 %. De las masas benignas 31 % eran endometriomas, 26,4 % teratomas, 19,5% cistoadenomas (11 serosos y 6 mucinosos), 4,6 % hidrosalpinx y 1,1 % fibroma ovárico.

Las estirpes en los tumores malignos fueron: 10 TBPM (6 serosos y 4 mucinosos), 3 primarios ováricos epiteliales de tipo invasivo (1 seroso, 1 mucinoso y 1 de células claras) y 2 tumores metastásicos al ovario (1 desde cáncer gástrico y 1 desde cáncer renal).

La edad media fue: 44,5 años (entre 25 y 72 años) para toda la muestra, siendo 51,2 años (entre 32 y 69 años) en malignas y 43,1 años (entre 25 y 72 años) en benignas.

Tabla 1 muestra aciertos, fallos y sin respuesta para los diferentes examinadores de diagnósticos ecográficos específicos según histopatología, lo que permite diferenciar el nivel de confianza para cada examinador al intentar acertar con imágenes y volúmenes ecográficos el diagnóstico histopatológico de las masas estudiadas.

Se observa rendimiento diagnóstico ecográfico específico para malignos en tabla 2 y para diferentes histopatologías de benignos en tabla 3.

En tabla 3 se agrupa hidrosalpinx y fibroma ovárico por la baja prevalencia de estas masas. Los examinadores A, B y C lograron identificar los 4 hidrosalpinx correctamente, en cambio examinador D solo detectó el 25 % (1/4) de estas masas. El examinador C fue el único que detectó el fibroma ovárico específicamente. En esta tabla fueron considerados como falsos negativos los fallos (diagnóstico escrito diferente a la masa correspondiente); la falta de respuesta diagnóstica fue cuantificada en casilla de falsos negativos o falsos positivos según correspondiese para cada masa y examinador.

En Tabla 4 se puede apreciar cómo cambia según LR+ la probabilidad de detección desde previo a posterior del diagnóstico ecográfico específico para cada examinador y por histopatología de masas.

Figura 2: para cada examinador se muestra el cambio de probabilidad post-test según acierto ecográfico para masas malignas y su discriminación de masas benignas. Nótese la evidente diferencia de área y vértice de ángulo al contrastar el cambio de probabilidad pre-test a post-test entre examinador con experiencia ≤ 5 años y aquellos de con mayor experiencia sobre todo al comparar con un examinador con más de 20 años de experiencia. Figuras 3, 4, 5 muestran proyección similar para diagnóstico específico en cistoadenoma, endometrioma y teratoma quístico maduro respectivamente.

Todos los cálculos mostraron significancia estadística según valor de $p \leq 0,05$.

DISCUSIÓN

Enfatizamos en expresar nuestros resultados con cálculos no influidos o menos modificados por la prevalencia del cáncer, como son ED, DOR y sobre todo LR+, parámetros más consistentes y útiles para utilizar en todas las situaciones y escenarios clínicos en torno a masas anexiales^{7-10,23,24,26-28,31}. A fue constante con ED siempre mayor a 95 % vs $D \leq 87\%$ a excepción del subtipo hidrosalpinx-fibroma, lo que puede ser explicado: 1) ED es la única de estas medidas que se modifica con la prevalencia pero proporcionalmente mucho menos que los valores predictivos; 2) al combinar sensibilidad y especificidad en un único indicador, se pierde el valor relativo de cada uno de éstos indicadores por separado y junto a prevalencia muy baja, a pesar de presentar una muy mala S, es enmascarada por ED; de igual forma también se explica que en el examinador D con S y E similar pero esta vez para masas malignas, la ED fue 10 puntos menos que en el subgrupo hidrosalpinx-fibroma. En este subgrupo consideramos como resultados útiles solo que examinadores con más de 5 años de experiencia son capaces de detectar todos lo hidrosalpinx vs solo 1 de 4 en examinadores con menor experiencia.

Difícil de precisar es cuánto tiempo de experiencia se requiere para efectuar informes confiables. Examinadores > 5 años siempre mostraron LR+ mayor a 10, ED mayor a 90 % y DOR mayor a 46, no así para examinador con menor tiempo de experiencia quien presentó resultados con mala utilidad clínica en base a LR-. Un salto cuantitativo en la capacidad diagnóstica específica parece estar en la experiencia en ecografía ginecológica con punto de corte a 5 años, con el mayor rendimiento constante para todos los subtipos de masas en examinador con más de 20 años de experiencia tanto para confirmar diagnóstico específico (LR+), como en su exclusión (LR-). Los modelos de puntuación, matemáticos y/o predictivos no son útiles en estas situaciones de alta exigencia diagnóstica³⁸, mostrando mayor número de masas con diagnóstico inconcluso³⁹, necesitando combinarse con caracterización subjetiva por expertos para disminuir diagnósticos no concluyentes y mejorar el rendimiento^{19,21,29,38,40}.

Entrenamientos específicos para ecografía de masas anexiales en base a volúmenes 3D y vídeos

han mostrado disminuir el tiempo necesario para lograr la experiencia deseada luego de 200 casos con alta prevalencia de cáncer en la serie¹⁰⁻¹³, pero lo que aún no se ha investigado es cuanto es necesario para que el operador mantenga un rendimiento diagnóstico óptimo en el tiempo.

Publicaciones²⁹ que estudiaron la capacidad diagnóstica específica de masas anexiales por ecografía, concluyeron que operadores nivel II (Europa)⁴¹ equivalentes al examinador C de nuestro estudio, no logran el rendimiento deseado²⁹. Nuestros resultados muestran un mejor rendimiento de detección específica que los ya reportados²⁹ tanto en operadores experimentados (nivel III)⁴¹, como con relativa experiencia (nivel II)⁴¹: para cistoadenomas 20-30 puntos porcentuales de diferencia, para endometriomas similar examinador C a nivel II y 10-20 puntos porcentuales mejor en expertos. Para teratomas los resultados de detección son similares. La diferencia anterior pudiese derivarse en que nuestra línea de estudio busca como principal meta la caracterización subjetiva para un diagnóstico específico, en cambio la otra investigación promueve mayormente los sistemas de puntuación menos operador-dependiente como la manera de resolver el diagnóstico. El rendimiento diagnóstico específico para los 5 subtipos de masas que estudiamos fue similar a publicaciones previas³, una de ellas con muestra 20 veces más numerosa⁴².

Un salto cuantitativo en la confianza para emitir diagnóstico específico considerando la falta de respuesta fue el corte de >10 años. Nuestros resultados son mejores con lo ya reportado (7 %) de ausencia diagnóstica en expertos⁴⁰ pero a expensas de mayor cantidad de subtipos de masas benignas por ser un gran estudio multicéntrico, siendo similares al 20 % de incertidumbre en examinadores con relativa experiencia³⁸. Sin embargo, no evaluamos el nivel de confianza para cada uno de los diagnósticos ecográficos emitidos por cada examinador, aunque la solicitud de emitir diagnóstico solo al sentir alta confianza puede solapar este punto y también explicar junto a masas con patrón ecográfico no típico las sensibilidades menores para diagnosticar cáncer, sorprendiéndonos que no ocurrió así para masas benignas, probablemente por lo importante de nuestra

línea en la caracterización subjetiva con diagnóstico ecográfico específico.

El grado de confianza diagnóstica es muy subjetivo y está estrictamente relacionado con la experiencia de los examinadores³⁸. En casos indeterminados, un examen de segundo nivel por un examinador de ultrasonido experimentado es beneficioso^{3,19,21,29,38,40}. Ningún modelo o marcador bioquímico de malignidad ovárica ha demostrado ser superior a la evaluación subjetiva ecográfica en escala de grises y Doppler^{3,19,21,29,38,40} y es bien conocido la implicancia que tiene para el bienestar de las pacientes^{1,7,8,21,43,44}.

El diagnóstico por imágenes falla por: 1) falta de reconocimiento de lesiones, 2) búsqueda fallida y mayormente 3) fracaso de decisión, factores íntimamente relacionados a la experiencia⁴⁵. Se ha demostrado que cuando la prevalencia cae del 50 % al 1 %, las tasas de detección también caen de 93 % a 70 %⁴⁵. Recientemente se ha redefinido el proceso de diagnóstico por expertos: reconocimiento inicial de una característica, seguido por una búsqueda y finalmente diagnosticar⁴⁵. Aunque ya en forma general se ha definido como experto aquel con más de 10 años en un área específica con dedicación preferible a tiempo completo^{46,38,41}, esto no es suficiente, siendo también factores la capacidad de memoria del operador⁴⁵ y el entrenamiento teórico y asesoramiento por un experto mientras se gana experiencia con la necesaria correlación ecográfica quirúrgica e histopatológica^{10,12,13}.

Los intervalos de confianza pudiesen estrecharse al aumentar la muestra de masas estudiadas. El empleo de p y de IC facilita la distinción entre significancia estadística y significancia clínica o importancia práctica³⁶. A mayor experiencia, el diagnóstico específico post-test es mayor. Nuestros resultados expresan importancia clínica.

La forma adecuada de hacer el diagnóstico es preparar a los operadores para mejorar su habilidad de discriminar las diversas alternativas de imágenes anexiales. En este campo las alternativas de exámenes en vivo (hands-on) son muy limitadas por la baja prevalencia del cáncer anexial en población no seleccionada. Por lo tanto, es fundamental diseñar un sistema de aprendizaje basado en bancos de imágenes, volúmenes 3D y sobre todo de vídeos que

acorten el periodo de entrenamiento, y que mejoren la calidad de los exámenes, así como también la confianza de los examinadores de emitir un diagnóstico específico presuntivo, que permita a sus médicos tratantes decidir según la situación clínica de cada paciente. Podrá ser el seguimiento conservador o la intervención quirúrgica por ginecólogo general o gineco-oncólogo, que aumenten las posibilidades de un adecuado manejo con la menor proporción de complicaciones por intervenciones médicas innecesarias^{9,43,44}.

La disponibilidad de expertos es de utilidad para emitir un informe adecuado que mejore la comunicación con los médicos tratantes, como muy recientemente se ha reportado en un metanálisis de GIRADS⁴⁷.

El diseño de esta investigación preliminar tiene algunas limitantes:

- Las conclusiones se han deducido de los resultados obtenidos por 4 evaluadores con diferente nivel de experiencia, en lugar de 4 grupos de evaluadores.
- Se analizaron cuatro imágenes fijas, en lugar de ver videos o mejor aún realizar la ecografía en tiempo real, el cual es un proceso dinámico que permite al operador de experiencia obtener el máximo del examen.
- Los examinadores carecieron de información clínica relevante, factores que muy probablemente incidirán significativamente en mejorar el diagnóstico ecográfico.

Como fortaleza y hasta donde sabemos, es el primer estudio con análisis estadístico que permite visualizar la relevancia en la práctica-clínica proponiendo la curva de aprendizaje necesaria para un rendimiento diagnóstico deseado. Sin embargo, serán necesarias investigaciones similares prospectivas con mayores muestras de masas estudiadas y preferiblemente en tiempo real, usando técnicas que permitan definir y acortar el cambio a experto para lograr un rendimiento óptimo que perdure en el tiempo.

CONCLUSIONES

Nuestros resultados sugieren que se necesitan más de 10 años de experiencia y con especial dedicación a ecografía ginecológica

avanzada para un rendimiento diagnóstico específico deseado junto con alta confianza en ecografía de masas anexiales. Examinadores entre > 5 y ≤ 10 años con entrenamiento por tutor experto en ecografía ginecológica avanzada muestran un rendimiento aceptable bueno con confianza moderada, en cambio examinadores con experiencia ≤ 5 años muestran rendimiento bajo y poca confianza para el diagnóstico ecográfico específico en masas anexiales.

REFERENCIAS

1. Timmerman D, Schwarzler P, Collins WP, Claerhout F, Coenen M, Amant F, et al. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. *Ultrasound Obstet Gynecol* 1999;13:11–6.
2. Liu JH, Zanotti KM. Management of the adnexal mass. *Obstet Gynecol*. 2011;117(6):1413-1428.
3. Coccia ME, Rizzello F, Romanelli C, Capezzuoli T. Adnexal masses: what is the role of ultrasonographic imaging?. *Arch Gynecol Obstet*. 2014;290(5):843-854.
4. Froyman W, Timmerman D. Methods of Assessing Ovarian Masses: International Ovarian Tumor Analysis Approach. *Obstet Gynecol Clin North Am*. 2019;46(4):625-641.
5. Andreotti RF, Timmerman D, Strachowski LM, Froyman W, Benacerraf B, Bennett G, et al. O-RADS US Risk Stratification and Management System: A Consensus Guideline from the ACR Ovarian-Adnexal Reporting and Data System Committee. *Radiology*. 2020;294(1):168-185.
6. Amor F, Alcázar JL, Vaccaro H, León M, Iturra A. GI-RADS reporting system for ultrasound evaluation of adnexal masses in clinical practice: a prospective multicenter study. *Ultrasound Obstet Gynecol*. 2011;38(4):450-455.
7. Amor F, Vaccaro H, Manukian B. US and adnexal pathology; looking at it the wrong way around. *Am J Obstet Gynecol*. 2016;215(5):676.
8. Diaz L, Zambrano B, Adami FJ, Alcazar JL. External Validation of Gynecological Imaging and Reporting Data System for Sonographic Evaluation of Adnexal Masses. *Donald School J Ultrasound Obstet Gynecol*. 2017;11(2):135-140.
9. Vaccaro H, Amor F, Rodriguez L, Díaz L. Eenie, meenie, minie, mo, What do the data really show?

- 35 años después. *Rev Chil Ultrasonog.* 2018;21(1): 50-52.
10. Alcázar JL, Díaz L, Flórez P, Guerriero S, Jurado M. Intensive training program for ultrasound diagnosis of adnexal masses: protocol and preliminary results. *Ultrasound Obstet Gynecol.* 2013;42(2):218-223.
 11. Zannoni L, Savelli L, Jokubkiene L, Di Legge A, Condous G, Testa A, et al. Intra- and interobserver reproducibility of assessment of Doppler ultrasound findings in adnexal masses. *Ultrasound Obstet Gynecol.* 2013;42(1):93-101.
 12. Ruiz de Gauna B, Sanchez P, Pineda L, Utrilla-Layna J, Juez L, Alcázar JL. Interobserver agreement in describing adnexal masses using the International Ovarian Tumor Analysis simple rules in a real-time setting and using three-dimensional ultrasound volumes and digital clips. *Ultrasound Obstet Gynecol.* 2014;44(1):95-99.
 13. Pineda L, Salcedo E, Vilhena C, Juez L, Alcázar JL. Interobserver agreement in assigning IOTA color score to adnexal masses using three-dimensional volumes or digital videoclips: potential implications for training. *Ultrasound Obstet Gynecol.* 2014;44(3):361-364.
 14. Ganesan A, Alakhras M, Brennan PC, Mello-Thoms C. A review of factors influencing radiologists' visual search behaviour. *J Med Imaging Radiat Oncol.* 2018;62(6):747-757.
 15. Beck MR, Martin BA, Smitherman E, Gaschen L. Eyes-on Training and Radiological Expertise: An Examination of Expertise Development and Its Effects on Visual Working Memory. *Hum Factors.* 2013;55(4):747-763.
 16. Valentin L, Hagen B, Tingulstad S, Eik-Nes S. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. *Ultrasound Obstet Gynecol* 2001;18(4):357–365.
 17. Yazbek J, Raju SK, Ben-Nagi J, Holland TK, Hillaby K, Jurkovic D. Effect of quality of gynaecological ultrasonography on management of patients with suspected ovarian cancer: a randomised controlled trial. *Lancet Oncol.* 2008;9(2):124-131.
 18. Levine D, Brown DL, Andreotti RF, et al. Management of asymptomatic ovarian and other adnexal cysts imaged at US: Society of Radiologists in Ultrasound Consensus Conference Statement. *Radiology.* 2010;256(3):943–954.
 19. Meys EM, Kaijser J, Kruitwagen RF, et al. Subjective assessment versus ultrasound models to diagnose ovarian cancer: A systematic review and meta-analysis. *Eur J Cancer.* 2016;58:17–29.
 20. American College of Obstetricians and Gynecologists' Committee on Practice Bulletins-Gynecology. Practice Bulletin No. 174: Evaluation and Management of Adnexal Masses. *Obstet Gynecol.* 2016;128:e210-e226.
 21. Glanc P, Benacerraf B, Bourne T, et al. First International Consensus Report on Adnexal Masses: Management Recommendations. *J Ultrasound Med* 2017;36(5):849–863.
 22. Stempel LE. Eenie, meenie, minie, mo . . . What do the data really show?. *Am J Obstet Gynecol.* 1982;144(7):745-752.
 23. Salech F, Mery V, Larrondo F, Rada G. Estudios que evalúan un test diagnóstico: interpretando sus resultados. *Rev Méd Chile* 2008; 136:1203-1208.
 24. Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol.* 2008;31(6):681-690.
 25. Eusebi P. Diagnostic accuracy measures. *Cerebrovasc Dis.* 2013;36(4):267-272.
 26. Aznar-Orovala E, Mancheño-Alvarob A, García-Lozanoa T, M. Sánchez-Yepes M. Razón de verosimilitud y nomograma de Fagan: 2 instrumentos básicos para un uso racional de las pruebas del laboratorio clínico. *Rev Calid Asist.* 2013;28(6):390-393.
 27. Bravo-Grau S, Cruz JP. Estudios de exactitud diagnóstica: Herramientas para su Interpretación. *Revista Chilena de Radiología.* 2015;21(4):158-164.
 28. Silva C, Molina M. Likelihood ratio (razón de verosimilitud): definición y aplicación en Radiología. *Rev Argent de Radiol.* 2017;81(3): 204-208.

29. Sayasneh A, Kaijser J, Preisler J, Smith A, Raslan F, Johnson S, et al. Accuracy of ultrasonography performed by examiners with varied training and experience in predicting specific pathology of adnexal masses. *Ultrasound Obstet Gynecol.* 2015;45(5):605-612.
30. Alcázar JL, Aubá M, Ruiz-Zambrana A, Olartecoechea B, Diaz D, Hidalgo J, et al. Ultrasound assessment in adnexal masses: an update, *Expert Rev Obstet Gynecol.* 2012;7(5):441-449.
31. Smith-Bindman R, Poder L, Johnson E, Miglioretti DL. Risk of Malignant Ovarian Cancer Based on Ultrasonography Findings in a Large Unselected Population. *JAMA Intern Med.* 2019;179(1):71-77.
32. Parazzini F, Frattaruolo MP, Chiaffarino F, Dridi D, Roncella E, Vercellini P. The limited oncogenic potential of unilocular adnexal cysts: A systematic review and meta-analysis. *Eur J Obstet Gynecol Reprod Biol.* 2018;225:101-109.
33. Hack K, Glanc P. The Abnormal Ovary: Evolving Concepts in Diagnosis and Management. *Obstet Gynecol Clin North Am.* 2019;46(4):607-624.
34. Agresti A, Coull B. (1998). Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician.* 1998;52(2):119-126.
35. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol.* 1991;44(8):763-770.
36. Armitage P, Berry G. *Estadística para la investigación biomédica.* 3ra ed. Madrid: Harcourt Brace; 1997.
37. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ.* 2016;6(11):e012799.
38. Szubert S, Szperek D, Wójtowicz A, Żywica P, Stukan M, Sajdak S, et al. Performance of Selected Models for Predicting Malignancy in Ovarian Tumors in Relation to the Degree of Diagnostic Uncertainty by Subjective Assessment With Ultrasound. *J Ultrasound Med.* 2020;39(5):939-947.
39. Van Calster B, Van Hoorde K, Valentin L, Testa AC, Fischerova D, Van Holsbeke C, et al. Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study. *BMJ.* 2014;349:g5920.
40. Valentin L, Ameye L, Savelli L, Fruscio R, Leone FP, Czekierdowski A, et al. Adnexal masses difficult to classify as benign or malignant using subjective assessment of gray-scale and Doppler ultrasound findings: logistic regression models do not help. *Ultrasound Obstet Gynecol.* 2011;38(4):456-465.
41. Education and Practical Standards Committee, European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB). Minimum training recommendations for the practice of medical ultrasound. *Ultraschall Med* 2006; 27: 79–105.
42. Alcázar JL, Guerriero S, Laparte C, Ajossa S, Ruiz-Zambrana A, Melis GB. Diagnostic performance of transvaginal gray-scale ultrasound for specific diagnosis of benign ovarian cysts in relation to menopausal status. *Maturitas.* 2011;68(2):182-188.
43. Alcázar JL, Olartecoechea B, Guerriero S, Jurado M. Expectant management of adnexal masses in selected premenopausal women: a prospective observational study. *Ultrasound Obstet Gynecol.* 2013;41(5):582-588.
44. Pascual MA, Graupera B, Pedrero C, Rodriguez I, Ajossa S, Guerriero S, et al. Long-term Results for Expectant Management of Ultrasonographically Diagnosed Benign Ovarian Teratomas. *Obstet Gynecol.* 2017;130(6):1244-1250.
45. Brunyé TT, Drew T, Weaver DL, Elmore JG. A review of eye tracking for understanding and improving diagnostic interpretation. *Cogn Res Princ Implic.* 2019;4(1):7
46. Ericsson KA, Krampe, RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychological Review.* 1993;100:363–406.
47. Rodríguez L, Díaz L, Vaccaro H, Amor F, Yamamoto M, Olid M, et al. GIRADS: Método de

clasificación de masas anexiales, revisión sistemática y metanálisis. 10 años de experiencia. Rev Chil Obstet Ginecol. 2020;85(5):468-485.

TABLAS Y FIGURAS

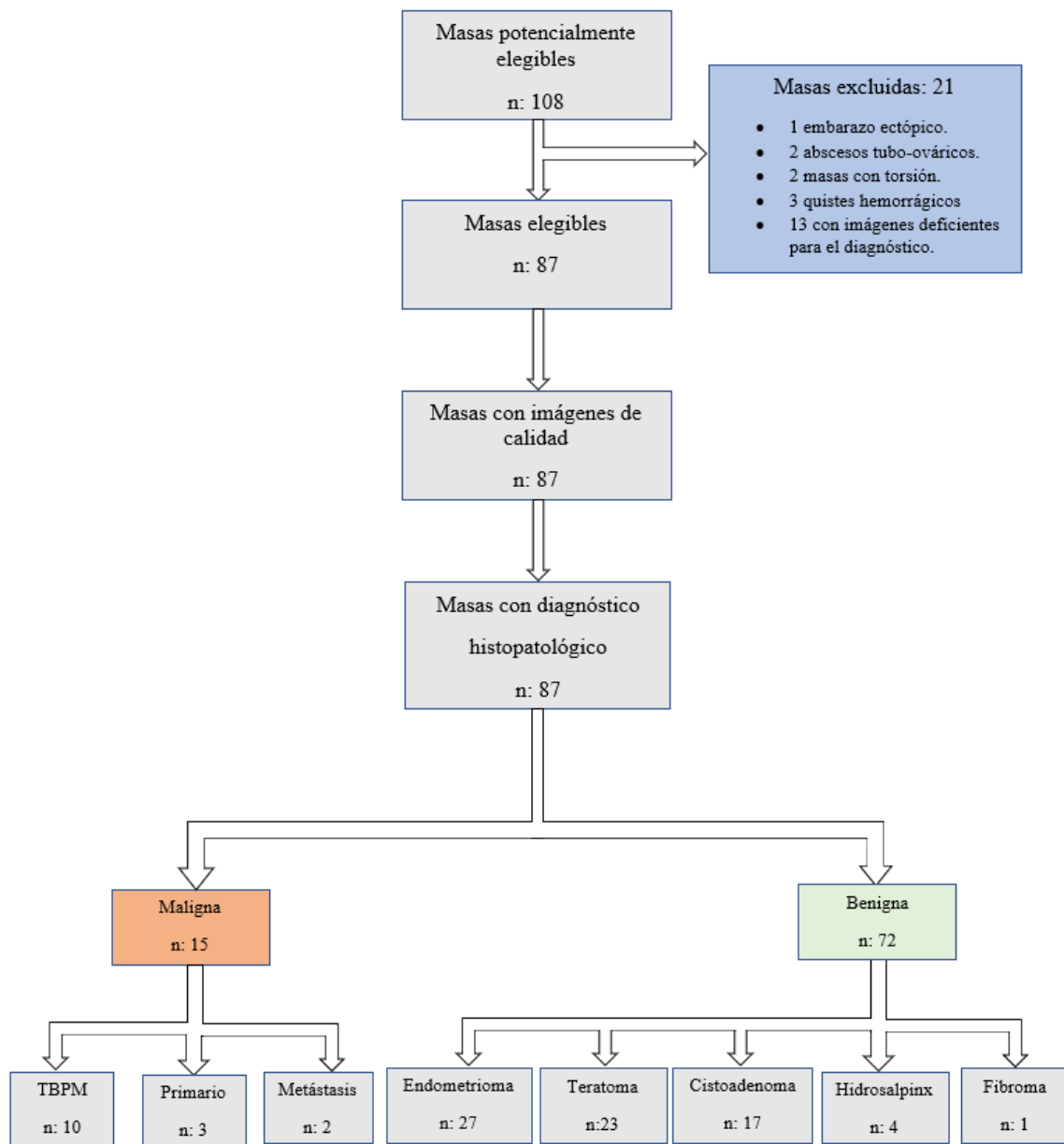


Figura 1. Flujograma de masas estudiadas.

Tabla 1. Aciertos, fallos y sin respuesta diagnóstica específica para cada examinador.

	A	B	C	D
Aciertos	75 (86.2 %)	74 (85.1 %)	64 (73.6 %)	35 (40.2 %)
Fallos	7 (8.1 %)	12 (13.8 %)	5 (5.7 %)	20 (23 %)
Sin respuesta	5 (5.7 %)	1 (1.1 %)	18 (20.7 %)	32 (36.8 %)
Total	87 (100 %)	87 (100 %)	87 (100 %)	87 (100 %)

Tabla 2: Rendimiento diagnóstico específico de patología maligna para cada examinador.

	S % (IC95 %)	E % (IC95 %)	LR + (IC95 %)	LR – (IC95 %)	ED % (IC95 %)	DOR (IC95 %)	P
A	67 (43 – 91)	99 (96 – 100)	48 (6.6 – 347.2)	0.3 (0.1 – 0.6)	93.1 (80.3 – 100)	142 (15 – 1342)	< 0,001
B	73 (51 – 96)	96 (91 – 100)	17.6 (5.5 – 55.5)	0.2 (0.1 – 0.6)	91.9 (78.1 – 100)	63.9 (12 – 321)	< 0,001
C	67 (43 – 91)	96 (91 – 100)	16.0 (5.0 – 51.2)	0.3 (0.1 – 0.7)	90.8 (76.2 – 100)	46.5 (10 – 222)	< 0,001
D	20 (0 – 40)	99 (96 – 100)	14.4 (1.6 – 129.1)	0.8 (0.6 – 1.0)	85.0 (67 – 100)	17.8 (2 – 185)	< 0,001

p según test exacto de Fisher. S: sensibilidad. E: especificidad. LR+: Likelihood ratio positivo. LR-: Likelihood ratio negativo. ED: exactitud diagnóstica. DOR: Odds ratio diagnóstica. IC95 %: intervalo de confianza al 95%.

Tabla 3. Rendimiento diagnóstico específico de patología benigna para cada examinador.

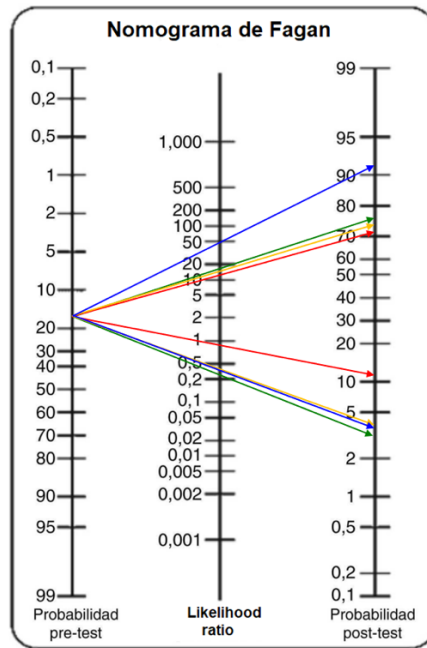
	S % (IC95 %)	E % (IC95 %)	LR + (IC95 %)	LR – (IC95 %)	ED % (IC95 %)	DOR	p
Cistoadenoma. Prevalencia: 19.5 % (17/87)							
A	88 (73 – 100)	97 (93 – 100)	30.8 (7.8 – 122.3)	0.1 (0.03 – 0.4)	95.4 (85.4 – 100)	255 (33 – 1957)	< 0,001
B	76 (56 – 97)	97 (93 – 100)	26.7 (6.6 – 107.5)	0.2 (0.1 – 0.5)	93.1 (81 – 100)	110 (18 – 667)	< 0,001
C	82 (64 – 100)	97 (93 – 100)	28.8 (7.2 – 114.9)	0.1 (0.06 – 0.5)	94.2 (83 – 100)	158 (24 – 1039)	< 0,001
D	82 (64 – 100)	89 (81 – 96)	7.2 (3.6 – 14.3)	0.2 (0.07 – 0.5)	87.3 (71.4 – 100)	36 (8 – 153)	< 0,001
Endometrioma Prevalencia: 31.0 % (27/87)							
A	96 (89 – 100)	97 (92 – 100)	28.8 (7.3 – 113)	0.04 (0.01 – 0.2)	96.5 (89.5 – 100)	754 (65 – 8690)	< 0,001
B	89 (77 – 100)	95 (89 – 100)	17.7 (5.8 – 54)	0.1 (0.04 – 0.3)	93.1 (83.5 – 100)	152 (28 – 807)	< 0,001
C	78 (62 – 93)	98 (95 – 100)	46.6 (6.6 – 329.3)	0.2 (0.1 – 0.4)	91.9 (81.9 – 100)	206 (23 – 1817)	< 0,001
D	44 (26 – 63)	97 (92 – 100)	13.3 (3.2 – 55.5)	0.5 (0.4 – 0.8)	80.4 (68.2 – 92.6)	23 (5 – 115)	< 0,001
Teratoma Prevalencia: 26.4 % (23/87)							
A	87 (73 – 100)	100 (94 – 100)	–	0.1 (0.04 – 0.3)	96.5 (89 – 100)	–	< 0,001
B	96 (87 – 100)	97 (92 – 100)	30.6 (7.8 – 120)	0.04 (0.01 – 0.3)	94.2 (84.7 – 100)	682 (59 – 7897)	< 0,001
C	61 (41 – 81)	100 (94 – 100)	–	0.3 (0.2 – 0.6)	90 (77.8 – 100)	–	< 0,001
D	22 (5 – 39)	98 (93 – 100)	13.9 (1.7 – 112.8)	0.8 (0.6 – 1.0)	78.1 (61.2 – 95)	18 (2 – 159)	< 0,001
Hidrosalpinx (4) + Fibroma (1): Prevalencia: 5.7 % (5/87)							
A	80 (45 – 100)	100 (96 – 100)	–	0.2 (0.03 – 1.0)	98.8 (89.3 – 100)	–	< 0,001
B	80 (45 – 100)	100 (96 – 100)	–	0.2 (0.03 – 1.0)	98.8 (89.3 – 100)	–	< 0,001
C	100 (60 – 100)	100 (96 – 100)	–	0.0 (0.0 – 0.0)	100 (94 – 100)	–	< 0,001
D	20 (0 – 55)	100 (96 – 100)	–	0.8 (0.5 – 1.0)	95.4 (77.1 – 100)	–	< 0,001

p según test exacto de Fisher. S: sensibilidad. E: especificidad. LR+: Likelihood ratio positivo. LR-: Likelihood ratio negativo. ED: exactitud diagnóstica. DOR: Odds ratio diagnóstica. IC95 %: intervalo de confianza al 95%.

Tabla 4. Cambio de probabilidad pre-test a post-test positivo según diagnóstico específico de masas para cada examinador.

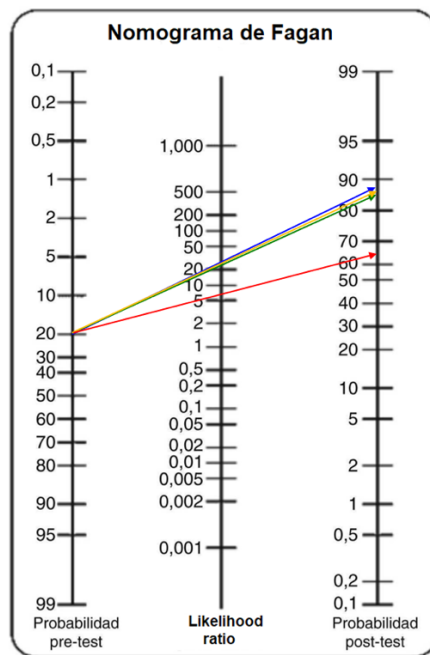
	Probabilidad pre-test %	LR+ (IC95 %)	Probabilidad post-test + %	p
Malignidad				
A	17.2	48 (6.6 – 347.2)	92	< 0,001
B	17.2	17.6 (5.5 – 55.5)	78	< 0,001
C	17.2	16.0 (5.0 – 51.2)	76	< 0,001
D	17.2	14.4 (1.6 – 129.1)	73	< 0,001
Cistoadenoma				
A	19.5	30.8 (7.8 – 122.3)	88	< 0,001
B	19.5	26.7 (6.6 – 107.5)	86	< 0,001
C	19.5	28.8 (7.2 – 114.9)	87	< 0,001
D	19.5	7.2 (3.6 – 14.3)	63	< 0,001
Endometrioma				
A	31	28.8 (7.3 – 113)	94	< 0,001
B	31	17.7 (5.8 – 54)	89	< 0,001
C	31	46.6 (6.6 – 329.3)	95	< 0,001
D	31	13.3 (3.2 – 55.5)	87	< 0,001
Teratoma				
A	26.4	–	–	< 0,001
B	26.4	30.6 (7.8 – 120)	92	< 0,001
C	26.4	–	–	< 0,001
D	26.4	13.9 (1.7 – 112.8)	80	< 0,001
Hidrosalpinx Fibroma				
A	5.7	–	–	< 0,001
B	5.7	–	–	< 0,001
C	5.7	–	–	< 0,001
D	5.7	–	–	< 0,001

p según test exacto de Fisher. LR+: Likelihood ratio positivo. IC95 %: intervalo de confianza al 95%.



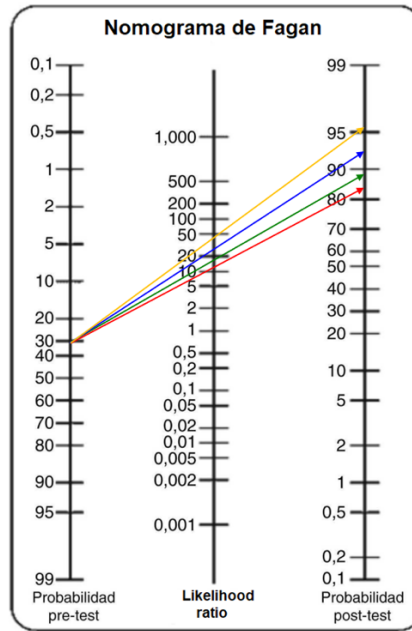
Examinador A: ----- Examinador B: ----- Examinador C: ----- Examinador D: -----

Figura 2. Cambio de probabilidad pre-test a post-test según diagnóstico de cáncer vs otras masas para todos los examinadores.



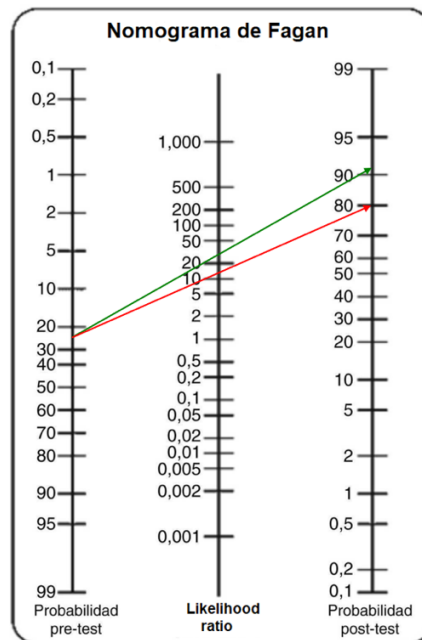
Examinador A: ----- Examinador B: ----- Examinador C: ----- Examinador D: -----

Figura 3. Cambio de probabilidad pre-test a post-test positivo según diagnóstico específico de cistoadenoma vs otras masas para todos los examinadores.



Examinador A: - - - - - Examinador B: - - - - - Examinador C: - - - - - Examinador D: - - - - -

Figura 4. Cambio de probabilidad pre-test a post-test positivo según diagnóstico específico de endometrioma vs otras masas para todos los examinadores.



Examinador B: - - - - - Examinador D: - - - - -

Figura 5. Cambio de probabilidad pre-test a post-test positivo según diagnóstico específico de teratoma vs otras masas para examinadores B vs D.