



clínica e investigación en ginecología y obstetricia

www.elsevier.es/gine



REVISIÓN DE CONJUNTO

Sistemas de clasificación y predicción de riesgo de malignidad de las lesiones anexiales



J.J. Hidalgo Mora^{a,c,*}, N. Rams Llop^b, F. Ros Bernal^c y J.L. Alcázar Zambrano^d

^a Servicio de Ginecología y Obstetricia, Hospital Comarcal Universitario de Vinaròs, Castellón, España

^b Servicio de Ginecología y Obstetricia, Hospital de la Santa Creu i Sant Pau, Barcelona, España

^c Unidad Predepartamental de Medicina, Facultad de Ciencias de la Salud, Universitat Jaume I, Castellón, España

^d Departamento de Obstetricia y Ginecología, Clínica Universidad de Navarra, Facultad de Medicina, Universidad de Navarra, Pamplona, España

Recibido el 25 de enero de 2017; aceptado el 27 de marzo de 2017

Disponible en Internet el 15 de mayo de 2017

PALABRAS CLAVE

Masas anexiales;
Ecografía;
Modelos predictivos;
Sistemas de
puntuación;
Triage

KEYWORDS

Adnexal masses;
Ultrasonography;
Predictive models;
Scoring systems;
Triage

Resumen El diagnóstico de una masa anexial es un problema frecuente en la consulta ginecológica. El principal objetivo de la evaluación de una masa anexial es el diagnóstico o exclusión de malignidad, dado que el cáncer de ovario es la neoplasia ginecológica más letal y que una cirugía inicial adecuada es uno de sus principales factores pronósticos. La evaluación ecográfica continúa siendo el mejor método para clasificar las masas anexiales. Si existe la sospecha de malignidad de la lesión, la paciente debe ser remitida a un centro especializado en ginecología oncológica. Sin embargo, si la sospecha es de benignidad, se puede optar por una actitud expectante o por un tratamiento quirúrgico conservador.

El objetivo de este artículo es revisar los sistemas de clasificación de las lesiones anexiales más importantes y más ampliamente empleados, analizando su metodología y los resultados de su aplicación en los principales estudios de validación publicados hasta el momento.

© 2017 Elsevier España, S.L.U. Todos los derechos reservados.

Classification systems and prediction of risk of malignancy of the adnexal masses

Abstract The diagnosis of an adnexal mass is a common problem in gynaecological consultation. The main objective of an adnexal mass evaluation is the diagnosis or exclusion of malignancy. This is the case because ovarian cancer is the most lethal gynaecological neoplasia and appropriate initial surgery is one of the main prognostic factors. Ultrasound scans continue to be the best method of classifying an adnexal mass. If there is any suspicion of a malignant tumour, the patient must be referred to a specialist gynaecological oncology centre.

* Autor para correspondencia.

Correo electrónico: hidalgomorajj@gmail.com (J.J. Hidalgo Mora).

However, if there is any suspicion of a benign tumour, watchful waiting or minimally invasive surgery may be indicated.

The objective of this article is to carry out a review of the most important and widely used classification systems of adnexal masses, analysing their methodology and the results of their application in the main validation studies published to date.

© 2017 Elsevier España, S.L.U. All rights reserved.

Introducción

Las masas anexiales constituyen un problema frecuente en la consulta ginecológica y una de las dolencias que más dudas genera a la hora de realizar un diagnóstico preciso y su posterior enfoque terapéutico. Se estima que un 13-21% de las pacientes con una masa pélvica presentarán una neoplasia ovárica epitelial invasiva¹. Además, el cáncer de ovario es uno de los más letales en los países desarrollados —es la sexta causa de muerte oncológica entre las mujeres europeas— por lo que resulta fundamental realizar un diagnóstico preciso con relación a su benignidad o malignidad².

Las lesiones con alta sospecha de benignidad pueden ser manejadas de forma expectante o con cirugía conservadora en centros no especializados, mientras que las probablemente malignas deben ser remitidas a centros especializados en ginecología oncológica, dado que un diagnóstico temprano y preciso y una adecuada cirugía inicial de estadificación y citorreducción se encuentran entre los factores pronósticos más importantes del cáncer de ovario. Sin embargo, menos de la mitad de estas pacientes son intervenidas inicialmente por ginecólogos oncológicos especializados, lo que en la mayoría de los casos se produce por un inadecuado diagnóstico preoperatorio³⁻⁸.

De los 3 parámetros empleados clásicamente en el diagnóstico diferencial de las masas anexiales: 1) la exploración física, 2) los niveles de marcadores tumorales y 3) la ecografía, los 2 primeros han demostrado un valor diagnóstico limitado, por lo que la valoración ecográfica sigue siendo la mejor herramienta para establecer una sospecha de benignidad o malignidad con una precisión adecuada. Aunque en muchos casos las características ecográficas de una masa anexial orientan hacia su naturaleza exacta, el objetivo de su estudio no debe ser tanto establecer este diagnóstico de certeza «anatomopatológico» como determinar la probabilidad de su benignidad o malignidad.

El análisis ecográfico de las lesiones anexiales puede realizarse con la simple valoración subjetiva del ecografista (*pattern recognition*), lo que puede superar el 90% de precisión diagnóstica en manos expertas, o con la aplicación de sistemas o modelos matemáticos predictivos basados en la presencia o ausencia de determinadas características morfológicas de las lesiones. Estos modelos de predicción, desarrollados desde principios de los años 90, no han demostrado hasta ahora mejorar el rendimiento diagnóstico de la valoración subjetiva por ecografistas expertos. Además, ninguno de ellos ha logrado extenderse en la práctica clínica diaria^{7,9-15}.

En este artículo nos planteamos realizar una revisión de los sistemas de clasificación de las lesiones anexiales más importantes y más utilizados en la práctica clínica, analizando su metodología y los resultados de su aplicación en los principales estudios de validación publicados hasta el momento.

Sistemas de clasificación de las lesiones anexiales

El análisis ecográfico de una masa anexial no debe limitarse a describir su apariencia morfológica, pero tampoco debe pretender sistemáticamente dar un diagnóstico específico e histológico de esta. Su objetivo fundamental debe ser realizar su clasificación o *triaje* para establecer su riesgo de malignidad, lo que puede realizarse de forma subjetiva o mediante sistemas objetivos.

Una crítica general al uso de la ecografía como único test diagnóstico para las lesiones anexiales es su subjetividad y su rendimiento dependiente de la experiencia y la habilidad del ecografista. Por eso, se han propuesto modelos predictivos basados en variables tanto ecográficas como epidemiológicas, clínicas o analíticas, con los objetivos de calcular individualmente el riesgo de malignidad de la lesión, apoyar a ecografistas menos experimentados de una forma objetiva y orientar a los clínicos en el manejo terapéutico de la paciente.

Desde que Grandberg et al.¹⁶ señalaran en 1989 que la valoración ecográfica por vía vaginal de una masa ovárica podía predecir la probabilidad de su malignidad, se han sucedido las propuestas de sistemas de puntuación y modelos predictivos basados en la ecografía con el objetivo de generar reglas y patrones de interpretación que ayuden a ecografistas menos experimentados a reproducir los resultados de los expertos para clasificar las masas ováricas con un nivel de precisión aceptable¹⁷. Sin embargo, ninguno de estos modelos ha sido aceptado de forma general para ser aplicado en la práctica clínica diaria. En 2 revisiones recientes se ha señalado la existencia de 116 diferentes modelos de predicción desarrollados para clasificar las masas anexiales como benignas o malignas, lo que da idea tanto de los esfuerzos realizados en este campo como de la falta de consistencia de los resultados^{11,15}.

En cualquier caso, en todo estudio ecográfico de una lesión anexial, además de su descripción morfológica, se debería incluir la valoración de su riesgo de malignidad, aun teniendo en cuenta que tanto para su diagnóstico

Tabla 1 Rendimiento diagnóstico de los principales sistemas de clasificación de las lesiones anexiales

	Sensibilidad (IC 95%)	Especificidad (IC 95%)
<i>Pattern recognition</i> (Valentin et al., 2001)	88-98	89-96
<i>Simple rules</i> (Timmerman et al., 2008)	93-95	90-91
RMI (Jacobs et al., 1990)	85	97
LR1 (Timmerman et al., 2005)	93	77
LR2 (Timmerman et al., 2005)	92	75
CA-125 + HE4 (Moore et al., 2008)	76	95
ROMA (Moore et al., 2009)	88	74
GI-RADS (Amor et al., 2009)	92	97
<i>Three steps strategy</i> (Ameje et al., 2012)	92	92
ADNEX (van Calster et al., 2014)	96	71

ADNEX: *Assessment of Different Neoplasias in the adnexa*; HE4: *human epididymis 4*; LR: *logistic regression model*; RMI: *Risk of Malignancy Index*; ROMA: *Risk of Ovarian Malignancy Algorithm*.

definitivo como para la posterior actitud terapéutica se han de considerar otros factores de la paciente como su edad, antecedentes personales y familiares, estado menopáusico, síntomas y situación clínica.

Clasificación subjetiva

El término *pattern recognition* fue introducido por Valentin en 1999 para referirse a la capacidad de un ecografista de establecer un diagnóstico específico o la probabilidad de benignidad o malignidad de una lesión ovárica con la simple valoración subjetiva de sus características ecográficas¹⁸. Diversos estudios han demostrado que este sistema, realizado por ecografistas experimentados, es el más preciso para caracterizar una masa anexial¹⁹⁻²¹; puede alcanzar una sensibilidad del 88-98% y una especificidad del 89-96% para establecer la probabilidad de su benignidad o malignidad, o superar el 90% de precisión al realizar su diagnóstico específico^{9,22}.

En algunos trabajos, el rendimiento diagnóstico del *pattern recognition* en el *triaje* de las masas anexiales ha superado al de los algoritmos y modelos objetivos, como los modelos de regresión logística de Taylor y Timmerman, el sistema de puntuación *Risk of Malignancy Index* (RMI) o el basado únicamente en niveles de biomarcadores *Risk of Ovarian Malignancy Algorithm* (ROMA)^{2,22,23} (tabla 1).

Sin embargo, a pesar de que el *pattern recognition* se ha señalado como la mejor forma de diagnóstico de cáncer ovárico, algunos estudios han mostrado que un 10% de los tumores son difíciles de clasificar incluso por ecografistas experimentados (*borderline*, cistoadenofibromas, cistoadenomas, fibromas y tumores benignos raros). En este grupo de tumores la precisión global del diagnóstico por un experto se sitúa alrededor del 68%, con pobres sensibilidad y

especificidad, lo que señala la conveniencia de contar con un test secundario seguro. En este sentido, se ha propuesto tanto la aplicación de los sistemas objetivos de *triaje* como la medición aislada de marcadores tumorales (como el CA125 y la proteína epididimal humana 4 [HE4]), o su incorporación a modelos predictivos como el ROMA, con resultados contradictorios hasta el momento²³.

Clasificación objetiva

El *triaje* subjetivo de las lesiones anexiales presenta una precisión diagnóstica y una reproducibilidad interobservador limitadas en caso de examinadores poco experimentados. Durante las últimas 3 décadas se han propuesto múltiples modelos predictivos con el fin de proporcionar a ecografistas y clínicos herramientas que contribuyan a realizar este *triaje* de forma más objetiva y con un rendimiento diagnóstico similar al alcanzado por la valoración subjetiva de ecografistas expertos. Algunos de estos modelos valoran exclusivamente las características ecográficas de las lesiones, otros emplean únicamente los niveles de marcadores tumorales y la mayoría combinan variables epidemiológicas, ecográficas, clínicas y analíticas, valoradas con reglas de presencia/ausencia, sistemas de puntuación multimodales, modelos de regresión logística o algoritmos matemáticos.

Para poder ser utilizados en la práctica clínica, los modelos predictivos deben poder proporcionar un diagnóstico preciso, ser de aplicación simple y resultar útiles en la práctica habitual. Además, requieren ser validados externamente ya que tienden a ser sobrevalorados al aplicarlos en la población con la que fueron desarrollados¹⁵.

Reglas ecográficas

El grupo *International Ovarian Tumour Analysis* (IOTA) propuso en 2008 las denominadas *simple rules*, consistentes en valorar la existencia o ausencia de determinadas características típicas de las lesiones benignas y malignas²⁴.

Para elaborar estas «reglas simples», el grupo IOTA aplicó retrospectivamente a 1.233 lesiones anexiales diferentes combinaciones de variables ecográficas testando su capacidad para predecir benignidad o malignidad mediante el cálculo de su sensibilidad, especificidad, valor predictivo positivo y negativo y curva ROC. Finalmente fueron elegidas 10 reglas en función de estos resultados, de la experiencia clínica de los miembros de IOTA y de su aplicabilidad al mayor número de tumores (tabla 2). Las *simple rules* pudieron ser aplicadas al 76% de los 1.233 tumores, con una sensibilidad del 93%, una especificidad del 90%, un valor predictivo positivo del 80% y un valor predictivo negativo del 97%, mientras que un 24% de los tumores no pudieron ser clasificados con ellas. Después, fueron evaluadas prospectivamente en 507 lesiones: pudieron ser aplicadas también al 76% de ellas, con una sensibilidad del 95% y una especificidad del 91%²⁴.

Teniendo en cuenta los buenos resultados, la aplicación de estas reglas podría ofrecer a ecografistas con poca experiencia una buena precisión diagnóstica, además de ser un método de aplicación sencilla en la práctica clínica. Sin embargo, una importante limitación de este sistema de clasificación es el alto porcentaje de lesiones (alrededor del 25%) en las que estas reglas no son aplicables, debido a que

Tabla 2 Parámetros ecográficos valorados en las *simple rules* (Timmerman et al., 2008)**Reglas para predecir malignidad (M-rules)**

- M1 Tumor sólido irregular
- M2 Presencia de ascitis
- M3 Al menos 4 estructuras papilares
- M4 Tumor multilocular sólido irregular con diámetro ≥ 100 mm
- M5 Flujo sanguíneo muy intenso (*score* de color 4)

Reglas para predecir benignidad (B-rules)

- B1 Unilocular
- B2 Presencia de componentes sólidos donde el mayor tiene un diámetro mayor a 7 mm
- B3 Presencia de sombras acústicas
- B4 Tumor multilocular liso con diámetro < 100 mm
- B5 Ausencia de flujo sanguíneo (*score* de color 1)

no todas las masas anexiales presentan características claramente predictivas de benignidad o malignidad. Cuando no son aplicables, las opciones podrían consistir en remitir a la paciente a un ecografista experto como un segundo test diagnóstico, considerar como malignas las lesiones no concluyentes o realizar una resonancia magnética en las masas difíciles de clasificar²⁵.

Sistemas de puntuación

Se han propuesto diversos sistemas de puntuación o *scores* que pretenden objetivar y hacer posible la reproducción interobservador de la valoración de las lesiones ováricas y que ecografistas menos expertos consigan una mayor precisión diagnóstica²⁶. De estos, el RMI, desarrollado por Jacobs et al. en 1990, es probablemente el más usado, dado que no es necesaria una gran experiencia para detectar las características ecográficas que deben ser puntuadas; incluso se ha propuesto en una revisión sistemática reciente como el sistema de clasificación con un mejor rendimiento^{12,15}. Además, es el modelo para el cual se han realizado más trabajos de validación externa¹⁵.

Este índice se calcula por un algoritmo basado en diferentes variables y se define como $RMI = U \times M \times CA125$, donde U es el *score* ecográfico, M el estado menopáusico y $CA125$ el nivel del marcador. U es calculado dando un punto a distintos parámetros ecográficos (multilocularidad, áreas sólidas, bilateralidad, ascitis y metástasis intraabdominales), de forma que un *score* de 0, 1 y ≥ 2 puntos otorga a la variable U los valores de 0, 1 y 3 respectivamente. La variable M adquiere valor de 1 en caso de premenopausia y de 3 en caso de posmenopausia. Estos valores, junto al de $CA125$ en U/ml, son introducidos en la ecuación, de forma que un valor ≥ 200 es sugestivo de malignidad^{2,12}.

En diversos estudios este modelo ha mostrado buenos resultados en cuanto a su rendimiento diagnóstico general y en su comparación con modelos totalmente objetivos como el ROMA, aunque resulta inferior al ser comparado con el *pattern recognition*^{1,2,15,27,28}.

Modelos de regresión logística

Los modelos de regresión logística tienen como objetivo estimar una probabilidad (un riesgo individualizado de malignidad) en función de factores o variables predictoras o

independientes. Presentan el inconveniente de que necesitan un *software* informático y que en ocasiones su aplicación puede resultar compleja. Además, se ha apuntado que su aplicación no mejora los resultados del *pattern recognition*.

Los primeros modelos de regresión logística desarrollados para el *triaje* de las lesiones anexiales fueron propuestos a finales de los años 90^{19,29}. Sin embargo, los resultados de los primeros estudios de validación externa no fueron demasiado prometedores, con sensibilidad y especificidad en alguno de ellos del 62 y 79%, respectivamente²². Para mejorar estos resultados, el grupo IOTA desarrolló en 2005 un modelo, el LR2 (*Logistic Regresión model*), en un estudio con 1.066 pacientes y con la valoración de más de 50 variables³⁰. Inicialmente fue desarrollada una versión con las 12 variables más importantes para el cálculo de probabilidad de malignidad (LR1) y se propuso posteriormente un modelo reducido con las 6 variables con mayor impacto (LR2): edad (1), ascitis (2), flujo sanguíneo en proyección papilar (3), diámetro máximo del componente sólido (4), paredes quísticas internas irregulares (5) y sombras acústicas (6).

LR2 estima la probabilidad de malignidad para una lesión anexial con la fórmula $1/(1+\exp(-z))$, donde $z = -5,3718 + 0,0354(1) + 1,6159(2) + 1,1768(3) + 0,0697(4) + 0,9586(5) - 2,9486(6)$, considerándose un punto de corte de probabilidad de 0,1 (10%) para clasificar las lesiones como benignas o malignas. Una ventaja de LR2 es que proporciona un riesgo absoluto de presentar un cáncer de ovario, lo cual puede contribuir a aconsejar a la paciente y tomar una decisión terapéutica compartida²⁵. Además, aunque la fórmula matemática anterior pueda parecer de cálculo complejo, para facilitar su aplicación práctica, el grupo IOTA ha desarrollado aplicaciones para dispositivos móviles que permiten obtener su resultado de forma inmediata para una lesión anexial sin necesidad de realizar cálculos matemáticos complejos.

En los estudios de validación interna y externa de este modelo los resultados han sido favorables, con una sensibilidad por encima del 90% en todos ellos y un mejor rendimiento diagnóstico en su comparación con otros sistemas como RMI o ROMA²⁹⁻³¹. De forma general, los modelos propuestos por el grupo IOTA han presentado mejores resultados que los modelos no desarrollados por IOTA: son LR1 y LR2 los que han mostrado el mejor rendimiento diagnóstico³².

Marcadores tumorales

Para dotar de la mayor objetividad posible a los modelos predictivos, se han desarrollado algoritmos que incluyen únicamente el uso de biomarcadores para realizar el *triaje* de las lesiones anexiales, como el ROMA o el OVA-1^{8,33}.

El marcador tumoral clásicamente empleado en el diagnóstico de las masas anexiales ha sido el CA125. Su valor es limitado al encontrarse elevado en menos de la mitad de las neoplasias ováricas en estadios tempranos, no ser expresado en un 20% de los tumores epiteliales de ovario y estar aumentado en enfermedades ginecológicas benignas en mujeres premenopáusicas y en enfermedades médicas en la posmenopausia, por lo que su uso sistemático no aporta mejor información clínica que la ecografía realizada por expertos^{8,21,34,35}.

El nuevo biomarcador que ha mostrado mejores resultados ha sido el HE4, secretado, entre otros tejidos, por el aparato genital femenino y sobreexpresado en cánceres ováricos serosos y endometrioides. Está elevado en más del 80% de las neoplasias epiteliales ováricas y menos frecuentemente que el CA125 en enfermedad benigna en mujeres premenopáusicas. Además, aumenta en más del 50% de los tumores que no expresan CA125, con lo que la suma de HE4 a CA125 permite la detección de neoplasia en pacientes con tumores que no expresan CA125, lo que aumenta la sensibilidad para la determinación del riesgo de malignidad de una masa anexial^{36,37}.

La combinación de HE4 y CA125 presenta la más alta sensibilidad entre las diversas combinaciones de marcadores tumorales³⁷, lo que fue empleado en 2009 en el desarrollo del ROMA, incorporando ambos marcadores a una serie de algoritmos matemáticos para clasificar las masas anexas en bajo o alto riesgo de malignidad⁸.

En estudios de validación posteriores los resultados han sido contradictorios. Mientras en algunos de ellos el ROMA se presenta con mejores resultados que otros modelos como el RMI, argumentando que los algoritmos basados en los biomarcadores presentan inherentemente una naturaleza objetiva y que la omisión de medidas subjetivas facilita la reproducibilidad entre centros⁸, en otros se señala que este algoritmo o el uso del HE4 no supera a otros test diagnósticos que emplean CA125 como biomarcador único^{31,38}, y que pueden incluso empeorar el rendimiento diagnóstico global en caso de emplearse tras la valoración ecográfica por un experto²³.

Nuevas estrategias de triaje

La utilidad relativa en la práctica clínica de los diferentes modelos predictivos propuestos y los resultados en muchos casos contradictorios de los estudios en los que se han evaluado han llevado a que se hayan propuestos nuevas estrategias para lograr una mayor precisión en la clasificación de las lesiones anexas.

Uno de estos nuevos modelos consiste en un sistema en 2 pasos. En un primer paso se realiza la valoración ecográfica subjetiva de la lesión anexial y posteriormente se mide el nivel sérico de HE4 o se aplica el algoritmo ROMA. Sin embargo, en un estudio de validación de este modelo, si en un primer paso el ecografista proponía un diagnóstico de benignidad o malignidad, la medición posterior del HE4 o la aplicación del ROMA mejoraba muy ligeramente, e incluso en algunos casos empeoraba, los resultados en términos de sensibilidad y especificidad. Incluso al aplicar el segundo paso a las lesiones difíciles de clasificar para un ecografista experimentado, se obtuvo una baja capacidad de clasificación con una pobre sensibilidad, con lo que esta estrategia en 2 pasos no parece útil para mejorar el rendimiento diagnóstico de la valoración ecográfica por expertos (tabla 3)²³.

Otra de estas nuevas estrategias consiste en un modelo de valoración ecográfica de las lesiones anexas que supone tanto una clasificación de su riesgo de malignidad como un nuevo sistema de comunicación de la información para los clínicos, denominado GI-RADS (*Gynecologic Imaging Reporting and Data System*) y basado en el sistema BI-RADS empleado para las lesiones mamarias. Clasifica las lesiones

Tabla 3 Rendimiento diagnóstico de la estrategia en 2 pasos para la clasificación de las lesiones anexas (Kaijser et al., 2014)

	Sensibilidad (IC 95%)	Especificidad (IC 95%)
<i>Masas anexas clasificadas como benignas</i>		
IS	97,2	89,8
IS + HE4 (<70 pmol/l)	97,9	79,6
IS + ROMA	97,9	75,9
<i>Masas anexas clasificadas como malignas</i>		
IS	97,2	89,8
IS + HE4 (≥70 pmol/l)	73,6	94,9
IS + ROMA	83,3	91,7

HE4: *human epididymis 4*; IS: impresión subjetiva; ROMA: *Risk of Ovarian Malignancy Algorithm*.

Tabla 4 Descriptores instantáneos en el sistema de diagnóstico en 3 pasos de IOTA (Ameye et al., 2012)

<i>Predictores de benignidad</i>
Tumor unilocular con ecogenicidad en vidrio esmerilado en mujer premenopáusica
Tumor unilocular con ecogenicidad mixta y sombras acústicas en mujer premenopáusica
Tumor unilocular anecoico con paredes regulares y máximo diámetro < 100 mm
Tumor unilocular restante con paredes regulares
<i>Predictores de malignidad</i>
Tumor con ascitis y al menos flujo vascular moderado con Doppler color en mujer posmenopáusica
Edad > 50 años y CA125 > 100 U/ml

anexas en GI-RADS 1 (definitivamente benigna), GI-RADS 2 (muy probablemente benigna), GI-RADS 3 (probablemente benigna), GI-RADS 4 (probablemente maligna) y GI-RADS 5 (muy probablemente maligna). Esta clasificación define el riesgo de malignidad basado en datos de estudios previos y orienta al clínico en el abordaje terapéutico: conducta expectante frente a cirugía por el ginecólogo general o cirugía por un ginecólogo oncólogo³⁹. El modelo ha sido validado internamente con resultados prometedores (sensibilidad del 92% y especificidad del 97%) y externamente por 2 grupos, que han obtenido también un buen rendimiento diagnóstico (sensibilidad del 99% y especificidad del 86-90%)³⁹⁻⁴².

Recientemente el grupo IOTA ha propuesto un sistema de triaje basado en la aplicación de 3 pasos (*three-step strategy*) que parece proporcionar mejores resultados que los modelos previos y una mayor aproximación a la realidad clínica. En un primer paso denominado «diagnóstico fácil inmediato» (*easy «instant» diagnosis*), a las lesiones anexas les son aplicados 6 descriptores (4 de benignidad y 2 de malignidad) basados en la información ecográfica y en los niveles del CA125 (tabla 4), de forma que, si ninguno de los 6 descriptores es aplicable a la lesión o se pueden aplicar uno benigno y uno maligno, el diagnóstico es considerado difícil y «no inmediato». En un segundo paso, los casos en los que no se puede realizar el diagnóstico inmediato son valorados a través de las *simple rules*. Finalmente, el tercer paso consiste en la valoración subjetiva por un ecografista

experto en los casos en los que no pueda realizarse un diagnóstico con los 2 pasos previos. En la estrategia en 3 pasos se apuntó en su estudio de validación interna una sensibilidad y especificidad global por encima del 92%: se observó que aproximadamente el 50% de las lesiones pueden ser diagnosticadas de forma inmediata, sin dificultad, por la mayoría de los clínicos (paso 1) y que el 80% son diagnosticadas con los 2 primeros pasos por ecografistas no expertos. Cuando los 2 primeros pasos no son concluyentes, la evaluación subjetiva por un experto (paso 3) es el test diagnóstico más preciso, mientras que RMI, LR1 y LR2 no resultan útiles⁴³. En los escasos estudios de validación externa de este sistema realizados hasta el momento, los resultados han sido todavía mejores que los iniciales, con una sensibilidad del 93-94% y una especificidad del 92-95%^{44,45}. De esta forma, esta valoración combinada podría cambiar la práctica clínica, al proporcionar una clasificación precisa e inmediata de la mayoría de las masas anexiales: es reducido el número de pacientes que requerirían ser derivados a un centro de referencia para el examen ecográfico por un experto.

Por último, también el grupo IOTA ha desarrollado recientemente el modelo predictivo de regresión logística multinomial *Assessment of Different Neoplasias in the adnexa* (ADNEX) con el objetivo de dar un paso más en la discriminación entre tumores ováricos benignos y malignos al intentar discriminar entre 5 tipos de tumores anexiales (benignos, *borderline*, invasivos en estadio I, invasivos en estadios II-IV y metastásicos). Para ello valora 3 variables clínicas (edad, nivel de CA125 y tipo de centro hospitalario) y 6 ecográficas (diámetro máximo de la lesión, proporción de tejido sólido, presencia de más de 10 lóculos, número de proyecciones papilares, presencia de sombra acústica y presencia de ascitis). En su estudio de validación interna, el rendimiento diagnóstico del modelo mostró en la discriminación entre lesiones benignas y malignas (con un punto de corte del 10% para predecir malignidad) una sensibilidad del 96,5% y una especificidad del 71,3%, con una buena capacidad para distinguir entre sí los 5 tipos de tumores, lo que es planteado por los autores como un sistema con una buena capacidad para mejorar las decisiones terapéuticas en mujeres con lesiones anexiales⁴⁶.

Limitaciones en la aplicación clínica de los modelos predictivos

Ninguno de los modelos predictivos propuestos para la clasificación de las lesiones anexiales ha logrado una aceptación generalizada ni su empleo sistemático en la práctica clínica habitual. Esto puede obedecer a las limitaciones que han mostrado los estudios tanto de propuesta como de validación de los modelos.

Existen pocos trabajos de validación externa para la mayoría de los modelos de predicción: hasta 2013, solo 19 de los 116 modelos propuestos habían sido validados externamente en al menos 2 estudios¹¹. Además la mayoría de ellos son de tipo retrospectivo y con sesgos en su diseño como la falta de la valoración ciega por parte de los investigadores o ecografistas. De hecho, en los trabajos de validación realizados prospectivamente la mayoría de los modelos presentaron peores resultados que en los estudios en los que fueron creados¹⁷.

Tampoco se han desarrollado muchos trabajos de comparación directa entre modelos, entre otros motivos porque los puntos de corte adoptados varían y porque en algunos casos cada uno emplea variables diferentes, lo que complica la interpretación y la comparación de resultados. Además, la mayoría de los estudios de validación no estratifican los resultados en función de variables importantes como el estado menopáusico, por lo que las conclusiones podrían no ser del todo correctas cuando no se produce esta estratificación o se realiza con posterioridad a la clasificación de las lesiones.

Además, aunque su objetivo es que ecografistas con un nivel bajo de competencia puedan proporcionar un diagnóstico de benignidad/malignidad con una precisión adecuada, muchos modelos han sido desarrollados y validados por ecografistas expertos que habrían podido determinar el diagnóstico mediante el *pattern recognition*, lo que habría contribuido positivamente a la precisión de estos modelos¹⁶.

Muchos estudios se han desarrollado en centros de referencia con unidades de ecografía especializada y con una alta prevalencia de lesiones malignas, lo que implica una influencia en los resultados de los valores predictivos. De esta forma, en centros de referencia, la mayor prevalencia de cáncer de ovario haría que el valor predictivo positivo tendiera a disminuir, mientras que en hospitales pequeños con baja prevalencia de cáncer ovárico, un test con la misma sensibilidad y especificidad tendría un valor predictivo positivo más bajo. Además, en los centros de referencia la experiencia de los ecografistas es mayor y, en consecuencia, mejor el rendimiento diagnóstico ecográfico.

Por otra parte, la mayoría de los trabajos están realizados con pacientes intervenidas quirúrgicamente, con el estudio histológico como *gold standard* de diagnóstico final. Sin embargo, existe poca información de pacientes en las que se haya optado por una conducta expectante y sobre los resultados de esta opción terapéutica en forma de errores de clasificación, complicaciones o transformación maligna. Además, en gran parte de los estudios solo se valora la variable dicotómica «benigno/maligno», sin considerar que existen diversos tipos de enfermedad maligna (*borderline*, estadio temprano/avanzado, metástasis de neoplasias no ováricas) cuyo manejo terapéutico es diferente.

Finalmente otro problema que ha hecho que estos sistemas predictivos no se hayan consolidado ha sido la falta de semejanza de las condiciones de los estudios con la práctica clínica habitual. Los estudios realizados para su desarrollo y validación han incluido todo tipo de lesiones anexiales, con independencia de si en la práctica clínica real estas lesiones podrían haber sido fácilmente diagnosticadas por los ecografistas. Como resultado, la precisión diagnóstica de estos test probablemente haya sido sobrestimada y su rendimiento puede resultar peor del esperado al ser aplicados en la clínica más allá del entorno de investigación⁴³.

Conclusiones

El diagnóstico preciso y precoz de la benignidad o malignidad de una lesión anexial resulta fundamental para la orientación terapéutica adecuada de las pacientes, con la ecografía como el método de elección en este proceso diagnóstico. Hasta el momento, el método de *triaje* que ha demostrado

mejores resultados ha sido la valoración subjetiva de ecografistas expertos, aunque presenta los inconvenientes de su gran dependencia de la subjetividad y competencia del ecografista y de la dificultad para su reproducibilidad inter-observador.

Para intentar lograr una mayor objetividad en la clasificación de las lesiones anexiales y un más alto nivel de precisión diagnóstica por ecografistas con menos nivel de experiencia, se han desarrollado múltiples modelos predictivos basados en el análisis de variables ecográficas, epidemiológicas y clínicas de las pacientes, así como en el valor de diferentes marcadores tumorales, sin que ninguno de ellos haya demostrado mejores resultados que la valoración subjetiva por expertos, lo que puede deberse en gran medida a la dificultad y a los sesgos en el planteamiento y diseño de los estudios para su desarrollo y validación.

Como nuevas estrategias de clasificación se han propuesto modelos consistentes en la aplicación sucesiva de diferentes sistemas de *triaje* complementarios, para intentar mejorar los resultados de la aplicación de cada uno de ellos de forma aislada. Aunque los buenos resultados obtenidos en la validación interna de algunos de ellos y su accesible aplicación práctica señalan que podrían proporcionar una buena aproximación a la realidad clínica, deben realizarse más estudios para su validación externa. Esto resulta de gran importancia, dado que se ha observado que los modelos diagnósticos predictivos presentan mejores resultados al aplicarlos en las poblaciones en las que fueron desarrollados, por lo que debe comprobarse si mantienen su rendimiento diagnóstico en diferentes poblaciones y situaciones clínicas.

Conflicto de intereses

Los autores firmantes de este artículo declaran no haber contado con fuentes de financiación para su elaboración ni presentar ningún tipo de conflicto de intereses que pueda ser relevante para su publicación.

Bibliografía

- Moore RG, Jabre-Raughley M, Brown AK, Robinson KM, Miller MC, Allard WJ, et al. Comparison of a novel multiple marker assay versus the Risk of Malignancy Index for the prediction of epithelial ovarian cancer in patients with a pelvic mass. *Am J Obstet Gynecol.* 2010;203:e1–6.
- Van Gorp T, Veldman J, van Calster B, Cadron I, Leunen K, Amant F, et al. Subjective assessment by ultrasound is superior to the risk of malignancy index (RMI) or the risk of ovarian malignancy algorithm (ROMA) in discriminating benign from malignant adnexal masses. *Eur J Cancer.* 2012;164:9–56.
- Paulsen T, Kjaerheim K, Kaern J, Tretli S, Tropé C. Improved short-term survival for advanced ovarian, tubal, and peritoneal cancer patients operated at teaching hospitals. *Int J Gynecol Cancer.* 2006;16 Suppl 1:11–7.
- Engelen MJ, Kos HE, Willemse PH, Aalders JG, de Vries EG, Schaapveld M, et al. Surgery by consultant gynecologic oncologists improves survival in patients with ovarian carcinoma. *Cancer.* 2006;106:589–98.
- Earle CC, Schrag D, Neville BA, Yabroff KR, Topor M, Fahey A, et al. Effect of surgeon specialty on processes of care and outcomes for ovarian cancer patients. *J Natl Cancer Inst.* 2006;98:172–80.
- Giede KC, Kieser K, Dodge J, Rosen B. Who should operate on patients with ovarian cancer? An evidence-based review. *Gynecol Oncol.* 2005;99:447–61.
- Alcázar JL, Royo P, Jurado M, Mínguez JA, García-Manero M, Laparte C, et al. Triage for surgical management of ovarian tumors in asymptomatic women: Assessment of an ultrasound-based scoring system. *Ultrasound Obstet Gynecol.* 2008;32:220–5.
- Moore RG, McMeekin DS, Brown AK, DiSilvestro P, Miller MC, Allard WJ, et al. A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol.* 2009;112:40–6.
- Sokalska A, Timmerman D, Testa AC, van Holsbeke C, Lissoni AA, Leone FPG. Diagnostic accuracy of transvaginal ultrasound examination for assigning a specific diagnosis to adnexal masses. *Ultrasound Obstet Gynecol.* 2009;34:462–70.
- Guerriero S, Alcazar JL, Ajossa S, Galvan R, Laparte C, Garcia-Manero M, et al. Transvaginal color Doppler imaging in the detection of ovarian cancer in a large study population. *Int J Gynecol Cancer.* 2010;20:781–6.
- Kaijser J, Sayasneh A, van Hoorde K, Ghaem-Maghami S, Bourne T, Timmerman D, et al. Presurgical diagnosis using mathematical models and scoring systems: A systematic review and meta-analysis. *Hum Reprod Update.* 2014;20:449–62.
- Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *Br J Obstet Gynaecol.* 1990;97:922–9.
- Sassone AM, Timor-Tritsch IE, Artner A, Westhoff C, Warren WB. Transvaginal sonographic characterization of ovarian disease: Evaluation of a new scoring system to predict ovarian malignancy. *Obstet Gynecol.* 1991;78:70–6.
- Bourne TH, Campbell S, Reynolds KM, Whitehead MI, Hampson J, Royston P, et al. Screening for early familial ovarian cancer with transvaginal ultrasonography and colour blood flow imaging. *BMJ.* 1993;306:1025–9.
- Geomini P, Kruitwagen R, Bremer GL, Cnossen J, Mol BW. The accuracy of risk scores in predicting ovarian malignancy: A systematic review. *Obstet Gynecol.* 2009;113:384–94.
- Grandberg S, Wikland M, Jansson I. Macroscopic characterization of ovarian tumors and the relation to the histological diagnosis: Criteria to be used for ultrasound evaluation. *Gynecol Oncol.* 1989;35:139–44.
- Nunes N, Yazbek J, Ambler G, Hoo W, Naftalin J, Jurkovic D. Prospective evaluation of the IOTA Logistic Regression Model (LR2) for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol.* 2012;40:355–9.
- Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: The contribution of Doppler ultrasound. *Ultrasound Obstet Gynecol.* 1999;14:338–47.
- Timmerman D, Bourne T, Taylor A, Collins WP, Verrelst H, Vandenberghe K, et al. A comparison of methods for the pre-operative discrimination between benign and malignant adnexal masses: The development of a new logistic regression model. *Am J Obstet Gynecol.* 1999;181:57–65.
- Timmerman D. The use of mathematical models to evaluate pelvic masses; can they beat an expert operator? *Best Pract Res Clin Obstet Gynaecol.* 2004;18:91–104.
- Valentin L, Jurkovic D, van Calster B, Testa A, van Holsbeke C, Bourne T, et al. Adding a single CA 125 measurement to ultrasound imaging performed by an experienced examiner does not improve preoperative discrimination between benign and malignant adnexal masses. *Ultrasound Obstet Gynecol.* 2009;34:345–54.
- 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;18:357–65.
23. Kaijser J, van Gorp T, Smet ME, van Holsbeke C, Sayasneh A, Epstein E, et al. Are serum HE4 or ROMA scores useful to experienced examiners for improving characterization of adnexal masses after transvaginal ultrasonography? *Ultrasound Obstet Gynecol.* 2014;43:89–97.
 24. Timmerman D, Testa A, Bourne T, Ameye L, Jurkovic D, van Holsbeke C. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol.* 2008;31:681–90.
 25. Sayasneh A, Wynants L, Preisler J, Kaijser J, Johnson S, Stalder C, et al. Multicenter external validation of IOTA prediction models and RMI by operators with varied training. *Br J Cancer.* 2013;108:2448–54.
 26. Lee TS, Kim JW, Park NH, Song YS, Kang SB, Lee HP. Assessing clinical performance of gynecology residents: Sonographic evaluation of adnexal masses based on morphological scoring systems. *Ultrasound Obstet Gynecol.* 2005;26:776–9.
 27. Bailey J, Tailor A, Naik R, Lopes A, Godfrey K, Hatem HM, et al. Risk of malignancy index for referral of ovarian cancer cases to a tertiary center: Does it identify the correct cases? *Int J Gynecol Cancer.* 2006;16 Suppl 1:30–4.
 28. Jacob F, Meier M, Caduff R, Goldstein D, Pochechueva T, Hacker N, et al. No benefit from combining HE4 and CA125 as ovarian tumor markers in a clinical setting. *Gynecol Oncol.* 2011;121:487–91.
 29. Tailor A, Jurkovic D, Bourne TH, Collins WP, Campbell S. Sonographic prediction of malignancy in adnexal masses using multivariate logistic regression analysis. *Ultrasound Obstet Gynecol.* 1997;10:41–7.
 30. Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, et al. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: A multicenter study by the International Ovarian Tumor Analysis Group. *J Clin Oncol.* 2005;23:8794–801.
 31. Kaijser J, van Gorp T, van Hoorde K, van Holsbeke C, Sayasneh A, Vergote I, et al. A comparison between an ultrasound based prediction model (LR2) and the risk of ovarian malignancy algorithm (ROMA) to assess the risk of malignancy in women with an adnexal mass. *Gynecol Oncol.* 2013;129:377–83.
 32. Van Holsbeke C, van Calster B, Bourne T, Ajossa S, Testa AC, Guerriero S, et al. External validation of diagnostic models to estimate the risk of malignancy in adnexal masses. *Clin Cancer Res.* 2012;18:815–25.
 33. Ueland FR, Desimone CP, Seamon LG, Miller RA, Goodrich S, Podzielinski I, et al. Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. *Obstet Gynecol.* 2011;117:1289–97.
 34. Timmerman D, van Calster B, Jurkovic D, Valentin L, Testa AC, Bernard JP, et al. Inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and malignant adnexal tumors. *J Clin Oncol.* 2007;25:4194–200.
 35. Van Calster B, Timmerman D, Bourne T, Testa AC, van Holsbeke C, Domali E, et al. Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125. *J Natl Cancer Inst.* 2007;99:1706–14.
 36. Drapkin R, von Horsten HH, Lin Y, Mok SC, Crum CP, Welch WR, et al. Human epididymis protein 4 (HE4) is a secreted glycoprotein that is overexpressed by serous and endometrioid ovarian carcinomas. *Cancer Res.* 2005;65:2162–9.
 37. Moore RG, Brown AK, Miller MC, Skates S, Allard WJ, Verch T, et al. The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass. *Gynecol Oncol.* 2008;108:402–8.
 38. Van Gorp T, Cadron I, Despierre E, Daemen A, Leunen K, Amant F, et al. HE4 and CA125 as a diagnostic test in ovarian cancer: Prospective validation of the risk of ovarian malignancy algorithm. *Br J Cancer.* 2011;104:863–70.
 39. Amor F, Vaccaro H, Alcázar JL, León M, Cgaig JM, Martínez J. Gynecologic imaging reporting and data system: A new proposal for classifying adnexal masses on the basis of sonographic findings. *J Ultrasound Med.* 2009;28:285–91.
 40. 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;38:450–5.
 41. Qiuyue C, Guorong L. The applicational value of GI-RADS ultrasonographic stratification in diagnosing adnexal masses. *Chinese J Ultrasound Med.* 2013;29:527–9.
 42. Rams N, Muñoz R, Soler C, Parra J. Resultados de la clasificación Gynecologic Imaging Reporting and Data System para la catalogación de masas anexiales. *Prog Obstet Ginecol.* 2015;58:125–9.
 43. Ameye L, Timmerman D, Valentin L, Paladini D, Zhang J, van Holsbeke C, et al. Clinically oriented three-step strategy for assessment of adnexal pathology. *Ultrasound Obstet Gynecol.* 2012;40:582–91.
 44. Sayasneh A, Kaijser J, Preisler J, Johnson S, Stalder C, Husicka R, et al. A multicenter prospective external validation of the diagnostic performance of IOTA simple descriptors and rules to characterize ovarian masses. *Gynecol Oncol.* 2013;130:140–6.
 45. Alcázar JL, Pascual MA, Graupera B, Auba M, Errasti T, Olartecoechea B, et al. External validation of IOTA simple descriptors and simple rules for classifying adnexal masses. *Ultrasound Obstet Gynecol.* 2016;48:397–402, <http://dx.doi.org/10.1002/uog.15854>.
 46. Van Calster B, van Hoorde K, Valentin L, Testa A, 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.