

## Trabajo Original

### GIRADS: método de clasificación de masas anexiales, revisión sistemática y metanálisis. 10 años de experiencia

### GIRADS: adnexal masses classification method, systematic review and metanalysis. 10 years of experience

Rodríguez G., Luis; MD<sup>1</sup>, Díaz C., Linder; MD<sup>2</sup>, Vaccaro C., Humberto; MD<sup>1</sup>, Amor L., Fernando; MD<sup>1</sup>, Yamamoto C., Masami; MD<sup>3</sup>, Olid M., Maria I.; MD<sup>4</sup>, Alcázar Z., Juan Luis; MD, PhD<sup>5</sup>.

<sup>1</sup> Centro Ecográfico Ultrasonic Panorámico, Santiago, Chile.

<sup>2</sup> Centro de Ultrasonografía AGB, Clínica Sanatorio Alemán, Concepción, Chile.

<sup>3</sup> Departamento de Obstetricia y Ginecología, Clínica Universidad de los Andes, Santiago, Chile.

<sup>4</sup> Departamento de Obstetricia y Ginecología, Hospital Mérida, Mérida, España.

<sup>5</sup> Departamento de Obstetricia y Ginecología, Clínica Universidad de Navarra; Pamplona, España.

#### Correspondencia

Nombre: Luis Armando Rodríguez Guzmán

Email: larodriguezg@hotmail.com

#### RESUMEN

**Objetivo:** evaluar la experiencia en la utilización del método GIRADS para clasificar masas anexiales a diez años de su primera publicación.

**Método:** Se realizó búsqueda de estudios que utilizan el sistema GIRADS: Medline (Pubmed), Google Scholar y Web of Science, desde enero de 2009 hasta diciembre de 2019. Se calculó la sensibilidad y especificidad agrupada, Likelihood ratio (LR) (+) y LR (-) y Odds ratio de diagnóstico (DOR). La calidad de los estudios se evaluó con QUADAS-2.

**Resultados:** Se identificaron 15 estudios y se incluyeron 13 de ellos con 4473 masas, 878 de ellas malignas. La prevalencia media de malignidad ovárica fue del 23 % y la agrupada de 19.6%. El riesgo de sesgo fue alto en cuatro estudios para el dominio "selección de pacientes" y fue bajo en todos en todos los estudios para los dominios "prueba índice" y "prueba de referencia". La sensibilidad, especificidad, LR (+) y LR (-) agrupadas y el DOR del sistema GIRADS para clasificar las masas anexiales fueron: 96.8% (intervalo de confianza [IC] 95% = 94% – 98%), 91.2 % (IC 95 % = 85% – 94%), 11.0 (IC 95% = 6.9 –13.4) y 0.035 (IC 95% = 0.02– 0.09), y 209 (IC 95% = 99-444), respectivamente. La heterogeneidad fue alta para la sensibilidad y especificidad. De acuerdo a la metaregresión, la heterogeneidad entre los estudios se explica por la prevalencia de malignidad, múltiples observadores y la ausencia de diagnóstico histopatológico para todos los casos incluidos en un determinado estudio.

**Conclusión:** el sistema GIRADS tiene un buen rendimiento diagnóstico para clasificar masas anexiales.

**Palabras claves:** GIRADS; neoplasias ováricas; diagnóstico de cáncer de ovario; masas anexiales, Gynecologic Imaging Reporting and Data System.

### ABSTRACT

**Objective:** to evaluate the experience of using GIRADS method to classify adnexal masses ten years after its publication.

**Method:** A search was carried out for studies reporting on the use of the GIRADS system in the Medline (Pubmed), Google Scholar and Web of Science databases, from January 2009 to December 2019. Pooled sensitivity and specificity, Likelihood ratio (LR) (+) and LR (-) and Diagnostic Odds ratio (DOR) were calculated. The quality of the studies was assessed by QUADAS-2.

**Results:** 15 studies were identified, and 13 of them were included with 4473 masses, of which 878 were malignant. The mean prevalence of ovarian malignancy was 23% and the prevalence pooled. of 19.6%. The risk of bias was high in four studies for the domain 'patient selection' and low for all studies for the domains 'index test' and 'reference test'. The sensitivity, specificity, pooled LR (+) and LR (-) and the DOR of the GIRADS system to classify adnexal masses were 96.8% (95% confidence interval [CI] = 94% -98%), 91.2 % (95% CI = 85% – 94%), 11.0 (95% CI = 6.9–13.4) and 0.035 (95% CI = 0.02–0.09), and 209 (95% CI = 99-444), respectively. Heterogeneity was high for both sensitivity and specificity. According to meta-regression, this heterogeneity was explained by the prevalence of malignancy, the use of multiple observers, and the absence of histopathological diagnosis for all cases included in a given study.

**Conclusion:** the GIRADS system has a good diagnostic performance to classify adnexal masses.

**Keywords:** GIRADS; ovarian neoplasms; diagnosis of ovarian cancer; adnexal masses, Gynecologic Imaging Reporting and Data System.

### Cuadro de abreviaturas

ABREVIATURA	ESPAÑOL	INGLÉS
GIRADS	Gynecologic Imaging Reporting and Data System	Gynecologic Imaging Reporting and Data System
DOR	Odds ratio diagnóstica	Diagnostic odds ratio
QUADAS-S	Evaluación de la calidad de los estudios de precisión diagnóstica	Quality Assessment of Diagnostic Accuracy Studies
PICOS	Pacientes, intervención, comparador, resultados, diseño del estudio	Patients, intervention, comparator, results, study design
LR	Likelihood ratio	Likelihood ratio
IC	Intervalo de confianza	Confidence interval
USG	Ultrasonido ginecológico	Gynecological ultrasound

## PUNTOS CLAVE

GIRADS es un sistema de notificación de riesgo de malignidad en masas ováricas / anexiales.

El metanálisis mostró un 96.8% y 91.2 % de sensibilidad y especificidad.

El reconocimiento subjetivo de las masas ováricas / anexiales es importante para un diagnóstico ecográfico correcto.

## INTRODUCCIÓN

Las masas anexiales son un problema común en Ginecología. Afortunadamente, la mayoría de ellos son benignas. Se estima que aproximadamente el 10 % de las mujeres tendrán en su vida una patología anexial, de las cuales la gran mayoría se resuelve espontáneamente. Por otra parte, el mayor porcentaje de las lesiones que se operan son benignas, en estas se reporta una incidencia de cáncer de un 11.7 %<sup>4</sup>. En esta patología, el cáncer de ovario es la neoplasia maligna ginecológica más letal. En el año 2008 se estimó que se diagnosticaron 225.500 cánceres de ovario en todo el mundo con aproximadamente 140.200 muertes por esta causa, con una incidencia y mortalidad mayor en países desarrollados respecto de los países en desarrollo<sup>1,2</sup>. La tasa de letalidad por cáncer de ovario ha ido en aumento mundialmente y en Chile se observó que la tasa de letalidad por cáncer de ovario aumentó del 3.4 al 4.9 por 100.000 desde el año 1997 al año 2010<sup>3</sup>. La discriminación precisa entre masas anexiales benignas y malignas es esencial para un manejo clínico adecuado, orientado especialmente a evitar los falsos positivos. Las lesiones clasificadas como benignas se deben observar o tratar mediante técnicas mínimamente invasivas<sup>5,6</sup>. Las masas anexiales clasificadas como malignas deben remitirse a las unidades de oncología ginecológica para una cirugía adecuada<sup>7</sup>.

La ecografía endocavitaria y/o transabdominal se considera la técnica de imagen de primera línea para evaluar las masas anexiales<sup>8</sup>. La impresión subjetiva de un examinador experto, que utiliza el llamado enfoque de reconocimiento de patrones (pattern recognition), se ha demostrado como el mejor método para discriminar entre masas anexiales benignas y malignas<sup>9</sup>. Sin embargo, este examinador no es quién habitualmente decide con respecto al manejo del paciente. Es por esto que la comunicación entre los

involucrados debe ser clara y estructurada para evitar errores de interpretación<sup>10,11</sup>.

En el año 2009 se desarrolló el Sistema de Generación de Informes de Imágenes Ginecológicas (GIRADS) para clasificar las masas anexiales y facilitar la comunicación entre los ecografistas y los médicos tratantes<sup>12</sup>. Desde entonces, se han publicado varios estudios que abordan el rendimiento diagnóstico de este sistema de clasificación.

## OBJETIVO

El objetivo de este estudio es realizar una revisión sistemática y un metanálisis del rendimiento del sistema GIRADS para clasificar las masas anexiales a diez años de su publicación.

## MÉTODOS

Protocolo y registro: Este metanálisis se realizó utilizando las recomendaciones de la Declaración PRISMA (<http://www.prisma-statement.org/>), así como siguiendo las pautas de la Evidencia de Síntesis de las Pruebas de Precisión Diagnósticas (SEDATe)<sup>13</sup>. El protocolo no fue registrado.

Se anuló la aprobación de la Junta de Revisión Institucional debido a la naturaleza y el diseño del estudio.

**Fuentes de datos y búsquedas:** Se examinaron tres bases de datos electrónicas, PubMed / MEDLINE, Web of Science y Google Scholar para identificar estudios potencialmente elegibles publicados entre enero de 2009 y diciembre de 2019. La estrategia de búsqueda incluyó los siguientes términos "masas anexiales", "cáncer de ovario", "ultrasonido transvaginal", "ultrasonido ginecológico", "GIRADS" y/o Gynecologic Imaging Reporting and Data System. No se estableció un límite de idioma.

Se identificaron estudios potencialmente elegibles aplicando los siguientes criterios:

1. Estudio de cohorte prospectivo o retrospectivo que incluye pacientes diagnosticados con al menos una masa anexial, lesión que se clasificó utilizando el sistema GIRADS después de la evaluación de ultrasonido transvaginal, transrectal o transabdominal como prueba índice.
2. Informe del diagnóstico histológico de la masa anexial después de la extirpación quirúrgica o la resolución espontánea después de la exploración de seguimiento ecográfico como estándar de referencia.
3. Presencia de datos informados que permitirían construir una tabla de 2x2 para estimar el rendimiento diagnóstico del sistema GIRADS.

Para evitar la inclusión de cohortes duplicadas, se examinó el período de estudio de cohortes informadas por los mismos autores. Si las fechas se superponían, se eligió el último estudio publicado. Se realizó búsqueda adicional en la lista de referencias de los estudios seleccionados para la lectura de texto completo. En caso de datos insuficientes, se contactó a los autores. Se registraron los aspectos relacionados con pacientes, intervención, comparación, variable de resultados, estudio (PICOS) utilizados para la inclusión y exclusión de estudios.

Los resultados de precisión diagnóstica de los estudios finalmente seleccionados fueron recuperados de forma independiente.

**Evaluación del riesgo de sesgo en estudios individuales:** La evaluación de la calidad de los estudios incluidos en el metanálisis se realizó utilizando la herramienta proporcionada por la "Evaluación de la Calidad de los Estudios de Precisión Diagnóstica-2" (QUADAS-2)<sup>14</sup>. El formato QUADAS-2 incluye cuatro dominios: 1) selección de pacientes, 2) prueba de índice, 3) estándar de referencia y 4) flujo y sincronización. Para cada dominio, el riesgo de sesgo y las preocupaciones sobre la aplicabilidad (este último no se aplica al dominio del flujo y el tiempo) se analizaron y calificaron como riesgo bajo, alto o poco claro. La evaluación de calidad se utilizó para proporcionar una evaluación general de los estudios e investigar posibles fuentes de heterogeneidad.

La evaluación de la calidad de cada trabajo se basó en la descripción del diseño del estudio, así como los criterios de inclusión y exclusión para el dominio de selección de pacientes; si el estudio informó sobre cómo se realizó e interpretó la prueba de índice para el dominio de prueba de índice; cuál era el estándar de referencia utilizado para el dominio estándar de referencia; y descripción del tiempo transcurrido desde la evaluación de la prueba de índice hasta el resultado estándar de referencia para el dominio de flujo y temporización.

#### **Análisis estadístico**

Se extrajo información sobre el rendimiento de diagnóstico del sistema GIRADS. Dado que GIRADS clasifica las masas anexiales en cinco grupos: GIRADS 1 (definitivamente benigno), GIRADS 2 (muy probablemente benigno), GIRADS 3 (probablemente benigno), GIRADS 4 (probablemente maligno) y GIRADS 5 (muy probablemente maligno), se utilizó la siguiente clasificación de dicotomía para construir tablas 2x2: los casos GIRADS 1, 2 y 3 se consideraron benignos y los casos GIRADS 4 y 5 se consideraron malignos.

Se usó un modelo de efectos aleatorios para estimar la sensibilidad, la especificidad, LR (+), LR (-) y Odds ratio diagnóstica (DOR) agrupadas. LRs se utilizan para caracterizar la utilidad clínica de una prueba y estimar la probabilidad de enfermedad posterior a la prueba<sup>15</sup>.

Utilizando la prevalencia media de malignidad ovárica (probabilidad previa a la prueba), las probabilidades posteriores a la prueba se calcularon utilizando LR (+) y (-) y se trazaron en el nomograma de Fagan.

La heterogeneidad para la sensibilidad y la especificidad se evaluó mediante el estadístico Q de Cochran y el índice I<sup>2</sup><sup>16</sup>. Un valor p <0.1 indica heterogeneidad. Se consideraría que los valores de I<sup>2</sup> de 25%, 50% y 75% indican heterogeneidad baja, moderada y alta, respectivamente<sup>16</sup>. Se trazaron graficas forest plot de sensibilidad y especificidad de todos los estudios. Se utilizó la metaregresión si existía heterogeneidad para evaluar las covariables que podrían explicar esta heterogeneidad. Las covariables analizadas fueron las siguientes: año de publicación, tamaño simple, diseño del estudio (prospectivo / retrospectivo), descripción de la prueba

índice (descrita / no descrita), prueba de referencia (histología para todos los casos incluidos / sin histología para todos los casos incluidos), prevalencia de malignidad ovárica, edad media de las pacientes y número de observadores (único / múltiple).

Las curvas resumen de las características operativas del receptor (sROC) se trazaron para ilustrar la relación entre sensibilidad y especificidad. Finalmente, el sesgo de publicación se evaluó utilizando el método de Deeks<sup>17</sup>.

Todos los análisis se realizaron con los comandos MIDAS y METANDI en STATA versión 12.0 para Windows (Stata Corporation, College Station, TX, EE. UU.). Un valor de  $p < 0.05$  se consideró estadísticamente significativo.

## RESULTADOS

### *Resultados de la búsqueda*

La búsqueda electrónica proporcionó 15 citas que reunían todos los criterios de selección. Se excluyeron 2 registros por no cumplir todos los criterios, especialmente el de la categorización según GIRADS<sup>19,24</sup>, quedando así 13 citas. Se leyó el texto completo de todos estos artículos<sup>12,18-31</sup>. No se encontraron estudios adicionales de las referencias citadas en estos 15 estudios. En la figura 1 se muestra un diagrama de flujo que resume la búsqueda bibliográfica.

### *Característica de los estudios incluidos.*

Se incluyeron trece estudios publicados entre enero de 2009 y diciembre de 2019, que informan 4473 masas anexiales en 4023 mujeres incluidos en los análisis finales para este trabajo. De estas 4473 masas anexiales, 878 resultaron malignas. No se informan masas indeterminadas en ningún estudio. La prevalencia media de malignidad ovárica fue del 23 % y la prevalencia agrupada fue de 19.6%, que oscila entre el 6,8% y el 41,8% (Tabla 1). Todos los estudios informaron las características clínicas de la cohorte. La edad media de los pacientes se informó en los 13 estudios, con edades promedio comprendidas entre 37.5 y 47.2 años. La tabla 2 muestra las características PICOS de los estudios incluidos.

### *Calidad metodológica de los estudios incluidos.*

La evaluación QUADAS-2 del riesgo de sesgo y las preocupaciones con respecto a la aplicabilidad de

los estudios seleccionados se muestra gráficamente en la figura 2.

El diseño del estudio se estableció claramente como prospectivo en ocho estudios<sup>12,18,20,23,26,27,29,30</sup> y retrospectivo en otros cinco estudios<sup>21,22,25,28,31</sup>. En el dominio de selección de pacientes hubo cinco estudios que consideraron solo los casos que se resolvieron quirúrgicamente y no las lesiones funcionales categorizadas como GIRADS 2<sup>20,21,22,26,28</sup>.

Con respecto al dominio "prueba de índice", los trece estudios describieron adecuadamente el método de texto de índice, así como también cómo se realizó (impresión subjetiva del examinador) y cómo se interpretó (clasificación GIRADS).

Para el dominio "estándar de referencia", todos los estudios se consideraron de bajo riesgo, ya que identificaron correctamente la condición objetivo por el estándar de referencia, ya sea histología definitiva después de la extirpación quirúrgica o seguimiento ecográfico con resolución espontánea o con tratamiento no quirúrgico de la masa.

Con respecto al dominio "flujo y tiempo", el tiempo transcurrido entre la prueba de índice y el estándar de referencia no estaba claro en todos los estudios.

Con respecto a la aplicabilidad, se consideró que todos los estudios tenían pocas preocupaciones con respecto a la prueba índice y los dominios estándar de referencia. Por la misma razón mencionada anteriormente con respecto al riesgo de sesgo, se consideró que cuatro estudios tenían una gran preocupación con respecto al dominio de selección de pacientes<sup>20,21,22,25</sup>.

### *Rendimiento diagnóstico del sistema GIRADS para clasificar masas anexiales*

En la tabla 3 se muestra como se distribuyen las distintas categorías de GIRADS en cada trabajo incorporado. Se incluyen el 100% de las lesiones estudiadas. Al analizar la distribución de cánceres en cada categoría se observa que en GIRADS 1-2 no hay cánceres, en GIRADS 3 un 1.9 %, en GIRADS 4 un 53.7% y en GIRADS 5 un 90 %. Por otra parte, si se analiza que porcentaje del total de cánceres corresponde a cada categoría, se observa que a GIRADS 3 corresponde el 4.4 %, a GIRADS 4 el 35 % y a GIRADS 5 el 60.6 %. Respecto de la prevalencia, si se promedian las prevalencias de cada trabajo esta

es del 23%, pero si se considera el conjunto de lesiones estudiadas, la prevalencia agrupada es de 19.63 %. (Tabla 3).

La sensibilidad, especificidad, LR (+) y LR (-) agrupadas y el DOR del sistema GIRADS para clasificar las masas anexiales fueron de 96.8% (IC 95%: 94%–98%), 91.2 % (IC 95%: 85%–94%), 11.0 (IC 95% = 6.9–13.4) y 0.035 (IC 95% = 0.02–0.09), y 209 (IC 95%: 99-444), respectivamente.

Se encontró heterogeneidad significativa para la sensibilidad ( $I^2 = 72.2\%$ ; Cochran  $Q = 43.15$ ;  $p < 0.001$ ) y para la especificidad ( $I^2 = 95.9\%$ ; Cochran  $Q = 294.02$ ;  $p < 0.001$ ). El forest plot se muestra en la figura 3.

La meta-regresión mostró que la prevalencia de malignidad ovárica, el uso de múltiples observadores e incluyendo masas con resolución espontánea explicaron la heterogeneidad observada en la especificidad. Mientras que múltiples observadores e incluyendo masas con resolución espontánea explicaron la heterogeneidad observada en la sensibilidad.

Las curvas ROC para el rendimiento diagnóstico del sistema GIRADS en clasificar masas anexiales se muestra en la figura 4. Área bajo la curva 0.98 (IC: 0.97-0.99).

El nomograma de Fagan muestra que un GIRADS 4-5 aumenta la probabilidad previa a la prueba de malignidad ovárica, de 22.6% a 76.0%; mientras que un GIRADS 1-3 disminuye la probabilidad previa a la prueba, de 22.6% a 1.0% (figura 5).

No se observó sesgo de publicación (figura 6).

## DISCUSIÓN

En el presente estudio, se ha realizado una revisión sistemática y metanálisis del sistema de clasificación GIRADS, con un universo de 4473 masas anexiales incluyendo 878 cánceres. La sensibilidad y especificidad agrupadas del sistema GIRADS para el diagnóstico del cáncer anexial fue del 96,8 % y el 91.2 %, respectivamente. Estas cifras indican que este sistema es un buen método para clasificar las masas anexiales. La distribución de cánceres por categoría está de acuerdo con lo reportado desde que se describió GIRADS en el año 2009 y 2011.

Tanto la sensibilidad como la especificidad fueron similares en la mayoría de los estudios analizados. De

hecho, el IC del 95% para ambas es relativamente estrecho, no obstante que estos estudios se realizaron en Europa, Asia, África, América del Norte y América del Sur. Considerando lo anterior el método GIRADS es potencialmente aplicable en cualquier población.

Se observa una heterogeneidad significativa para la sensibilidad y especificidad. Esto implica que los estudios podrían no ser comparables. En el análisis de metaregresión se encuentran como factores asociados a esta heterogeneidad, el número de examinadores y tipo de test de referencia (solo histología versus histología y seguimiento hasta la resolución).

El objetivo principal de GIRADS es informar y facilitar la comunicación entre el ecografista y el médico tratante. De los trece estudios analizados, solo dos abordaron este tema. En ambos estudios (uno con 15 y otro con 20 médicos tratantes) <sup>18,23</sup>, todos los médicos tratantes encontraron este sistema muy útil para el proceso de toma de decisiones sobre el manejo de sus pacientes, que es el objetivo principal de GIRADS.

Un problema recurrente en la evaluación de las masas anexiales por ultrasonido es la variabilidad y la calidad de los informes <sup>32</sup>. El consenso del grupo IOTA estableció los términos y definiciones para la descripción de los hallazgos ultrasonográficos <sup>33</sup>. Este fue un paso significativo para utilizar un "lenguaje común". Sin embargo, el cómo informar no se resolvió <sup>34</sup>.

GIRADS no es una forma para interpretar el hallazgo ecográfico, sino un sistema estructurado de informe para mejorar la comunicación entre los médicos tratantes y los ecografistas, proponiendo un diagnóstico etiológico y/o un riesgo estimado de malignidad para orientar el manejo clínico<sup>12</sup>. Se basa en la caracterización subjetiva por un examinador experto, que hasta la fecha, es la forma más eficiente para discriminar entre masas anexiales benignas y malignas <sup>9</sup>. Por lo tanto, el sistema GIRADS no es un método alternativo a IOTA para clasificar las lesiones anexiales.

IOTA ha desarrollado varios algoritmos de clasificación para masas ováricas, como los modelos logísticos LR1 y LR2 <sup>35</sup>, Reglas simples <sup>36</sup>, y el modelo ADNEX <sup>37</sup>. La mayoría de ellos proporcionan una estimación del riesgo de malignidad <sup>35, 37,38</sup>.

Debe considerarse que en el análisis de las cifras de los diversos sistemas de IOTA en un 20 a 25 % de lesiones estos modelos no se pueden aplicar y deben ser enviados a un experto para su informe definitivo<sup>39,40,41,42</sup>.

Recientemente, el Colegio Americano de Radiólogos ha propuesto un Sistema de Datos de Informe Anexial Ovárico (O-RADS)<sup>43</sup>. Este sistema, O-RADS, describe un léxico estandarizado para lesiones ováricas y anexiales basado en la terminología IOTA y el objetivo final será aplicarlo a una clasificación de estratificación de riesgo para el seguimiento y manejo clínico. Las cifras que acompañan la publicación O-RADS indican una probabilidad de malignidad de hasta 10, bajo 50 y sobre 50 % para las categorías 3, 4 y 5 respectivamente. Cabe hacer notar que estos valores son una extrapolación de trabajos anteriores, sin casuística propia, por lo tanto, sin validación interna o externa actual. En GIRADS, estos porcentajes fueron de 2%, 54% y 90% para las categorías 3,4 y 5 (tabla 3).

El sistema IOTA, que es la base de la propuesta de O-RADS, fue concebido para que operadores de poca experiencia obtuvieran resultados diagnósticos semejantes a los de expertos. Futuros trabajos deberán confirmar o no esta hipótesis una vez que se disponga de casuística y seguimiento suficiente. Además, O-RADS incorpora la Resonancia Nuclear Magnética como elemento de segunda línea en caso de duda, esquema difícilmente aplicable en forma masiva en muchas realidades incluyendo la nuestra. Aun así, los autores, en la discusión, califican como decepcionante la especificidad del sistema propuesto<sup>43,44</sup>. La baja especificidad, se repite en otros trabajos y es una limitación que a menudo, se pasa por alto. Sin embargo, esta se traduce en un aumento de los falsos positivos, intervenciones innecesarias y eventuales complicaciones<sup>45,46</sup>, lo que se traduce en un perjuicio más que en un beneficio para las pacientes.

La principal conclusión de este metanálisis es que, en poblaciones diversas, operadores con diferencias socio culturales importantes, obtienen resultados comparables con una alta precisión diagnóstica en toda la patología anexial.

Debe tenerse presente que GIRADS está concebido para ser utilizado por personal con conocimiento y experiencia en el tema.

#### Declaración conflictos de Interés

Este trabajo de investigación se ha realizado con un esfuerzo personal del grupo, sin apoyo, reconocimiento ni financiamiento de la industria farmacéutica, de empresas de ultrasonografía, del Estado, de ONGs, de universidades o de cualquier otra institución en especial. Tres autores han participado en algunos de los trabajos revisados en este Metaanálisis. Esta situación no inferiría en la revisión objetiva de las publicaciones incluidas, por lo que declaramos no tener conflictos de interés.

#### REFERENCIAS

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics [published correction appears in CA Cancer J Clin. 2011 Mar-Apr;61(2):134]. CA Cancer J Clin. 2011;61(2):69–90. doi:10.3322/caac.20107.
2. Lowe, Kimberly A. et al. An international assessment of ovarian cancer incidence and mortality Gynecologic Oncology, Volume 130, Issue 1, 107 – 114 doi.org/10.1016/j.ygyno.2013.03.026
3. Cuello M. GES en cáncer de ovario epitelial: un avance sanitario necesario, pero no exento de riesgos y dificultades futuras. Rev Chil Obstet Ginecol 2013; 78(3): 161 – 166.
4. Cuello M et al. Distribución de la patología anexial en mujer chilena: experiencia de la Universidad Católica de Chile. Rev Chilena Obstet Ginecol 2004; 69(6): 429-440
5. 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. doi:10.1002/uog.12369
6. Canis M, Rabischong B, Houille C, et al. Laparoscopic management of adnexal masses: a gold standard? Curr Opin Obstet Gynecol. 2002;14(4):423–428. doi:10.1097/00001703-200208000-00010
7. Vernooij, Flora et al. The outcomes of ovarian cancer treatment are better when provided by gynecologic oncologists and in specialized

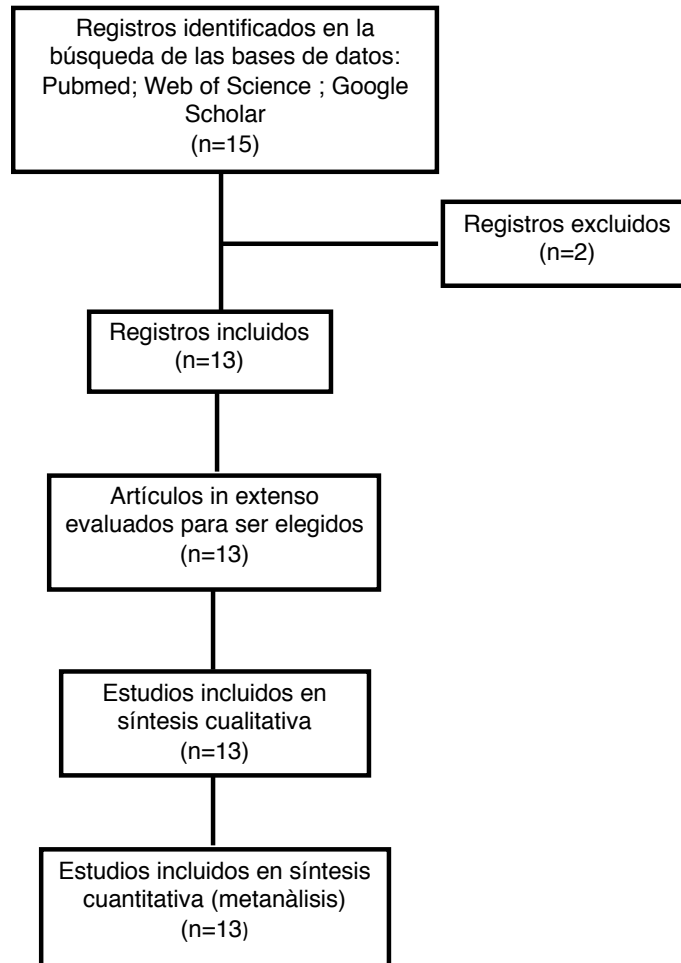
- hospitals: A systematic review *Gynecologic Oncology*, 2007, Volume 105, Issue 3, 801 – 812 <https://doi.org/10.1016/j.ygyno.2007.02.030>
8. American College of Obstetricians and Gynecologists' Committee on Practice Bulletins—Gynecology. Practice Bulletin No. 174: Evaluation and Management of Adnexal Masses. *Obstet Gynecol.* 2016;128(5):e210–e226. doi:10.1097/AOG.0000000000001768
  9. 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. <https://doi.org/10.1016/j.ejca.2016.01.007>
  10. Le T, Fayadh RA, Menard C, Hicks-Boucher W, Faught W, Hopkins L, Fung-Kee-Fung M. Variations in ultrasound reporting on patients referred for investigation of ovarian masses. *J Obstet Gynaecol Can.* 2008;30:902-6. [https://doi.org/10.1016/S1701-2163\(16\)32970-X](https://doi.org/10.1016/S1701-2163(16)32970-X)
  11. Brown DL, Dudiak KM, Laing FC. Adnexal masses: US characterization and reporting. *Radiology.* 2010;254(2):342–354. doi:10.1148/radiol.09090552
  12. Amor F, Vaccaro H, Alcázar JL, León M, Craig 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(3):285–291. doi:10.7863/jum.2009.28.3.285.
  13. Sotiriadis A, Papatheodorou SI, Martins WP. Synthesizing Evidence from Diagnostic Accuracy TEsts: the SEDATE guideline. *Ultrasound Obstet Gynecol.* 2016;47(3):386–395. doi:10.1002/uog.15762
  14. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529–536. doi:10.7326/0003-4819-155-8-201110180-00009
  15. European Network for Health Technology Assessment (EUnetHTA). Methodological guideline on “Meta-analysis of diagnostic test accuracy studies”. January 1, 2014/ EUnetHTA Joint Action 2 (2012-15), JA2-WP7, Outputs
  16. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327(7414):557–560. doi: <https://doi.org/10.1136/bmj.327.7414.557>
  17. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol.* 2005;58(9):882–893. doi: 10.1016/j.jclinepi.2005.01.016
  18. 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. doi:10.1002/uog.9012
  19. Moszynski R, Szubert S, Szpurek D, Michalak S, Sajdak S. Role of osteopontin in differential diagnosis of ovarian tumors. *J Obstet Gynaecol Res.* 2013;39(11):1518–1525. doi:10.1111/jog.12097
  20. Qiuyue C, Guorong L. The applicational value of GI-RADS ultrasonographic stratification in diagnosing adnexal masses. *Chinese J Ultrasound Med.* 2013; 29: 527-30.
  21. Rams N, Muñoz R, Soler C, Parra J. Results of the Gynecologic Imaging Reporting and Data System classification in adnexal masses. *Prog Obstet Gynecol.* 2015; 58(3): 125-129. doi.org/10.1016/j.pog.2014.11.020
  22. Fernández RO, Rama AP, Llanos Llanos MC, Mendoza AM, Linde FM and Díaz AN . Clinical Application of the Gynecologic Imaging Reporting and Data System (Gi-Rads) for the Evaluation of Adnexal Masses. *SM J Gynecol Obstet.* 2015; 1(2): 1009.
  23. Díaz L, Zambrano B, Adami FJ, Alcázar JL. External Validation of Gynecological Imaging and Reporting Data System for Sonographic Evaluation of Adnexal Masses. *Donald School J Ultrasound Obstet Gynecol.* 2017;11:135-40. DOI: 10.5005/jp-journals-10009-1514
  24. Koneczny J, Czekierdowski A, Florczak M, Poziemski P, Stachowicz N, Borowski D. The use of sonographic subjective tumor assessment, IOTA logistic regression model 1, IOTA Simple Rules and GI-RADS system in the preoperative prediction of malignancy in women with adnexal



- masses. *Ginekol Pol.* 2017; 88:647-653. <https://doi.org/10.5603/gp.a2017.0116>
25. Zhang T, Li F, Liu J, Zhang S. Diagnostic performance of the Gynecology Imaging Reporting and Data System for malignant adnexal masses. *Int J Gynaecol Obstet.* 2017;137:325-31. <https://doi.org/10.1002/ijgo.12153>
  26. Migda M, Bartosz M, Migda MS, Kierszk M, Katarzyna G, Maleńczyk M. Diagnostic value of the gynecology imaging reporting and data system (GI-RADS) with the ovarian malignancy marker CA-125 in preoperative adnexal tumor assessment. *J Ovarian Res.* 2018; 11:92. <https://doi.org/10.1186/s13048-018-0465-1>
  27. Basha MAA, Refaat R, Ibrahim SA, Madkour NM, Awad AM, Mohamed EM, El Sammak AA, Zaitoun MMA, Dawoud HA, Khamis MEM, Mohamed HAE, El-Maghraby AM, Abdalla AAEM, Assy MM, Nada MG, Obaya AA, Abdelbary EH. Gynecology Imaging Reporting and Data System (GI-RADS): diagnostic performance and interreviewer agreement. *Eur Radiol.* 2019 Apr 16. doi: 10.1007/s00330-019-06181-0.
  28. Zheng H, Tie Y, Wang X, Yang Y, Wei X, Zhao X. Assessment of the diagnostic value of using serum CA125 and GI-RADS system in the evaluation of adnexal masses. *Medicine (Baltimore).* 2019;98(7):e14577. doi:10.1097/MD.00000000000014577
  29. Khalaf, L.M.R., Desoky, H.H.M., Seifeldin, G.S. et al. The diagnostic efficacy of Gynecology Imaging Reporting and Data System (GI-RADS): single-center prospective cross-sectional study. *Egypt J Radiol Nucl Med* 50, 61 (2019). doi.org/10.1186/s43055-019-0071-2
  30. Behnamfar F, Adibi A, Khadra H, Moradi M. Diagnostic accuracy of gynecology imaging reporting and data system in evaluation of adnexal lesions. *J Res Med Sci* 2019; 24:57.
  31. Sonnemann MSI, Torres HVE. Diagnostic performance of the GI-RADS classification system for anexial masses. *Rev Sanid Milit Mex.* 2019;73(1):41-45.
  32. Le T, Fayadh RA, Menard C, Hicks-Boucher W, Faught W, Hopkins L, Fung-Kee-Fung M. Variations in ultrasound reporting on patients referred for investigation of ovarian masses. *J Obstet Gynaecol Can.* 2008; 30:902-906. [https://doi.org/10.1016/s1701-2163\(16\)32970-x](https://doi.org/10.1016/s1701-2163(16)32970-x)
  33. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I; International Ovarian Tumor Analysis (IOTA) Group. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group. *Ultrasound Obstet Gynecol.* 2000; 16:500-5. doi.org/10.1046/j.1469-0705.2000.00287.x
  34. Brown DL, Dudiak KM, Laing FC. Adnexal masses: US characterization and reporting. *Radiology.* 2010; 254:342-54 <https://doi.org/10.1148/radiol.09090552>
  35. Timmerman D, Testa AC, Bourne T, 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(34):8794–8801. doi:10.1200/JCO.2005.01.7632
  36. Timmerman D, Testa AC, Bourne T, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol.* 2008;31(6):681–690. doi:10.1002/uog.5365
  37. Van Calster B, Van Hoorde K, Valentin L, 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. Published 2014 Oct 15. doi:10.1136/bmj. g5920.
  38. Timmerman D, Van Calster B, Testa A, et al. Predicting the risk of malignancy in adnexal masses based on the Simple Rules from the International Ovarian Tumor Analysis group. *Am J Obstet Gynecol.* 2016;214(4):424–437. doi: 10.1016/j.ajog.2016.01.007
  39. Tinnangwattana D, Vichak-Ururrote L, Tontivuthikul P, Charoenratana C, Lerthiranwong T, Tongsong T. IOTA Simple Rules in Differentiating between Benign and Malignant Adnexal Masses by Non-expert Examiners. *Asian Pac J Cancer Prev.* 2015;16(9):3835–3838. doi:10.7314/apjcp.2015.16.9.3835

40. Meys E, Rutten I, Kruitwagen R, et al. Simple Rules, Not So Simple: The Use of International Ovarian Tumor Analysis (IOTA) Terminology and Simple Rules in Inexperienced Hands in a Prospective Multicenter Cohort Study. Simple Rules“ – nicht so einfach: Anwendung der „International Ovarian Tumor Analysis“(IOTA)-Terminologie und der „Simple Rules“ in unerfahrenen Händen in einer prospektiven multizentrischen Kohortenstudie. *Ultraschall Med.* 2017;38(6):633–641. doi:10.1055/s-0043-113819
41. Knafel A, Banas T, Nocun A, Wiechec M, Jach R, Ludwin A, Kabzinska-Turek M, Pietrus M, Pitynski K. The Prospective External Validation of International Ovarian Tumor Analysis (IOTA) Simple Rules in the Hands of Level I and II Examiners. *Ultraschall Med.* 2016; 37:516-523. doi:10.1055/s-0034-1398773
42. Ning CP, Ji X, Wang HQ, et al. Association between the sonographer's experience and diagnostic performance of IOTA simple rules. *World Journal of Surgical Oncology.* 2018 Sep;16(1):179. DOI: 10.1186/s12957-018-1479-2.
43. Andreotti RF, Timmerman D, Benacerraf BR, et al. Ovarian-Adnexal Reporting Lexicon for Ultrasound: A White Paper of the ACR Ovarian-Adnexal Reporting and Data System Committee [published correction appears in *J Am Coll Radiol.* 2019 Mar;16(3):403-406]. *J Am Coll Radiol.* 2018;15(10):1415–1429. doi: 10.1016/j.jacr.2018.07.004.
44. Andreotti RF, Timmerman, D 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; 294:168–185. doi.org/10.1148/radiol.2019191150
45. Buys SS, Partridge E, Black A, et al. Effect of Screening on Ovarian Cancer Mortality: The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA.*2011;305(22):2295–2303. doi:10.1001/jama.2011.766
46. Menon U, Gentry-Maharaj A et al., Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncol* 2009; 10: 327–40. DOI:10.1016/S1470- 2045(09)70026-9 .

## TABLAS Y FIGURAS

**Figura 1.** Diagrama de flujo del proceso de selección de los 15 estudios seleccionados para este metaanálisis.

**Tabla 1.** Registro de los resultados de cada trabajo incluido en el análisis final.

ESTUDIO/AÑO	Pacientes (n)	Masas (n)	Malignos (n)	Benignos (n)	V(+)	F(+)	F(-)	V(-)	Prevalencia malignidad %
Amor/2009	171	187	25	162	25	16	0	146	13.7
Amor/2011	372	432	112	320	111	45	1	275	25.9
Qiuye/2013	1149	1449	99	1350	98	38	1	1312	6.8
Rams/2015	87	98	14	84	14	9	0	75	14.3
Orozco/2015	387	387	66	321	56	12	10	309	17.1
Diaz/2017	257	281	35	246	34	17	1	229	12.5
Zhang/2017	242	263	110	153	106	24	4	129	41.8
Migda/2018	215	215	53	162	50	45	3	117	24.7
Zheng/2019	325	325	129	196	117	18	12	178	39.7
Basha/2019	308	325	127	198	124	12	3	186	39.1
Khalaf/2019	116	156	53	103	52	3	1	100	33.3
Behnamfar/2019	197	158	12	146	11	28	1	118	7.6
Sonnemann/2019	197	197	43	154	42	55	1	99	21.8
<b>TOTALES</b>	<b>4023</b>	<b>4473</b>	<b>878</b>	<b>3595</b>	<b>840</b>	<b>322</b>	<b>38</b>	<b>3273</b>	

(N): numero; V(+): verdaderos positivos; V(-): verdadero negativo; F(+): falso positivo; F(-): falso negativo.

**Tabla 1a.** Registro de los resultados de sensibilidad, especificidad, valor predictivo positivo y negativo, Likelihood ratio (+) y (-) para cada trabajo incluido en el análisis final.

ESTUDIO/AÑO	SENSIBILIDAD (IC:95%)	ESPECIFICIDAD (IC:95%)	VP (+)	VP (-)	LR (+)	LR (-)
Amor/2009	100.0 (86.3 - 100.0)	90.1(84.5 - 94.2)	61.0	100.0	10.13	0.000
Amor/2011	99.1(95.1 - 100.0)	85.9 (81.6 - 89.6)	71.2	99.6	7.03	0.010
Qiuye/2013	99.0 (93.0 - 99.8)	97.2 (96.2 - 98.0)	72.1	99.9	35.17	0.010
Rams/2015	100.0 (76.8 - 100.0)	89.3 (80.6 - 95.0)	60.9	100.0	9.33	0.000
Orozco/2015	84.8 (73.9 - 92.5)	96.3 (93.6 - 98.1)	82.4	96.9	22.70	0.157
Diaz/2017	97.1(85.1 - 99.9)	93.1(88.4 - 95.6)	66.7	99.6	14.06	0.031
Zhang/2017	96.4 (91.0 - 99.0)	84.3 (77.6 - 89.7)	81.5	97.0	6.14	0.043
Migda/2018	94.3 (84.3 - 98.8)	72.2 (64.7 - 79.0)	52.6	97.5	3.40	0.078
Zheng/2019	90.7 (84.3 - 95.1)	90.8 (85.9 - 94.5)	86.7	93.7	9.88	0.102
Basha/2019	97.6 (93.3 - 99.5)	93.9 (89.7 - 96.8)	91.2	98.4	16.11	0.025
Khalaf/2019	98.1(89.9 - 100.0)	97.1(91.7 - 99.4)	94.5	99.0	33.69	0.019
Behnamfar/2019	91.7 (61.5 - 99.8)	80.8 (73.5 - 86.9)	28.2	99.2	4.78	0.103
Sonnemann/2019	97.7 (87.7 - 99.9)	64.3 (56.2 - 71.8)	43.3	99.0	2.7	0.036

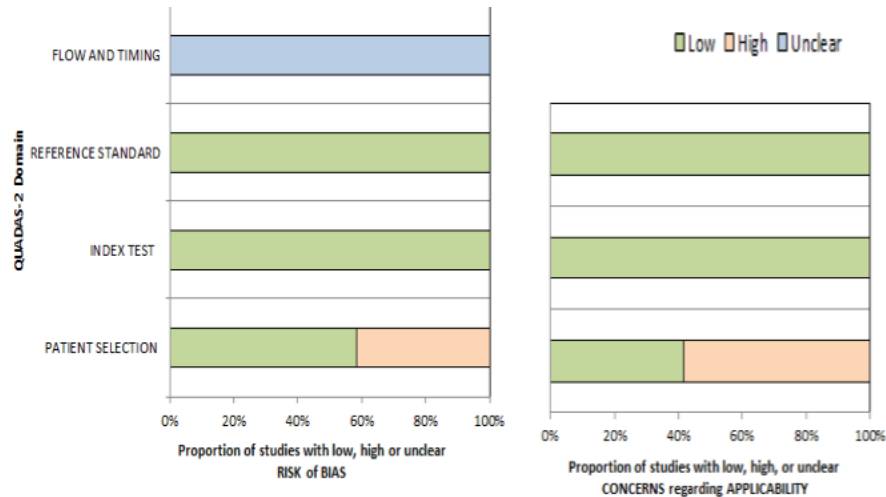
IC: Intervalo de confianza; VP: Valor predictivo; LR: Likelihood ratio

**Tabla 2.** Criterios de selección de cada trabajo incorporado en la selección.

Estudio/ año	País	Número			Edad	Centros	Diseño	Test		
		Mujeres	Masas	CA	(Años Prom.)			Indice	Ref.	Exam.
Amor / 2009	Chile	171	187	25	39	U	P	USG – GIRADS	H + RE	M
Amor / 2011	Chile /España	372	432	112	44	M	P	USG – GIRADS	H + RE	M
Qiuye / 2013	China	1149	1449	99	39.5	U	P	USG – GIRADS	H + RE	M
Rams / 2015	España	87	98	14	NA	U	R	USG – GIRADS	H	M
Orozco / 2015	España	387	387	66	43	U	R	USG – GIRADS	H	M
Díaz / 2017	Venezuela	257	281	35	40.3	U	P	USG – GIRADS	H + RE	U
Zhang / 2017	China	242	263	110	47.2	U	R	USG – GIRADS	H + RE	M
Migda / 2018	Polonia	215	215	53	40	U	P	USG – GIRADS	H	U
Zheng / 2019	China	325	325	129	47.2	U	R	USG – GIRADS	H	M
Basha / 2019	Egipto	308	325	127	41	M	P	USG – GIRADS	H + RE	M
Khalaf / 2019	Egipto	116	156	53	42	U	P	USG – GIRADS	H + RE	M
Behnamfar / 2019	Iran	197	158	12	37.5	U	P	USG – GIRADS	H	U
Sonnemann / 2019	Mexico	197	197	43	45.3	U	R	USG – GIRADS	H	M
<b>Totales</b>		<b>4023</b>	<b>4473</b>	<b>878</b>						

CA: Cánceres; Prom: Promedio; U: único; M: múltiple; P: Prospectivo; R: Retrospectivo; USG: Ultrasonido ginecológico; Ref.: Referencia; H: Histología; R.E: Resolución Espontánea; Exam.: Examinador

**Figura 2.** Gráfico de histograma que muestra la evaluación de calidad (riesgo de sesgo y preocupaciones sobre la aplicabilidad) para todos los estudios incluidos en el metanálisis.

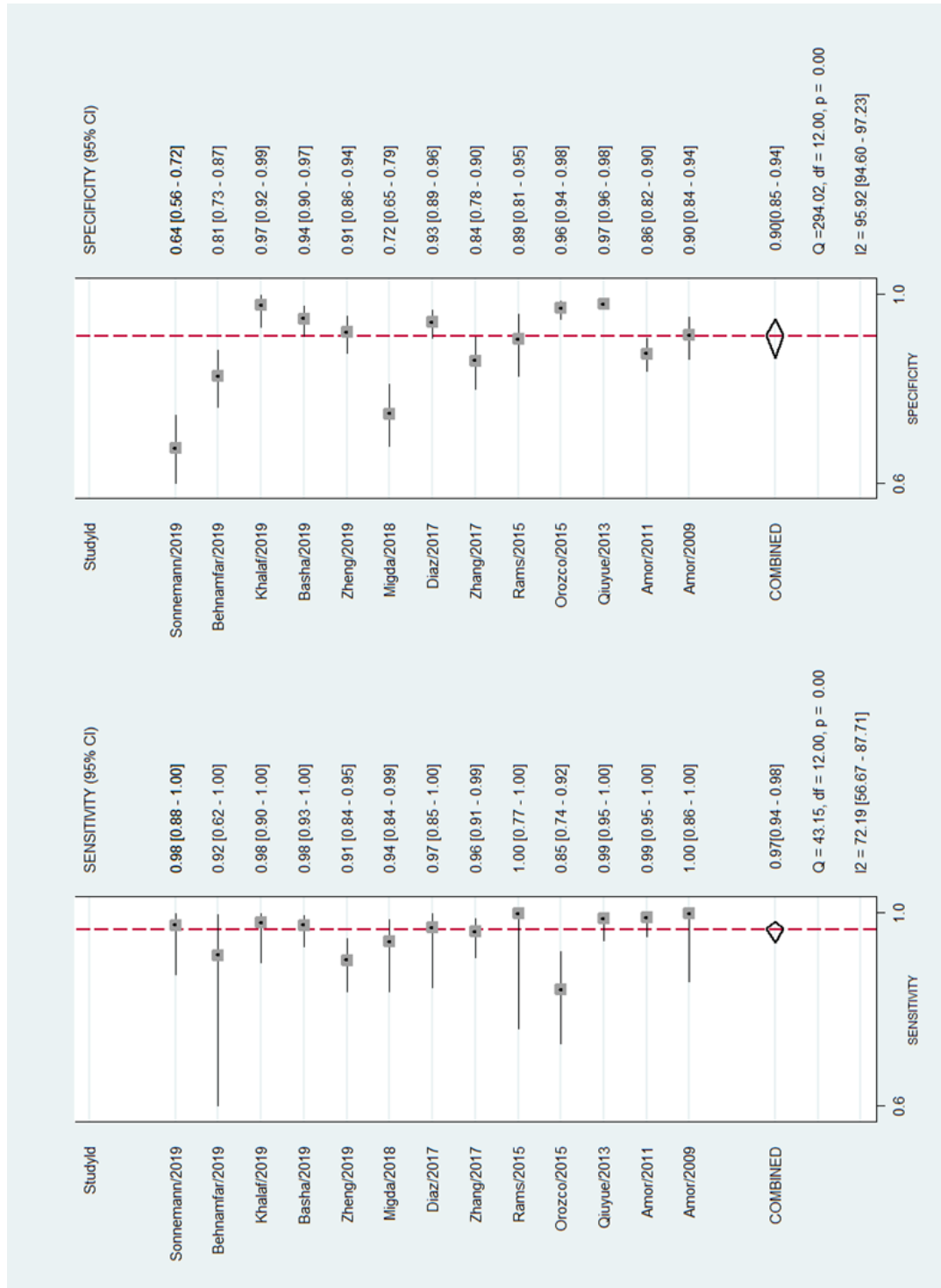


**Tabla 3.** Distribución de tipos de lesiones en cada trabajo incorporado en el análisis final.

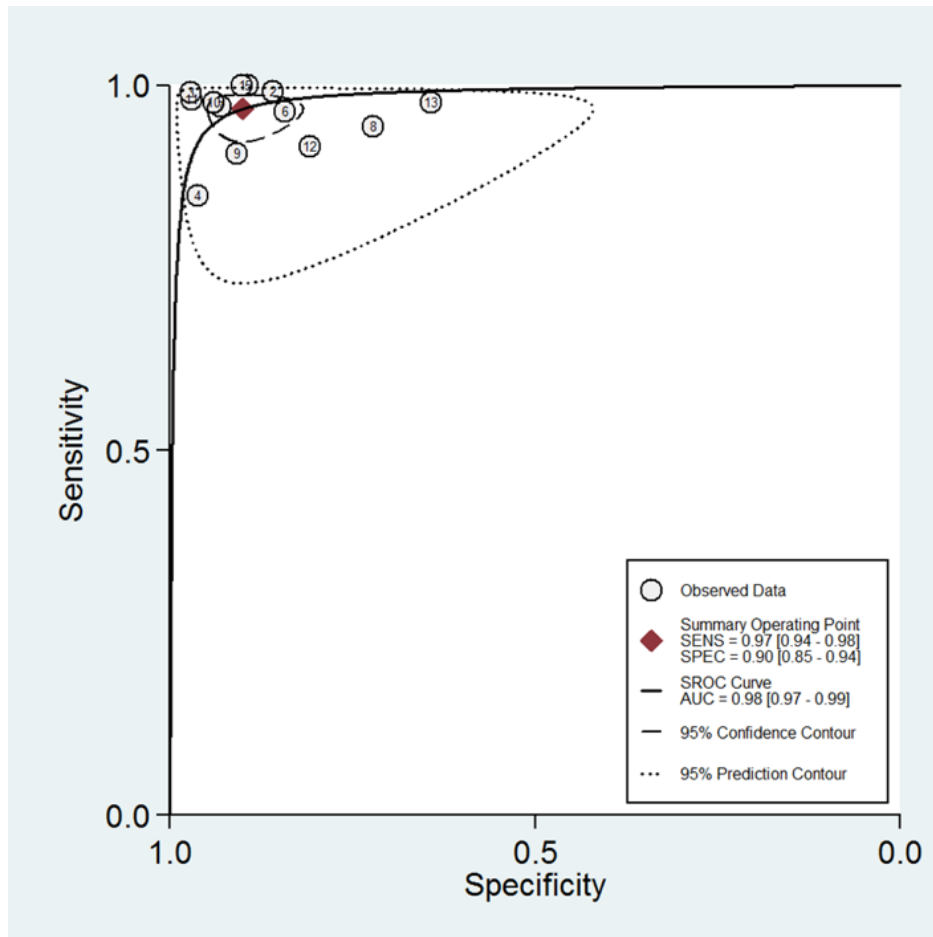
	G1 - G2		G3		G4		G5		TOTAL MAL. (N)	TOTAL MASAS (N)	PREV. %
	BEN. (N)	MAL. (N)	BEN. (N)	MAL. (N)	BEN. (N)	MAL. (N)	BEN. (N)	MAL. (N)			
<b>Amor/2009</b>	56	0	90	0	11	2	5	23	25	187	13.4
<b>Amor/2011</b>	92	0	183	1	32	8	13	103	112	432	25.9
<b>Qiuye/2013</b>	731	0	581	1	32	28	6	70	99	1449	6.8
<b>Rams/2015</b>	0	0	75	0	8	3	1	11	14	98	14.3
<b>Orozco/2015</b>	3	0	306	10	8	12	4	44	66	387	17.1
<b>Diaz/2017</b>	56	0	173	1	11	8	6	26	35	281	12.5
<b>Zhang/2017</b>	65	0	64	4	24	77	0	29	110	263	41.8
<b>Migda/2018</b>	2	0	114	4	45	41	0	9	54	215	25.1
<b>Zheng/2019</b>	95	0	91	3	7	6	5	118	127	325	39.1
<b>Basha/2019</b>	0	0	178	12	18	88	0	29	129	325	39.7
<b>Khalaf/2019</b>	67	0	32	1	4	11	1	40	52	156	33.3
<b>Behnamfar/ 2019</b>	59	0	59	1	24	6	4	5	12	158	7.6
<b>Sonnemann/2019</b>	1	0	98	1	41	17	14	25	43	197	21.8
<b>TOTALES</b>	<b>1227</b>	<b>0</b>	<b>2044</b>	<b>39</b>	<b>265</b>	<b>307</b>	<b>59</b>	<b>532</b>	<b>878</b>	<b>4473</b>	<b>19.63</b>
<b>% CA por categoría</b>			<b>1.9</b>		<b>53.7</b>		<b>90</b>		<b>PREV. PROMEDIO</b>		<b>23</b>
<b>% CA por categoría en relación al total de cánceres</b>			<b>4.4</b>		<b>35</b>		<b>60.6</b>				

G1: GIRADS 1; G2: GIRADS 2; G3: GIRADS 3; G4: GIRADS 4; G5: GIRADS 5; BEN: Beningno; MAL: Maligno; CA: Cáncer; PREV: Prevalencia

**Figura 3:** Forest plot para sensibilidad y especificidad para cada estudio y combinada de GIRADS.

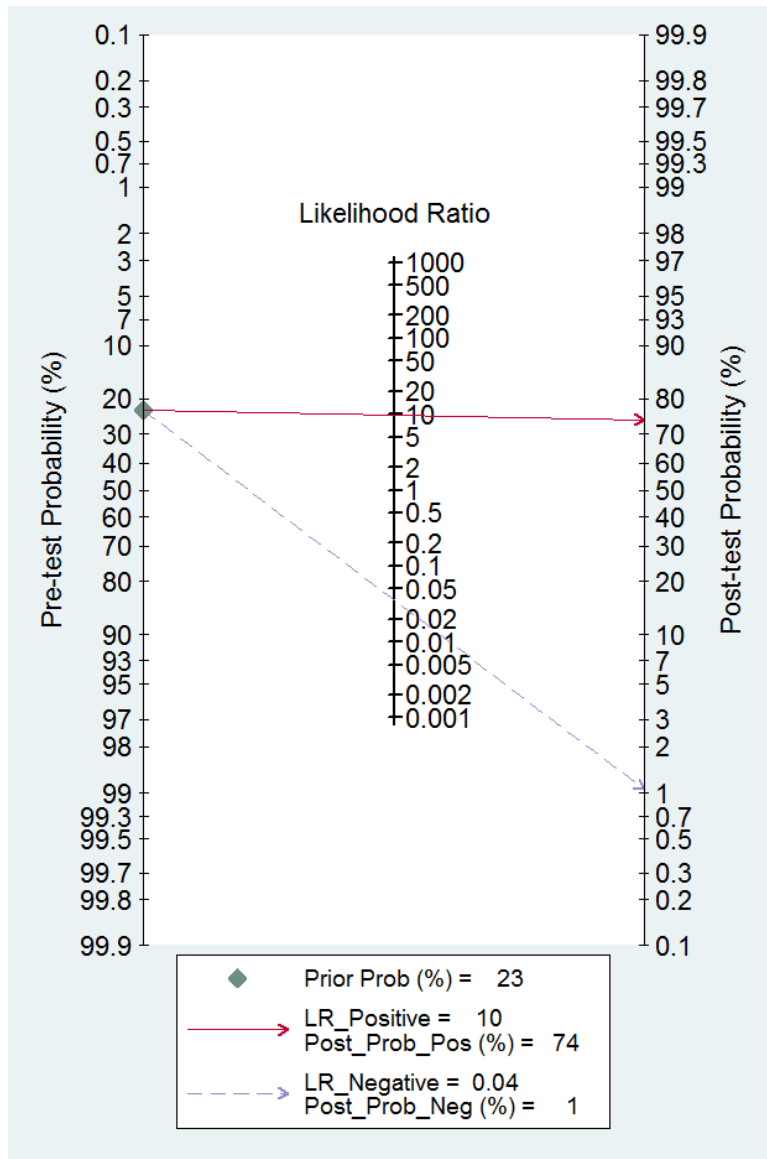


**Figura 4:** Curva ROC resumen para GIRADS.





**FIGURA 5.** Nomograma de Fagan para detectar lesión maligna anexial utilizando la metodología GIRADS.



**FIGURA 6.** Gráfico utilizando método Deeks para evaluar el sesgo de publicación (p=0.27).

