



## OPEN ACCESS

## EDITED BY

Emanuele Perrone,  
Agostino Gemelli University Polyclinic  
(IRCCS), Italy

## REVIEWED BY

Giulia Zinicola,  
Policlinico Agostino Gemelli, Italy  
Angelo Finelli,  
ULSS2 Marca Trevigiana, Italy  
Jacques Abramowicz,  
The University of Chicago, United States  
Mario Palumbo,  
Federico II University Hospital, Italy

## \*CORRESPONDENCE

Wei Hu

✉ submitpaper2023@163.com

RECEIVED 14 December 2023

ACCEPTED 15 April 2024

PUBLISHED 02 May 2024

## CITATION

Han J, Wen J and Hu W (2024) Comparison of O-RADS with the ADNEX model and IOTA SR for risk stratification of adnexal lesions: a systematic review and meta-analysis. *Front. Oncol.* 14:1354837. doi: 10.3389/fonc.2024.1354837

## COPYRIGHT

© 2024 Han, Wen and Hu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Comparison of O-RADS with the ADNEX model and IOTA SR for risk stratification of adnexal lesions: a systematic review and meta-analysis

Jing Han<sup>1</sup>, Jing Wen<sup>2</sup> and Wei Hu<sup>3\*</sup>

<sup>1</sup>Department of Radiology, Suzhou Hospital, Affiliated Hospital of Medical School, Nanjing University, Suzhou, China, <sup>2</sup>Department of Medical Imaging, Jiangsu Vocational College of Medicine, Yancheng, China, <sup>3</sup>Department of Radiology, Yixing Traditional Chinese Medicine Hospital, Yixing, China

**Purpose:** This study aims to systematically compare the diagnostic performance of the Ovarian-Adnexal Reporting and Data System with the International Ovarian Tumor Analysis Simple Rules and the Assessment of Different NEoplasias in the adneXa model for risk stratification of ovarian cancer and adnexal masses.

**Methods:** A literature search of online databases for relevant studies up to July 2023 was conducted by two independent reviewers. The summary estimates were pooled with the hierarchical summary receiver-operating characteristic model. The quality of the included studies was assessed with the Quality Assessment of Diagnostic Accuracy Studies–2 and the Quality Assessment of Diagnostic Accuracy Studies-Comparative Tool. Metaregression and subgroup analyses were performed to explore the impact of varying clinical settings.

**Results:** A total of 13 studies met the inclusion criteria. The pooled sensitivity and specificity for eight head-to-head studies between the Ovarian-Adnexal Reporting and Data System and the Assessment of Different NEoplasias in the adneXa model were 0.96 (95% CI 0.92–0.98) and 0.82 (95% CI 0.71–0.90) vs. 0.94 (95% CI 0.91–0.95) and 0.83 (95% CI 0.77–0.88), respectively, and for seven head-to-head studies between the Ovarian-Adnexal Reporting and Data System and the International Ovarian Tumor Analysis Simple Rules, the pooled sensitivity and specificity were 0.95 (95% CI 0.93–0.97) and 0.75 (95% CI 0.62–0.85) vs. 0.91 (95% CI 0.82–0.96) and 0.86 (95% CI 0.76–0.93), respectively. No significant differences were found between the Ovarian-Adnexal Reporting and Data System and the Assessment of Different NEoplasias in the adneXa model as well as the International Ovarian Tumor Analysis Simple Rules in terms of sensitivity ( $P = 0.57$  and  $P = 0.21$ ) and specificity ( $P = 0.87$  and  $P = 0.12$ ). Substantial heterogeneity was observed among the studies for all three guidelines.

**Conclusion:** All three guidelines demonstrated high diagnostic performance, and no significant differences in terms of sensitivity or specificity were observed between the three guidelines.

#### KEYWORDS

O-RADS, IOTA, ADNEX model, ovarian cancer, diagnostic

## Introduction

Ovarian carcinoma is the leading cause of mortality from gynecological malignancy in the USA, where approximately more than 13,000 deaths are from ovarian carcinoma in 2023 and the 5-year survival rate is no more than 50% (1). The early diagnosis of ovarian carcinoma is associated with a significantly higher 5-year survival rate, which is increased to >90% for stage I (2). Therefore, it is important to accurately differentiate malignant tumors from benign tumors, thereby optimizing patient triaging and reducing unnecessary surgeries without missing cancer. Although several imaging modalities such as MRI and CT play a role in the assessment and management of adnexal lesions, ultrasound (US) is still the first-line preoperative differential diagnosis method for ovarian masses (3, 4).

Several risk stratification systems have been developed to standardize the assessment of adnexal masses with US to improve accuracy and interreader agreement. The International Ovarian Tumor Analysis (IOTA) group proposed terminology and definitions to describe ultrasound features of adnexal lesions in 2008, aiming to provide a standardized tool for differentiating benign and malignant adnexal lesions (5, 6). The IOTA Simple Rules (IOTA SR) includes five descriptions for benignity (benign features) and five for malignancy (malignant features), and adnexal masses are classified as benign, malignant, and inconclusive. Previous studies showed that the IOTA SR has high performance, with a pooled sensitivity of 0.93 and specificity of 0.80 (7). However, the IOTA SR is unable to classify all adnexal masses, leaving as much as 25% inconclusive lesions; when both malignant and benign features were present, or if none of the features were present, the simple rules were inconclusive (6).

In 2014, the IOTA group developed a new scoring system named the Assessment of Different Neoplasias in the adnexa (ADNEX) model, which used three clinical variables and six ultrasound variables to calculate the risk of an adnexal lesion (benign or malignant), distinguishing four types of malignant ovarian tumors: borderline, stage I cancer, stage II–IV cancer, and secondary metastatic cancer (8–10). Additionally, other standardized guidelines or risk stratification systems were proposed such as the Gynecologic Imaging Reporting and Data System (GI-RADS), the Risk of Malignancy Index 4 (RMI4), and the logistic regression model 2 (LR2) (11–13). Nonetheless, many of

these standardized models were found inferior to subjective expert assessment (14). In 2018, based on IOTA terms and data sets, the American College of Radiology (ACR) introduced the Ovarian-Adnexal Reporting and Data System (O-RADS) US risk stratification and management (15). With O-RADS US, an adnexal mass is stratified to a 1–5 category (1, physiologic; 2, almost certainly benign; 3, low risk of malignancy; 4, intermediate of malignancy; 5, high risk of malignancy) according to its sonographic features. Since the publication of O-RADS US, a number of studies evaluating this scoring system have been published. Additionally, some of them had performed head-to-head comparisons between O-RADS US with other guidelines. Although several meta-analyses or systematic reviews have summarized the diagnostic accuracy of O-RADS US, a comparison with other guidelines has not been reported systematically. Therefore, in this study, we aimed to systematically compare the performance of O-RADS US with IOTA SR and the ADNEX model.

## Materials and methods

This meta-analysis and systematic review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (16). The primary outcome of this study was the direct comparison between O-RADS and IOTA SR along with the ADNEX model. Furthermore, the overall diagnostic performance of O-RADS for all the included studies was calculated.

### Search strategy and selection criteria

An electronic search of PubMed, EMBASE, Cochrane Library, Web of Science, and Google Scholar online scientific publication databases was conducted to identify relevant studies that were published up to 31 July 2023, with language restricted to English only. The following terms in combination with abbreviations were used for the literature search: (“O-RADS” OR “Ovarian-Adnexal Reporting and Data System”) AND [(“IOTA SR” OR “SR” OR “simple rules” OR “International Ovarian Tumor Analysis SR” OR “IOTA simple rules”) OR (“ADNEX” OR “ADNEX models” OR

“IOTA ADNEX”]. An additional literature search was supplemented by manually screening the bibliographies among the included studies and reviews to prohibit missing potential eligible studies. Two reviewers (H.J. and W.J.) independently assessed the search results, and any disagreements were resolved through discussion until a consensus was reached.

## Inclusion and exclusion criteria

Studies that met all of the following criteria were included: 1) used O-RADS and IOTA SR and/or the ADNEX model for the risk stratification of adnexal lesions, with head-to-head comparisons of diagnostic accuracy; 2) provided sufficient details to construct  $2 \times 2$  contingency tables for determining diagnostic accuracy; and 3) had surgical pathology results or at least 1-year follow-up as the reference standard. Studies that met any of the following criteria were excluded: 1) had no direct comparison between O-RADS with the other guidelines; 2) did not report sufficient data to assess the diagnostic performance; and 3) were meta-analyses, guidelines, editorials, reviews, conference abstracts, and letters.

## Data extraction and quality assessment

A predefined standardized form was employed to extract the following data from the included studies: 1) clinical and demographic characteristics, e.g., number of patients and lesions, patient age, and tumor size; and 2) study characteristics, e.g., first author, study design (prospective or retrospective), publication year, location of the study and period, number and experience of radiologists, cutoff values, guidelines, and the reference standard. We used the Quality Assessment of Diagnostic Accuracy Studies–2 (QUADAS-2) to perform quality assessment of the included studies (17), with each study categorized as having either low, unclear, or high risk of bias according to the following four domains: patient selection, method of the index test, reference standard, and flow and timing. The quality of the studies that included a head-to-head comparison of O-RADS US with either the IOTA SR or the ADNEX model was assessed with the QUADAS-Comparative (QUADAS-C), an extension of QUADAS-2 designed for comparative diagnostic performance studies. Two reviewers (H.J. and W.J.) independently conducted the data extraction and quality assessment, with discrepancies resolved through a discussion with a third reviewer (H.W.).

## Data synthesis and statistical analysis

In this meta-analysis, we used the hierarchical summary receiver-operating characteristic (HSROC) model to summarize the estimates of sensitivity, specificity, and their 95% confidence intervals (CIs) (18). Forest plots and HSROC curves were used to graphically present the results. For studies that provided at least two results, we chose the most accurate; for studies that provided validation results of internal

validation and external validation, we chose the latter. The Cochran Q statistics and Higgins  $I^2$  value were employed to measure the degree of heterogeneity among the studies:  $I^2$  value between 0% and 40%, not important;  $I^2$  value between 30% and 60%, moderate;  $I^2$  value between 50% and 90%, substantial; and  $I^2$  value between 75% and 100%, considerable (19). To explore the source of heterogeneity, the following covariates were used to perform metaregressions: the country where the study was conducted, publication year, number of patients, number of malignancies, and malignant rate. The Deeks' funnel plot was used to assess the publication bias, and the statistical significance was tested with the Deeks' funnel plot asymmetry test. All analyses were performed with STATA (version 15.1) and R statistical software (version 3.6.1), with a  $P$ -value  $<0.05$  indicating statistical significance.

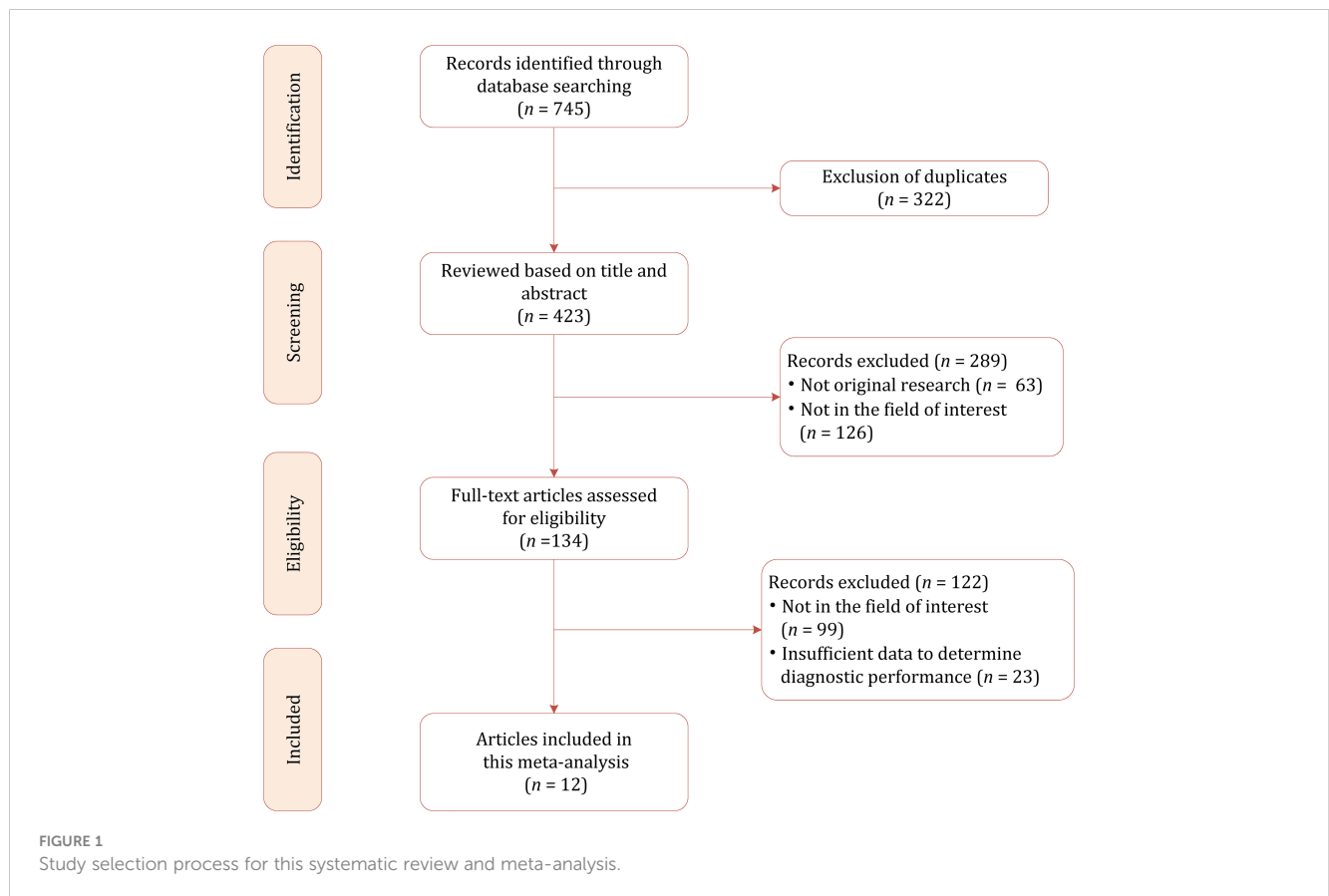
## Results

### Literature search and data extraction

Based on our literature search strategy, a total of 745 references were identified initially, of which 322 were excluded for duplicates. After examining the titles and abstracts, 289 results were excluded because they were not relevant to this meta-analysis. We reviewed the remaining 134 full-text articles, and 122 were excluded for reasons as follows: insufficient data to determine diagnostic performance ( $n = 23$ ) and not in the field of interest ( $n = 99$ ). Finally, a total of 12 studies were included in this meta-analysis (20–31). The flowchart of the literature selection process is demonstrated in Figure 1.

### Characteristics of the included studies

For this meta-analysis, all the studies included had a retrospective study design, with eight studies reporting a head-to-head comparison between O-RADS and the ADNEX model (21, 23–28, 31) and seven studies reporting a direct comparison between O-RADS and IOTA SR (including three studies performing a comparison between all three guidelines) (20, 22, 23, 25, 29–31). The study sample ranged from 122 to 1,179 patients, with an average age of 35–52.3. The average size of the adnexal mass lesion in seven studies was 60–190 mm. In most studies, surgical pathology results were used as the reference; however, in two studies, a follow-up of 12–24 months was also used when histopathological results were not available (20, 31). Borderline lesions were reported in 11 studies, and all of these studies classified those masses as malignant. Ten studies reported the experience of radiologists, with most of them having at least 5 years of experience. In two studies, readers had partial knowledge of patients' clinical information (22, 23). The reported kappa values were substantial to almost perfect for the three guidelines: 0.62–0.93 for O-RADS, 0.85–0.86 for the ADNEX model, and 0.73–0.90 for IOTA SR. The most used cutoff values for O-RADS and the ADNEX model were  $\geq 4$  and  $\geq 10\%$ , respectively. For IOTA SR, six of seven studies



reported details on indeterminate cases, with a prevalence of 5.4%–24.7%. Of these cases, the malignant rate ranged from 43.1% to 48.6%. Details on demographic characteristics and study characteristics are presented in [Tables 1, 2](#).

### Quality assessment

The overall quality assessment using QUADAS-2 is presented in [Figure 2](#). For the patient selection domain, five studies had an unclear

**TABLE 1** Demographic characteristics of the included studies.

First author	Country	Year	No. of patients	No. of lesions	Malignant	Borderline Tumors	Age (year, mean ± SD/median)	Tumor size (mm, mean ± SD/range)	Reference
Basha et al. (20)	Egypt	2021	609	647	178	32	48 ± 13.7	NA	Pathological or 2-year follow-up
Chen et al. (21)	Taiwan	2022	322	322	264	8	44 (20–83)	NA	Pathological
Guo et al. (22)	China	2022	575	592	145	22	36.6 ± 13.9/46.5 ± 13.9 <sup>a</sup>	NA	Pathological
Hiett et al.	USA	2021	150	150	40	12	48.2 ± 1.7/47.5 ± 3.1 <sup>a</sup>	83.7 ± 5.2/101.4 ± 7.7 <sup>a</sup>	Pathological
Lai et al. (24)	China	2021	734	734	170	69	35 (29–46)/48 (36–55.3) <sup>a</sup>	NA	Pathological
Pelayo et al. (25)	Spain	2023	122	122	41	NA	51.4 ± 15.7	94.2 ± 52.1	Pathological
Poonyakanok et al. (26)	Thailand	2023	357	357	61	11	43 (35–53)	60 (42–83.8) 190 (159–277.5) <sup>a</sup>	Pathological

(Continued)

TABLE 1 Continued

First author	Country	Year	No. of patients	No. of lesions	Malignant	Borderline Tumors	Age (year, mean $\pm$ SD/median)	Tumor size (mm, mean $\pm$ SD/range)	Reference
Spagnol et al. (27)	Italy	2023	514	514	89	25	51 (41–63)	NA	Pathological
Wang et al. (28)	China	2023	445	445	180	31	40.19 $\pm$ 15.96/ 52.34 $\pm$ 13.15 <sup>a</sup>	95 $\pm$ 285/605 $\pm$ 492 <sup>a</sup>	Pathological
Xie et al. (29)	China	2022	453	453	269	48	48.8 $\pm$ 13.4	105.9 $\pm$ 64	Pathological
Yang et al. (30)	China	2023	1,179	1,179	213	67	38 (29–50)	69 (51–94)	Pathological
Yoeli-Bik et al. (31)	USA	2023	511	511	81	15	45.4 $\pm$ 14.8	NA	Pathological (341) or 1-year follow-up

NA, not available; SD, standard deviation.

<sup>a</sup>Benign/malignant.

risk of bias because of two high malignancy rates (20, 23, 25, 28, 29). In three studies, the details on blinding were not provided or reported whether readers have partial knowledge of patient information, thus were assigned an unclear risk of bias in terms of index domain (22, 23, 28). Supplementary Tables S1, S2 show the details of the quality assessment using QUADAS-2 and QUADAS-C.

## Diagnostic performance

In terms of individual studies, the sensitivity and specificity for O-RADS, the ADNEX model, and IOTA SR were 0.88–1.00 and 0.46–0.94, 0.88–0.98 and 0.64–0.91, and 0.80–1.00 and 0.52–0.83, respectively. The pooled sensitivity and specificity for 12 studies using O-RADS at a cutoff value of  $\geq 4$  were 0.96 (95% CI 0.92–0.98) and 0.82 (95% CI 0.71–0.90), with an area under the HSROC of 0.97 (95% CI 0.95–0.98); for 8 studies using ADNEX at a cutoff value of  $\geq 10\%$ , the pooled sensitivity and specificity were 0.94 (95% CI 0.91–0.95) and 0.83 (95% CI 0.77–0.88), with an area under the HSROC of 0.95 (0.93–0.96); and for 7 studies using IOTA SR, the pooled sensitivity and specificity were 0.91 (95% CI 0.82–0.96) and 0.86 (95% CI 0.76–0.93), with an area under the HSROC of 0.95 (95% CI 0.92–0.96). The coupled forest plots for O-RADS, the ADNEX model, and IOTA SR are presented in Figure 3, and the head-to-head comparisons between the O-RADS and the ADNEX model as well as IOTA SR are presented in Figure 4.

Substantial heterogeneity was observed with respect to sensitivity and specificity for all three guidelines. Specifically, the  $I^2$  values for O-RADS were 62.4% (95% CI 38.9%–85.9%) and 96.3% (95% CI 95.1%–97.5%) for sensitivity and specificity; for ADNEX, 14.6% (95% CI 0.0%–74.2%) and 90.4% (95% CI 85.2%–95.6%); and for IOTA SR, 88.4% (95% CI 81.3%–95.6%) and 97.0% (95% CI 95.8%–98.3%), respectively. The space between the 95% confidence region and the 95% prediction region also suggested heterogeneity among the studies (Supplementary Figure S1). Metaregression analyses revealed that for O-RADS, the country where the studies were conducted (China vs. other countries) was a significant factor

for the heterogeneity of sensitivity ( $P = 0.03$ ), and the number of patients was the significant factor for the ADNEX model ( $P = 0.01$ ) and IOTA SR ( $P = 0.01$ ).

We compared the sensitivity and specificity between guidelines as used in the studies providing direct comparisons. Our analyses demonstrated that no significant differences were found between O-RADS and the ADNEX model, with  $P = 0.57$  for sensitivity and  $P = 0.87$  for specificity. Likewise, no significant differences were observed between O-RADS and IOTA SR, with  $P = 0.21$  for sensitivity and  $P = 0.12$  for specificity. The Deeks' funnel plots demonstrated that there was no publication bias for all three guidelines, with  $P$ -values of 0.88, 0.22, and 0.87 for O-RADS, the ADNEX model, and IOTA SR.

## Discussion

In this meta-analysis, we systematically compared three guidelines for the risk stratification of ovarian carcinoma. Based on 12 studies, our findings demonstrated that all three risk stratification systems had high diagnostic performance, with the area under the HSROC of 0.97, 0.95, and 0.95 for O-RADS, the ADNEX model, and IOTA SR. No significant differences were found between O-RADS and the ADNEX model ( $P = 0.85$ ) as well as IOTA SR ( $P = 0.15$ ) using the respective eight and seven head-to-head comparison studies. In addition to overall accuracy, we compared the pooled sensitivity and specificity of O-RADS with the ADNEX model and the IOTA SR; however, no significant difference was found between these three guidelines. In the current study, the pooled sensitivity and specificity from 12 studies of O-RADS were 0.95 and 0.82. In two recent meta-analyses evaluating the overall accuracy of O-RADS US, the pooled sensitivity and specificity based on 15 and 10 studies, respectively, were 0.95 and 0.82 and 0.96 and 0.77 at a cutoff value of  $\geq 4$ , which is comparable with our findings (32, 33). As for IOTA SR and the ADNEX model, the reported pooled sensitivity and specificity from previous meta-analyses or systematic reviews were 0.93 and 0.80 from 5 studies (7) and 0.92 and 0.82 from 10

TABLE 2 Study characteristics of the included studies.

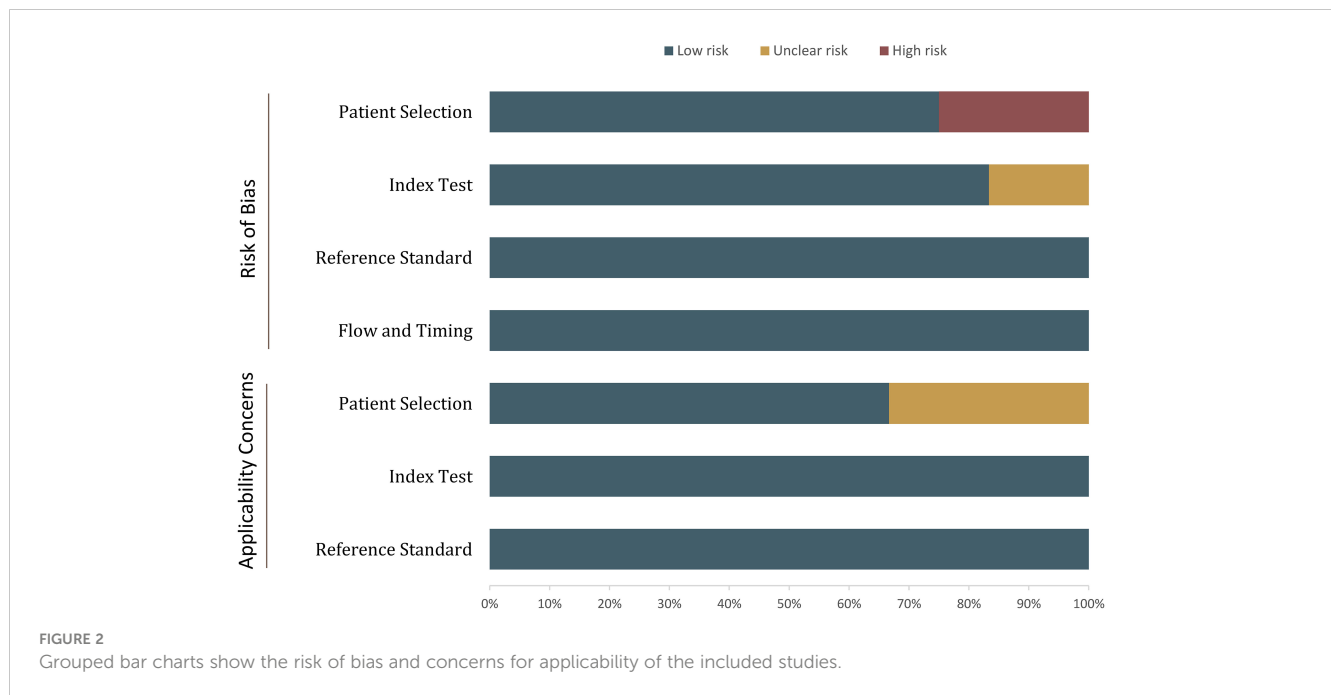
First author	Publication year	Study design	Analysis	Period	No. of readers	Experience (years)	Blinded	Guideline	$\kappa$ value	Cutoff value
Basha et al. (20)	2021	Retrospective	Per lesion	2016.5–2019.12	5	15	Yes	O-RADS/IOTA SR	0.77/0.63	$\geq 4$ /NA
Chen et al. (21)	2022	Retrospective	Per person	2020.1–2020.10	5	$\geq 5$	Yes	O-RADS/ADNEX	0.78 <sup>a</sup>	$\geq 4$ / $>10\%$
Guo et al. (22)	2022	Retrospective	Per lesion	2017–2020	4	$\geq 10$ (2 readers)/1 (2 readers)	Yes <sup>b</sup>	O-RADS/IOTA SR	0.71/0.77	$\geq 4$ /NA
Hiett et al.	2021	Retrospective	Per person	2018.3–2021.2	2	25	Yes <sup>b</sup>	O-RADS/ADNEX/IOTA SR	NA	$\geq 4$ / $>10\%$ /NA
Lai et al. (24)	2021	Retrospective	Per person	2017.1–2020.11	2	5	Yes	O-RADS/ADNEX	0.83/0.86	$\geq 4$ / $>10\%$
Pelayo et al. (25)	2023	Retrospective	Per person	2021.01–2022.12	2	$\geq 15$	Yes	O-RADS/ADNEX/IOTA SR	NA	$\geq 4$ / $>10\%$ /NA
Poonyakanok et al. (26)	2023	Retrospective	Per person	2018.05–2019.05	6	5–8	Yes	O-RADS/ADNEX	NA	$\geq 4$ / $>10\%$
Spagnol et al. (27)	2023	Retrospective	Per person	2018.01–2021.12	2	NA	Yes	O-RADS/ADNEX	NA	$\geq 4$ / $>10\%$
Wang et al. (28)	2023	Retrospective	Per person	2020.01–2021.12	2	$>5$	NA	O-RADS/ADNEX	0.87/0.85	$\geq 4$ / $>10\%$
Xie et al. (29)	2022	Retrospective	Per person	2017.1–2020.9	2	5	Yes	O-RADS/IOTA SR	0.62/0.73	$\geq 4$ /NA
Yang et al. (30)	2023	Retrospective	Per person	2021.09–2022.02	2	10/NA	Yes	O-RADS/IOTA SR	0.93/0.90	$\geq 4$ /NA
Yoeli-Bik et al. (31)	2023	Retrospective	Per person	2017.1–2022.10	3	20/40/NA	Yes	O-RADS/ADNEX/IOTA SR	NA	$\geq 4$ / $>10\%$ /NA

ADNEX, Assessment of Different NEoplasia in the adneXa model; NA, not available. O-RADS, Ovarian-Adnexal Reporting and Data System; IOTA SR, Ovarian Tumor Analysis Simple Rules.

<sup>a</sup>Only for O-RADS.

<sup>b</sup>Known partial clinical information of the patients.





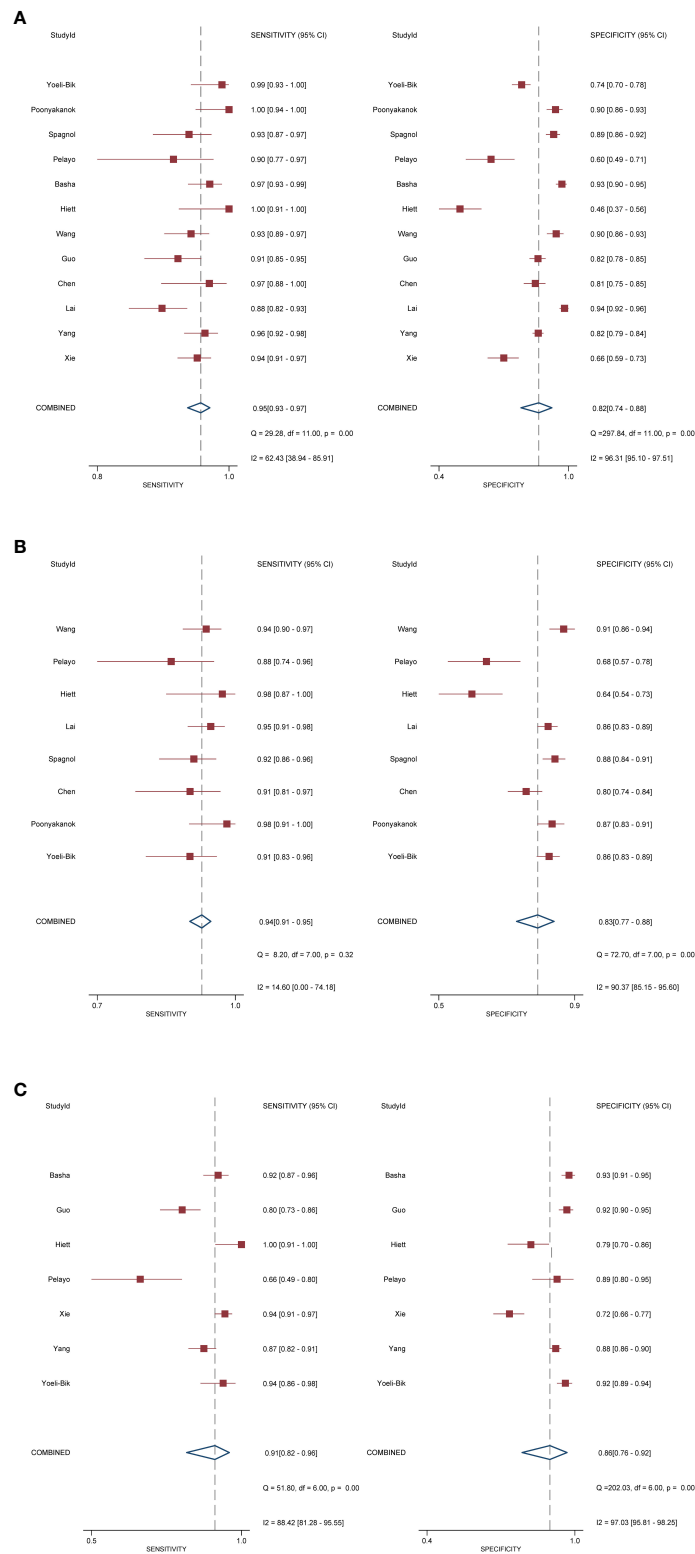
studies (at the cutoff value of 15%) (34), respectively. In addition to overall diagnostic performance, all three guidelines reported high interreader agreement between radiologists. However, the kappa values were provided only in three studies for IOTA SR and in two studies for the ADNEX model. Therefore, it is unfeasible to perform a meta-analysis and compare the interreader agreement between studies.

The O-RADS US risk stratification and management tool is another effort for the standardization of the risk stratification of adnexal masses, which is modeled on the IOTA rules and based on IOTA data that included 5,905 patients with adnexal masses. Even though IOTA SR had high diagnostic performance, which may result in up to one-quarter of indeterminate lesions, it is suggested that clinicians with less experience need to be assisted by senior clinicians in using diagnostic models to correctly diagnose these lesions (5). In the current study, the reported inconclusive adnexal masses ranged from 5.4% to 24.7%, and nearly half of them were malignant. For unclassified lesions, IOTA SR recommends referring the patient to experts for subjective assessment of US findings, which could provide the most accurate diagnosis. In an earlier study, subjective assessment of adnexal masses using IOTA SR yielded a sensitivity of 89% and a specificity of 80% (5). To address the issue regarding the absence of experienced US examiners, the Simple Rules Risk (SRR) model was developed as a solution, which is a logistic regression model that utilizes TV-US features based on the SR. Its primary objective is to provide an estimated risk of malignancy for any type of adnexal masses, thereby eliminating inconclusive classification. However, because there were only three studies in our meta-analysis that reported the results of SRR, it is unfeasible to pool the data (23, 25, 27).

Compared with IOTA SR, the relatively lower specificity of O-RADS may lead to overtreatment of adnexal masses. However, for the O-RADS indeterminate adnexal masses (categories 3 and 4), the use

of O-RADS MRI is suggested for further evaluation of these masses in order to better characterize their nature (32). Compared with the ADNEX model which used three clinical variables and six ultrasound variables, the O-RADS classification only employs ultrasound characteristics to classify ovarian tumors (24, 35). One shortcoming that should be addressed in the present O-RADS is that two variables (bilocular for cystic lesions and shadowing for solid smooth lesions) were not taken into consideration; therefore, it exhibited a higher sensitivity but a lower specificity than the ADNEX model, as reported in various studies (36). Some studies demonstrated that by considering acoustic shadowing as an indication of benign lesions, the overall diagnostic performance was improved significantly, with AUC increased from 0.91 to 0.94 ( $P = 0.01$ ) (37). These findings suggested that acoustic shadowing is an important US feature for classifying ovarian tumors, especially in solid lesions, and should be included. In the updated O-RADS US v2022, the addition of the descriptors bilocular for cystic lesions and acoustic shadowing for solid smooth lesions, along with the expanded lexicon descriptors for the typical appearance of some classic benign lesions, may be beneficial for reducing overtreatment (38).

Although O-RADS, IOTA SR, and the ADNEX model all demonstrated high diagnostic performance, in clinical practice, subjective assessment of pelvic ultrasound images by clinicians with considerable experience in gynecologic ultrasound has demonstrated a high degree of accuracy in differentiating between benign and malignant pelvic lesions (7, 39). In fact, subjective assessment appears to be the best method to predict the likelihood of a pelvic malignancy (40). However, clinicians with this level of expertise may not be universally available, presenting a challenge to accurate diagnosis and patient management. Transferring the expertise of experienced ultrasound examiners to less experienced ones poses a significant challenge in the field of gynecologic US. While scoring systems and risk calculation models can potentially



**FIGURE 3** Coupled forest plot of pooled sensitivity and specificity. **(A)** Ovarian-Adnexal Reporting and Data System; **(B)** Assessment of Different NEoplasias in the adneXa; **(C)** International Ovarian Tumor Analysis Simple Rules.

assist less experienced examiners in characterizing pelvic lesions, there are valid criticisms regarding the complexity of US information required by some ultrasound-based risk calculation models, particularly outside of specialist centers. One of the primary

criticisms of these models is their reliance on sophisticated ultrasonic features and measurements that may be challenging to obtain consistently and accurately by less experienced examiners. Moreover, the interpretation of ultrasound findings can be



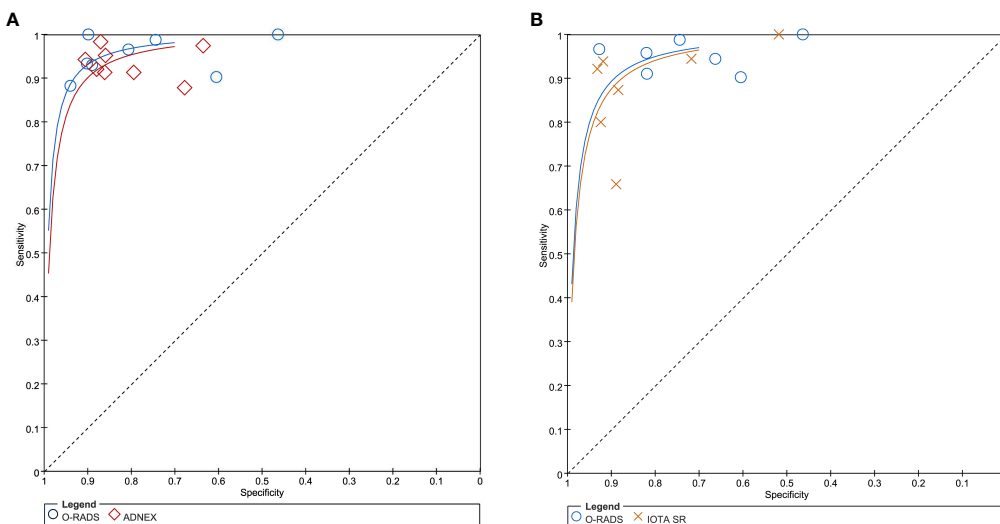


FIGURE 4

(A) Comparison between O-RADS and ADNEX; (B) comparison between O-RADS and IOTA SR. ADNEX, Assessment of Different NEoplasias in the adneXa; O-RADS, Ovarian-Adnexal Reporting and Data System; IOTA SR, International Ovarian Tumor Analysis Simple Rules.

subjective and may vary among examiners, leading to potential discrepancies in risk assessment and diagnostic accuracy. Considering the low incidence but high mortality rate, risk stratification of adnexal masses is a trade-off between sensitivity and specificity, which should take into consideration a number of factors such as risk tolerance for missing cancer and surgery risk (31). Therefore, the physician and the patient have to contemplate the risks and benefits of any procedure and determine the individual cutoff in specific circumstances in which the adnexal mass is evaluated.

The main strength of our study is that we systematically summarized currently available evidence on the comparison between O-RADS US with the IOTA SR and the ADNEX model. However, our study has some limitations that must be taken into consideration. First, all studies included in this meta-analysis had a retrospective study design, which was subjected to a selection bias, emphasizing the need for prospective validation. Second, substantial heterogeneity was observed among the studies, which affected the general applicability of our study. To investigate the heterogeneity, we performed metaregression analysis using several potential covariates. Nevertheless, these analyses only accounted for the partial source of heterogeneity, and a portion remains unexplained. Third, comparisons between O-RADS and the ADNEX model as well as the IOTA SR were based on nine and seven studies, respectively; thus, our conclusions and results should be regarded with caution and future large, prospective studies are needed to compare these different guidelines.

## Conclusion

The O-RADS US, the ADNEX model, and IOTA SR showed favorable diagnostic accuracy for risk stratification of adnexal masses, and these three guidelines demonstrated comparable performance. However, O-RADS US yielded a slightly higher

sensitivity but a lower specificity than the ADNEX model and IOTA SR.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Author contributions

JH: Data curation, Resources, Writing – original draft. JW: Formal analysis, Methodology, Validation, Writing – review & editing. WH: Supervision, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin.* (2023) 73:17–48. doi: 10.3322/caac.21763
- Veneziani AC, Gonzalez-Ochoa E, Alqaisi H, Madariaga A, Bhat G, Rouzbahman M, et al. Heterogeneity and treatment landscape of ovarian carcinoma. *Nat Rev Clin Oncol.* (2023) 20:820–42. doi: 10.1038/s41571-023-00819-1
- Sadowski EA, Rockall A, Thomassin-Naggara I, Barroilhet LM, Wallace SK, Jha P, et al. Adnexal lesion imaging: past, present, and future. *Radiology.* (2023) 307:e223281. doi: 10.1148/radiol.223281
- Sisodia RC, del Carmen MG. Lesions of the ovary and fallopian tube. *N Engl J Med.* (2022) 387:727–36. doi: 10.1056/NEJMr2108956
- Timmerman D, Amey L, Fischerova D, Epstein E, Melis GB, Guerriero S, et al. Simple ultrasound rules to distinguish between benign and Malignant adnexal masses before surgery: prospective validation by IOTA group. *BMJ.* (2010) 341:c6839. doi: 10.1136/bmj.c6839
- Timmerman D, Testa AC, Bourne T, Amey L, Jurkovic D, Van Holsbeke C, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol.* (2008) 31:681–90. doi: 10.1002/uog.5365
- Meys EMJ, Kaijser J, Kruiwagen RPFM, Slangen BFM, Van Calster B, Aertgeerts B, et al. Subjective assessment versus ultrasound models to diagnose ovarian cancer: A systematic review and meta-analysis. *Eur J Cancer Oxf Engl 1990.* (2016) 58:17–29. doi: 10.1016/j.ejca.2016.01.007
- Van Calster B, Valentin L, Van Holsbeke C, Testa AC, Bourne T, Van Huffel S, et al. Polytomous diagnosis of ovarian tumors as benign, borderline, primary invasive or metastatic: development and validation of standard and kernel-based risk prediction models. *BMC Med Res Methodol.* (2010) 10:96. doi: 10.1186/1471-2288-10-96
- Timmerman D, Van Calster B, Jurkovic D, Valentin L, Testa AC, Bernard J-P, et al. Inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and Malignant adnexal tumors. *J Clin Oncol Off J Am Soc Clin Oncol.* (2007) 25:4194–200. doi: 10.1200/JCO.2006.09.5943
- Van Calster B, Valentin L, Van Holsbeke C, Zhang J, Jurkovic D, Lissoni AA, et al. A novel approach to predict the likelihood of specific ovarian tumor pathology based on serum CA-125: a multicenter observational study. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol.* (2011) 20:2420–8. doi: 10.1158/1055-9965.EPI-11-0422
- 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 Off J Int Soc Ultrasound Obstet Gynecol.* (2012) 40:355–9. doi: 10.1002/uog.11088
- 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 Off J Int Soc Ultrasound Obstet Gynecol.* (2011) 38:450–5. doi: 10.1002/uog.9012
- Yamamoto Y, Yamada R, Oguri H, Maeda N, Fukaya T. Comparison of four Malignancy risk indices in the preoperative evaluation of patients with pelvic masses. *Eur J Obstet Gynecol Reprod Biol.* (2009) 144:163–7. doi: 10.1016/j.ejogrb.2009.02.048
- Timmerman D, Planchamp F, Bourne T, Landolfo C, du Bois A, Chiva L, et al. ESGO/ISUOG/IOTA/ESGE Consensus Statement on preoperative diagnosis of ovarian tumors. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol.* (2021) 58:148–68. doi: 10.1002/uog.23635
- Andreotti RF, Timmerman D, Strachowski LM, Froyman W, Benacerraf BR, Bennett GL, et al. O-RADS US risk stratification and management system: A consensus guideline from the ACR ovarian-adnexal reporting and data system committee. *Radiology.* (2019) 294:191150. doi: 10.1148/radiol.2019191150
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Epidemiol Biostat Public Health.* (2009) 6:e1–e34. doi: 10.1016/j.jclinepi.2009.06.006
- Whiting PF. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* (2011) 155:529. doi: 10.7326/0003-4819-155-8-201110180-00009
- Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med.* (2001) 20:2865–84. doi: 10.1002/sim.942
- Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* (2011) 343:889–93. doi: 10.1136/bmj.d5928
- Basha MAA, Metwally MI, Gamil SA, Khater HM, Aly SA, El Sammak AA, et al. Comparison of O-RADS, GI-RADS, and IOTA simple rules regarding Malignancy rate, validity, and reliability for diagnosis of adnexal masses. *Eur Radiol.* (2021) 31:674–84. doi: 10.1007/s00330-020-07143-7
- Chen G-Y, Hsu T-F, Chan I-S, Liu C-H, Chao W-T, Shih Y-C, et al. Comparison of the O-RADS and ADNEX models regarding Malignancy rate and validity in evaluating adnexal lesions. *Eur Radiol.* (2022) 32:7854–64. doi: 10.1007/s00330-022-08803-6
- Guo Y, Zhao B, Zhou S, Wen L, Liu J, Fu Y, et al. A comparison of the diagnostic performance of the O-RADS, RMI4, IOTA LR2, and IOTA SR systems by senior and junior doctors. *Ultrasound Seoul Korea.* (2022) 41:511–18. doi: 10.14366/usg.21237
- Hielt AK, Sonek JD, Guy M, Reid TJ. Performance of IOTA Simple Rules, Simple Rules risk assessment, ADNEX model and O-RADS in differentiating between benign and Malignant adnexal lesions in North American women. *Ultrasound Obstet Gynecol.* (2022) 59:668–76. doi: 10.1002/uog.24777
- Lai H-W, Lyu G-R, Kang Z, Li L-Y, Zhang Y, Huang Y-J. Comparison of O-RADS, GI-RADS, and ADNEX for diagnosis of adnexal masses: an external validation study conducted by junior sonologists. *J Ultrasound Med Off J Am Inst Ultrasound Med.* (2021) 41:6:1497–507. doi: 10.1002/jum.15834
- Pelayo M, Pelayo-Delgado I, Sancho-Sauco J, Sanchez-Zurdo J, Abarca-Martinez L, Corraliza-Galán V, et al. Comparison of ultrasound scores in differentiating between benign and Malignant adnexal masses. *Diagn Basel Switz.* (2023) 13:1307. doi: 10.3390/diagnostics13071307
- Poonyakanok V, Tanmahasamut P, Jaishuen A. Prospective comparative trial comparing O-RADS, IOTA ADNEX model, and RMI score for preoperative evaluation of adnexal masses for prediction of ovarian cancer. *J Obstet Gynaecol Res.* (2023) 49:1412–7. doi: 10.1111/jog.15624
- Spagnol G, Marchetti M, Tommasi OD, Vitagliano A, Cavallin F, Tozzi R, et al. Simple rules, O-RADS, ADNEX and SRR model: Single oncologic center validation of diagnostic predictive models alone and combined (two-step strategy) to estimate the risk of Malignancy in adnexal masses and ovarian tumors. *Gynecol Oncol.* (2023) 177:109–16. doi: 10.1016/j.ygyno.2023.08.012
- Wang R, Yang Z. Evaluating the risk of Malignancy in adnexal masses: validation of O-RADS and comparison with ADNEX model, SA, and RMI. *Ginekol Pol.* (2023) 94:799–806. doi: 10.5603/GP.a2023.0019
- Xie WT, Wang YQ, Xiang ZS, Du ZS, Huang SX, Chen YJ, et al. Efficacy of IOTA simple rules, O-RADS, and CA125 to distinguish benign and Malignant adnexal masses. *J Ovarian Res.* (2022) 15:15. doi: 10.1186/s13048-022-00947-9
- Yang Y, Ju H, Huang Y. Diagnostic performance of IOTA SR and O-RADS combined with CA125, HE4, and risk of Malignancy algorithm to distinguish benign and Malignant adnexal masses. *Eur J Radiol.* (2023) 165:110926. doi: 10.1016/j.ejrad.2023.110926
- Yoeli-Bik R, Longman RE, Wroblewski K, Weigert M, Abramowicz JS, Lengyel E. Diagnostic performance of ultrasonography-based risk models in differentiating between benign and Malignant ovarian tumors in a US cohort. *JAMA Netw Open.* (2023) 6:e2323289. doi: 10.1001/jamanetworkopen.2023.23289
- Zhang Q, Dai X, Li W. Systematic review and meta-analysis of O-RADS ultrasound and O-RADS MRI for risk assessment of ovarian and adnexal lesions. *Am J Roentgenol.* (2023) 221:21–33. doi: 10.2214/AJR.22.28396
- Lee S, Lee JE, Hwang JA, Shin H. O-RADS US: A systematic review and meta-analysis of category-specific Malignancy rates. *Radiology.* (2023) 308:e223269. doi: 10.1148/radiol.223269
- Huang X, Wang Z, Zhang M, Luo H. Diagnostic accuracy of the ADNEX model for ovarian cancer at the 15% Cut-off value: A systematic review and meta-analysis. *Front Oncol.* (2021) 11:684257. doi: 10.3389/fonc.2021.684257
- Van Calster B, Van Hoorde K, Froyman W, Kaijser J, Wynants L, Landolfo C, et al. Practical guidance for applying the ADNEX model from the IOTA group to discriminate between different subtypes of adnexal tumors. *Facts Views Vis ObGyn.* (2015) 7:32–41.
- Strachowski LM, Jha P, Chawla TP, Davis KM, Dove CK, Glanc P, et al. O-RADS for ultrasound: A user's guide, from the AJR special series on radiology reporting and data systems. *AJR Am J Roentgenol.* (2021) 216:1150–65. doi: 10.2214/AJR.20.25064
- Hack K, Gandhi N, Bouchard-Fortier G, Chawla TP, Ferguson SE, Li S, et al. External validation of O-RADS US risk stratification and management system. *Radiology.* (2022) 304:114–20. doi: 10.1148/radiol.211868

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2024.1354837/full#supplementary-material>

38. Strachowski LM, Jha P, Phillips CH, Blanchette-Porter MM, Froyman W, Glanc P, et al. O-RADS US v2022: an update from the American College of radiology's ovarian-adnexal reporting and data system US committee. *Radiology*. (2023) 308:e230685. doi: 10.1148/radiol.230685

39. Timmerman D, Schwärzler 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 Off J Int*

*Soc Ultrasound Obstet Gynecol*. (1999) 13:11–6. doi: 10.1046/j.1469-0705.1999.13010011.x

40. Valentin L. Prospective cross-validation of Doppler ultrasound examination and gray-scale ultrasound imaging for discrimination of benign and Malignant pelvic masses. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol*. (1999) 14:273–83. doi: 10.1046/j.1469-0705.1999.14040273.x