Original Article

🛡 Thieme

Evaluation of IOTA-ADNEX Model and Simple Rules for Identifying Adnexal Masses by Operators with Varying Levels of Expertise: A Single-Center Diagnostic Accuracy Study

Authors
Maria Giourga[1] , Abraham Pouliakis[2], Panagiotis Vlastarakos[1] , Sofoklis Stavrou[3], Maria Tsiriva[1], Angeliki Gerede[4], Georgios Daskalakis[1,5], Charalampos Voros[1], Petros Drakakis[6], Ekaterini Domali[1]

Affiliations
1 1st Department of Obstetrics and Gynecology, National and Kapodistrian University of Athens, Alexandra Hospital, Athens, Greece
2 2nd Department of Pathology, National and Kapodistrian University of Athens School of Medicine, Athens, Greece
3 first department of obstetrics and gynecology, National and Kapodistrian University of Athens Faculty of Medicine, Athens, Greece
4 3rd Department of Obstetrics and Gynecology, Aristotle University of Thessaloniki School of Medicine, Kavala, Greece
5 First Department of Obstetrics and Gynaecology, University of Athens, Greece, National and Kapodistrian University of Athens School of Medicine, Athens, Greece
6 Third Department of Obstetrics and Gynecology, Attikon Hospital, Athens, Greece, National and Kapodistrian University of Athens School of Medicine, Athens, Greece

Correspondence
Maria Giourga, MD
1st Department of Obstetrics and Gynecology, National and Kapodistrian University of Athens, Alexandra Hospital
Leof. Vasilissis Sofias 80
11528 Athens
Greece
giourga.m@gmail.com

ABSTRACT

Objectives   The discrimination of ovarian lesions presents a significant problem in everyday clinical practice with ultra-sonography appearing to be the most effective diagnostic technique. The aim of our study was to externally evaluate the performance of different diagnostic models when applied by examiners with various levels of experience.

Methods   This was a diagnostic accuracy study including women who were admitted for adnexal masses, between July 2018 and April 2021, to a Greek tertiary oncology center. Preoperatively sonographic data were evaluated by an expert gynecologist, a 6th and a 1st year gynecology resident, who applied the International Ovarian Tumor Analysis (IOTA) Simple Rules (SR) and Assessment of Different NEoplasias in the adneXa (ADNEX) model to discriminate between benign and malignant ovarian tumors. The explant pathology report was used as the reference diagnosis. Kappa statistics were used for the investigation of the level of agreement between the examined systems and the raters.

Results   We included 66 women, 39 with benign and 27 with malignant ovarian tumors. ADNEX (with and without "CA-125") had high sensitivity (96–100%) when applied by all raters but a rather low specificity (36%) when applied by the 1st year resident. SR could not be applied in 6% to 17% of the cases. It had slightly lower sensitivity, higher specificity, and higher overall accuracy, especially when applied by the 1st year resident (61% vs. 92%), compared to ADNEX.

Conclusion   Both ADNEX and SR can be utilized for screening in non-oncology centers since they offer high sensitivity even when used by less experienced examiners. In the hands of inexperienced examiners, SR appears to be the best model for assessing ovarian lesions.

Introduction

In women with ovarian malignancies, early detection is critical as it can lead to high cure and survival rates [1]. When a malignancy is suspected, referral to an expert oncology center for further management is crucial for attaining the best therapeutic results. Ultrasonography appears to be the most effective diagnostic technique, which assists in the differentiation between benign and malignant adnexal masses prior to surgery, but only when conducted by experienced examiners [2, 3]. However, women seeking evaluation aren't always assessed by physicians with expertise. In an effort to improve care, the International Ovarian Tumor Analysis (IOTA) group has developed specific ultrasound criteria ("Simple Rules", "SR"), which include five features typical for benign tumors (B-features) and five features typical for malignant tumors (M-features) that can classify the majority of adnexal masses as probably benign or probably malignant [4]. Moreover, the IOTA group has created a logistic tool, the Assessment of Different NEoplasias in the AdneXa (ADNEX) model, which calculates the preoperative risk of a lesion being benign, borderline, stage I, II–IV or secondary metastatic cancer, using three clinical and six ultrasonographic variables [5]. The aim of our study was to evaluate the diagnostic accuracy of the IOTA's SR and ADNEX model in the preoperative discrimination of adnexal masses, when used by three raters with varying levels of experience.

Methods

Women of all ages who underwent biopsy or surgical removal of adnexal masses at our gynecological oncology center were included in the study if they had available ultrasonographic data of the mass up to 120 days prior to surgery. As per our hospital protocol, all women undergo ultrasound assessment prior to surgery. The study was approved by our Hospital's Ethics Committee and all included women provided written consent. The STARD statement was followed for reporting the study [6].

All ultrasound examinations were performed by the same obstetrics and gynecology assistant professor, an expert in ultrasonography. Images and videos were obtained, stored, and then assessed consecutively by three different raters with varying levels of experience. Rater 1 was an expert gynecologist, rater 2 was a 6th year obstetrics and gynecology resident, and rater 3 was a 1st year resident. The expert gynecologist had performed more than 300 gynecological ultrasound examinations per year. The 6th year resident had performed approximately 100 ultrasound examinations per year for the past 4 years under supervision. The 1st year resident had no prior experience in gynecological ultrasound. Each examiner evaluated the same images and videos and then applied the ADNEX model with and without the incorporation of "CA-125" (ADNEX 125), as well as SR, to categorize the adnexal masses. When multiple masses were present, the one with the most complex morphology was chosen, as indicated by the literature [7]. After surgery, the histopathology reports were collected and used as the reference standard for the diagnosis of each mass. The patients' age, history, and the serum levels of "CA-125" were retrieved and obtained from their medical records. All ratings were triple blinded, with the raters being unaware of the patient they were rating, the results of the other raters, and the histology reports.

Data analysis was performed with SAS 9.4 for Windows (SAS Institute Inc. NC, USA) [8, 9]. We calculated the sensitivity, specificity, positive and negative predictive values (PPV and NPV), false-positive and false-negative rates (FPR and FNR, respectively). The detailed diagnostic results and performance metrics are cumulatively presented for all raters and for all scoring systems in ▶ **Table 2** and **3**, respectively. We applied kappa (κ) statistics for the investigation of the level of agreement between the examined systems in a pairwise comparison. According to the κ value, the agreement is characterized as: < 0 no, 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1 as almost perfect agreement [10]. We also examined the level of agreement for the same rater when using different scoring systems. The first comparison was between benign and malignant cases, the second comparison was for the staging. For this step the agreement was estimated using the Kendall's test that takes into account information related to the proximity of the test results and introduces a metric (W) ranging from no agreement (W ≤ 0) to complete agreement (W = 1) [11].

Results

Study population and baseline characteristics

Out of 227 women admitted to our center due to adnexal masses, 121 were operated on or submitted to biopsy. Out of these, 29 did not have sufficient ultrasonographic data stored, in 21 cases ultrasound was performed more than 120 days prior to surgery and 5 did not consent to participating in our study. In total, 66 women were included. 60 of them underwent surgery and 6 underwent biopsy (▶ **Fig. 1** – Flowchart). All women were evaluated using the three systems ADNEX model, ADNEX 125, and SR by three observers. 39 women (59.1%) had benign lesions and 27 (40.9%) had malignant lesions, based on their histopathology reports. Women with malignant lesions had a higher age (median: 58 years) than women with benign lesions (median: 46 years), which was as expected but without statistical significance (p = 0.15). The histopathology of the included ovarian tumors is depicted in ▶ **Table 1**.

Analysis of the performance of the scoring systems for each observer

For the ADNEX model, a cut-off of 10% was used in order to classify a lesion as benign or malignant. Before analyzing the level of agreement among the raters, the performance of each scoring system and each rater was evaluated. The histopathology report was used as the reference standard.

Inter-rater agreement

There was agreement among all three raters in 43 of the cases (65%) for ADNEX, in 40 of the cases (61%) for ADNEX 125, and in 50 cases (76%) including inconclusive cases for SR. According to κ statistics, for ADNEX, raters 1 and 2 had almost perfect agreement, raters 2 and 3 had fair agreement, and raters 1 and 3 had moderate agreement, with an overall κ (Fleiss) coefficient of 54.2% (95%CI: 39.0–69.3%), indicating an overall moderate agreement among the three raters. For ADNEX 125, rater 3 had fair and mod-

▶ **Table 1** Histopathology of the tumors.

Histopathology	Frequency	Frequency
Benign	Serous cystadenoma	12
	Mucinous cystadenoma	6
	Mature cystic teratoma	6
	Simple serous cyst	5
	Endometrioma	5
	Hydrosalpinx	3
	Brenner tumor	1
	Luteal cyst	1
BOT	Mucinous BOT	4
	Serous BOT	2
Malignant	Serous carcinoma	11
	Endometrioid carcinoma	2
	Mucinous carcinoma	2
	Immature cystic teratoma	2
	Yolk sac tumor	1
	Dysgerminoma	1
	Metastatic	2
Total		66

BOT: Borderline ovarian tumor

▶ **Table 2** Cumulative results for the three raters and for all prediction models.

		TP	TN	FP	FN	Total
Rater 1	ADNEX	27	28	11	0	66
	ADNEX 125	27	31	8	0	66
	SR	20	31	1	6	58
Rater 2	ADNEX	25	27	12	2	66
	ADNEX 125	26	27	12	1	66
	SR	21	28	3	3	55
Rater 3	ADNEX	26	14	25	1	66
	ADNEX 125	26	14	25	1	66
	SR	23	34	3	2	62

ADNEX: Assessment of Different NEoplasias in the AdneXa model, SR: Simple Rules, TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative. For SR, inconclusive results were excluded.

▶ **Table 3** Performance metrics for the three raters and for all prediction models.

	Performance metric	Rater 1	Rater 2	Rater 3
ADNEX	Sensitivity	100.00%	92.59%	96.30%
	Specificity	71.79%	69.23%	35.90%
	PPV	71.05%	67.57%	50.98%
	NPV	100.00%	93.10%	93.33%
	FPR	28.21%	30.77%	64.10%
	FNR	0.00%	7.41%	3.70%
	OA	83.33%	78.79%	60.61%
ADNEX 125	Sensitivity	100.00%	96.30%	96.30%
	Specificity	79.49%	69.23%	35.90%
	PPV	77.14%	68.42%	50.98%
	NPV	100.00%	96.43%	93.33%
	FPR	20.51%	30.77%	64.10%
	FNR	0.00%	3.70%	3.70%
	OA	87.88%	80.30%	60.61%
SR	Sensitivity	76.92%	87.50%	92.00%
	Specificity	96.88%	90.32%	91.89%
	PPV	95.24%	87.50%	88.46%
	NPV	83.78%	90.32%	94.44%
	FPR	3.13%	9.68%	8.11%
	FNR	23.08%	12.50%	8.00%
	OA	87.93%	89.09%	91.94%

ADNEX: Assessment of Different NEoplasias in the AdneXa model, SR: Simple Rules, PPV: Positive Predictive Value, NPV: Negative Predictive Value, FPR: False Positive Rate, FNR: False Negative Rate, OA: Overall Accuracy. For SR, inconclusive results were excluded.

raters 1 and 2 had substantial agreement, raters 2 and 3 had almost perfect, and raters 1 and 3 moderate. The overall level of agreement was κ = 69.5 % (95 % CI: 56.9–82.0 %) and was characterized as substantial. When inconclusive cases were excluded, the best level of agreement was achieved, which was almost perfect for all pairs of comparisons (κ > 81 %) and the overall agreement (for the 50 conclusive cases for all raters) was also almost perfect [κ = 94.5 % (95 % CI: 77.3 %–100 %)] (▶ **Fig. 2**). In conclusion, the model with which most raters had a higher level of agreement was SR, followed by ADNEX and ADNEX 125 (▶ **Table 4**). Notably, SR was the system with the highest overall accuracy (see ▶ **Table 2**) but not the highest sensitivity.

Intra-rater agreement

For rater 1, there was total agreement for all three scoring systems for 43 cases (65 %), for rater 2 for 46 cases (70 %), and for rater 3 for 40 cases (61 %). Inconclusive cases in the SR system were considered disagreements. According to κ statistics, for rater 1 there was substantial agreement between ANDEX and ADNEX 125 (κ = 78.6 %), as well as ADNEX 125 and SR (κ = 66.2 %) but moderate

erate agreement with raters 1 and 2, respectively, while raters 1 and 2 had substantial agreement. For ADNEX 125, the overall κ coefficient was 50.4 % (95 %CI: 34.9–65.8 %), indicating an overall moderate agreement. For SR, the agreement was evaluated by a) including and b) excluding the inconclusive outcomes. For SR,
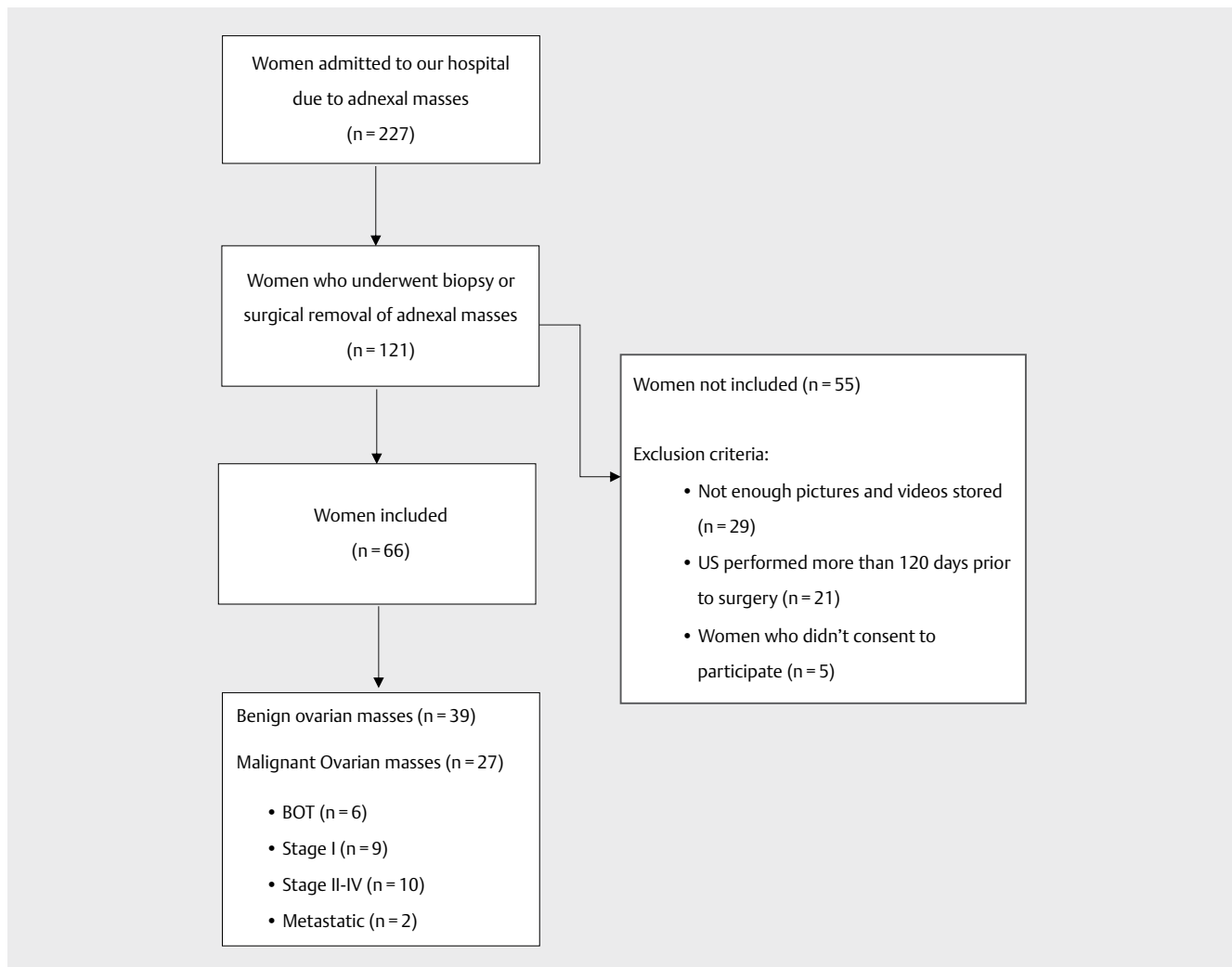
▶ **Fig. 1** Flowchart summarizing the inclusion of patients with adnexal masses in the study. US: Ultrasound; BOT: Borderline Ovarian Tumor.

agreement for ADNEX and SR (κ = 57.2%). After excluding 8 inconclusive cases, the overall agreement was κ = 67.8% (95% CI: 53.3%–82.3%), indicating substantial agreement. For rater 2, there was almost perfect agreement between ADNEX and ADNEX 125 and substantial agreement for the two models compared to SR. When considering all three systems together (after excluding 11 inconclusive cases), the overall agreement was κ = 76.2% (95% CI: 59.7%–92.7%), indicating substantial agreement. Finally, for rater 3, despite the fact that there was almost perfect agreement between ADNEX and ADNEX 125 (κ = 91.4%), the agreement between the two models with the SR approach was fair (κ = 37.5%). The overall agreement for the three systems (after excluding 4 inconclusive cases) was κ = 48.3% (95% CI = 34.0%–62.7%), indicating moderate agreement. In conclusion, for all raters, ADNEX and ADNEX 125 had the highest degree of agreement, while a great variation was observed among the three raters for the agreement of ADNEX and ANDEX 125 with SR. Notably, rater 3 who had the smallest percentage of inconclusive cases, had the smallest agreement between SR and ADNEX (and ADNEX 125) (p < 0.001) and the highest degree

of agreement between ADNEX and ADNEX 125. The comparisons can be seen in ▶ **Table 5**.

Analysis of the level of agreement for staging (ADNEX and ADNEX 125)

Subsequent analysis involves only ADNEX and ADNEX 125. The evaluation was done in two levels a) for each individual classification system among the three raters (i.e., inter-rater agreement) and b) for each individual rater between the two classification systems (intra-rater agreement). Benign cases were not excluded since a benign case for one system may be malignant for the other.

Inter-rater agreement for staging

For ADNEX all three raters had exact agreement for 36 cases (55%). The comparison of raters per pairs is presented in ▶ **Table 6**. Raters 1 and 2 had substantial agreement between each other, while rater 3 had moderate agreement with raters 1 and 2. The Kendall's W was 72.4%, indicating substantial agreement. The individual κ agreement values for the subcategories were: 54.2% ± 7.1%,
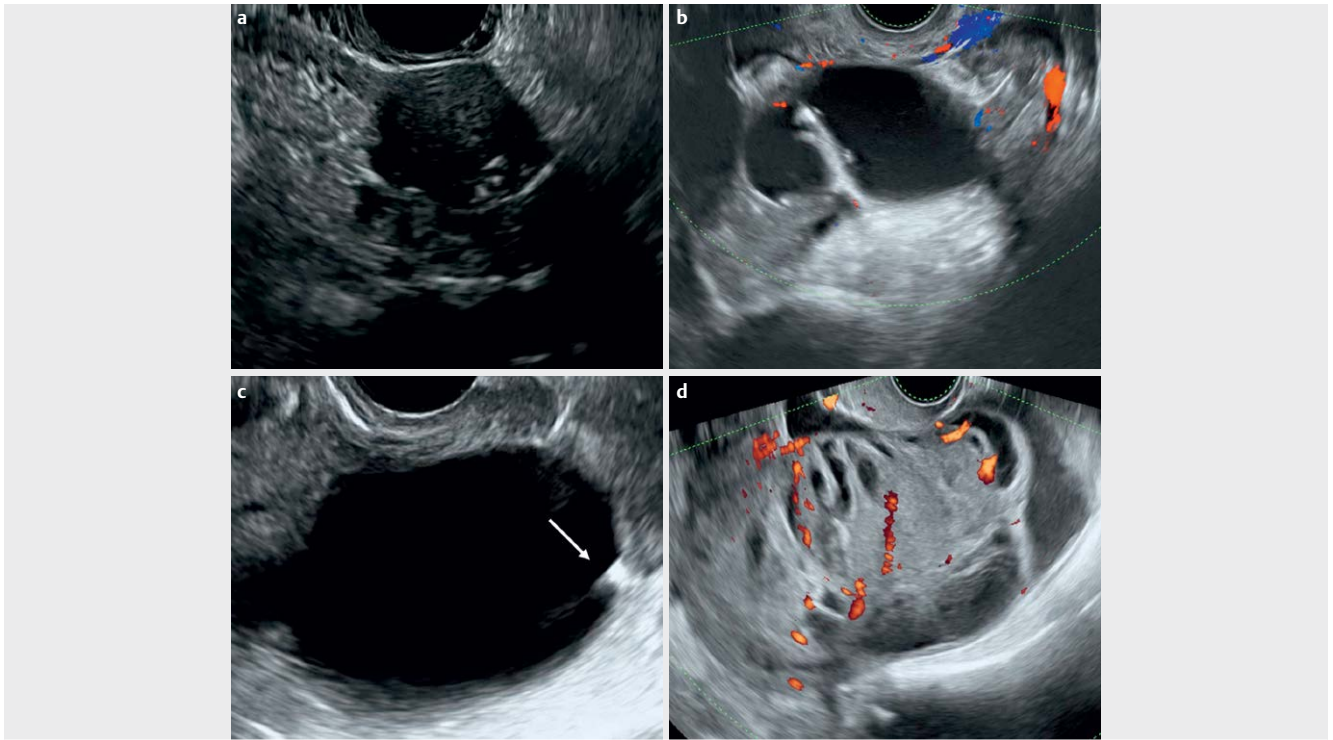
▶ **Fig. 2** Ultrasound examples of disagreement between raters. **a** Grayscale image of a solid ovarian lesion (benign Brenner tumor) with almost perfect inter-rater agreement between the experienced gynecologist and the 4th year resident and no agreement between the previous raters and the 1st year resident. It was classified as a benign solid lesion with acoustic shadows by the first two raters and as a solid malignant lesion without acoustic shadows by rater 3, using the three scoring systems. **b** Color Doppler image of an ovarian multilocular mucinous cystadenoma which had almost perfect inter-rater agreement between the experienced gynecologist and the 4th year resident and was classified as benign using all scoring systems. The 1st year resident classified the mass as a malignant multilocular solid cyst with 2 papillary projections, thus misinterpreting parts of the inner cystic walls as papillary projections. **c** Grayscale image of an ovarian unilocular serous cystadenoma. The expert gynecologist classified the mass as benign using the three scoring systems, while the 4th year resident and the 1st year resident were not able to classify the mass using Simple Rules. When ADNEX and ADNEX 125 was applied by the residents, the mass was classified as malignant. The two raters interpreted the part of the inner cystic wall marked by the white arrow, as a papillary projection a finding that couldn't be confirmed by the histopathology report. **d** Power Doppler image of a high-grade serous carcinoma with moderate inter-rater agreement for staging using ADNEX and ADNEX 125 and almost perfect inter-rater agreement for Simple Rules. All raters classified the mass as malignant using the three scoring systems.

52.2 % ± 7.1 %, 17.9 % ± 7.1, 57.4 % ± 7.1 % for the benign, borderline, stage I and stages II–IV categories, respectively, indicating that ADNEX had very poor agreement for stage I. For ADNEX 125 all three raters had exact agreement for 40 cases (61 %). The paired comparisons are presented in ▶ **Table 6**. For this system there was improved agreement (compared to ADNEX). Specifically, rater 3 had substantial agreement with raters 1 and 2 and raters 1 and 3 had almost perfect agreement with each other. The Kendall's W was 83.3 %, indicating perfect agreement. The individual κ agreement values for the subcategories were: 50.4 %, 48.8 %, 26.0 %, 92.4 %, and 74.5 % (SE: 7.1 % for all cases), for the benign, borderline, stage I, stages II–IV, and metastatic categories, respectively, indicating that ADNEX 125 had better agreement for stage I and great improvement for stages II–IV (▶ **Table 6**). In conclusion, for all pairs of raters, the level of agreement was higher for ADNEX 125, and the overall agreement was also better.

Intra-rater agreement for staging

When considering a single rater, the level of agreement between ADNEX and ADNEX 125 for staging was 90.2 % (95 % CI: 59.9 % -

86.4 %) for rater 1, 92.3 % for rater 2, and 89.4 % for rater 3, indicating nearly perfect agreement for all raters. Moreover, when considering all three raters together and estimating the level of agreement, the overall Kendall's W was 90.9 % and the individual agreement values were 84.5 % (SE: 3.9 %), 71.0 % (SE: 5.5 %), 31.3 % (SE:17.9 %), 61.0 % (SE: 6.2 %) and 0 % (SE: 0.5 %) for the benign, borderline, stage I, stages II–IV, and metastatic stages, respectively, showing that overall, the same person (when using the two systems) has almost perfect agreement for the benign cases, substantial agreement for borderline cases and stages II–IV, fair agreement for stage I, and no agreement for the metastatic cases. Note that our study cannot provide reliable results for the latter category, since only two metastatic cases were present.

Discussion

Ovarian masses are frequent in both premenopausal and postmenopausal women [12–14]. They can be difficult to classify, particularly by physicians with limited training, while gynecologists with a high level of expertise in ultrasonography can preoperatively dis-

▶ **Table 4** Pairwise comparisons for the level of agreement between the three raters for each individual scoring system.

		Rater 1	Rater 3
ADNEX	Rater 2	84.6 % (71.6–97.6)	35.1 % (14.2–56)
	Rater 3	43.8 % (23.3–64.4)	
ADNEX 125	Rater 2	72.5 % (55.8–89.1)	37.2 % (16.1–58.4)
	Rater 3	43.6 % (24.4–62.8)	
SR (with inconclusive)	Rater 2	74.3 % (60.7–87.8)	84.2 % (75–93.4)
	Rater 3	77.8 % (64.2–91.4)	
SR (without inconclusive)	Rater 2	88 % (74.9–100)	100 % (100–100) (N = 53)
	Rater 3	85.5 % (71.9–99.1)	

ADNEX: Assessment of Different NEoplasias in the AdneXa model, SR: Simple Rules.

▶ **Table 5** Pairwise comparisons for the level of agreement between the three systems for each individual rater.

		ADNEX 125	SR
Rater 1	ADNEX	78.6 % (63.7–93.5)	57.2 % (38.7–75.7)
	ADNEX 125		66.2 % (48.2–84.1)
Rater 2	ADNEX	84.6 % (71.6–97.6)	78.4 % (62.5–94.3)
	ADNEX125		71.5 % (54–89)
Rater 3	ADNEX	91.4 % (79.6–100)	37.5 % (20.5–54.4)
	ADNEX 125		37.5 % (20.5–54.4)

ADNEX: Assessment of Different NEoplasias in the AdneXa model, SR: Simple Rules.

▶ **Table 6** Pairwise comparisons for the level of agreement between the three raters for the stage of ADNEX and ADNEX 125.

		Rater 1	Rater 3
ADNEX	Rater 2	74.4 % (60.5–88.4)	41.5 % (23.2–59.9)
	Rater 3	56.8 % (41–72.7)	
ADNEX 125	Rater 2	83.4 % (72.5–94.3)	65.6 % (51.4–79.9)
	Rater 3	69.9 % (57.1–82.8)	

ADNEX: Assessment of Different NEoplasias in the AdneXa model, SR: Simple Rules.

sults show that when the models are utilized by an inexperienced 1st year resident, their performance decreases significantly. The 1st year resident's specificity was 36 % (with or without the "CA-125"), which is even lower than in the existing literature [19]. This could be explained by the fact that ultrasonography was not conducted by the raters, resulting in a higher level of difficulty, especially for the rater with limited experience. For the intermediate and experienced raters, the specificity was 69 % and 72 %, respectively (without the "CA-125" levels), which is comparable to the findings of Van Calster et al. [5]. Additionally, our data indicated that the addition of "CA-125" to the model led to an even better sensitivity for the rater with intermediate experience and a rise in specificity for the most experienced examiner who had already achieved perfect (100 %) sensitivity. The overall accuracy increased as well in parallel with sensitivity and specificity of the model. Lastly, our results showed that the use of "CA-125" improved the staging of malignant masses, especially for stages I and II–IV, but didn't affect the differentiation between benign and malignant tumors. These results are in accordance with the existing literature [5, 7].

Simple Rules

SR appears to be the best model to assess adnexal masses, especially for the less experienced user. Studies have shown that it is conclusive for up to 80 % of women, even when used by inexperienced examiners, while inconclusive results need to be assessed by an experienced gynecologist [20, 21]. In our study, the experienced gynecologist, the 6th year resident, and the 1st year resident were able to categorize 87.8 %, 83.3 %, and 93.9 % of the masses, respectively. The overall accuracy of SR, when applied by the 1st year resident, was significantly higher compared to ADNEX (with or without "CA-125"). Our results showed that the 1st year resident, using the SR, had the highest sensitivity (92 %), compared to the resident (87.5 %), and the most experienced rater (77 %). On the other hand, the most experienced rater had the highest specificity (97 %), using SR, compared to the 6th year resident (90 %) and the 1st year resident (92 %). The unexpected high sensitivity in less experienced raters could be explained by the tendency of inexperienced physicians to overdiagnose especially when using tools with subjective criteria [22]. The specificity and sensitivity we found for the most experienced rater are similar to the ones mentioned in the literature, thus affirming our results and confirming the dependence of

tinguish benign from malignant lesions with high sensitivity [2]. IOTA has developed validated logistical models and rules for describing and characterizing ovarian masses [7]. The goal of our study was to compare the performance of ADNEX, ADNEX 125, and SR models, when applied by physicians with limited, intermediate, and significant experience. Our goal was to evaluate the performance of each model, primarily based on its sensitivity and secondarily on the overall accuracy, as our aim was to identify tools fit to be used by inexperienced examiners with minimum false-negative results.

ADNEX model

Our results indicate that the ADNEX model with the incorporation of "CA-125" has a sensitivity of 96–100 %, irrespective of the rater's experience, which is higher compared to previous studies, which have demonstrated that the ADNEX model can achieve sensitivity of 89–97 % and specificity of 54–94 % [5, 15–18]. While experienced sonographers attain high rates of specificity and sensitivity, our re-

SR on the rater's experience [23, 24]. In the hands of inexperienced examiners, SR appears to be more efficient as a screening tool compared to other models [19].

Limitations

Our study admittedly has certain limitations. It was a single-center study including a relatively small number of patients. Nevertheless, patient histopathology results were proportionally distributed, between benign and malignant cases, allowing comparability concerning diagnostic accuracy. For the same reason, we were unable to reliably evaluate the sub-classification of the malignant masses. Lastly, ultrasound examinations were not performed by the raters themselves, but they were conducted by a highly experienced medical professional and the use of stored anonymized recourses, allowed us to avoid detection and reporting bias.

Conclusion

The aim of our study was to demonstrate the performance of the ADNEX model (with and without "CA-125") and Simple Rules in diagnosing ovarian cancer, when applied by sonographers with different levels of experience. Both methods offer high sensitivity when used by inexperienced examiners although Simple Rules is easier to apply. The ADNEX model has good to excellent performance in categorizing adnexal masses only when applied by raters with a moderate to high level of experience, while SR cannot predict the stage in malignant cases. Both models can be used in non-oncology centers for screening, but patients with suspicious findings or inconclusive results must be evaluated in specialized facilities.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Forstner R. *Early detection of ovarian cancer*. Eur Radiol 2020; 30: 5370–5373

[2] Meys EM et al. *Subjective assessment versus ultrasound models to diagnose ovarian cancer: A systematic review and meta-analysis*. Eur J Cancer 2016; 58: 17–29

[3] Van Holsbeke C et al. *Prospective internal validation of mathematical models to predict malignancy in adnexal masses: results from the international ovarian tumor analysis study*. Clin Cancer Res 2009; 15: 684–691

[4] Timmerman D et al. *Simple ultrasound-based rules for the diagnosis of ovarian cancer*. Ultrasound Obstet Gynecol 2008; 31: 681–690

[5] Van Calster B 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

[6] Bossuyt PM et al. *STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies*. Radiology 2015; 277: 826–32

[7] Van Calster B 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

[8] DiMaggio C. SAS for Epidemiologists: Applications and Methods. New York: Springer; 2013: xvii, 258

[9] SAS Institute. *SAS Home Page*. 2014; Available from: http://www.sas.com.

[10] Landis JR, Koch GG. *An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers*. Biometrics 1977; 33: 363–374

[11] Raghavachari M. Measures of Concordance for Assessing Agreement in Ratings and Rank Order Data, in Advances in Ranking and Selection, Multiple Comparisons, and Reliability, Balakrishnan N, Nagaraja HN, and Kannan N, Editors. Birkhäuser; Boston: 2005; 245–263

[12] Christensen JT, Boldsen JL, Westergaard JG. *Functional ovarian cysts in premenopausal and gynecologically healthy women*. Contraception 2002; 66: 153–157

[13] Dørum A et al. *Prevalence and histologic diagnosis of adnexal cysts in postmenopausal women: an autopsy study*. Am J Obstet Gynecol 2005; 192: 48–54

[14] Pérez-López FR, Chedraui P, Troyano-Luque JM. *Peri- and post-menopausal incidental adnexal masses and the risk of sporadic ovarian malignancy: new insights and clinical management*. Gynecol Endocrinol 2010; 26: 631–643

[15] Sayasneh A et al. *Evaluating the risk of ovarian cancer before surgery using the ADNEX model: a multicentre external validation study*. Br J Cancer 2016; 115: 542–548

[16] Araujo KG et al. *Performance of the IOTA ADNEX model in preoperative discrimination of adnexal masses in a gynecological oncology center*. Ultrasound Obstet Gynecol 2017; 49: 778–783

[17] Hiett AK et al. *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: 668–676

[18] Jeong SY et al. *Validation of IOTA-ADNEX model in discriminating characteristics of adnexal masses: A comparison with subjective assessment*. J Clin Med 2020; 9: 6

[19] Tavoraitė I, Kronlachner L, Opolskienė G et al. Ultrasound Assessment of Adnexal Pathology: Standardized Methods and Different Levels of Experience. Medicina (Kaunas) 2021; 57: 708. DOI: 10.3390/medicina 57070708

[20] Ning CP et al. *Association between the sonographer's experience and diagnostic performance of IOTA simple rules*. World J Surg Oncol 2018; 16: 179

[21] Nunes N et al. *Use of IOTA simple rules for diagnosis of ovarian cancer: meta-analysis*. Ultrasound Obstet Gynecol 2014; 44: 503–514

[22] Lam JH et al. *Why clinicians overtest: development of a thematic framework*. BMC Health Services Research 2020; 20: 1

[23] Fathallah K et al. *External validation of simple ultrasound rules of Timmerman on 122 ovarian tumors*. Gynecol Obstet Fertil 2011; 39: 477–481

[24] Timmerman D et al. *Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group*. BMJ 2010; 341: c6839