



# The learning curve and difficult points of the O-RADS ultrasound risk stratification system in 54 trainees

Shan Zhou\*, Yuyang Guo\*, Lieming Wen, Baihua Zhao, Minghui Liu

Second Xiangya Hospital, Central South University, Changsha, China

**Purpose:** This study aimed to evaluate the learning curve and explore the difficult points of the Ovarian-Adnexal Reporting and Data System (O-RADS) ultrasound risk stratification system.

**Methods:** One hundred adnexal masses (AMs) were randomly selected for five tests as training data. Two experienced trainers had an inter-rater agreement of 0.95 for the O-RADS scores. Fifty-four trainees (26 level I practitioners [group 1], 17 level II practitioners [group 2], and 11 experienced level II practitioners [group 3]) attended the training. Every trainee received assessment and feedback after 20 scored cases. The outcomes of the five tests were compared among the three groups using repeated-measurements analysis of variance.

**Results:** Of the 100 AMs, 52 were pathologically benign and 48 were malignant; the O-RADS scores were 2, 3, 4, and 5 in 22, 11, 48, and 19 AMs, respectively. The between-subjects effects test showed no significant differences between groups 1, 2, and 3 for the five tests ( $P=0.501$ ). For each group, the differences among the five tests were significant ( $P<0.001$ ,  $P=0.006$ , and  $P=0.044$  for groups 1, 2, and 3, respectively). Test 2 was the worst. In 23 cases, more than 40% of trainees gave incorrect answers, which mainly related to classic benign lesions, the color flow score, and solid-appearing masses.

**Conclusion:** After training, junior doctors at different levels can reach a coincident O-RADS ultrasound risk stratification. The difficulties primarily related to subjective judgments of classic benign lesions, the color flow score, and solid-appearing masses. More experience is needed to improve the applicability of the system.

**Keywords:** O-RADS US system; International Ovarian Tumour Analysis; Learning curve; Repeated measurement analysis

**Key points:** After training, junior doctors at different levels can reach a coincident O-RADS ultrasound risk stratification. The difficult points focused on the subjective judgment of classic benign lesions, color flow score, and solid-appearing mass. More experiences are needed to reach a better application of the system.

# ULTRA SONO GRAPHY

## ORIGINAL ARTICLE

<https://doi.org/10.14366/usg.21158>  
pISSN: 2288-5919 • eISSN: 2288-5943  
Ultrasonography 2022;41:365-372

Received: July 31, 2021  
Revised: October 26, 2021  
Accepted: November 1, 2021

**Correspondence to:**  
Lieming Wen, MD, PhD, Department of  
Ultrasound Diagnosis, Second Xiangya  
Hospital of Central South University,  
139 Renmin Road (M), Changsha,  
Hunan, 410011, China  
Tel. +086-731-8529-2140  
E-mail: wen\_lieming@csu.edu.cn,  
wen\_lieming@163.com

\*These authors contributed equally to  
this work.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2022 Korean Society of  
Ultrasound in Medicine (KSUM)



**How to cite this article:**  
Zhou S, Guo Y, Wen L, Zhao B, Liu M. The learning curve and difficult points of the O-RADS ultrasound risk stratification system in 54 trainees. Ultrasonography. 2022 Apr;41(2):365-372.

## Introduction

Ultrasound (US) imaging is the first choice to describe ovarian adnexal masses (AMs) and estimate their malignancy risk [1]. US is low-cost and easily accessible, but highly operator-dependent. To

improve the malignancy risk estimate and the management of AMs, many guidelines and structured reporting systems have been established, using subjective assessments, simple scoring, or statistically derived scoring [2–10].

The Ovarian-Adnexal Reporting and Data System (O-RADS) ultrasound risk stratification and management system is the only lexicon and classification system encompassing all risk categories of AMs, with a management recommendation for each risk category [10]. It may be the most complex AM diagnosis system, including six categories (O-RADS 0–5) and at least 21 detailed combined lexicon descriptors for scoring. Meanwhile, it is the most effective US system, as it improved the accuracy of assessments of the malignancy risk of AMs by providing a standardized reporting tool describing masses in terms of echogenicity, size, cystic wall, internal septum, boundary, shape, and blood flow [1,9,11].

The learning curve can reflect the difficulties and important steps of clinical diagnosis and treatment methods, and then strengthen clinicians' cumulative experience. Analyses of the learning curve are widely used in various fields of medical imaging diagnosis [12–15].

The aim of the present study was to conduct an O-RADS US system training for junior doctors, draw the corresponding learning curve, and explore the difficult points of this system to provide a reference for clinical training.

## Materials and Methods

### Compliance with Ethical Standards

This study was approved by the Human Research Ethics Committee of Second Xiangya Hospital with a waiver of informed consent (No. 2021-038).

### Patients

The diagnostic US images, clinical records, and pathological information of 642 women who underwent adnexal tumor resection at the Second Xiangya Hospital between June 2018 and June 2020 were collected.

The inclusion criteria were (1) a clear pathological diagnosis; (2) an intact clinical, ultrasonographic, and surgical record; (3) US images showing enough diagnostic signs without artifacts; and (4) an interval of less than 1 month between ultrasonography and surgery.

In total, 100 AMs were randomly selected for five average groups as the training data. The authors Wen and Zhao, two senior doctors with more than 10 years of gynecological US experience, read all the images blinded to pathological information. The intraclass correlation coefficients for inter-rater agreement were 0.95 (95% confidence interval, 0.93 to 0.96) for the O-RADS US score. The two

authors determined all the O-RADS US scores together with lexicon descriptors.

Fifty-four doctors from 18 hospitals participated in the training in May 2021. Of the 54 doctors, 26 who had finished their second-year training for residents were included in group 1; 17 who had completed their 1-year attending doctor training in gynecological US were group 2, and 11 experienced attending doctors comprised group 3. The doctors in group 1 can be seen as level I practitioners, those in group 2 as level II practitioners, and those in group 3 as experienced level II practitioners, according to the European standard training requirements for gynecological US practice published by the European Federation of Societies for Ultrasound in Medicine and Biology, including standards for theoretical knowledge and practical skills [16,17]. Two trainers (the authors Wen and Zhao) were equivalent to level III practitioners (experts). All trainees consented to the use of their data for this research.

The author Wen conducted the training. First, the definition of all terms was explained in detail, including (1) normal ovaries; (2) simple cysts, unilocular cysts, and multilocular cysts; (3) typical benign lesions; (4) smooth or irregular inner margins or walls; (5) papillary projections, solid components, and solid-appearing masses; (6) ascites and peritoneal nodules; and (7) color scores of 1–4 [9,11]. After receiving the feedback that all doctors understood the above terms, the specific rules of O-RADS US scoring and classification were further explained with corresponding legends. All the legends used in the explanation of theoretical knowledge did not appear in the subsequent assessment.

In the subsequent image reading test and training, every trainee read the diagnostic images of each 20 cases independently. All O-RADS US scores with lexicon descriptors were listed on the answer sheet. The trainee only needed to tick the correct answer for each case. After the test, the trainee received feedback from the trainer, had sufficient communication with the trainer, and continued to the next 20 cases. All tests and training were finished within 1 week. All the answers were reviewed. One point was assigned for a correct answer and 0 for a wrong answer. The maximum possible score for each test was 20.

### Statistical Analysis

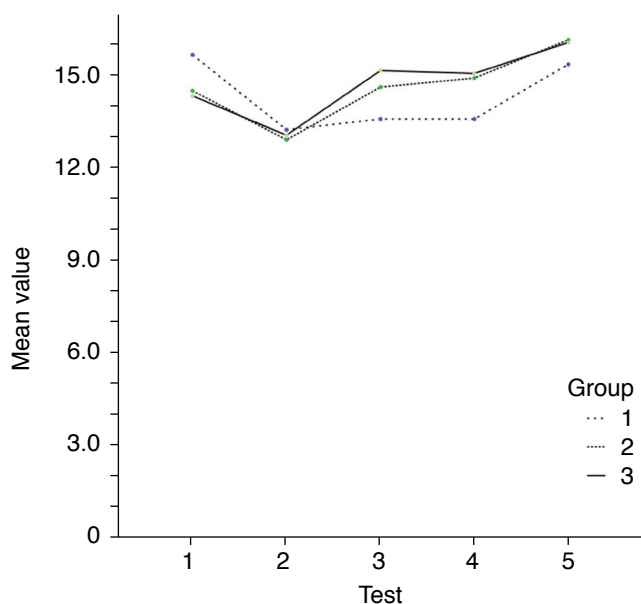
Statistical analysis was performed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). Repeated-measurements analysis of variance was used to test the differences among the five tests and three comparison groups and to produce the learning curve. The Mauchly test was used to evaluate the distribution of variation in terms of sphericity, followed by the between-subjects effects test. The within-subjects effects test was used to analyze the interaction between the two factors, followed by the simple-effect test if the

interaction was significant. A P-value <0.05 was considered to indicate statistical significance.

**Table 1.** Differences among the three groups for the five tests (n=54)

Test	Group	Mean	SD	SE	95% CI	Simple-effect test	
						F-value	P-value
1	1	14.7	1.8	0.4	14.0–15.4	2.967	0.060
	2	13.5	1.8	0.5	12.6–14.4		
	3	13.4	2.1	0.6	12.2–14.5		
2	1	12.3	2.6	0.5	11.3–13.2	0.094	0.910
	2	11.9	2.4	0.6	10.8–13.1		
	3	12.1	2.1	0.7	10.6–13.6		
3	1	12.6	2.2	0.4	11.7–13.5	2.246	0.116
	2	13.7	2.7	0.6	12.6–14.7		
	3	14.2	1.5	0.7	12.8–15.5		
4	1	12.6	2.3	0.4	11.8–13.4	2.933	0.062
	2	13.9	2.0	0.5	12.9–15.0		
	3	14.1	1.8	0.6	12.8–15.4		
5	1	14.4	2.7	0.5	13.4–15.3	0.682	0.510
	2	15.2	2.2	0.6	14.0–16.3		
	3	15.1	1.9	0.7	13.6–16.5		

SD, standard deviation; SE, standard error; CI, confidence interval.



**Fig. 1.** The learning curve for the three groups showing no significant difference among the groups for the five tests. Group 1, 26 who had finished their second-year training for residents; group 2, 17 who had completed their 1-year attending doctor training in gynecological US; group 3, 11 experienced attending doctors.

## Results

Of the 100 AMs, the pathological findings showed that 52 were benign and 48 were malignant. Twenty-two had an O-RADS score of 2, 11 had an O-RADS score of 3, 48 had an O-RADS score of 4, and 19 had an O-RADS score of 5.

The outcomes of the five tests for groups 1 to 3 are shown in Table 1. The Mauchly test for sphericity yielded a P-value of 0.219 (F=11.921). The between-subjects effects test showed there was

**Table 2.** Differences among the five tests in the three groups (n=54)

Group	Pairwise comparison				Simple-effect test					
	Test (I)	Test (J)	Difference of mean (I-J)	Standard error	P-value	F-value	P-value			
Group 1	1	2	2.4	0.5	<0.001	8.853	<0.001			
		3	2.1	0.5	<0.001					
		4	2.1	0.5	0.004					
	2	3	-0.3	0.6	>0.99					
		4	-0.3	0.7	>0.99					
		5	-2.1	0.6	0.015					
	3	4	-0.1	0.6	>0.99					
		5	-1.8	0.6	0.038					
		4	5	-1.8	0.6			0.041		
		5	5	-1.8	0.6			0.041		
Group 2	1	2	-1.6	0.7	0.175	4.172	0.006			
		3	-0.1	0.6	>0.99					
		4	-0.4	0.7	>0.99					
	2	3	-1.6	0.7	0.175					
		4	-2.0	0.9	0.252					
		5	-3.2	0.8	<0.001					
	3	4	-0.3	0.7	>0.99					
		5	-1.5	0.7	0.331					
		4	5	-1.2	0.7			0.638		
	Group 3	1	2	1.2	0.8			0.735	2.663	0.044
			3	-0.8	0.8			0.973		
			4	-0.7	0.8			0.993		
		2	3	-2.1	0.9			0.189		
			4	-2.0	1.1			0.546		
			5	-3.0	1.0			0.031		
3		4	0.1	0.9	>0.99					
		5	-0.9	0.9	0.978					
		4	5	-1.0	0.9	0.960				

NS, a P-value of ≥0.05, no significant difference.

no significant difference between groups 1, 2, and 3 at the five test times ( $F=0.708$ ,  $P=0.501$ ), as shown on the learning curve (Fig. 1). The subsequent within-subjects effects test demonstrated significant differences among the five tests ( $F=10.849$ ,  $P<0.001$ ). No significant interaction was found between the test times and the comparison groups ( $F=1.944$ ,  $P=0.059$ ).

The simple-effect test showed no significant differences among the three groups for all five tests ( $P=0.060-0.910$ ) (Table 1). For each test, no significant difference was found in the pairwise comparison between groups. For each group, the differences among the five tests were significant ( $P=0.001$ ,  $P=0.006$ , and  $P=0.044$  for groups 1, 2, and 3, respectively) (Table 2). The outcome of test 2 was the worst and was significantly poorer than test 5 for all groups.

More than 40% of the trainees failed to give a correct answer in 23 cases, which were reviewed to find the difficulty of the O-RADS US system (Table 3, Fig. 2). The main difficult points were (1) eight cases of classic benign lesions were wrongly read as unilocular or multilocular cystic masses; (2) there was a failure to differentiate unilocular and multilocular cysts in five cases; (3) solid-appearing masses, which had solid components of more than 80%, were confused with multilocular/unilocular cysts with a solid component in four cases; (4) the color scores of five multilocular cysts with solid

components and one solid mass were incorrect.

An incorrect color score was a common mistake in tests 1 and 2. The cases in test 2 had the most difficult points. In test 5, the main points of difficulty were distinguishing classic benign lesions from unilocular/multilocular cysts.

## Discussion

The O-RADS US system is the only lexicon and classification system that encompasses six risk categories (O-RADS 0–5), incorporating the range of normal to high risk of malignancy [1,10]. The system provides the necessary lexicon descriptors for AM malignancy risk stratification. Understanding the lexicon descriptors is the key to reaching an accurate and consistent interpretation for doctors at different levels [9,10]. The application needs to be tested through extensive clinical practice [11]. In this study, for the first time, the authors explored the difficult points for junior doctors in practice by drawing a learning curve.

The most common difficulty point in the training was the subjective color flow grading in the system. The color flow score seems to provide a quantitative assessment of the blood flow of AMs, but it is a subjective parameter, with grades of minimal,

**Table 3.** Difficult points of 23 cases with many incorrect judgments

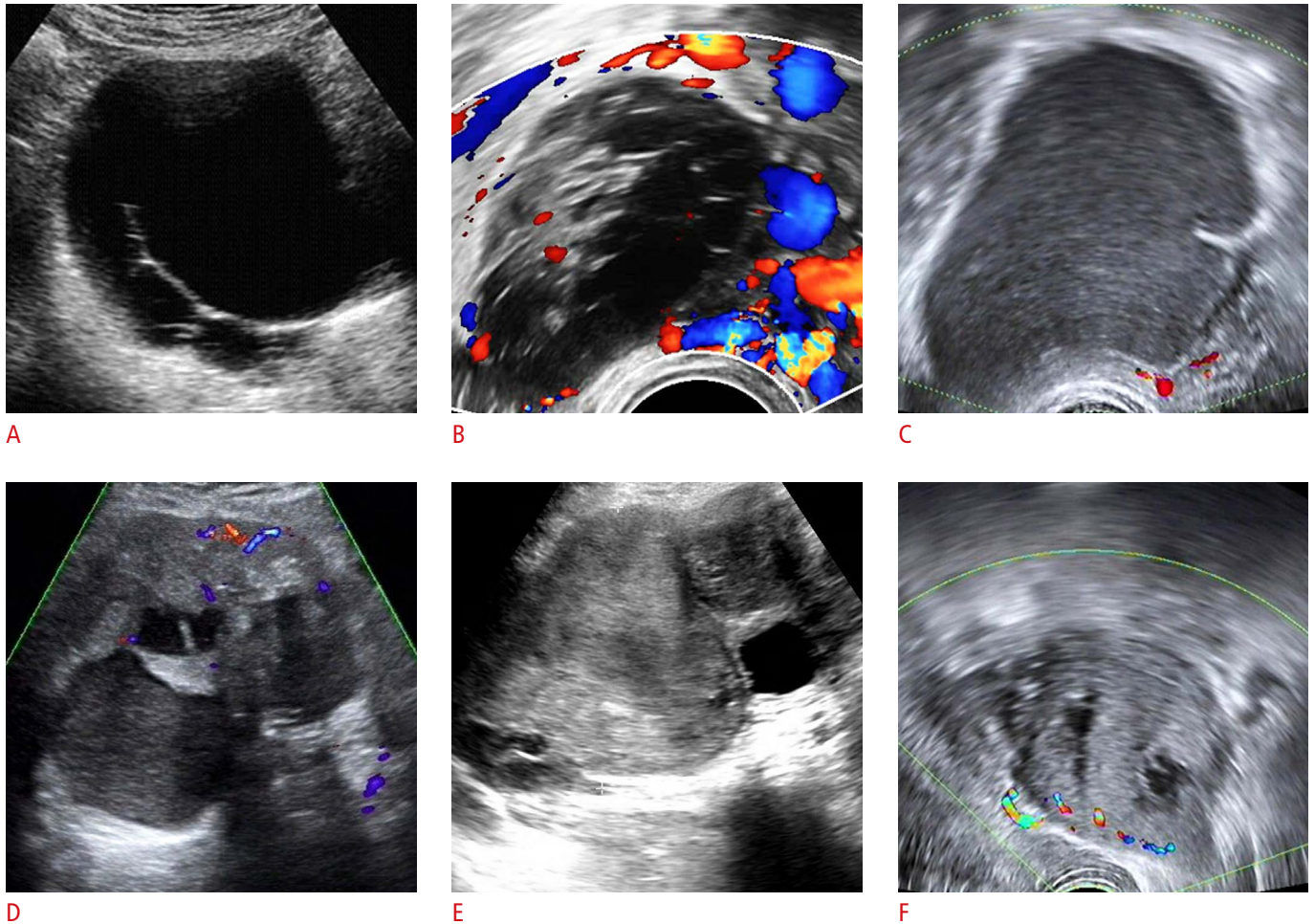
Test No.	O-RADS US scoring				Difficult point
	Correct answer and PD	No. (%)	Incorrect answers	No. (%)	
1 51	S2: Classic benign lesion <10 cm	18 (33.3)	S2: Simple cyst <10 cm	16 (29.6)	Classic benign lesions
	PD: Endometriotic cyst		S2: Non-simple unilocular cyst, smooth inner margin, <10 cm	14 (25.9)	
	194	S5: Multilocular cyst with solid component, CS=3–4	14 (25.9)	S4: Multilocular cyst with solid component, CS=1–2	
	PD: Serous cyst adenocarcinoma				
395	S3: Solid smooth, CS=1	28 (51.8)	S4: Solid smooth, CS=2–3	23 (53.5)	CS
	PD: Ovarian thecoma				
2 110	S5: Multilocular cyst with solid component, CS=3–4	25 (85.2)	S5: Solid irregular, CS=any	10 (18.5)	Solid-appearing mass, multilocular cyst, CS
	PD: Yolk-sac tumor		C4: Unilocular cyst with solid component, CS=any	8 (14.8)	
			S4: Multilocular cyst with solid component, CS=1–2	5 (16.7)	
146	S2: Classic benign lesions <10 cm	11 (20.4)	S2: Simple cyst <10 cm	24 (44.4)	Classic benign lesions
	PD: Endometriotic cyst		S3: Unilocular cyst, any size with irregular inner wall <3 mm in height	8 (14.8)	
			S2: Non-simple unilocular cyst, smooth inner margin, <10 cm	5 (9.3)	
191	S2: Classic benign lesions <10 cm	7 (13.0)	S5: Solid irregular, CS=any	14 (25.9)	Classic benign lesions
	PD: Endometriotic cyst		S4: Solid smooth, CS=2–3	9 (16.6)	
			S4: Multilocular cyst with solid component, CS=1–2	18 (33.3)	

Continued

Table 3. Continued

Test No.	O-RADS US scoring				Difficult point
	Correct answer and PD	No. (%)	Incorrect answers	No. (%)	
514	S4: Multilocular cyst with solid component, CS=1–2 PD: Yolk-sac tumor	32 (59.3)	S5: Multilocular cyst with solid component, CS=3–4	16 (29.6)	CS
557	S4: Multilocular cyst with solid component, CS=1–2 PD: Mucinous cystadenoma	26 (48.1)	S5: Multilocular cyst with solid component, CS=3–4	15 (27.7)	CS
633	S4: Multilocular cyst with solid component, CS=1–2 PD: Serous papillary cystadenocarcinoma	18 (33.3)	S4: Unilocular cyst with solid component, CS=any	32 (59.3)	Multilocular cyst
3	18 S3: Classic benign lesions >10 cm PD: Hydrosalpinx	10 (18.5)	S4: Multilocular cyst, no solid component, >10 cm S3: Multilocular cyst, smooth inner wall, CS=1–3	23 (42.6)	Classic benign lesions
49	S2: Classic benign lesions <10 cm PD: Endometriotic cyst	26 (48.1)	S4: Unilocular cyst with solid component, CS=any S3: Multilocular cyst <10 cm, smooth inner wall, CS=1–3	14 (25.9) 6 (11.1)	Classic benign lesions
125	S5: Solid irregular, CS=any PD: Serous adenocarcinoma	26 (48.1)	S4: Multilocular cyst with solid component, CS=1–2	19 (35.2)	Solid-appearing mass
139	S2: Simple cyst <10 cm PD: Ovarian cyst	26 (48.1)	S2: Classic benign lesions <10 cm S2: Non-simple unilocular cyst, smooth inner margin, <10 cm	9 (16.7) 9 (16.7)	Simple cyst
245	S3: Multilocular cyst <10 cm, smooth inner wall, CS=1–3 PD: Mucinous cystadenoma	25 (46.3)	S2: Non-simple unilocular cyst, smooth inner margin, <10 cm	23 (42.6)	Multilocular cyst
631	S5: Multilocular cyst with solid component, CS=3–4 PD: Endometrioid adenocarcinoma	30 (55.6)	S4: Multilocular cyst with solid component, CS=1–2	20 (37.0)	CS
4	179 S3: Unilocular cyst, any size with irregular inner wall <3 mm in height PD: Ovarian cystadenoma	6 (11.1)	S3: Multilocular cyst, smooth inner wall, CS=1–3 S2: Simple cyst <10 cm	15 (27.8) 12 (22.2)	Multilocular cyst
471	S4: Solid smooth, CS=2–3 PD: Serous adenocarcinoma	14 (25.9)	S5: Solid irregular, CS=any S4: Multilocular cyst with solid component, CS=1–2 S5: Multilocular cyst with solid component, CS=3–4	19 (35.2) 9 (16.7) 9 (16.7)	Solid-appearing mass
559	S5: Solid irregular, CS=any PD: Serous cystadenocarcinoma	29 (53.7)	S4: Solid smooth, CS=2–3	19 (35.2)	Smooth contour
577	S5: Solid irregular, CS=any PD: Sclerosing stromal tumor	6 (11.1)	S4: Solid smooth, CS=2–3 S4: Multilocular cyst with solid component, CS=1–2	32 (59.3) 6 (11.1)	Solid-appearing mass
5	1 S2: Classic benign lesions <10 cm PD: Hemorrhagic cyst	4 (7.4)	S4: Multilocular cyst with solid component, CS=1–2 S4: Unilocular cyst with solid component, CS=any	22 (40.7) 20 (37.0)	Classic benign lesions
175	S2: Classic benign lesions <10 cm PD: Mature cystic teratoma	17 (31.5)	S4: Unilocular cyst with solid component, CS=any	20 (37.0)	Classic benign lesions
466	S2: Classic benign lesions >10 cm PD: Mesosalpinx cyst	10 (18.5)	S4: Unilocular cyst with solid component, CS=any S4: Multilocular cyst with solid component, CS=1–2	20 (37.0) 9 (16.7)	Classic benign lesions
556	S4: Multilocular cyst with solid component, CS=1–2 PD: Serous adeno-carcinoma	32 (59.3)	S4: Unilocular cyst with solid component, CS=any	18 (33.3)	Multilocular cyst

O-RADS, Ovarian-Adnexal Reporting and Data System; US, ultrasonography; PD, pathological diagnosis; S1-5, score 1-5; CS, color score.



**Fig. 2.** Images exemplifying the main difficult points and incorrect judgments.

**A.** This is the image of a 47-year-old woman with a surgical and pathologically proven hydrosalpinx (case 18 in test 3). The lesion was categorized as O-RADS 3, a classic benign lesion with a maximum diameter  $\geq 10$  cm. Approximately 60% of the trainees read it as a multilocular cystic mass of O-RADS 3 or 4. **B.** This is the image of a 26-year-old woman with a pathologically proven hemorrhagic cyst (case 1 in test 5). The lesion was categorized as O-RADS 2, a classic benign lesion with a maximum diameter  $< 10$  cm. Approximately 76% of the trainees scored it as O-RADS 4, a multilocular/unilocular cyst with solid components. **C.** This is the image of a 35-year-old woman with a pathologically proven endometriotic cyst (case 51 in test 1). The lesion was categorized as O-RADS 2, a classic benign lesion with a maximum diameter  $< 10$  cm. Approximately 56% of the trainees answered that it was simple cyst or a non-simple unilocular cyst of O-RADS 2. **D.** This is the image of a 45-year-old woman with a pathologically proven serous cystadenocarcinoma (case 194 in test 1). The lesion was categorized as O-RADS 5, a multilocular cyst with solid components, with a color score of 3–4. Approximately 56% of the trainees decreased the score to O-RADS 4 because they assigned it a color score of 1–2. **E.** This is a solid-appearing irregular mass with a pathological diagnosis of serous adenocarcinoma in a postmenopausal woman (case 125 in test 3). The lesion was categorized as O-RADS 5. Approximately 35% of the trainees read it as O-RADS 4, a multilocular cyst with solid components. The color score was 1–2. **F.** This is a solid-appearing mass with a pathological diagnosis of sclerosing stromal tumor in a 16-year-old girl (case 577 in test 4). The lesion was categorized as a solid irregular mass of O-RADS 5. Approximately 11% of the trainees read it as O-RADS 4, a multilocular cyst with solid components. The color score was 1–2. Furthermore, 59% of the trainees categorized it as O-RADS 4 (a smooth solid mass with a color score of 2–3). O-RADS, Ovarian-Adnexal Reporting and Data System.

moderate, and strong flow [9,10]. In addition, "0" is generally used to represent "nothing." To use this color score system, more practice is needed to change the familiar idiom. In this study, this difficult point disappeared after three tests involving practice with 60 cases.

In the O-RADS US system, there are many detailed lexicon descriptors of cystic masses and their walls [10]. These terms were not easy to apply in the initial practice. After four tests (practice with 80 cases), wrong answers became rare for distinguishing between a

simple cyst and a non-simple unilocular cyst or multilocular cyst.

The definition of a solid-appearing mass is that "the lesion should be at least 80% solid when assessed subjectively in perpendicular two-dimensional planes" [9]. This lexicon descriptor was poorly understood and applied in training. Incorrect answers were present in all five tests.

In the final test, some trainees still could not correctly distinguish classic benign lesions from other unilocular and multilocular cystic tumors. The lexicon of "classic benign lesions" represented multiple kinds of lesions [10]. Each kind of lesion has varied shapes and echoes, which were difficult to cover by the images in the references. An endometriotic cyst or hydrosalpinx may resemble a unilocular or multilocular cyst with a smooth inner wall. For less experienced doctors, the content of "classic benign lesions" was not as clear as that of other lexicon items. Improving the recognition of classic benign lesions may need a lot of practice over a long time. A significant increase in the learning curve may be far in the future for this task.

In summary, if the definition of a lexicon descriptor needs subjective judgment, it was a difficult point for junior doctors. More experience was needed to better the understanding of these lexicon descriptors, such as the color score, solid-appearing masses, and classic benign lesions.

There are many limitations in this study that need to be acknowledged. First, the test with 20 cases was not enough to test all the lexicon descriptors in the O-RADS US system. The tests had various difficult points due to the random selection of cases. Second, no remarkable improvement was observed in the learning curve with five tests. More training data or a more effective training modality is needed for future studies. Third, the effect of experience could not be fully evaluated because no senior doctors attended the training. A large study of interobserver variability would be needed to validate the use of the system by experts as well as less experienced observers [11,18–20].

In conclusion, after training, junior doctors at different levels can reach a coincident O-RADS US risk stratification. The difficulties focused on the subjective judgment of classic benign lesions, the color flow score, and solid-appearing masses. More experience is needed to improve doctors' understanding of the system and ability to apply it in real-world circumstances.

ORCID: Shan Zhou: <https://orcid.org/0000-0001-5795-8277>; Yuyang Guo: <https://orcid.org/0000-0001-5630-6912>; Lieming Wen: <https://orcid.org/0000-0003-3283-888X>; Baihua Zhao: <https://orcid.org/0000-0001-7185-292X>; Minghui Liu: <https://orcid.org/0000-0002-6548-3264>

### Author Contributions

Conceptualization: Wen L, Zhao B, Liu M. Data acquisition: Zhou S, Guo Y, Wen L, Zhao B. Data analysis or interpretation: Zhou S. Drafting of the manuscript: Zhou S, Guo Y. Critical revision of the manuscript: Wen L, Zhao B, Liu M. Approval of the final version of the manuscript: all authors.

### Conflict of Interest

No potential conflict of interest relevant to this article was reported.

## References

1. 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* 2021;58:148-168.
2. Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol* 2008;31:681-690.
3. Timmerman D, Van Calster B, Testa A, Savelli L, Fischerova D, Froyman W, 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;214:424-437.
4. Van Calster B, Van Hoorde K, Valentin L, Testa AC, Fischerova D, Van Holsbeke C, et al. Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study. *BMJ* 2014;349:g5920.
5. Van Calster B, Timmerman D, Valentin L, McIndoe A, Ghaem-Maghani S, Testa AC, et al. Triaging women with ovarian masses for surgery: observational diagnostic study to compare RCOG guidelines with an International Ovarian Tumour Analysis (IOTA) group protocol. *BJOG* 2012;119:662-671.
6. Amor F, Vaccaro H, Alcazar JL, Leon M, Craig JM, Martinez J. Gynecologic imaging reporting and data system: a new proposal for classifying adnexal masses on the basis of sonographic findings. *J Ultrasound Med* 2009;28:285-291.
7. Amor F, Alcazar JL, Vaccaro H, Leon M, Iturra A. GI-RADS reporting system for ultrasound evaluation of adnexal masses in clinical practice: a prospective multicenter study. *Ultrasound Obstet Gynecol* 2011;38:450-455.
8. Kaijser J, Sayasneh A, Van Hoorde K, Ghaem-Maghani S, Bourne T, Timmerman D, et al. Presurgical diagnosis of adnexal tumours using mathematical models and scoring systems: a systematic review and meta-analysis. *Hum Reprod Update* 2014;20:449-462.
9. Andreotti RF, Timmerman D, Benacerraf BR, Bennett GL, Bourne T, Brown DL, et al. Ovarian-adnexal reporting lexicon for ultrasound: a white paper of the ACR Ovarian-Adnexal Reporting and Data

- System Committee. *J Am Coll Radiol* 2018;15:1415-1429.
10. 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* 2020;294:168-185.
  11. 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;31:674-684.
  12. Guerriero S, Pascual MA, Ajossa S, Rodriguez I, Zajicek M, Rolla M, et al. Learning curve for ultrasonographic diagnosis of deep infiltrating endometriosis using structured offline training program. *Ultrasound Obstet Gynecol* 2019;54:262-269.
  13. Meng X, Rosenkrantz AB, Huang R, Deng FM, Wysock JS, Bjurlin MA, et al. The institutional learning curve of magnetic resonance imaging-ultrasound fusion targeted prostate biopsy: temporal improvements in cancer detection in 4 Years. *J Urol* 2018;200:1022-1029.
  14. Ong J, Leonardi M, Espada M, Stamatopoulos N, Georgousopoulou E, Condous G. Ureter visualization with transvaginal ultrasound: a learning curve study. *J Ultrasound Med* 2020;39:2365-2372.
  15. Patel M, Agius S, Wilkinson J, Patel L, Baker P. Value of supervised learning events in predicting doctors in difficulty. *Med Educ* 2016;50:746-756.
  16. Yazbek J, Raju SK, Ben-Nagi J, Holland TK, Hillaby K, Jurkovic D. Effect of quality of gynaecological ultrasonography on management of patients with suspected ovarian cancer: a randomised controlled trial. *Lancet Oncol* 2008;9:124-131.
  17. Education Practical Standards Committee; European Federation of Societies for Ultrasound in Medicine Biology. Minimum training recommendations for the practice of medical ultrasound. *Ultraschall Med* 2006;27:79-105.
  18. Alcazar JL, Pascual MA, Graupera B, Auba M, Errasti T, Olartecoechea B, et al. External validation of IOTA simple descriptors and simple rules for classifying adnexal masses. *Ultrasound Obstet Gynecol* 2016;48:397-402.
  19. Ruiz de Gauna B, Sanchez P, Pineda L, Utrilla-Layna J, Juez L, Alcazar JL. Interobserver agreement in describing adnexal masses using the International Ovarian Tumor Analysis simple rules in a real-time setting and using three-dimensional ultrasound volumes and digital clips. *Ultrasound Obstet Gynecol* 2014;44:95-99.
  20. Basha MA, Refaat R, Ibrahim SA, Madkour NM, Awad AM, Mohamed EM, et al. Gynecology Imaging Reporting and Data System (GI-RADS): diagnostic performance and inter-reviewer agreement. *Eur Radiol* 2019;29:5981-5990.