European Radiology (2019) 29:5981–5990
https://doi.org/10.1007/s00330-019-06181-0

ULTRASOUND

Gynecology Imaging Reporting and Data System (GI-RADS): diagnostic performance and inter-reviewer agreement

Mohammad Abd Alkhalik Basha[1] · Rania Refaat[2] · Safaa A. Ibrahim[3] · Nadia M. Madkour[3] · Awad Mahmoud Awad[4] · Elshaimaa Mohamed Mohamed[1] · Ahmed A. El Sammak[1] · Mohamed M. A. Zaitoun[1] · Hitham A. Dawoud[1] · Mai E. M. Khamis[1] · Heba A. E. Mohamed[1] · Ahmed Mohamed El-Maghraby[1] · Ahmed A. El-Hamid M. Abdalla[1] · Mostafa Mohamad Assy[1] · Mohamad Gamal Nada[1] · Ahmed Ali Obaya[5] · Eman H. Abdelbary[6]

Received: 11 December 2018 / Revised: 12 March 2019 / Accepted: 19 March 2019 / Published online: 16 April 2019

Abstract

Objective To evaluate diagnostic performance and inter-reviewer agreement (IRA) of the Gynecologic Imaging Reporting and Data System (GI-RADS) for diagnosis of adnexal masses (AMs) by pelvic ultrasound (US).

Patients and methods A prospective multicenter study included 308 women (mean age, 41 ± 12.5 years; range, 15–73 years) with 325 AMs detected by US. All US examinations were analyzed, and AMs were categorized into five categories according to the GI-RADS classification. We used histopathology and US follow-up as the reference standards for calculating diagnostic performance of GI-RADS for detecting malignant AMs. The Fleiss kappa (κ) tests were applied to evaluate the IRA of GI-RADS scoring results for predicting malignant AMs.

Results A total of 325 AMs were evaluated: 127 (39.1%) were malignant and 198 (60.9%) were benign. Of 95 AMs categorized as GI-RADS 2 (GR2), none was malignant; of 94 AMs categorized as GR3, three were malignant; of 13 AMs categorized as GR4, six were malignant; and of 123 AMs categorized as GR5, 118 were malignant. On a lesion-based analysis, the GI-RADS had a sensitivity, a specificity, and an accuracy of 92.9%, 97.5%, and 95.7%, respectively, when regarding only those AMs classified as GR5 for predicting malignancy. Considering combined GR4 and GR5 as a predictor for malignancy, the sensitivity, specificity, and accuracy of GI-RADS were 97.6%, 93.9%, and 95.4%, respectively. The IRA of the GI-RADS category was very good ($\kappa = 0.896$). The best cutoff value for predicting malignant AMs was >GR3.

Conclusions The GI-RADS is very valuable for improving US structural reports.

Key Points

- *There is still a lack of a standard in the assessment of AMs.*
- *GI-RADS is very valuable for improving US structural reports of AMs.*
- *GI-RADS criteria are easy and work at least as well as IOTA.*

Keywords Adnexal diseases · Neoplasms · Gynecology · Ultrasonography

✉ Mohammad Abd Alkhalik Basha
drmohammad_basha@yahoo.com

[1] Department of Radiodiagnosis, Zagazig University, Zagazig, Egypt
[2] Department of Radiodiagnosis, Ain Shams University, Cairo, Egypt
[3] Department of Obstetrics & Gynecology, Zagazig University, Zagazig, Egypt
[4] Department of Obstetrics & Gynecology, Al-Azhar University, Cairo, Egypt
[5] Department of Clinical Oncology, Zagazig University, Zagazig, Egypt
[6] Department of Pathology, Zagazig University, Zagazig, Egypt

Abbreviations

AMs	Adnexal masses
AUC	Area under the ROC curve
BI-RADS	Breast Imaging Reporting and Data System
CI	Confidence interval
FIGO	Federation of Gynaecology and Obstetrics
GI-RADS	Gynecologic Imaging Reporting and Data System
IOTA	International Ovarian Tumor Analysis
IRA	Inter-reviewer agreement
NPV	Negative predictive value
PPV	Positive predictive value
ROC	Receiver operating characteristic

TV Transvaginal
US Pelvic ultrasound

Introduction

Adnexal masses (AMs) are clinical and imaging-wise common problem [1]. AMs represent a wide variety of causes, and so they can signify a diagnostic dilemma [2]. Imaging is a cornerstone in the diagnosis of AMs, from the early detection to categorization [3]. Pelvic ultrasound (US) is still the most frequently used imaging method for detecting and characterizing AMs [4].

Improper reporting might cause undue worry by the patient and referring physician and can result in unwarranted extra tests or surgery [4]. Several groups have performed substantial attempts in designating definitions and terms for US features in AMs [5]. A consensus conference of the society of radiologists in the US which organized a team of specialists from radiology, gynecology, and pathology established that "investigation into designed reporting of AMs to permit for better communication of findings and suggestions for follow-up" is required [6]. Amor et al [7] in 2009 designed the Gynecology Imaging Reporting and Data System (GI-RADS) resembling the Breast Imaging Reporting and Data System (BI-RADS) used for breast masses as an approach to enable structured reporting of AMs. This system stands on criteria and recognition patterns suggested by the International Ovarian Tumor Analysis (IOTA) [8]. The GI-RADS is based on subjective characterization of the adnexal images, by a trained operator, a method currently validated by several publications [7–10]. The lexicon of GI-RADS is aimed to offer a unified language for US reporting and for preventing misinterpretation in communication between the physician and the sonographer [6]. In contrast to the IOTA models, GI-RADS does not involve objective criteria for AM evaluation and depends on the subjective assessment of the sonographer [9].

Numerous studies have examined the diagnostic ability of US in the evaluation of AMs, but a few of these studies have been used GI-RADS [7–10]. As a standardized and consensus-based reporting is the first step in the process of developing reproducible US features, we performed a multicenter prospective study to assess the diagnostic performance and inter-reviewer agreement (IRA) of US in reference to GI-RADS to classify AMs. We additionally made a simple comparison between using GI-RADS and using pattern recognition analysis without standardized diagnostic algorithm for classification of AMs to detect whether the use of GI-RADS would improve the diagnostic accuracy of US for predicting malignant AMs.

Patients and methods

Ethical statement

This prospective study was conducted according to international guidelines approved by the Research Ethics Committee. Informed consents were obtained from all patients prior to the study. We followed the ethical principles of the Declaration of Helsinki during the preparation of this study.

Study population

The study was carried out between January 2017 and August 2018. We initially collected 335 consecutive women from three institutions.

Inclusion criterion was women with suspected AMs based on clinician request.

Exclusion criteria were as follows:

– Patients categorized as GI-RADS 1 (GR1), e.g., normal ovaries at US ($n = 8$).
– Patients lost during follow-up ($n = 13$).
– Pregnant patients at the time of examination ($n = 6$).

This resulted in a final cohort of 308 women (mean age of 41 ± 12.5 years and age range of 15–73 years). All participants underwent a full history, a complete general examination, and US examination. The flow chart of our study is illustrated in Fig. 1.

Ultrasound examination

All US examinations were performed using the same equipment (SonoScape S40 Exp/S40 Pro/S40/S35 Digital Color Doppler ultrasound system) to avoid inter-ultrasound equipment effect on the results. The transvaginal (TV) US was performed using a real-time sector scanner with a high-frequency TV probe (5/7.5 MHz). The patient lied in the lithotomy position after voiding the urine. The transabdominal US was also done for virgin patients ($n = 18$) or with large tumors that cannot be completely seen by the TV route ($n = 54$). Transabdominal US scanning was done using a real-time scanner with a low-frequency probe (3/3.5 MHz).

Image analysis

All US examinations were performed centrally. Three highly experienced reviewers (with over 15 years of US experience and have performed > 1000 US examinations per year) from three different institutions blinded to patients' clinical data independently performed all US examinations and assigned GI-RADS category for each AM. Before the beginning of

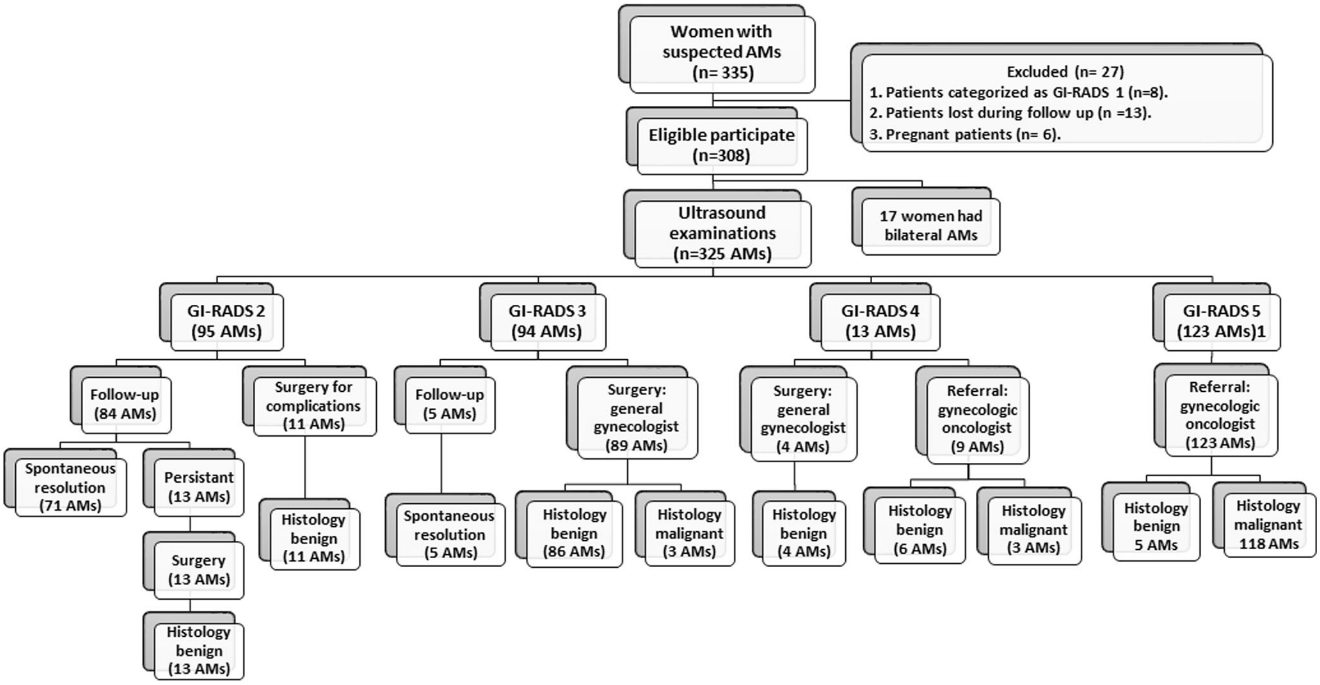

Fig. 1 Flowchart of the study population. AMs, adnexal masses

the study, the reviewers were provided 1 h of lecture-based and hands-on instruction that explained GI-RADS in detail.

The following morphological features obtained at US examination were individually evaluated for each AM:

1. Laterality (unilateral or bilateral).
2. Maximum diameter.
3. Echogenicity.
4. Wall thickness.
5. Cystic content.
6. Solid areas.
7. Septations.
8. Solid papillary projections.
9. The presence of ascites.
10. The Doppler evaluation of AMs was done using the color Doppler gate to detect color signals. Peripheral blood flow was defined if color signals were seen in the periphery of the mass; while central blood flow was defined if color signals were seen in the central part of the mass, solid areas, septa, or papillary projections. If the mass exhibited central and peripheral blood flow, the central blood flow only was used for analysis [11].

The morphological features suspicious of malignancy included thick wall and septum, solid papillary projection, solid areas, the presence of ascites, and central blood flow [5, 12]. Each radiologist appointed a GI-RADS category to all detected AMs using GI-RADS classification developed by Amor et al [7]. The GI-RADS categories are:

– *GR1*: Definitively benign. Normal ovaries.
– *GR2*: Very probably benign. Functional AMs, e.g., corpora lutea, follicles, and hemorrhagic cysts.
– *GR3*: Probably benign. Benign neoplastic AMs, e.g., paraovarian cyst, teratoma, endometrioma, pedunculated myoma, peritoneal pseudocyst, hydrosalpinx, and pelvic inflammatory disease.
– *GR4*: Probably malignant. AMs that cannot be included in the above three groups and with one or two of the morphological features suspicious of malignancy.
– *GR5*: Very probably malignant. AMs with three or more of the morphological features suspicious of malignancy.

The IRA was performed among three independent reviewers regarding diagnostic morphological features and GI-RADS categorization of AMs. In order to estimate the diagnostic performance of GI-RADS for predicting malignant AMs, the category for each AM scaled by three independent reviewers was merged into a final category. During this procedure, any disagreement between reviewers in categorizing GI-RADS features of the AM was reviewed until consensus was reached.

Another fourth highly experienced reviewer (with over 15 years of US experience and have performed > 1000 US examinations per year) independently performed US examination for all patients and categorized each AM as malignant or benign using pattern recognition analysis provided by IOTA (Table 1) [5], but did not use GI-RADS. The rationale for this comparison was the question of whether the use of GI-

Table 1 IOTA simple rules for identifying a benign or malignant tumor

Rules for predicting a malignant tumor (M-rules)		Rules for predicting a benign tumor (B-rules)	
M1	Irregular solid tumor	B1	Unilocular
M2	Presence of ascites	B2	Presence of solid components where the largest solid component has a largest diameter < 7 mm
M3	At least four papillary structures	B3	Presence of acoustic shadows
M4	Irregular multilocular solid tumor with largest diameter ≥ 100 mm	B4	Smooth multilocular tumor with largest diameter < 100 mm
M5	Very strong blood flow (color score 4)	B5	No blood flow (color score 1)

If one or more M-rules apply in the absence of a B-rule, the mass is classified as malignant. If one or more B-rules apply in the absence of an M-rule, the mass is classified as benign. If both M-rules and B-rules apply, the mass cannot be classified. If no rule applies, the mass cannot be classified

IOTA, International Ovarian Tumor Analysis

RADS would improve the diagnostic accuracy of US for predicting malignant AMs.

Patient management

Our management protocol followed that of Amor et al [9]:

- *GR1 patients*: Were excluded from the study.
- *GR2 patients*: Were followed up by US. Follow-up was performed every 3 months up to 1 year. If the AMs did not resolve or increased in size, they were submitted to surgery.
- *GR3 patients*: Were submitted to surgery.
- *GR4 and GR5 patients*: Were referred to oncologists for further diagnostic workup and surgical management.

Reference standard

The final diagnoses of AMs were confirmed based on the followings:

1. Post-operative histopathological results ($n = 249$ AMs). All specimens were reviewed by two experienced pathologists and the findings were gained by consensus. We classified tumors as stated by the criteria recommended by the International Federation of Gynaecology and Obstetrics (FIGO) [13]. Borderline AMs were regarded as malignant.
2. The remaining 76 AMs (55 follicular cysts, 16 hemorrhagic cysts, 5 corpus luteum cysts) were resolved spontaneously during follow-up and were considered to be benign.

Statistical analysis

The collected data were computerized and statistically analyzed using MedCalc program (version 11.1). Categorical variables were described as number and percentage. Continuous variables were compared using an independent sample t test. The Fleiss kappa (κ) statistics for multiple reviewers were applied to evaluate the IRA of morphological features and GI-RADS scoring results for predicting malignancy of AMs. The κ values were interpreted as follows: 0.01–0.20 = poor agreement; 0.21–0.40 = fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.80 = good agreement; and 0.81–1.0 = very good agreement. On a lesion-based analysis, we used a fourfold table test to evaluate the diagnostic performance of GI-RADS to categorize AMs using histopathology and US follow-up as reference standards. We used McNemar test to assess significance of GI-RADS in comparison with pattern recognition analysis without standardized diagnostic algorithm. The cutoff value and the area under the curve (AUC) were determined using receiver operating characteristic (ROC) curve. A p value ≤ 0.05 indicated statistically significant results.

Results

Patients

Three hundred eight women with at least one AM on US examination were enrolled in our study. We found a total of 325 AMs (17 patients (5.5%) had bilateral AMs). The patients' data are summarized in Table 2. The mean presenting age for malignant AMs was 51.5 ± 11.3 years. No significant difference between benign and malignant AMs as regards patient age ($p = 0.064$). Malignant AMs were more common in postmenopausal women (62.2%) than in premenopausal women (37.8%) ($p = 0.002$).

Table 3 shows the definitive final diagnoses of our patients. The final diagnoses in 325 AMs were 127 malignant (39.1%) and 198 benign (60.9%). We found that the follicular cyst was the most common benign AMs (18%), and the serous

Table 2 General characteristics of the patients with adnexal masses ($n = 308$)

Characteristic		No. of patients	Final diagnosis		p value
			Benign ($n = 189$)	Malignant ($n = 119$)	
Age, years, mean ± SD		308	41.47 ± 14.45	51.46 ± 11.31	0.064
Postmenopausal	Yes	177 (57.5)	103 (54.5)	74 (62.2)	0.002
	No	131 (42.5)	86 (45.5)	45 (37.8)	
Married	Yes	290 (94.2)	175 (92.6)	115 (96.6)	0.111
	No	18 (5.8)	14 (7.4)	4 (3.4)	
Gravidity ≥ 1	Yes	281 (91.2)	176 (93.1)	105 (88.2)	0.163
	No	27 (8.8)	13 (6.9)	14 (11.8)	
Bilateral involvement	Yes	17 (5.5)	9 (4.8)	8 (6.7)	0.481
	No	291 (94.5)	180 (95.2)	111 (93.3)	

Unless otherwise indicated, data are number with the percentage in parentheses

SD, standard deviation

cystadenocarcinoma was the most common malignant AMs (18.6%).

Assignment of GI-RADS categories

Categorization of benign and malignant AMs based on GI-RADS classification system is presented in Table 3. Of the 325 AMs assessed, 95 (29.2%) were classified as GR2, 94 (28.9%) as GR3, 13 (4%) as GR4, and 123 (37.8%) as GR5. Of the 95 AMs categorized as GR2, none was malignant; of the 94 AMs categorized as GR3, three were malignant; of the 13 AMs categorized as GR4, six were malignant; and of the 123 AMs categorized as GR5, 118 were malignant.

Table 3 Final diagnosis by GI-RADS categorization ($n = 325$)

Histologic diagnosis	GI-RADS 2	GI-RADS 3	GI-RADS 4	GI-RADS 5	Total
Benign adnexal masses	95 (29.2)	91 (28)	7 (2.2)	5 (1.5)	198 (60.9)
Follicular cyst	58	1			59 (18.2)
Hemorrhagic cyst	26	2			28 (8.6)
Endometrioma	4	19	1		24 (7.4)
Mature cystic teratoma		19	1	1	21 (6.5)
Tubo-ovarian abscess		15		1	16 (4.9)
Serous cystadenoma	2	10	2	1	15 (4.6)
Mucinous cystadenoma		7	1	2	10 (3.1)
Paraovarian cyst		9			9 (2.8)
Corpus luteum cyst	5	3	1		9 (2.8)
Lutein cyst		2	1		3 (0.9)
Peritoneal inclusion cyst		2			2 (0.6)
Fibroma		2			2 (0.6)
Malignant adnexal masses	0	3 (0.9)	6 (1.8)	118 (36.3)	127 (39.1)
Serous cystadenocarcinoma		1	2	58	61 (18.8)
Mucous cystadenocarcinoma			2	28	30 (9.2)
Borderline tumor				11	11 (3.4)
Germ cell tumor		1		8	9 (2.8)
Metastatic carcinoma				6	6 (1.8)
Malignant stromal tumors			1	5	6 (1.8)
Endometrioid carcinoma				2	2 (0.6)
Granulosa cell tumor		1	1		2 (0.6)
Total	95 (29.2)	94 (28.9)	13 (4)	123 (37.9)	325 (100)

Unless otherwise indicated, data are number with the percentage in parentheses

GI-RADS, Gynecology Imaging Reporting and Data System

Table 4 Inter-reviewer agreement for morphological diagnostic features and GI-RADS categorization for diagnosis of adnexal masses

Feature	Inter-reviewer	
	κ coefficients	95% CI
Maximum diameter of lesion (mean ± SD)	0.952	0.908 to 0.995
Wall thickness	0.639	0.353 to 0.925
Septum	0.714	0.458 to 0.971
Solid papillary projection	0.850	0.650 to 1.000
Solid area	0.650	0.337 to 0.963
Cystic content	0.750	0.491 to 1.000
CDFI	0.381	0.0303 to 0.732
Ascites	1.000	1.000 to 1.000
GI-RADS categorization	0.896	0.775 to 1.000

GI-RADS, Gynecology Imaging Reporting and Data System; *CDFI*, color Doppler flow imaging

IRA for morphological diagnostic features and GI-RADS categorization for predicting malignant AMs

The morphological features of AMs detected by the US are reported in Table 4. The IRA was very good for the identification of ascites ($\kappa = 1.0$), the maximum diameter of lesion ($\kappa = 0.952$), and solid papillary projection ($\kappa = 0.850$). The IRA was good for the identification of cystic content ($\kappa = 0.750$), septum ($\kappa = 0.714$), a solid area ($\kappa = 0.650$), and wall thickness ($\kappa = 0.639$). The IRA was fair for the detection of color Doppler flow ($\kappa = 0.381$). The IRA of the GI-RADS category was very good ($\kappa = 0.896$).

Diagnostic performance of GI-RADS for predicting malignant AMs

On a lesion-based analysis, the diagnostic performance of GI-RADS for predicting malignant AMs is summarized in Table 5. Regarding only those AMs classified as GR5 for predicting malignancy, the GI-RADS had a sensitivity, specificity, PPV, NPV and accuracy of 92.9%, 97.5%, 95.9%, 95.5%, and 95.7%, respectively. Considering combined GR4 and GR5 as a predictor for malignancy, the sensitivity, specificity, PPV, NPV, and accuracy were 97.6%, 93.9%, 91.2%, 98.4%, and 95.4%, respectively.

By using pattern recognition analysis in the diagnosis of AMs, the diagnostic performance was 91.3% sensitivity, 92.9% specificity, 89.2% PPV, 94.36% NPV, and 92.3% accuracy. According to the McNemar test, we did not find a significant difference between using GI-RADS and using pattern recognition analysis without standardized diagnostic algorithm for improving the diagnostic accuracy of US for predicting malignant AMs ($p = 0.109$; 95% CI = −0.36% to 3.06%).

ROC analyses

We analyzed the data set of the diagnostic performance of GI-RADS to determine the cutoff value for predicting malignant AMs using the ROC curve. ROC analyses showed an AUC of 0.965 (95% CI 0.939 to 0.982, $p < 0.0001$) with the optimal cutoff value for predicting malignant AMs was >GR3. The use of this cutoff value was associated with a sensitivity of 97.6% (95% CI 93.3 to 99.5), and a specificity of 93.3% (95% CI 89.7 to 96.8). On comparing the ROC areas of combined GR4 and GR5 and GR5 alone, it was found that combined GR4 and GR5 was significantly superior to GR5 alone in predicting malignancy of AMs ($p = 0.0473$; 95% CI 0.0002 to 0.0484) (Fig. 2).

The GI-RADS revealed 100% sensitivity and 48% specificity in very probably benign AMs (GR2), and 92.9% sensitivity and 97.5% specificity in probably malignant AMs (GR4).

Representative cases of our study are shown in (Figs. 3, 4, 5, and 6).

Table 5 The diagnostic performance of GI-RADS for diagnosis of malignant adnexal masses

	GR5		GR4+5	
	%	95% CI	%	95% CI
Accuracy	95.69		95.38	
Sensitivity	92.91	86.97 to 96.71	97.64	93.25 to 99.51
Specificity	97.47	94.21 to 99.18	93.94	89.65 to 96.83
Positive likelihood ratio	36.79	15.47 to 87.54	16.11	9.30 to 27.90
Negative likelihood ratio	0.07	0.04 to 0.14	0.03	0.01 to 0.08
Disease prevalence	39.08	33.74 to 44.62	39.08	33.74 to 44.62
Positive predictive value	95.93	90.77 to 98.67	91.18	85.09 to 95.36
Negative predictive value	95.54	91.71 to 97.94	98.41	95.43 to 99.67

GI-RADS, Gynecology Imaging Reporting and Data System; *CI*, confidence interval

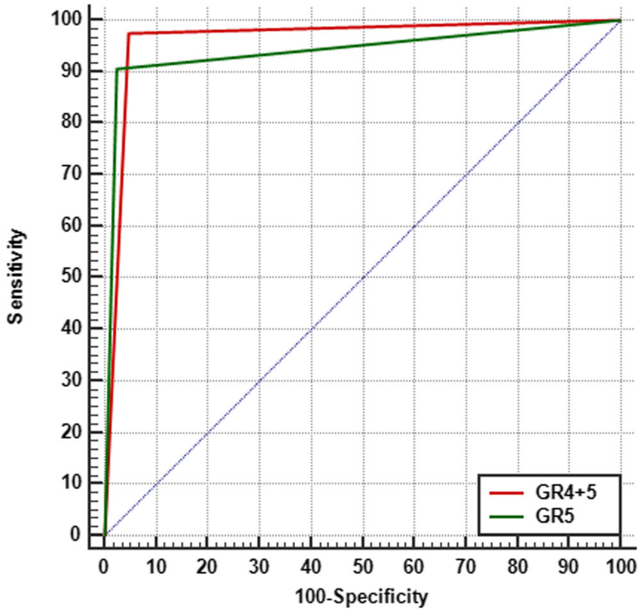

Fig. 2 Comparison of the ROC areas of GR4+5 and GR5 alone for predicting malignancy of adnexal masses as evidenced by histopathology as the reference standard ($p = 0.0473$; 95% CI 0.0002 to 0.0484)

Discussion

Since its development in 2009, attempts have been made to establish the true value of the GI-RADS and support its external validation as a useful classification for US examination of malignant AMs. However, external validations are still lacking the pros and cons that may support its use. In the current multicenter study, we prospectively examined 325 AMs in 308 women by US according to the GI-RADS classification system to evaluate diagnostic performance and IRA of GI-RADS for diagnosis of AMs by US. As regards a lesion-based analysis, the overall results confirmed the high diagnostic performance of GI-RADS. We found that when respecting only those AMs categorized as GR5 for predicting malignancy, the GI-RADS had a sensitivity and a specificity of 92.9% and 97.5%, respectively. Our results are congruent with those mentioned in the previous attempts [7–10, 14, 15], which reported sensitivity ranging from 84.9 to 100% and specificity ranging from 84.3 to 97%, respectively.

Our study revealed that combined GR4 and GR5 was significantly better than GR5 alone in predicting malignancy of AMs. The ROC areas of combined GR4 and GR5 were significantly superior to GR5 alone ($p = 0.047$). Moreover, with this combination, the sensitivity and NPV were significantly increased (97.6% and 98.4%, respectively), whereas the specificity and PPV were minimally decreased (93.9% and 91.2%, respectively). Thus, we should combine both GR4 and GR5 for the diagnosis of malignant AMs because if we consider GR5 alone as conclusive for malignant AMs diagnosis, the GI-RADS would miss a relevant number of malignant AMs.

A remarkable finding in our study was the small number of GR4 lesions, accounted for 13/325 (4%) of all AMs. This could be explained by the complexity of some of the assessed AMs. However, as reported in previous studies [8–10, 14], this category represented a significant source of false-positive patients, where seven patients (53.8%) in this category were actually benign. Thus, this category should be subjected to future studies with additional markers and subclassifications in order to define new US features that reduce the false-positive rate of this category and improve its specificity.

Sajdak et al [16] reported that some malignant tumors could still be found among those considered benign tumors especially within the unexpected population, e.g., premenopausal patients. In our study, we found nine false-negative patients and five false-positive patients. The false-negative patients were as follows: three serous cystadenocarcinomas and two mucinous cystadenocarcinomas diagnosed as suspected benign tumors (the vegetation could not be seen in two patients, and the obesity reduced sonographic accessibility in other three patients), two granulosa cell tumors diagnosed as tubo-ovarian abscess and fibroma, one germ cell

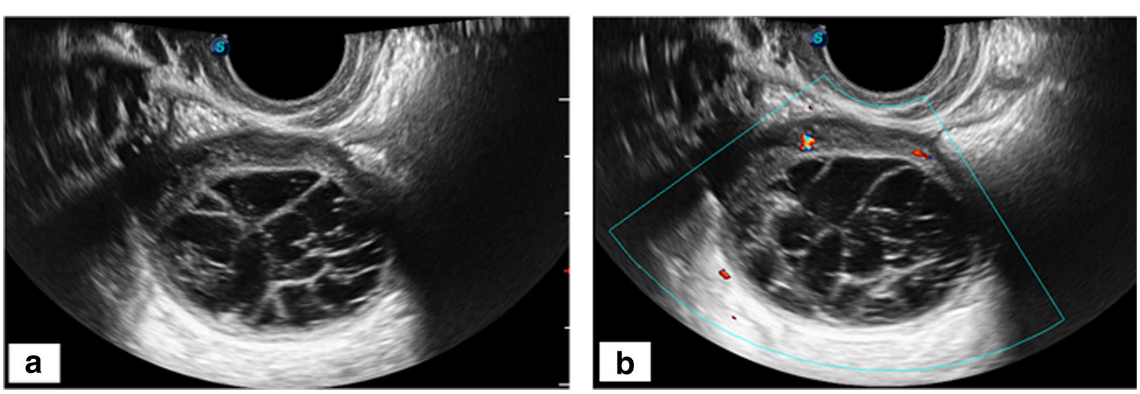

Fig. 3 **a** Transvaginal ultrasound reveals a left adnexal complex cystic lesion with fine internal reticulation (web-like appearance) and scattered echoes within the fluid. **b** Color Doppler examination reveals no internal vascularity. The lesion was diagnosed as a hemorrhagic cyst and classified as GR2. The patient was followed up, and the cyst resolved spontaneously after 50 days

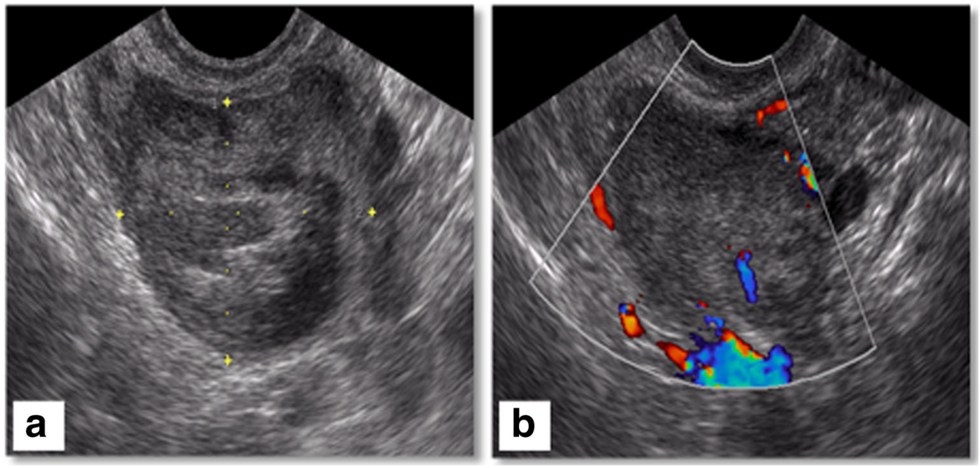

Fig. 4 **a** Transvaginal ultrasound shows a mixed cystic and solid adnexal mass. **b** Color Doppler study shows scanty vascularity. The mass was diagnosed as a suspicious malignant neoplastic lesion and classified as GR4. Surgery was performed, and the diagnosis was confirmed on histopathology as mucinous cystadenocarcinoma

tumor diagnosed as a suspected benign tumor, and one malignant stromal tumor diagnosed as a fibroma (fibroma demonstrated vessels as did malignant stromal cell tumor and could not be differentiated by US). On the other hand, the five false-positive patients were as follows: two mucinous cystadenomas and one serous cystadenoma with relevant internal vascularity diagnosed as carcinomas, one mature cystic teratoma with large solid component diagnosed as carcinoma, and one tubo-ovarian abscess diagnosed as a borderline tumor. All the false-positive patients were showing patterns mentioned by Alcázar et al [17] in an article published about diagnostic performance of US for specific diagnosis of AMs.

MRI with the functional imaging techniques such as diffusion-weighted imaging or dynamic contrast-enhanced imaging can partially overcome the limitations that faced US in diagnosis of AMs [18]. However, the problem is a lack of what defines a mass as indeterminate at US. The application of the good sonographic GI-RADS criteria to characterize AMs reduces the number of truly US-indeterminate examination findings and limits the need for additional imaging with other modalities.

The reporting of the GI-RADS category of AMs in our study showed very good IRA ($\kappa = 0.896$). This result is very similar to that of Amor et al [9] who found very good IRA ($\kappa = 0.846$) for GI-RADS categorization of AMs. Additionally, we reported the IRA for morphological features of AMs. The IRA was very good for the identification of ascites ($\kappa = 1.0$), the maximum diameter of lesion ($\kappa = 0.952$), and solid papillary projection ($\kappa = 0.850$). The IRA was good for the identification of cystic content ($\kappa = 0.750$), septum ($\kappa = 0.714$), a solid area ($\kappa = 0.650$), and wall thickness ($\kappa = 0.639$). The IRA was fair for the detection of color Doppler flow ($\kappa = 0.381$). The overall reproducibility values are higher than those of previously published studies, possibly due to the greater experience of our reviewers, the use of the same US device in all examinations, and the real-time images. However, to date, there are no studies assessing the IRA of such features for GI-RADS. Thus, our results are not directly comparable to the literature.

Using the ROC curve, we found the optimal cutoff value of GI-RADS for predicting malignant AMs was >GR3. The use of this cutoff value was associated with a sensitivity of 97.6%, and a specificity of 93.3%. However, we were not capable to

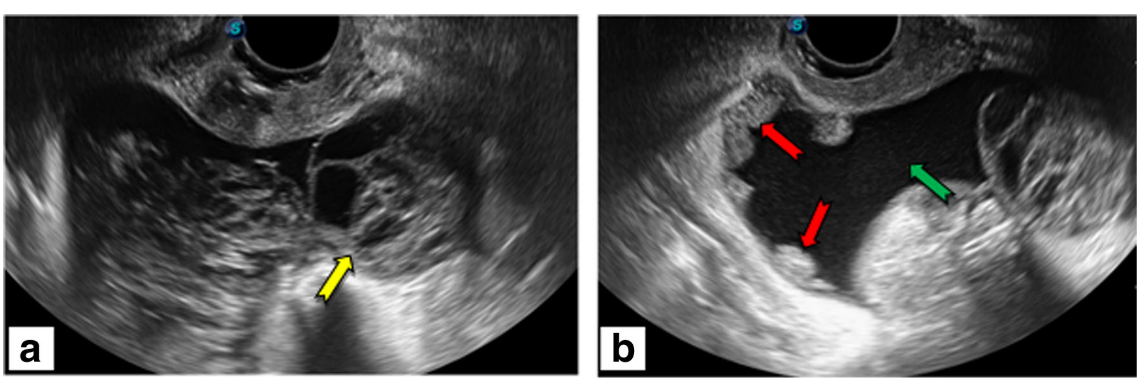

Fig. 5 **a** Transvaginal ultrasound demonstrates a left adnexal complex cystic mass (yellow arrows) of solid and cystic components and multiple thick internal septations. **b** Turbid ascites (green arrow) and multiple peritoneal deposits (red arrows) were present in the abdomen. The mass was diagnosed as a malignant neoplastic lesion and classified as GR5. Surgery was performed, and the diagnosis was confirmed on the histopathology as serous cystadenocarcinoma

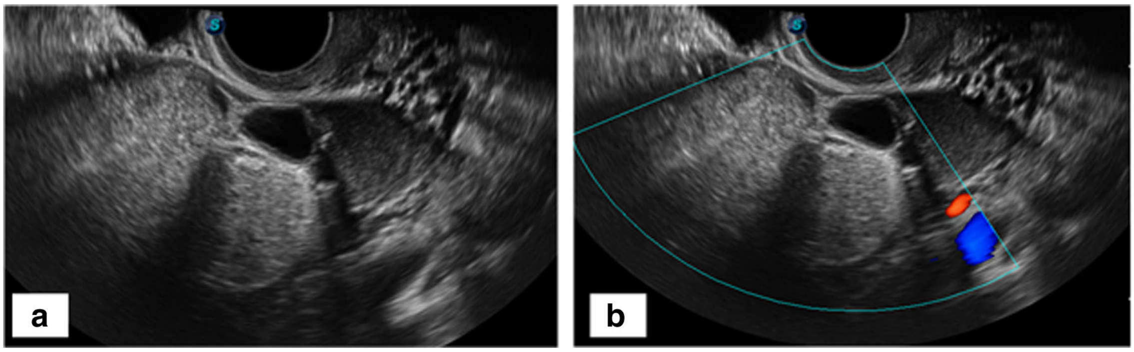

Fig. 6 **a** Transvaginal ultrasound demonstrates a right adnexal complex mass. The mass shows highly echogenic content with posterior shadowing. **b** Color Doppler study demonstrates no vascularity. The mass was diagnosed as a dermoid cyst and classified as GR3. Surgery was performed, and the diagnosis was confirmed on the histopathology as a mature cystic teratoma

compare our cutoff value as no available previous studies provided cutoff value for the GI-RADS. Thus, we recommend further similar studies with a larger population to confirm or refute our cutoff value.

Subjective US assessment is the best approach for diagnosis of AMs, no other method has proven to be superior in prospective testing, nor can any outperform an expert sonographer [19, 20]. Our results proved that the GI-RADS increased the sensitivity of US for categorization of AMs compared with pattern recognition analysis without standardized diagnostic algorithm (97.6 vs 93.1%). Although it was not significant ($p = 0.109$), this is actually a very good result and gives additional weight to the GI-RADS system. The GI-RADS criteria are easier to learn than IOTA criteria and have less category 4. So, they might be easier to teach and to be implemented in daily practice than the pattern recognition system. Moreover, the GI-RADS is a categorized reporting of pattern recognition which can reduce mistakes of important data from US reports by standardizing report structure and content, decrease inconsistency in interpretation of AMs, and optimize patient management. Additionally, the systematic use of GI-RADS system will help in the future to monitor performance, quality control, and possibly patient outcomes.

Finally, in keeping with our results, several authors [7–10, 14, 15] have tested GI-RADS and considered this system as an effective classification for US examination of malignant AMs. Moreover, some authors [21] are trying to enhance the diagnostic power of GI-RADS with the addition of CA-125. Therefore, the GI-RADS can stand alone and works like BI-RADS as a universal system that helps the clinician to go from one imaging technique such as US towards CT and MRI. However, the GI-RADS needs further modification to become accurate, useful, and comprehensive of all pertinent descriptors and definitions.

Very recently, during the time of this study, the American College of Radiology (ACR) developed another classification system called Ovarian-Adnexal Reporting and Data System (O-RADS) and published a white paper describing the consensus process in the creation of a standardized lexicon for ovarian and adnexal lesions and the resultant lexicon [22]. Future large studies comparing this new classification with GI-RADS and IOTA criteria are recommended.

The current study has some strong points. It is a prospective, large, multicenter study with no selection bias of retrospective study. However, our study has limitations. First, the GI-RADS depends only on the US in all classifications, which is an operator-dependent modality. Second, all examinations were performed by highly experienced sonographers. Third, the GI-RADS is still uncommon and unfamiliar by many clinicians. Finally, the GI-RADS categorization system is still under modification.

Conclusion

In conclusion, our results support the high reliability of GI-RADS classification system for diagnosis of AMs by US. Additionally, the GI-RADS criteria are easy to learn and work at least as well as IOTA.

Acknowledgements The authors thank all staff members and colleagues in Radiology department-Zagazig University for their helpful cooperation.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Dr. Mohammad Abd Alkhalik Basha.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Ethical approval Institutional Review Board approval was obtained.

Informed consent Written informed consent was obtained from all patients.

Statistics and biometry The corresponding author has great statistical expertise.

Methodology
• Prospective
• Diagnostic or prognostic study
• Multicenter study

References

1. Harris RD, Javitt MC, Glanc P et al (2013) ACR appropriateness Criteria® clinically suspected adnexal mass. Ultrasound Q 29:79–86
2. Hall TR, Randall TC (2015) Adnexal masses in the premenopausal patient. Clin Obstet Gynecol 58:47–52
3. Perera DS, Prabhakar HB (2015) Imaging of the adnexal mass. Clin Obstet Gynecol 58:28–46
4. Brown DL, Dudiak KM, Laing FC (2010) Adnexal masses: US characterization and reporting. Radiology 254:342–354
5. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I (2000) Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group. Ultrasound Obstet Gynecol 16:500–505
6. Levine D, Brown DL, Andreotti RF et al (2010) Management of asymptomatic ovarian and other adnexal cysts imaged at US: Society of Radiologists in Ultrasound Consensus Conference Statement. Radiology 256:943–954
7. Amor F, Vaccaro H, Alcázar JL, León M, Craig JM, Martinez J (2009) Gynecologic imaging reporting and data system: a new proposal for classifying adnexal masses on the basis of sonographic findings. J Ultrasound Med 28:285–291
8. Orozco Fernández R, Peces Rama A, Llanos Llanos MC, Martinez Mendoza A, Machado Linde F, Nieto Diaz A (2015) Clinical application of the gynecologic imaging reporting and data system (GI-RADS) for the evaluation of adnexal masses. SM J Gynecol Obstet 1:1009–1012
9. Amor F, Alcázar JL, Vaccaro H et al (2011) GI-RADS reporting system for ultrasound evaluation of adnexal masses in clinical practice: a prospective multicenter study. Ultrasound Obstet Gynecol 38:450–455
10. Zhang T, Li F, Liu J, Zhang S (2017) Diagnostic performance of the Gynecology Imaging Reporting and Data System for malignant adnexal masses. Int J Gynaecol Obstet 137:325–331
11. Alcazar JL, Pascual MA, Graupera B et al (2016) External validation of IOTA simple descriptors and simple rules for classifying adnexal masses. Ultrasound Obstet Gynecol 48:397–402
12. Alcázar JL, Errasti T, Laparte C, Jurado M, López-García G (2001) Assessment of a new logistic model in the preoperative evaluation of adnexal masses. J Ultrasound Med 20:841–848
13. Heintz AP, Odicino F, Maisonneuve P et al (2003) Carcinoma of the ovary. Int J Gynecol Obstet 83:135–166
14. Rams N, Muñoz R, Soler C, Parra J (2015) Resultados de la clasificación Gynecologic Imaging Reporting and Data System para la catalogación de masas anexiales. Prog Obstet Ginecol 58:125–129
15. Koneczny J, Czekierdowski A, Florczak M, Poziemski P, Stachowicz N, Borowski D (2017) The use of sonographic subjective tumor assessment, IOTA logistic regression model 1, IOTA Simple Rules and GI-RADS system in the preoperative prediction of malignancy in women with adnexal masses. Ginekol Pol 88:647–653
16. Moszynski R, Szpurek D, Szubert S, Sajdak S (2013) Analysis of false negative results of subjective ultrasonography assessment of adnexal masses. Ginekol Pol 84:102–107
17. Alcázar JL, Aubá M, Ruiz-Zambrana Á et al (2012) Ultrasound assessment in adnexal masses: an update. Expert Rev Obstet Gynecol 7:441–449
18. Forstner R, Thomassin-Naggara I, Cunha TM et al (2017) ESUR recommendations for MR imaging of the sonographically indeterminate adnexal mass: an update. Eur Radiol 27:2248–2257
19. Timmerman D (2004) The use of mathematical models to evaluate pelvic masses; can they beat an expert operator? Best Pract Res Clin Obstet Gynaecol 18:91–104
20. Moszyński R, Zywica P, Wojtowicz A et al (2014) Menopausal status strongly influences the utility of predictive models in differential diagnosis of ovarian tumors: an external validation of selected diagnostic tools. Ginekol Pol 85:892–899
21. Migda M, Bartosz M, Migda MS, Kierszk M, Katarzyna G, Maleńczyk M (2018) Diagnostic value of the gynecology imaging reporting and data system (GI-RADS) with the ovarian malignancy marker CA-125 in preoperative adnexal tumor assessment. J Ovarian Res 11:92
22. Andreotti RF, Timmerman D, Benacerraf BR et al (2018) Ovarian-adnexal reporting lexicon for ultrasound: a white paper of the ACR Ovarian-Adnexal Reporting and Data System Committee. J Am Coll Radiol 5:1415–1429