ORIGINAL ARTICLE

CrossMark

Ultrasound-based logistic regression model LR2 versus magnetic resonance imaging for discriminating between benign and malignant adnexal masses: a prospective study

Kanane Shimada[1,4] · Koji Matsumoto[1] · Takashi Mimura[1] · Tetsuya Ishikawa[1] · Jiro Munechika[2] ·
Yoshimitsu Ohgiya[2] · Miki Kushima[3] · Yusuke Hirose[1] · Yuka Asami[1] · Chiaki Iitsuka[1] · Shingo Miyamoto[1] ·
Mamiko Onuki[1] · Hajime Tsunoda[4] · Ryu Matsuoka[1] · Kiyotake Ichizuka[1] · Akihiko Sekizawa[1]

Abstract

Background The diagnostic performances of the International Ovarian Tumor Analysis (IOTA) ultrasound-based logistic regression model (LR2) and magnetic resonance imaging (MRI) in discriminating between benign and malignant adnexal masses have not been directly compared in a single study.

Methods Using the IOTA LR2 model and subjective interpretation of MRI findings by experienced radiologists, 265 consecutive patients with adnexal masses were preoperatively evaluated in two hospitals between February 2014 and December 2015. Definitive histological diagnosis of excised tissues was used as a gold standard.

Results From the 265 study subjects, 54 (20.4%) tumors were histologically diagnosed as malignant (including 11 borderline and 3 metastatic tumors). Preoperative diagnoses of malignant tumors showed 91.7% total agreement between IOTA LR2 and MRI, with a kappa value of 0.77 [95% confidence interval (CI), 0.68–0.86]. Sensitivity of IOTA LR2 (0.94, 95% CI, 0.85–0.98) for predicting malignant tumors was similar to that of MRI (0.96, 95% CI, 0.87–0.99; $P = 0.99$), whereas specificity of IOTA LR2 (0.98, 95% CI, 0.95–0.99) was significantly higher than that of MRI (0.91, 95% CI, 0.87–0.95; $P = 0.002$). Combined IOTA LR2 and MRI results gave the greatest sensitivity (1.00, 95% CI, 0.93–1.00) and had similar specificity (0.91, 95% CI, 0.86–0.94) to MRI.

Conclusions The IOTA LR2 model had a similar sensitivity to MRI for discriminating between benign and malignant tumors and a higher specificity compared with MRI. Our findings suggest that the IOTA LR2 model, either alone or in conjunction with MRI, should be included in preoperative evaluation of adnexal masses.

Keywords IOTA · Ultrasound · Magnetic resonance imaging (MRI) · Ovarian tumor

✉ Koji Matsumoto
matsumok@mui.biglobe.ne.jp

1 Department of Obstetrics and Gynecology, Showa University School of Medicine, 1-5-8 Hatanodai, Shinagawa-ku, Tokyo 142-8666, Japan

2 Department of Radiology, Showa University School of Medicine, 1-5-8 Hatanodai, Shinagawa-ku, Tokyo 142-8666, Japan

3 Department of Pathology, Koto Toyosu Hospital, Showa University School of Medicine, 5-1-38 Toyosu, Koto-ku, Tokyo 135-8577, Japan

4 Department of Obstetrics and Gynecology, NTT Medical Center Tokyo, 5-9-22 Higashi-Gotanda, Shinagawa-ku, Tokyo 141-8625, Japan

Introduction

In therapeutic decision making for a patient with an adnexal mass, characterizing the mass as malignant or benign is important because therapeutic approaches differ considerably according to this status. Benign masses can be managed conservatively or with laparoscopic surgery, whereas malignant tumors need referral to a gynecological oncologist for proper staging and debulking surgery. Imaging techniques, such as ultrasonography (US) and magnetic resonance imaging (MRI), have a central role in the preoperative assessment of adnexal tumors when planning surgery for adnexal masses [1]. Although US is accepted as an initial modality for characterizing adnexal masses, the ability to characterize adnexal

masses as benign or malignant with US correlates with the level of the US examiner's experience [2].

To improve the diagnostic performance and reproducibility of US evaluation for adnexal masses, the International Ovarian Tumor Analysis (IOTA) group developed US-based logistic regression models and predictive rules (LR1, LR2, and simple rules) to discriminate malignant from benign adnexal tumors [1, 3]. A total of 47 centers from 17 countries have participated in IOTA validation studies so far [3]. The LR2 protocol, which uses only six variables, has performed very well across validation studies, with a pooled sensitivity and specificity of 0.93 [95% confidence interval (CI), 0.89–0.95] and 0.84 (95% CI, 0.78–0.89), respectively [4]. In addition, good intra- and interobserver agreement was observed [5, 6].

MRI is also useful in an assessment of adnexal masses before treatment. However, the cost of MRI limits its use in initial adnexal tumor screening. In the European Society of Urogenital Radiology (ESUR) guideline, MRI is recommended as a secondary test for US-indeterminate adnexal masses [7]. A systematic review of 18 MRI studies on adnexal tumors showed conventional MRI (1.5 T) to have a pooled sensitivity of 0.92 (95% CI, 0.89–0.94) and a specificity of 0.85 (95% CI, 0.82–0.87) for predicting the malignancy of adnexal masses [8].

The present study was designed to compare diagnostic performances of the IOTA LR2 model and MRI in discriminating between benign and malignant adnexal masses in a Japanese study population. Although data from pooled analyses suggest that the IOTA LR2 model and MRI give comparable results [4, 8], the diagnostic performances of the IOTA LR2 model and MRI have not been directly compared in a single study. We also examined the diagnostic performance of combining IOTA LR2 and MRI results.

Methods

Study design

We prospectively recruited 265 consecutive women patients who visited either of two hospitals (Showa University Hospital or NTT Medical Center Tokyo) between February 2014 and December 2015 for diagnosis and treatment of adnexal masses. Their adnexal masses were preoperatively evaluated using US and MRI.

Inclusion criteria were presentation with at least one adnexal mass, and having undergone an US examination by a principal investigator and MRI imaging at one of participating centers. In the case of bilateral adnexal masses, the final analysis included the mass with the most complex morphology or the largest mass. We excluded patients who were pregnant; who refused to undergo transvaginal US or

MRI; or who failed to undergo surgical removal of the mass within 120 days of the US examination.

Before each US examination, personal and family histories of ovarian or breast cancer were taken by the US examiner. Demographic data collected from each patient included the patient's age, menopausal status, days of menstrual cycle (if appropriate), previous hormonal therapy, and surgical history. Women aged 50 years or more who had undergone hysterectomies were defined as postmenopausal.

The institutional ethical and research review boards of Showa University Hospital and NTT Medical Center Tokyo approved the study protocol. All patients entered the study only after voluntarily giving signed informed consent.

Ultrasound examination

We used the IOTA logistic regression model (LR2) for US evaluation of adnexal masses. In all cases, transvaginal US examinations were performed by two gynecologists [the Japan Society of Ultrasonics in Medicine (JSUM) certified sonologist T.M. at Showa University Hospital and non-certified operator K.S. at NTT Medical Center Tokyo], before MRI. The ProSound α7 (Hitachi Aloka Medical, Tokyo, Japan) or Voluson P8 (GE Healthcare Ultrasound, Milwaukee, WI, USA) ultrasound machines were used with transvaginal probe frequencies ranging between 5 and 12 MHz for standardized gray-scale US examinations. Transabdominal US was used to examine large masses that could not be seen in their entirety using transvaginal probes. Operators also used color Doppler US to obtain blood flow information to characterize adnexal masses. The US information was prospectively recorded and not changed after surgery. Serum CA125 levels were not available at the time of the US examinations.

The following six variables were used for the calculation: (1) patient's age in years; (2) presence of ascites (yes = 1, no = 0); (3) presence of blood flow within a solid papillary (yes = 1, no = 0); (4) maximal diameter of the solid component [expressed in millimeters (mm), but with no increase >50 mm]; (5) irregular internal cyst walls (yes = 1, no = 0); and (6) presence of acoustic shadows (yes = 1, no = 0). The estimated probability of malignancy (POM) for an adnexal tumor equaled $1/(1 + e^{-Z})$, where $Z = -5.3718 + 0.0354 \times (1) + 1.6159 \times (2) + 1.1768 \times (3) + 0.0697 \times (4) + 0.9586 \times (5) - 2.9486 \times (6)$, as described in the original IOTA study [9]. On the basis of the IOTA Phase I study, we set 0.1 of the POM as the cutoff value: i.e., an adnexal mass was regarded as malignant when POM > 0.1 [9].

MRI examination

Preoperative diagnoses based on MRI data were made according to subjective assessments of radiologists, who

were blinded to the IOTA results. All MRI examinations were performed using a 1.5-T MR scanner (Magnetom Avanto; Siemens Medical Solutions, Erlangen, Germany) with a phase-array body coil. All images were obtained with the parallel imaging technique. The imaging protocol included axial and sagittal T_2-weighted fast-spin echo [FSE: repetition time (TR)/echo time (TE), 4000/97 ms; flip angle (FA), 170°], axial T_1-weighted gradient recall echo (GRE) (TR/TE, 7.46/2.39 ms; FA, 15°), and diffusion-weighted images (DWI: TR/TE, 3600/87 ms; FA, 90°). The MRI sequences were performed with slice thicknesses of 3.5–5 mm, intersectional gaps of 1.0 mm, matrices of 128 × 128 to 220 × 220, and fields of view (FOV) of 25–35 cm. Dynamic contrast-enhanced MRIs were obtained with axial fat-saturated T_1-weighted GRE imaging (TR/TE, 4.81/1.77 ms; FA, 15°) before and 30, 60, or 120 s after intravenous bolus administration of contrast media using 0.1 mmol/kg meglumine gadoterate (Magnescope; Guerbet Japan, Tokyo, Japan). For complex cystic or hemorrhagic masses, subtracted dynamic contrast-enhanced T_1-weighted imaging was used to detect possible solid or mural nodules. Additional sequences were added depending on indication. According to a diagnostic algorithm recommended by the ESUR guideline [7], the Japan Radiological Society (JRS) certified radiologists (J.M. and Y.O. at Showa University Hospital and two radiologists at NTT Medical Center Tokyo) gave subjective assessments of whether each adnexal mass was likely to be malignant or benign. Each patient's MRI was read by one or two radiologists. When disagreement occurred between two radiologists, the final MRI diagnoses were based on their discussion.

Outcome measures

The final outcome measure of the present study was the histological diagnosis. Surgery was performed by laparoscopy or laparotomy, according to the surgeon's judgment. The excised specimens were examined histologically at each hospital, and classified according to the classification criteria of the World Health Organization (WHO) [10]. Definitive histological diagnosis of excised tissues was used as the gold standard.

Statistical analysis

The present study assessed the diagnostic performance of the IOTA LR2 model and MRI subjective assessment in characterizing preoperatively whether adnexal masses were malignant or benign. Diagnostic accuracy of the IOTA LR2 and MRI was assessed by calculating sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), with 95% Wilson score confidence intervals. Using analyses of agreement (kappa coefficient and percent total agreement), the US IOTA L2 model was compared with MRI subjective assessment for the prediction of malignant tumors. Agreement by kappa values was considered poor at 0.0–0.20; fair at 0.21–0.40; moderate at 0.41–0.60; good at 0.61–0.80; and very good at 0.81–1.00. McNemar's exact χ^2 test was used to determine whether differences in diagnosis between the two methods or institutions were statistically significant. We also examined the diagnostic performance of combining IOTA LR2 and MRI results. In the analysis of combined results, discrepant cases were regarded as malignant. In univariate analyses of the six variables included in the IOTA LR2 model, the Mann–Whitney U test was used to compare continuous variables between benign and malignant tumors, and Pearson's χ^2 test was used for 2 × 2 contingency table analyses. All analyses were carried out using R version 3.3.2 (the R Foundation for Statistical Computing, Vienna, Austria) at a statistical laboratory (P4 Statistics, Tokyo, Japan). Two-sided P values were calculated, and $P < 0.05$ was considered significant.

Results

Patient characteristics

Imaging data from 265 women with adnexal masses were analyzed with histological diagnoses. Their histological profiles are summarized in Table 1. Overall, 211 (79.6%) adnexal masses were histologically diagnosed as benign, and 54 (20.4%) tumors were classified as malignant (including 11 borderline and 3 metastatic tumors). Malignant tumors were seen in 16.1% (31/192) of patients treated at Showa University Hospital and in 31.5% (23/73) at NTT Medical Center Tokyo.

Validation of IOTA LR2 model

Data of the six variables included in the IOTA LR2 model are shown in Table 2. Women with malignant tumors were significantly older than those with benign tumors (median age: 55 years vs. 39 years, $P < 0.001$). Ultrasound-based variables were also strongly associated with the nature of adnexal masses. Maximal diameter of the solid component was significantly greater in malignancies than in benign tumors (median diameter: 38 mm vs. 0 mm, $P < 0.001$). Risk of malignancy was strongly associated with presence of ascites [odds ratio (OR), 9.76, 95% confidential interval (CI), 3.47–27.5, $P < 0.001$], papillations with blood flow (OR, 283.0, 95% CI, 36.9–2171, $P < 0.001$) and irregular cyst walls (OR, 35.5, 95% CI, 12.6–100.1, $P < 0.001$). By contrast, acoustic shadows were significantly associated with benign nature in adnexal masses (OR, 0.08, 95% CI, 0.01–0.58, $P = 0.01$). Accordingly, POM calculated from

Table 1 Histological findings of adnexal tumors by participating centers

Histology	All (n = 265)	By center	
		Showa University Hospital (n = 192)	NTT Medical Center Tokyo (n = 73)
Benign	211(79.6%)	161 (83.9%)	50 (68.5%)
Endometrioma	90 (34.0%)	78 (40.6%)	12 (16.4%)
Serous cystadenoma	17 (6.4%)	12 (6.2%)	5 (6.8%)
Mucinous cystadenoma	17 (6.4%)	9 (4.7%)	8 (11.0%)
Teratoma	69 (26.0%)	51 (26.6%)	18 (24.7%)
Fibroma	4 (1.5%)	2 (1.0%)	2 (2.7%)
Functional cyst	8 (3.0%)	3 (1.6%)	5 (6.8%)
Abscess	0 (0.0%)	0 (0.0%)	0 (0.0%)
Hydrosalpinx	0 (0.0%)	0 (0.0%)	0 (0.0%)
Parasalpingeal cyst	4 (1.5%)	4 (2.1%)	0 (0.0%)
Rare benign	2 (0.8%)	2 (1.0%)	0 (0.0%)
Peritoneal pseudocyst	0 (0.0%)	0 (0.0%)	0 (0.0%)
Malignant	54 (20.4%)	31 (16.1%)	23 (31.5%)
Primary invasive	40 (15.1%)	26 (13.5%)	14 (19.2%)
Serous adenocarcinoma	9 (3.4%)	6 (3.1%)	3 (4.1%)
Endometrioid adenocarcinoma	16 (6.0%)	9 (4.7%)	7 (9.6%)
Clear cell adenocarcinoma	10 (3.8%)	7 (3.6%)	3 (4.1%)
Mucinous adenocarcinoma	3 (1.1%)	2 (1.0%)	1 (1.4%)
Others	2 (0.8%)	2 (1.0%)	0 (0.0%)
Borderline	11 (4.2%)	4 (2.1%)	7 (9.6%)
Serous borderline tumor	1 (0.4%)	0 (0.0%)	1 (1.4%)
Mucinous borderline tumor	5 (1.9%)	3 (1.6%)	2 (2.7%)
Other	5 (1.9%)	1 (0.5%)	4 (5.5%)
Metastatic	3 (1.1%)	1 (0.5%)	2 (2.7%)

Table 2 Univariate analyses of six variables included in IOTA LR2 model

Variables included in IOTA LR2 modeling	All n = 265	Benign n = 211	Malignant n = 54	P
Median age in years (range)	41 (19–88)	39 (19–87)	55 (31–88)	<0.001
10–19 years	1 (0.4%)	1 (0.5%)	0 (0.0%)	
20–29 years	42 (15.8%)	42 (19.9%)	0 (0.0%)	
30–39 years	75 (28.3%)	66 (31.3%)	9 (16.7%)	
40–49 years	78 (29.4%)	65 (30.8%)	13 (24.1%)	
50–59 years	29 (10.9%)	13 (6.2%)	16 (29.6%)	
60–69 years	19 (7.2%)	10 (4.7%)	9 (16.7%)	
70+ years	21 (7.9%)	14 (6.6%)	7 (13.0%)	
Median maximal diameter of solid component in mm (range)	0 (0–65)	0 (0–65)	38 (0–50)	<0.001
Ascites	18 (6.8%)	6 (2.8%)	12 (22.2%)	<0.001
Papillations with blood flow	32 (12.1%)	1 (0.5%)	31 (57.4%)	<0.001
Irregular cyst walls	30 (11.3%)	5 (2.4%)	25 (46.3%)	<0.001
Acoustic shadows	42 (15.8%)	41 (19.4%)	1 (1.9%)	0.003

Data shown as n (%) unless otherwise described

these six variables was higher in malignant tumors than in benign tumors [mean (± standard deviation), 0.56 (± 0.32) vs. 0.03 (± 0.03), $P < 0.001$]. When the receiver operating characteristic (ROC) curve was constructed, the area

of the curve (AUC) was 0.99 (95% CI, 0.98–1.00). Using a cutoff of 0.10 (original cutoff derived from IOTA phase I study) to predict malignancy, IOTA LR2 had a sensitivity of 0.94 (95% CI, 0.85–0.98), specificity of 0.98 (95% CI, 0.95–0.99), PPV of 0.91 (95% CI, 0.81–0.96), and NPV of 0.99 (95% CI, 0.96–1.00; Table 3).

Of 265 study subjects, the IOTA LR2 modeling classified only 8 (3.0%) adnexal tumors incorrectly: 5 benign tumors as malignant and 3 primary invasive tumors as benign. Histology of these tumors included endometriotic cyst ($n = 2$), mucinous cystadenoma ($n = 1$), thecoma ($n = 1$), struma ovarii ($n = 1$), endometrioid adenocarcinoma ($n = 2$), and mucinous adenocarcinoma ($n = 1$).

MRI results

Diagnostic performance of MRI based on expert radiologists' subjective assessment had a sensitivity of 0.96 (95% CI, 0.87–0.99) for discriminating malignant from benign adnexal tumors, specificity of 0.91 (95% CI, 0.87–0.95),

PPV of 0.74 (95% CI, 0.63–0.83), and NPV of 0.99 (95% CI, 0.96–1.00; Table 3).

Of 265 subjects, MRI classified 20 (7.5%) adnexal tumors incorrectly: 18 benign tumors as malignant and 2 malignant tumors as benign, including endometriotic cyst ($n = 4$), mucinous cystadenoma ($n = 9$), serous cystadenoma ($n = 4$), struma ovarii ($n = 1$), endometrioid adenocarcinoma ($n = 1$), and mucinous borderline tumor ($n = 1$).

Comparison of IOTA LR2 and MRI

Preoperative diagnoses of malignant tumors had good agreement between IOTA LR2 and MRI, with 91.7% total agreement and a kappa value of 0.77 (95% CI, 0.68–0.86). In comparison of test performance characteristics, sensitivity of IOTA L2 (0.94; 95% CI, 0.85–0.98) for predicting malignant tumors was similar to that of MRI (0.96; 95% CI, 0.87–0.99; $P = 0.99$, exact McNemar's test; Table 3). However, the specificity of IOTA LR2 (0.98; 95% CI, 0.95–0.99) was significantly greater than for MRI (0.91; 95% CI, 0.87–0.95; $P = 0.002$, exact McNemar's test). Accuracy of the IOTA

Table 3 Comparison of diagnostic performances of IOTA LR2 model and MRI subjective assessment

Imaging method	Correctly classified		Incorrectly classified		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
	Malignant	Benign	Malignant as benign	Benign as malignant					
Ultrasound-based IOTA LR2 model									
All	51	206	3	5	0.94 (0.85–0.98)	0.98 (0.95–0.99)	0.91 (0.81–0.96)	0.99 (0.96–1.00)	0.97 (0.94–0.98)
Showa University Hospital	29	156	2	5	0.94 (0.79–0.98)	0.97 (0.93–0.99)	0.85 (0.70–0.94)	0.99 (0.96–1.00)	0.96 (0.93–0.98)
NTT Medical Center Tokyo	22	50	1	0	0.96 (0.79–1.00)	1.00 (0.93–1.00)	1.00 (0.85–1.00)	0.98 (0.90–1.00)	0.99 (0.93–1.00)
MRI subjective assessment									
All	52	193	2	18	0.96 (0.87–0.99)	0.91 (0.87–0.95)	0.74 (0.63–0.83)	0.99 (0.96–1.00)	0.92 (0.89–0.95)
Showa University Hospital	29	147	2	14	0.94 (0.79–0.98)	0.91 (0.86–0.95)	0.67 (0.53–0.80)	0.99 (0.95–1.00)	0.92 (0.87–0.95)
NTT Medical Center Tokyo	23	46	0	4	1.00 (0.86–1.00)	0.92 (0.81–0.97)	0.85 (0.68–0.94)	1.00 (0.92–1.00)	0.95 (0.87–0.98)
Combination of IOTA and MRI[a]									
All	54	191	0	20	1.00 (0.93–1.00)	0.91 (0.86–0.94)	0.73 (0.62–0.82)	1.00 (0.98–1.00)	0.92 (0.89–0.95)
Showa University Hospital	31	145	0	16	1.00 (0.89–1.00)	0.90 (0.84–0.94)	0.66 (0.52–0.78)	1.00 (0.97–1.00)	0.92 (0.87–0.95)
NTT Medical Center Tokyo	23	46	0	4	1.00 (0.86–1.00)	0.92 (0.81–0.97)	0.85 (0.68–0.94)	1.00 (0.92–1.00)	0.95 (0.87–0.98)

[a]Cases of disagreement were classified as malignant

LR2 model (0.97; 95% CI, 0.94–0.98) was also greater than that of the MRI subjective assessment (0.92; 95% CI, 0.89–0.95). When the IOTA LR2 results were combined with the MRI subjective assessment (with cases of disagreement regarded as malignant), sensitivity increased to 1.00 (95% CI, 0.93–1.00), although specificity decreased to 0.91 (95% CI, 0.86–0.94).

Table 4 shows histologically specific agreement between the IOTA LR2 model and MRI subjective assessment. Good agreement between IOTA LR2 and MRI was observed for endometriotic cysts (total agreement, 95.6%), mature cystic teratomas (100.0%), and malignant tumors (90.7%). However, MRI subjective assessment was more likely to classify benign mucinous tumors as malignant (total agreement, 52.9%). Three benign tumors were classified incorrectly as malignant by both IOTA LR2 and MRI: endometriotic cyst ($n = 1$), mucinous cystadenoma ($n = 1$), and struma ovarii ($n = 1$). However, no malignant adnexal tumors were classified incorrectly as benign by either imaging method.

Imaging data were obtained independently from the two centers by different investigators. Similar patterns between the diagnostic performances of IOTA LR2 and MRI were observed at the two centers, which suggests that our results can be generalized (Table 3).

Discussion

We confirmed that the IOTA LR2 model worked very well in a Japanese population. In the present study, US data were collected by an expert operator in a university hospital, and by a non-expert operator in a medical center. Additionally, the prevalence of malignant tumors differed considerably between the two institutions. However, the test performance of the IOTA regression model LR2 was similar between two providers; and diagnostic performance did not significantly differ between the expert and non-expert US operators. Furthermore, the diagnostic performance for the IOTA LR2 model was comparable to that of the subjective assessment by expert US examiners in the IOTA collaboration study of 3511 adnexal masses (sensitivity, 0.94 vs. 0.91; specificity, 0.98 vs. 0.96) [11]. These observations suggest that the IOTA LR2 protocol is generally applicable across populations and institutions to assist gynecologists with varied training backgrounds and levels of expertise.

The present study showed good agreement between IOTA LR2 and MRI subjective assessment. The sensitivity of IOTA LR2 was similar to that of MRI. However, the specificity was significantly higher in IOTA LR2, suggesting that IOTA LR2 modeling provides fewer false-positive cases (benign tumors classified as malignant) than does MRI. Among six histological categories of endometriomas, teratomas, benign mucinous tumors, other benign tumors, primary

Table 4 Diagnostic agreement between IOTA LR2 and MRI by adnexal tumor histology

Histology	n	Discriminating malignant from benign masses				Agreement analysis				Agreement (%)
		IOTA LR2 model		MRI subjective assessment		IOTA correct		IOTA incorrect		
		Correctly classified	Incorrectly classified	Correctly classified	Incorrectly classified	MRI correct	MRI incorrect	MRI correct	MRI incorrect	
All	265	257	8	245	20	240	17	5	3	91.7%
Benign	211	206	5	193	18	191	15	2	3	91.9%
Endometrioma	90	88	2	86	4	85	3	1	1	95.6%
Teratoma	69	69	0	69	0	69	0	0	0	100.0%
Mucinous	17	16	1	8	9	8	8	0	1	52.9%
Others	35	33	2	30	5	29	4	1	1	85.7%
Malignant	54	51	3	52	2	49	2	3	0	90.7%
Primary invasive	40	37	3	39	1	36	1	3	0	90.0%
Borderline	11	11	0	10	1	10	1	0	0	90.9%
Metastatic	3	3	0	3	0	3	0	0	0	100.0%

invasive tumors, borderline tumors, and metastatic tumors, the % total agreement between IOTA LR2 and MRI was the lowest for benign mucinous tumors (only 52.9%). The IOTA LR2 modeling had no histologically specific weak points in discriminating between benign and malignant tumors and was more than 90% accurate for each histological category. However, MRI subjective assessments were more likely to incorrectly classify benign mucinous tumors as borderline. Although several MRI studies have addressed characteristic morphological features of mucinous borderline tumors, few data are available on MRI findings that can reliably differentiate borderline from benign mucinous tumors [12, 13]. As MRI can facilitate differentiation of the components (blood, fat, simple fluids, and solid tissues) of adnexal masses, MRI could reliably diagnose endometriomas and fat-containing cystic teratomas as benign. When mucinous tumors were excluded from analysis, very good agreement was observed between IOTA LR2 and MRI, with 94.4% total agreement and a kappa value of 0.84 (95% CI, 0.76–0.92).

In the current study, the IOTA LR2 modeling classified 8 (3.0%) adnexal tumors incorrectly. A collaborative analysis of IOTA studies reported that only a small portion (approximately 7%) of adnexal masses cannot be confidently classified as benign or malignant presurgically, even when using subjective assessment by experienced US examiners or logistic regression models [11]. Serum CA125 and human epididymis protein 4 were evaluated as predictive biomarkers to discriminate between malignant and benign adnexal tumors, but this choice did not improve the performance of IOTA logistic regression models in discriminating between benign and malignant tumors [14].

Adding MRI to the IOTA LR2 model may improve preoperative assessment of adnexal tumors. In the present study, three malignant tumors were classified incorrectly as benign by IOTA LR2 but correctly as malignant by MRI. Of these, two tumors were early-stage endometrioid adenocarcinoma arising from endometrioma. The IOTA LR2 model may work less well for malignant transformation in benign cysts (endometrioma and mature teratoma, etc.) because the LR2 calculation protocol is likely to underestimate the risk of malignancy caused by the benign parts of the tumors. MRI may serve as a secondary test to classify these difficult cases. In the present study, the greatest sensitivity was reached by combining IOTA LR2 and MRI results; no malignant adnexal tumors were classified incorrectly as benign by either method. Because intraoperative rupture of a stage I ovarian cancer during laparoscopic surgery increases the risk of recurrence, maximized diagnostic test sensitivity is arguably more important than specificity in preoperative assessment of adnexal tumors. To minimize the risk of misclassifying malignant tumors (classified incorrectly as benign), algorithms that use both imaging methods may be recommended for preoperative evaluation of adnexal masses, although this may slightly increase the number of benign tumors removed with laparotomy because of reduced specificity.

The present study had several limitations. First, the main limitations were its small numbers of subjects, participating institutions, US examiners, and radiologists. Although the LR2 showed significantly greater specificity than did MRI in this study, these limitations may have affected the diagnostic data of both modalities. Second, differences in the prevalence of ovarian histology between present and previous studies may have influenced the results because some benign pathology and advanced cancers are relatively easy to characterize. In the present study, the majority of benign tumors consisted of teratomas and endometriomas, which are usually easy for accurate diagnosis. Third, the lack of central pathological review in the present study might have led to misclassified histology of adnexal tumors. Because of the relevance to clinical practice, however, we did not review histological specimens that were used as a gold standard for the final outcome measure.

In conclusion, our data indicate that the IOTA LR2 model is as sensitive as MRI subjective assessment in discriminating between a malignant and benign tumor, and has a higher specificity compared with MRI, with the greatest sensitivity obtained by combining IOTA LR2 and MRI results. These findings recommend the addition of the IOTA LR2 model, either alone or in conjunction with MRI, to preoperative evaluation of adnexal masses. However, the current study used a rather small cohort. To confirm our results, further prospective multicenter studies with larger cohorts are warranted.

Compliance with ethical standards

Conflict of interest All authors report no conflict of any financial interest.

References

1. Kaijser J, Vandecaveye V, Deroose CM et al (2014) Imaging techniques for the pre-surgical diagnosis of adnexal tumours. Best Pract Res Clin Obstet Gynaecol 28:683–695
2. Timmerman D, Schwärzler P, Collins WP et al (1999) Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. Ultrasound Obstet Gynecol 13:11–16
3. Kaijser J (2015) Towards an evidence-based approach for diagnosis and management of adnexal masses: findings of the International Ovarian Tumour Analysis (IOTA) studies. Facts Views Vis Obgyn 7:42–59

4. Meys EM, Kaijser J, Kruitwagen RF et al (2016) Subjective assessment versus ultrasound models to diagnose ovarian cancer: a systematic review and meta-analysis. Eur J Cancer 58:17–29

5. Sladkevicius P, Valentin L (2015) Interobserver agreement in describing the ultrasound appearance of adnexal masses and in calculating the risk of malignancy using logistic regression models. Clin Cancer Res 21:594–601

6. Zannoni L, Savelli L, Jokubkiene L et al (2014) Intra- and interobserver agreement with regard to describing adnexal masses using International Ovarian Tumor Analysis terminology: reproducibility study involving seven observers. Ultrasound Obstet Gynecol 44:100–108

7. Spencer JA, Forstner R, Cunha TM et al (2010) ESUR guidelines for MR imaging of the sonographically indeterminate adnexal mass: an algorithmic approach. Eur Radiol 20:25–35

8. Medeiros LR, Freitas LB, Rosa DD et al (2011) Accuracy of magnetic resonance imaging in ovarian tumor: a systematic quantitative review. Am J Obstet Gynecol 204:67 (e1–10)

9. Timmerman D, Testa AC, Bourne T et al (2005) Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group. J Clin Oncol 23:8794–8801

10. Tavassoli FA, Devilee P (eds) (2003) World Health Organization Classification of Tumors. Pathology and Genetics. Tumors of the breast and female genital organs. IARC Press, Lyon, pp 117–145

11. Valentin L, Ameye L, Savelli L et al (2011) Adnexal masses difficult to classify as benign or malignant using subjective assessment of gray-scale and Doppler ultrasound findings: logistic regression models do not help. Ultrasound Obstet Gynecol 38:456–465

12. Okamoto Y, Tanaka YO, Tsunoda H et al (2007) Malignant or borderline mucinous cystic neoplasms have a larger number of loculi than mucinous cystadenoma: a retrospective study with MR. J Magn Reson Imaging 26:94–99

13. Zhao SH, Qiang JW, Zhang GF et al (2014) MRI in differentiating ovarian borderline from benign mucinous cystadenoma: pathological correlation. J Magn Reson Imaging 39:162–166

14. Kaijser J, Van Gorp T, Van Hoorde K et al (2013) A comparison between an ultrasound based prediction model (LR2) and the risk of ovarian malignancy algorithm (ROMA) to assess the risk of malignancy in women with an adnexal mass. Gynecol Oncol 129:377–383