

# Diagnostic accuracy of transvaginal ultrasound examination for assigning a specific diagnosis to adnexal masses

A. SOKALSKA\*, D. TIMMERMAN†, A. C. TESTA‡, C. VAN HOLSBEKE†, A. A. LISSONI§, F. P. G. LEONE¶, D. JURKOVIC\*\* and L. VALENTIN‡

\*Division of Infertility and Reproductive Endocrinology, Department of Gynecology, Obstetrics and Gynecological Oncology, Karol Marcinkowski University of Medical Sciences, Poznan, Poland, †Department of Obstetrics and Gynecology, University Hospitals KU Leuven, Leuven, Belgium, ‡Istituto di Clinica Ostetrica e Ginecologica, Università Cattolica del Sacro Cuore, Rome, §Clinica Ostetrica e Ginecologica, Ospedale S. Gerardo, Università di Milano Bicocca, Monza and ¶Clinical Sciences Institute L. Sacco, University of Milan, Milan, Italy, \*\*Department of Obstetrics and Gynaecology, King's College Hospital, London, UK and ‡Department of Obstetrics and Gynecology, Malmö University Hospital, Lund University, Malmö, Sweden

**KEYWORDS:** ovarian neoplasm; ultrasonography

## ABSTRACT

**Objectives** To determine the sensitivity and specificity of subjective evaluation of gray-scale and Doppler ultrasound findings (here called pattern recognition) when used by experienced ultrasound examiners with regard to making a specific diagnosis of adnexal masses.

**Methods** Within the framework of a European multi-center study, the International Ovarian Tumor Analysis study, comprising nine ultrasound centers, women with at least one adnexal mass were examined with gray-scale and color Doppler ultrasonography by experienced ultrasound examiners. A standardized examination technique, and standardized terms and definitions were used. Using pattern recognition the examiners classified each mass as benign or malignant and suggested a specific diagnosis (e.g. dermoid cyst or endometrioma). The reference standard was the histology of the surgically removed adnexal tumors.

**Results** A total of 1066 women were included, of whom 800 had a benign mass and 266 a malignant mass. A specific diagnosis based on ultrasound findings was suggested in 899 (84%) tumors. The specificity was high for all diagnoses (range, 94–100%). The sensitivity was highest for benign teratoma/dermoid cysts (86%, 100/116), hydrosalpinges (86%, 18/21), peritoneal pseudocysts (80%, 4/5) and endometriomas (77%, 153/199), and lowest for functional cysts (17%, 4/24), paraovarian/parasalpingeal cysts (14%, 3/21), benign rare tumors (11%, 1/9), adenofibromas (8%, 3/39), simple cysts (6%, 1/18) and struma ovarii (0%, 0/5). The positive and negative likelihood ratios

of pattern recognition with regard to dermoid cysts, hydrosalpinges and endometriomas were 68.2 and 0.14, 38.9 and 0.15, and 33.3 and 0.24, respectively. Dermoid cysts, hydrosalpinges, functional cysts, paraovarian cysts, peritoneal pseudocysts, fibromas/fibrothecomas and simple cysts were never misdiagnosed as malignancies by the ultrasound examiner, whereas more than 10% of inflammatory processes, adenofibromas and rare benign tumors including struma ovarii were misdiagnosed as malignancies.

**Conclusions** Using subjective evaluation of gray-scale and Doppler ultrasound findings it is possible to make an almost conclusive diagnosis of a dermoid cyst, endometrioma and hydrosalpinx. Many other adnexal pathologies can be recognized but not confidently confirmed or excluded. Copyright © 2009 ISUOG. Published by John Wiley & Sons, Ltd.

## INTRODUCTION

An adnexal tumor may be found in women presenting with gynecological complaints or it may be an incidental finding. The finding of an adnexal mass usually raises anxiety because of the possibility of malignancy. Imaging methods, particularly ultrasound imaging, are almost always used to determine the nature of a mass<sup>1</sup>. In the hands of an experienced ultrasound examiner subjective evaluation of ultrasound findings, i.e. pattern recognition, is an excellent method for discrimination between benign and malignant masses<sup>2,3</sup>, but pattern recognition may also

Correspondence to: Prof. L. Valentin, Department of Obstetrics and Gynecology, Malmö University Hospital, SE 205 02 Malmö, Sweden (e-mail: lil.valentin@med.lu.se)

Accepted: 6 March 2009

be used to make a specific diagnosis, for example dermoid cyst, endometrioma or hydrosalpinx<sup>4</sup>. In the original publication in which the term pattern recognition was first mentioned, it was used to describe subjective evaluation of gray-scale ultrasound findings<sup>4</sup>. However, the term has now been extended to include subjective evaluation of Doppler ultrasound findings<sup>3</sup>, even though it has been shown that subjective evaluation of Doppler ultrasound findings adds very little to subjective evaluation of gray-scale ultrasound findings<sup>4</sup>.

A correct indication of the nature of an adnexal mass is important for choosing appropriate treatment. Some masses are probably best treated expectantly, if they do not cause any symptoms (e.g. functional cysts) and are not associated with reproductive dysfunction (e.g. hydrosalpinx and uterine myoma), some are possibly best treated expectantly if they do not cause symptoms (e.g. small dermoid cysts<sup>5</sup>), others might be treated by cyst puncture (e.g. peritoneal cysts<sup>6</sup>), and others by surgery (e.g. borderline tumors and invasive malignancies).

The aim of this study was to determine the sensitivity and specificity of subjective evaluation of gray-scale and Doppler ultrasound findings (here called pattern recognition) when used by experienced ultrasound examiners with regard to making a specific diagnosis in adnexal masses, and to determine which histological diagnoses are most likely to be confused with each other at ultrasound examination.

## METHODS

We used the information in the database of the International Ovarian Tumor Analysis (IOTA) study Phase 1. Phase 1 of the IOTA study is a multicenter study including nine European ultrasound centers. It was approved by the local ethics committees and has been described in detail in another publication<sup>7</sup>. All participants gave written informed consent to participate. The design of the study is outlined only briefly below.

Women with at least one adnexal mass were recruited into the study from 1 June 1999 to 30 June 2002. They were examined with gray-scale and color Doppler ultrasonography by experienced ultrasound examiners who used high-end ultrasound systems equipped with high-frequency vaginal probes, a standardized examination technique, and standardized terms and definitions<sup>8</sup>. About 80% of the examinations were carried out by examiners defined as Level III examiners using the terminology of the European Federation of Ultrasound in Medicine and Biology, i.e. the examiners worked in tertiary referral centers, had an academic record, and a high level of experience and expertise<sup>9</sup>. Moreover, all examiners had a special interest in ultrasound diagnosis of adnexal masses, and some had performed up to 16 000 gynecological ultrasound examinations by the start of the study.

All women underwent a transvaginal scan; transabdominal sonography was added when large masses could not be fully visualized via the transvaginal route. On the basis of subjective evaluation of gray-scale and color

Doppler findings (pattern recognition<sup>4,10</sup>) the ultrasound examiner classified each mass as: certainly benign, probably benign, difficult to classify as benign or malignant (complete uncertainty), probably malignant or certainly malignant. Even when the examiner found the mass difficult to classify, he/she was obliged to state whether the mass was more likely to be benign or malignant. On the basis of published descriptions of typical gray-scale and color Doppler ultrasound findings of various specific diagnoses<sup>4,10</sup> the examiner also suggested a specific histological diagnosis (e.g. endometrioma, dermoid cyst or hydrosalpinx) whenever this was possible. The ultrasound examiner had no knowledge of the patient's serum CA 125 values when suggesting a diagnosis but was aware of the patient history.

The reference standard was the histology of the surgically removed adnexal tumors, the tumors being classified and malignant tumors staged according to the criteria recommended by the International Federation of Gynecology and Obstetrics<sup>11</sup>. Both the ultrasound diagnoses and the histological diagnoses were grouped retrospectively into 22 categories (Table 1). For subjective assessment of the mass on the basis of ultrasound findings,

Table 1 Histopathological diagnoses (*n* = 1066)

Diagnosis	n (%)
Benign	
Endometrioma	199 (18.7)
Serous cyst/cystadenoma	149 (14.0)
Teratoma	116 (10.9)
Mucinous cyst/cystadenoma	86 (8.1)
Adenofibroma	39 (3.7)
Two or more diagnoses*	35 (3.3)
Functional cyst	24 (2.2)
Hydrosalpinx, hematosalpinx	21 (2.0)
Paraovarian/parasalpingeal cyst	21 (2.0)
Inflammatory process†	20 (1.9)
Fibroma/fibrothecoma	19 (1.8)
Simple cyst	18 (1.7)
Other‡	13 (1.2)
Leiomyoma	12 (1.1)
Rare benign tumor (excluding struma ovarii)§	9 (0.8)
Torsion of lesion	9 (0.8)
Peritoneal pseudocyst	5 (0.5)
Struma ovarii	5 (0.5)
Malignant	
Primary invasive tumor	144 (13.5)
Stage 1	42
Stage 2, 3 or 4	102
Borderline tumor	55 (5.1)
Serous	24
Mucinous	25
Other	6
Metastatic tumor	42 (3.9)
Rare malignant tumor¶	25 (2.3)

\*E.g. endometrioma or other ovarian lesion plus a hemorrhagic corpus luteum cyst, or ovarian plus tubal pathology, or two neoplasms in the same ovary (e.g. dermoid cyst plus mucinous cystadenoma). †Including abscess and pyosalpinx. ‡E.g. tuberculous granuloma, granulation tissue, vestigial cyst, mucinous histiocytoma. §E.g. Brenner tumor. ¶E.g. granulosa cell tumor or dysgerminoma.

in addition to these 22 categories, a 23<sup>rd</sup> category was added, i.e. 'don't know'. Only patients operated on  $\leq 120$  days after the ultrasound examination were included and, in the case of bilateral or multiple masses in the same patient, only the most complex mass – or, if all masses had similar ultrasound morphology, the largest one or the one most easily accessible by ultrasound examination – was used in our analysis.

In this analysis, a woman was considered to be postmenopausal if she reported no menstruation for at least 1 year after the age of 40 years, provided that the amenorrhea was not explained by pregnancy, medication or disease. Women who were 50 years or older and had undergone a hysterectomy, so that the time of menopause could not be determined, were also defined as postmenopausal.

Statistical analyses were carried out using the Statview 4.5<sup>TM</sup> statistical program (Abacus Concepts, Inc., Berkeley, CA, USA). The diagnostic performance of pattern recognition was expressed as the accuracy, sensitivity, specificity, and positive and negative likelihood ratios. The 95% CIs of sensitivity, specificity, and likelihood ratios were calculated using the Evidence-Based Medicine (EBM) Calculator Version 1.2 (<http://www.cebm.utoronto.ca/palm/ebmcalc/>).

## RESULTS

The IOTA Phase 1 database contains 1066 patients with an adnexal mass: 639 (60%) premenopausal patients (mean age, 37 (range, 17–59) years) and 427 (40%) postmenopausal patients (mean age, 63 (range, 41–94) years). There were 800 benign tumors and 266 (25%) malignant tumors: 15% (98/639) of the premenopausal patients had a malignant tumor vs. 39% (168/427) of the postmenopausal patients. Bilateral tumors were found in 20% (216/1066) of the patients. The histological diagnoses of the 1066 tumors included are shown in Table 1.

### Prediction of malignancy

Pattern recognition correctly predicted malignancy in 97% (95% CI, 93–99%) (140/144) of the primary invasive malignancies (93% of Stage 1 primary invasive tumors vs. 99% of Stage 2–4 primary invasive tumors), in 93% (95% CI, 80–97%) (39/42) of the metastatic tumors, in 88% (95% CI, 69–95%) (22/25) of the rare malignancies and in 60% (95% CI, 47–72%) (33/55) of the borderline tumors.

### Sensitivity and specificity with regard to specific diagnoses

A specific diagnosis based on ultrasound findings was suggested in 899/1066 (84%) tumors. In the remaining 167 (16%) tumors the ultrasound examiner stated that it was not possible to suggest a specific diagnosis ('don't know'). The accuracy, sensitivity, specificity,

and positive and negative likelihood ratios of pattern recognition with regard to specific histological diagnoses are shown in Tables 2 and 3. The specificity was high for all specific diagnoses (range, 94–100%). Sensitivity was highest for benign teratomas/dermoid cysts (86%), hydrosalpinges (86%), peritoneal pseudocysts (80%) and endometriomas (77%), and lowest for functional cysts (17%), paraovarian/parasalpingeal cysts (14%), rare benign tumors (11%), adenofibromas (8%), simple cysts (6%) and struma ovarii (0%).

### Confusion of specific diagnoses

Dermoid cysts, hydrosalpinges, functional cysts, paraovarian cysts, peritoneal pseudocysts, fibromas/fibrothecomas and simple cysts were never misdiagnosed as malignancies by the ultrasound examiner, whereas more than 10% of inflammatory processes, adenofibromas and benign rare tumors including struma ovarii were misdiagnosed as malignancies.

The ultrasound examiner often misdiagnosed mucinous cysts as serous cysts (23%, 20/86); often misdiagnosed adenofibromas as serous cysts (21%, 8/39), simple cysts (10%, 4/39) or hydrosalpinx (10%, 4/39); often misdiagnosed functional cysts as endometriomas (21%, 5/24), serous cysts (17%, 4/24) or simple cysts (17%, 4/24); often misdiagnosed paraovarian/parasalpingeal cysts as serous cysts (29%, 6/21) or hydrosalpinx (5/21, 24%); and often misdiagnosed rare benign tumors as fibromas/fibrothecomas (56%; 5/9). Endometriomas and dermoid cysts were confused with a variety of other conditions with no particular pathology being over-represented among the misdiagnoses. Images illustrating an example of overlapping ultrasound morphology of serous and mucinous cystadenomas are shown in Figure 1, and images illustrating an example of overlapping ultrasound morphology of cystadenoma and cystadenofibroma are shown in Figure 2.

The ultrasound examiner often misdiagnosed both metastatic tumors and rare malignancies as primary invasive ovarian tumors (36% (15/42) and 40% (10/25), respectively), whereas borderline tumors were often misdiagnosed either as primary invasive tumors (24%, 13/55) or as benign mucinous cysts (16%, 9/55).

The benign histopathological diagnoses that were most frequently confused with other diagnoses were rare benign tumors (89%, 8/9), functional cysts (79%, 19/24), and adenofibromas (67%, 26/39). The malignant histopathological diagnoses that were most frequently confused with other diagnoses were rare malignant tumors (64%, 16/25) and borderline tumors (62%, 34/55). Serous cyst/cystadenoma was the most frequent incorrect diagnosis constituting 17% of all incorrect diagnoses (50/302).

### Differences in results between centers

We compared the sensitivity and specificity with regard to specific diagnoses between centers that had contributed

**Table 2.** Accuracy, sensitivity, specificity and positive (LR+) and negative (LR-) likelihood ratios of 'pattern recognition' with regard to specific diagnosis, and diagnosis suggested by the ultrasound examiner

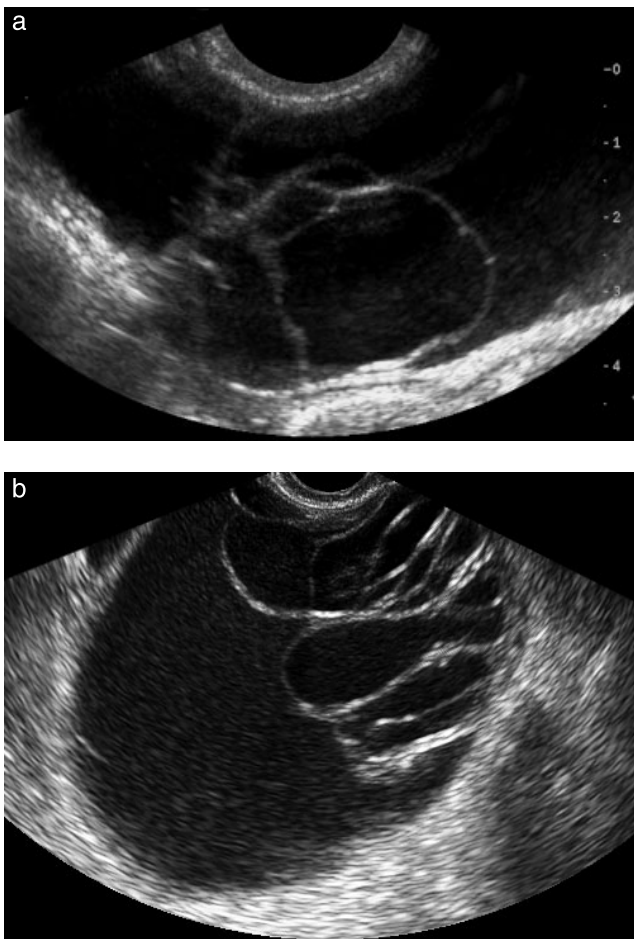
Tumor	n	Accuracy (% (n))	Sensitivity (% (95% CI) (n))	Specificity (% (95% CI) (n))	LR+ (95% CI)	LR- (95% CI)	Diagnosis suggested by ultrasound examiner			
							Incorrect benign diagnosis* (n (%))	Incorrect malignant diagnosis† (n (%))	'Don't know' (n (%))	
<b>Benign tumors</b>										
Teratoma	116	97 (1038/1066)	86 (79-91) (100/116)	99 (98-99) (938/950)	68.2 (38.7-120.3)	0.14 (0.09-0.22)	12 (10)	0 (0)	4 (3)	
Hydrosalpinx-hematosalpinx	21	98 (1040/1066)	86 (65-95) (18/21)	98 (97-99) (1022/1045)	38.9 (25.1-60.5)	0.15 (0.05-0.42)	0 (0)	0 (0)	3 (14)	
Endometrioma	199	94 (1000/1066)	77 (77-82) (153/199)	98 (97-99) (847/867)	33.3 (21.5-51.7)	0.24 (0.18-0.31)	27 (14)	3 (2)	16 (8)	
Serous cyst/cystadenoma	149	88 (940/1066)	54 (46-62) (80/149)	94 (92-95) (860/917)	8.6 (6.5-11.6)	0.49 (0.42-0.59)	36 (24)	4 (3)	29 (19)	
Inflammatory process	20	98 (1047/1066)	45 (26-66) (9/20)	99 (99-100) (1038/1046)	58.8 (25.3-136.7)	0.55 (0.37-0.82)	3 (15)	3 (15)	5 (25)	
Mucinous cyst/cystadenoma	86	92 (979/1066)	36 (27-47) (31/86)	97 (95-98) (948/980)	11.0 (7.1-17.2)	0.66 (0.56-0.78)	30 (35)	4 (5)	21 (24)	
Functional cyst	24	98 (1040/1066)	17 (7-36) (4/24)	99 (99-100) (1036/1042)	28.9 (8.7-96.0)	0.84 (0.70-1.00)	19 (79)	0 (0)	1 (4)	
Paraovarian/parasalpingeal cyst	21	98 (1041/1066)	14 (5-35) (3/21)	99 (99-100) (1038/1045)	21.3 (5.9-76.8)	0.86 (0.72-1.03)	13 (62)	0 (0)	5 (24)	
Adenofibroma	39	95 (1015/1066)	8 (3-20) (3/39)	98 (98-99) (1012/1027)	5.3 (1.6-17.5)	0.94 (0.86-1.03)	21 (54)	5 (13)	10 (26)	
<b>Malignant tumors</b>										
Primary invasive tumor	144	91 (974/1066)	72 (64-78) (104/144)	94 (93-96) (870/922)	12.8 (9.7-17.0)	0.29 (0.23-0.38)	4 (3)	13 (9)	23 (16)	
Borderline tumor	55	95 (1008/1066)	29 (18-42) (16/55)	98 (97-99) (992/1011)	15.2 (8.3-27.9)	0.73 (0.62-0.86)	21 (38)	13 (24)	5 (9)	
Metastatic tumor	42	97 (1030/1066)	26 (15-41) (11/42)	99 (99-100) (1019/1024)	53.6 (19.5-147.4)	0.74 (0.62-0.89)	2 (5)	20 (48)	9 (21)	
Rare malignant tumor	25	97 (1037/1066)	16 (6-35) (4/25)	99 (99-100) (1033/1041)	20.8 (6.7-64.6)	0.85 (0.71-1.00)	6 (24)	10 (40)	5 (20)	

Results are shown only for diagnoses represented by at least 20 cases. \*Incorrect benign diagnosis means either that a benign tumor was correctly classified as benign but assigned an incorrect specific benign diagnosis, or that a malignant tumor was incorrectly classified as benign. †Incorrect malignant diagnosis means either that a malignant tumor was correctly classified as malignant but assigned an incorrect specific malignant diagnosis, or that a benign tumor was incorrectly classified as malignant.

**Table 3** Accuracy, sensitivity, specificity and positive (LR+) and negative (LR-) likelihood ratios of 'pattern recognition' with regard to specific diagnosis, and diagnosis suggested by the ultrasound examiner

Tumor	n	Accuracy (% (n))	Sensitivity (% (95% CI) (n))	Specificity (% (95% CI) (n))	LR+ (95% CI)	LR- (95% CI)	Diagnosis suggested by ultrasound examiner		
							Incorrect benign diagnosis* (n (%))	Incorrect malignant diagnosis† (n (%))	'Don't know' (n (%))
Peritoneal pseudocyst	5	99 (1052/1066)	80 (38-99) (4/5)	99 (98-99) (1048/1061)	65.3 (32.6-130.9)	0.20 (0.04-1.17)	1 (20)	0 (0)	0 (0)
Torsion of lesion	9	99 (1058/1066)	56 (27-81) (5/9)	100 (99-100) (1053/1057)	146.8 (47.0-458.8)	0.45 (0.22-0.93)	2 (22)	0 (0)	2 (22)
Fibroma/fibrothecoma	19	98 (1045/1066)	42 (23-64) (8/19)	99 (98-100) (1037/1047)	44.0 (19.6-99.2)	0.58 (0.40-0.90)	6 (32)	0 (0)	5 (26)
Leiomyoma	12	99 (1053/1066)	25 (9-53) (3/12)	100 (99-100) (1050/1054)	65.9 (16.5-263.1)	0.75 (0.54-1.04)	5 (42)	1 (8)	3 (25)
Rare benign tumor (excluding struma ovarii)	9	99 (1056/1066)	11 (1-44) (1/9)	100 (99-100) (1055/1057)	58.7 (5.8-591.1)	0.89 (0.71-1.1)	7 (78)	1 (11)	0 (0)
Simple cyst	18	96 (1024/1066)	6 (0-26) (1/18)	98 (97-98) (1023/1048)	2.3 (0.3-16.3)	0.97 (0.87-1.08)	5 (28)	0 (0)	12 (67)
Struma ovarii	5	0	(0/5)	100 (1066/1066)	0	0	2 (40)	3 (60)	0 (0)

Results are shown for diagnoses represented by fewer than 20 cases. \*Incorrect benign diagnosis means either that a benign tumor was correctly classified as benign but assigned an incorrect specific benign diagnosis, or that a malignant tumor was incorrectly classified as benign. †Incorrect malignant diagnosis means either that a malignant tumor was correctly classified as malignant but assigned an incorrect specific malignant diagnosis, or that a benign tumor was incorrectly classified as malignant.

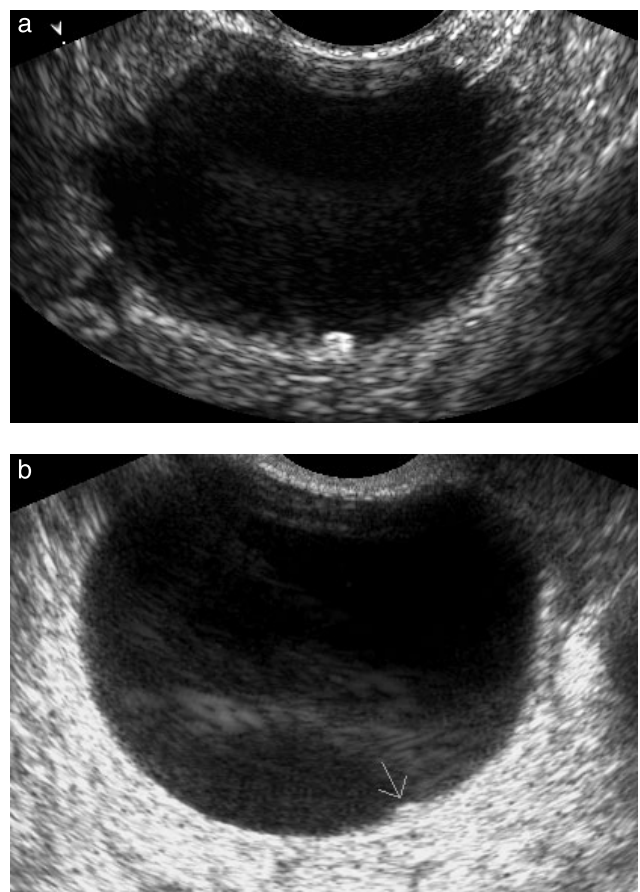


**Figure 1** Ultrasound images illustrating overlapping ultrasound morphology of serous (a) and mucinous (b) cystadenoma. The layering in (a) is explained by artifacts. Both lesions are multilocular cysts with a conglomerate of small cysts in one area.

at least 10 patients with the specific diagnosis in question, provided that there were at least three centers that had contributed at least 10 cases each (Table S1 online). The sensitivity and specificity varied between the centers, but the specificity was almost always >90%. For most of the diagnoses, the sensitivity with regard to the specific diagnosis in question was similar in all centers, i.e. either high (arbitrarily defined as 75–100%), intermediate (arbitrarily defined as 50–74%) or low (arbitrarily defined as <50%). Many false-negative diagnoses were explained by the ultrasound examiner not having suggested a specific diagnosis ('don't know') (Tables 2 and 3). This was particularly common in one of the participating centers. The single examiner in this center contributed 315/1066 (30%) cases to the study but was responsible for 105/167 (63%) cases where a diagnosis of 'don't know' had been noted.

## DISCUSSION

To the best of our knowledge this is the first multicenter study evaluating the ability of ultrasound examiners using subjective evaluation of gray-scale and Doppler ultrasound findings (pattern recognition) to make a



**Figure 2** Ultrasound images illustrating overlapping ultrasound morphology of serous cystadenoma (a) and serous cystadenofibroma (b). The layering in (b) is explained by artifacts. Both lesions are unilocular cysts with anechogenic cyst contents and a very small papillary projection (arrow).

specific diagnosis of an adnexal mass. It is a strength that our study is a multicenter study involving many examiners and many tumors, because this makes our results likely to be more generalizable than small single-center studies. On the other hand, despite our study being large, the number of some specific diagnoses (e.g. peritoneal pseudocyst, rare benign tumor, struma ovarii, fibroma/fibrothecoma, leiomyoma and simple cyst) is too small for a precise estimate of sensitivity and specificity with regard to these diagnoses to be possible (see CIs for sensitivity, specificity, and positive and negative likelihood ratios).

Our results agree well with those from single centers in that the sensitivity and specificity of pattern recognition (i.e. subjective evaluation of gray-scale ultrasound findings with or without subjective evaluation of Doppler ultrasound findings) with regard to benign teratoma/dermoid cyst, endometrioma and hydrosalpinx were high<sup>4,12–25</sup>.

Some studies have evaluated the sensitivity and specificity of pattern recognition with regard to paraovarian cysts<sup>4,26</sup> and mucinous and serous cystadenomas<sup>15,27</sup> and all included rather few cases (three and 17 cases of paraovarian cyst, 18 and 38 cases of serous cystadenoma, and 21 cases of mucinous cystadenoma). The sensitivity with regard to these three diagnoses was

substantially higher in the studies cited than in ours, whereas the specificity was similar. The low sensitivity of pattern recognition with regard to serous and mucinous cystadenomas and adenofibromas in our study is explained either by an incorrect specific diagnosis having been assigned to the tumor or by no specific diagnosis ('don't know') having been suggested by the ultrasound examiner. As illustrated in Figure 1, the ultrasound morphology of serous and mucinous cystadenomas may overlap extensively. The ultrasound morphology of serous and mucinous cystadenomas also overlaps to some extent with that of adenofibromas (Figure 2). Indeed, serous cyst/cystadenoma was the most frequent incorrect ultrasound diagnosis in our series. While endometriomas and dermoid cysts were confused – albeit very rarely – with a variety of other conditions (with no particular pathology being over-represented among the misdiagnoses), serous cysts, adenofibromas, simple cysts, hydrosalpinx, functional cysts and paraovarian/parasalpingeal cysts were often confused with each other. This illustrates that many of the latter pathologies do not have a pathognomonic appearance at ultrasound examination.

We know of only one study reporting on the sensitivity and specificity of pattern recognition with regard to peritoneal cysts and fibromas/fibrothecomas. In that study there were only three cases of peritoneal cyst and nine cases of fibroma/fibrothecoma<sup>4</sup>. The reported sensitivity and specificity with regard to peritoneal cyst in the study cited were 100% and 99% vs. 80% and 99% in our study, and the sensitivity and specificity with regard to fibroma/fibrothecoma were 56% and 100% vs. 42% and 99% in our study. We have found no study reporting on the sensitivity and specificity of pattern recognition with regard to ovarian adenofibroma.

Typical ultrasound findings in different types of ovarian malignancies and of metastases in the ovaries of tumors of different primary origin have been described<sup>28,29</sup> but, to the best of our knowledge, there are no published reports on the sensitivity and specificity of pattern recognition with regard to a diagnosis of primary invasive ovarian cancer, rare malignant ovarian tumors or metastatic tumors in the ovary. However, one study reported the sensitivity and specificity of pattern recognition with regard to ovarian borderline tumors<sup>30</sup>. That study comprised 35 borderline tumors, 99 benign tumors and 32 invasive ovarian malignancies. The reported sensitivity with regard to borderline tumor was 69% and the specificity 94%. Our sensitivity with regard to borderline tumor was much lower (29%) and our specificity was only marginally higher (98%), with 9% of our borderline tumors not having been assigned a specific diagnosis, 38% having been classified as a benign cyst (vs. 29% in the study cited) and 24% as invasive malignancies (vs. 3% in the study cited). Differences in tumor population are likely to contribute to the discrepant results, even though the proportion of serous and mucinous borderline tumors was similar in the two studies.

It is important to be aware that the apparent sensitivity and specificity of pattern recognition with regard to

specific diagnoses are affected not only by the skill of the ultrasound examiner and the quality of the ultrasound equipment used. The mix of tumor type is clearly very important. Our study population comprises only adnexal tumors that were removed surgically (a prerequisite for obtaining the true diagnosis). This means that masses with atypical or complex ultrasound morphology are likely to be more common in our study population than in a total population of adnexal masses, because an unequivocal ultrasound diagnosis of, for example, uterine leiomyoma, peritoneal cyst, paraovarian cyst, hydrosalpinx, simple cyst or functional cyst is less likely to result in the mass being removed surgically than an equivocal diagnosis. For example, in our series most functional cysts had been misdiagnosed as non-functional cysts, e.g. as endometriomas or cystadenomas, whereas in reality most functional cysts are recognized as such and are not removed surgically. Moreover, two of the centers contributing cases to this study were tertiary referral centers affiliated to cancer centers. For this reason, too, tumors with equivocal, unusual or complicated ultrasound morphology are likely to be over-represented in our study. All of the above means that both the sensitivity and specificity of pattern recognition with regard to specific diagnoses may have been underestimated in our study. The much higher sensitivity with regard to serous and mucinous cystadenoma reported in other studies (sensitivity with regard to serous cystadenoma 70–78% vs. 54% in our study; sensitivity 95% with regard to mucinous cystadenoma vs. 36% in our study) may be explained by a rather 'artificial' mix of tumors in the other studies<sup>15,27</sup>. The apparent sensitivity and specificity are also affected by the willingness of the ultrasound examiner to suggest a specific diagnosis if he/she is not completely certain about the diagnosis. As clearly illustrated by our results, this willingness is likely to differ between examiners (a single examiner who contributed 30% of the cases to our study was responsible for 63% of the 'don't know' diagnoses). Some of the ultrasound examiners might have guessed a diagnosis and entered their guess into the study protocol, whereas they might not have been confident enough to suggest the same diagnosis in a clinical report. Moreover, there is almost certainly some interobserver disagreement between pathologists when they assign a diagnosis to an adnexal mass. All the above complicates not only the estimation of sensitivity and specificity but also the determination of which diagnoses are likely to be confused with each other at a scan.

From a clinical point of view, it is most important to be able to distinguish benign from malignant adnexal tumors<sup>31</sup>, and there are numerous studies showing that pattern recognition can do this<sup>2,3,32–35</sup>. It is also of value to be able to reliably discriminate between primary ovarian malignancies, borderline tumors and metastatic tumors in the ovaries, because the management of these diagnoses differs. Our study suggests that this was not possible with the knowledge available at the time of our data collection. However, studies describing the typical ultrasound morphology of rare adnexal tumors have been published recently<sup>36–38</sup>, and it is reasonable to

believe that the skill in discriminating between different types of ovarian pathology will increase, if more work describing the typical ultrasound morphology of specific types of adnexal tumor is published. It is definitely clinically useful to be able to make a specific diagnosis of dermoid cyst, hydrosalpinx, endometrioma, peritoneal cyst and hemorrhagic corpus luteum cyst, because it seems reasonable to treat these conditions expectantly if they do not cause any symptoms and are not associated with subfertility. Pattern recognition is likely to be able to reliably distinguish not only dermoid cysts, endometriomas and hydrosalpinges (as shown in this study) but also peritoneal cysts and hemorrhagic corpus luteum cysts from other adnexal masses, but there are too few peritoneal cysts and corpus luteum cysts in our study for us to be able to state this with any certainty. The most likely explanation for the low number of these pathologies is that they were treated expectantly if the diagnosis was certain. If screening for ovarian cancer is to become a reality<sup>39</sup>, expert ultrasound imaging with pattern recognition will probably be used as a secondary test to evaluate the need for surgery, and our results support the contention that pattern recognition is likely to be able to provide an exact diagnosis in many screen-positive cases.

To sum up, using subjective evaluation of gray-scale and Doppler ultrasound findings it is possible to make an almost conclusive diagnosis of dermoid cyst, endometrioma and hydrosalpinx (and possibly peritoneal pseudocysts), but currently it does not seem to be possible to make any other histological diagnosis conclusively. Even though it is possible to recognize many other adnexal pathologies on the basis of their ultrasound characteristics they cannot be confidently confirmed or excluded. Serous cysts, cystadenofibromas, simple cysts, hydrosalpinx, functional cysts and paraovarian/parasalpingeal cysts are diagnoses that are often confused with each other. Our results are generalizable only to tumor populations, ultrasound examiners (level of experience) and ultrasound equipment similar to those in this study.

## ACKNOWLEDGMENTS

This work was supported by the Swedish Medical Research Council (grants numbers K2001-72X-11605-06A, K2002-72X-11605-07B, K2004-73X-11605-09A and K2006-73X-11605-11-3), by funds administered by Malmö University Hospital, and by GOA-AMBioRICS, CoE EF/05/006, Belgian network DYSCO, BIOPATTERN (FP6-2002-IST 508803), ETUMOUR (FP6-2002-LIFESCIHEALTH 503094).

## APPENDIX

### IOTA Steering Committee

Dirk Timmerman, Lil Valentin, Tom Bourne, William P. Collins, Sabine Van Huffel, Ignace Vergote

### IOTA principal investigators (alphabetical order)

Jean-Pierre Bernard, Maurepas, France  
 Tom Bourne, London, UK  
 Enrico Ferrazzi, Milan, Italy  
 Davor Jurkovic, London, UK  
 Fabrice Lécuru, Paris, France  
 Andrea Lissoni, Monza, Italy  
 Ulrike Metzger, Paris, France  
 Dario Paladini, Naples, Italy  
 Antonia Testa, Rome, Italy  
 Dirk Timmerman, Leuven, Belgium  
 Lil Valentin, Malmö, Sweden  
 Sabine Van Huffel, Leuven, Belgium  
 Dominique Van Schoubroeck, Leuven, Belgium  
 Ignace Vergote, Leuven, Belgium  
 Gerardo Zanetta (deceased), Monza, Italy

## REFERENCES

1. Valentin L. Imaging in gynecology. *Best Pract Res Clin Obstet Gynaecol* 2006; **20**: 881–906.
2. Valentin L. Prospective cross-validation of Doppler ultrasound examination and gray-scale ultrasound imaging for discrimination of benign and malignant pelvic masses. *Ultrasound Obstet Gynecol* 1999; **14**: 273–283.
3. Van Calster B, Timmerman D, Bourne T, Testa AC, Van Holsbeke C, Domali E, Jurkovic D, Neven P, Van Huffel S, Valentin L. Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125. *J Natl Cancer Inst* 2007; **99**: 1706–1714.
4. Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. *Ultrasound Obstet Gynecol* 1999; **14**: 338–347.
5. Caspi B, Appelman Z, Rabinerson D, Zalel Y, Tulandi T, Shoham Z. The growth pattern of ovarian dermoid cysts: a prospective study in premenopausal and postmenopausal women. *Fertil Steril* 1997; **68**: 501–505.
6. Tsai CC, Shen CC, Changchien CC, Hsu TY, Kung FT, Chang SY, Chang MY, Huang FJ. Ultrasound-guided transvaginal cyst aspiration for the management of pelvic pseudocyst: a preliminary experience. *Chang Gung Med J* 2002; **25**: 751–757.
7. Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, Van Calster B, Collins WP, Vergote I, Van Huffel S, Valentin L. 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* 2005; **23**: 8794–8801.
8. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I. 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* 2000; **16**: 500–505.
9. European Federation of Societies in Ultrasound in Medicine and Biology (EFSUMB). Minimum training recommendations for the practice of medical ultrasound. *Ultrasound Med* 2005; **26**: 79–105.
10. Valentin L. Use of morphology to characterize and manage common adnexal masses. *Best Pract Res Clin Obstet Gynaecol* 2004; **18**: 71–89.
11. Pecorelli S, Benedet JL, Creasman WT, Shepherd JH. FIGO staging of gynecologic cancer. 1994–1997 FIGO Committee on Gynecologic Oncology. International Federation of Gynecology and Obstetrics. *Int J Gynaecol Obstet* 1999; **65**: 243–249.
12. Jain KA, Friedman DL, Pettinger TW, Alagappan R, Jeffrey RB Jr, Sommer FG. Adnexal masses: comparison of specificity of



- endovaginal US and pelvic MR imaging. *Radiology* 1993; **186**: 697–704.
13. Mais V, Guerriero S, Ajossa S, Angiolucci M, Paoletti AM, Melis GB. Transvaginal ultrasonography in the diagnosis of cystic teratoma. *Obstet Gynecol* 1995; **85**: 48–52.
  14. Ekici E, Soysal M, Kara S, Dogan M, Gokmen O. The efficiency of ultrasonography in the diagnosis of dermoid cysts. *Zentralbl Gynakol* 1996; **118**: 136–141.
  15. Guerriero S, Mallarini G, Ajossa S, Risalvato A, Satta R, Mais V, Angiolucci M, Melis GB. Transvaginal ultrasound and computed tomography combined with clinical parameters and CA-125 determinations in the differential diagnosis of persistent ovarian cysts in premenopausal women. *Ultrasound Obstet Gynecol* 1997; **9**: 339–343.
  16. Mais V, Guerriero S, Ajossa S, Angiolucci M, Paoletti AM, Melis GB. The efficiency of transvaginal ultrasonography in the diagnosis of endometrioma. *Fertil Steril* 1993; **60**: 776–780.
  17. Guerriero S, Mais V, Ajossa S, Paoletti AM, Angiolucci M, Labate F, Melis GB. The role of endovaginal ultrasound in differentiating endometriomas from other ovarian cysts. *Clin Exp Obstet Gynecol* 1995; **22**: 20–22.
  18. Volpi E, De Grandis T, Zuccaro G, La Vista A, Sismondi P. Role of transvaginal sonography in the detection of endometriomata. *J Clin Ultrasound* 1995; **23**: 163–167.
  19. Guerriero S, Ajossa S, Paoletti AM, Mais V, Angiolucci M, Melis GB. Tumor markers and transvaginal ultrasonography in the diagnosis of endometrioma. *Obstet Gynecol* 1996; **88**: 403–407.
  20. Guerriero S, Ajossa S, Mais V, Risalvato A, Lai MP, Melis GB. The diagnosis of endometriomas using colour Doppler energy imaging. *Hum Reprod* 1998; **13**: 1691–1695.
  21. Alcazar JL, Laparte C, Jurado M, Lopez-Garcia G. The role of transvaginal ultrasonography combined with color velocity imaging and pulsed Doppler in the diagnosis of endometrioma. *Fertil Steril* 1997; **67**: 487–491.
  22. Dogan MM, Ugur M, Soysal SK, Soysal ME, Ekici E, Gokmen O. Transvaginal sonographic diagnosis of ovarian endometrioma. *Int J Gynaecol Obstet* 1996; **52**: 145–149.
  23. Ubaldi F, Wisanto A, Camus M, Tournaye H, Clasen K, Devroey P. The role of transvaginal ultrasonography in the detection of pelvic pathologies in the infertility workup. *Hum Reprod* 1998; **13**: 330–333.
  24. Guerriero S, Ajossa S, Lai MP, Mais V, Paoletti AM, Melis GB. Transvaginal ultrasonography associated with colour Doppler energy in the diagnosis of hydrosalpinx. *Hum Reprod* 2000; **15**: 1568–1572.
  25. Brazert J, Pietryga M, Jasinski P, Szablonski W, Biczysko R. [Diagnostic value of transvaginal ultrasound in the detection of ovarian endometriosis.] *Ginekol Pol* 2001; **72**: 358–363.
  26. Guerriero S, Ajossa S, Piras S, Angiolucci M, Marisa O, Melis GB. Diagnosis of paraovarian cysts using transvaginal sonography combined with CA 125 determination. *Ultrasound Obstet Gynecol* 2006; **28**: 856–858.
  27. Buy JN, Ghossain MA, Scioc C, Bazot M, Guinet C, Prevot S, Hugol D, Laromiguiere M, Truc JB, Poitout P, Vadrot D, Ecoffier J. Epithelial tumors of the ovary: CT findings and correlation with US. *Radiology* 1991; **178**: 811–818.
  28. Valentin L, Ameye L, Testa A, Lecuru F, Bernard JP, Paladini D, Van Huffel S, Timmerman D. Ultrasound characteristics of different types of adnexal malignancies. *Gynecol Oncol* 2006; **102**: 41–48.
  29. Testa AC, Ferrandina G, Timmerman D, Savelli L, Ludovisi M, Van Holsbeke C, Malaggesi M, Scambia G, Valentin L. Imaging in gynecological disease (1): ultrasound features of metastases in the ovaries differ depending on the origin of the primary tumor. *Ultrasound Obstet Gynecol* 2007; **29**: 505–511.
  30. Yazbek J, Raju KS, Ben-Nagi J, Holland T, Hillaby K, Jurkovic D. Accuracy of ultrasound subjective ‘pattern recognition’ for the diagnosis of borderline ovarian tumors. *Ultrasound Obstet Gynecol* 2007; **29**: 489–495.
  31. Yazbek J, Helmy S, Ben-Nagi J, Holland T, Sawyer E, Jurkovic D. Value of preoperative ultrasound examination in the selection of women with adnexal masses for laparoscopic surgery. *Ultrasound Obstet Gynecol* 2007; **30**: 883–888.
  32. Timmerman D, Schwärzler P, Collins WP, Claerhout F, Coenen M, Amant F, Vergote I, Bourne TH. Subjective assessment of adnexal masses using ultrasonography: an analysis of interobserver variability and experience. *Ultrasound Obstet Gynecol* 1999; **13**: 11–16.
  33. Valentin L, Hagen B, Tingulstad S, Eik-Nes S. Comparison of ‘pattern recognition’ and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. *Ultrasound Obstet Gynecol* 2001; **18**: 357–365.
  34. Stein SM, Laifer-Narin S, Johnson MB, Roman LD, Muder-spach LI, Tyszka JM, Ralls PW. Differentiation of benign and malignant adnexal masses: relative value of gray-scale, color Doppler, and spectral Doppler sonography. *AJR Am J Roentgenol* 1995; **164**: 381–386.
  35. Buy JN, Ghossain MA, Hugol D, Hassen K, Scioc C, Truc JB, Poitout P, Vadrot D. Characterization of adnexal masses: combination of color Doppler and conventional sonography compared with spectral Doppler analysis alone and conventional sonography alone. *AJR Am J Roentgenol* 1996; **166**: 385–393.
  36. Demidov VN, Lipatenkova J, Vikhareva O, Van Holsbeke C, Timmerman D, Valentin L. Imaging of gynecological disease (2): clinical and ultrasound characteristics of Sertoli cell tumors, Sertoli–Leydig cell tumors and Leydig cell tumors. *Ultrasound Obstet Gynecol* 2008; **31**: 85–91.
  37. Van Holsbeke C, Domali E, Holland TK, Achten R, Testa AC, Valentin L, Jurkovic D, Moerman P, Timmerman D. Imaging of gynecological disease (3): clinical and ultrasound characteristics of granulosa cell tumors of the ovary. *Ultrasound Obstet Gynecol* 2008; **31**: 450–456.
  38. Savelli L, Testa AC, Timmerman D, Paladini D, Ljungberg O, Valentin L. Imaging in gynecological disease (4): clinical and ultrasound characteristics of struma ovarii. *Ultrasound Obstet Gynecol* 2008; **32**: 210–219.
  39. Menon U. Ovarian cancer: challenges of early detection. *Nat Clin Pract Oncol* 2007; **4**: 498–499.

#### SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:

**Table S1** Sensitivity, specificity and positive and negative likelihood ratios for each ultrasound center with regard to those specific diagnoses for which at least three centers contributed at least 10 cases each