ULTRASOUND in Obstetrics & Gynecology isuog

Vessel morphology depicted by three-dimensional power Doppler ultrasound as a second stage test in difficult adnexal tumors: a prospective diagnostic accuracy study

Journal:	*Ultrasound in Obstetrics and Gynecology*
Manuscript ID	UOG-2020-0698.R1
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	20-Jul-2020
Complete List of Authors:	Sladkevicius, Povilas; Malmö University Hospital, Lund University, Department of Obstetrics and Gynecology Jokubkiene, Ligita; Lund University, Department of Obstetrics and gynecology Timmerman, Dirk; Katholieke Universiteit Leuven, Obstetrics & gynecology Fischerova, Daniela; Gynecological Oncology Center, Department of Obstetrics and Gynecology, First Faculty of Medicine and General University Hospital, Charles University Van Holsbeke, Caroline; Ziekenhuis Oost-Limburg, 6Department of Obstetrics and Gynecology Franchi, Dorella; IEO, Gynecology Savelli, Luca; University of Bologna, Obstetrics and Gynecology; Epstein, Elisabeth; Department of Clinical Science and Education, Södersjukhuset , Obstetrics and Gyencology; Department of Women´s and Children´s health Karolinska Institutet, , Fruscio, Robert; San Gerardo Hospital, University of Milan-Bicocca, Clinic of Obstetrics and Gynecology, Department of Medicine and Surgery Kaijser, Jeroen; Ikazia Hospital, Department of Obsterics and Gynecology Czekierdowski, Artur; First Department of Gynecological Oncology and Gynecology, Medical University of Lublin., Gynecological Oncology and Gynecology Guerriero, Stefano; University of Cagliari, Department of Obstetrics and Gynecology Pascual, M.Angela; Hospital Universitari Dexeus-Quiron, Departament of Obstetrics and Gynecology Testa, Antonia Carla; Istituto di Ginecologia e Ostetricia, Università Cattolica del Sacro Cuore, Roma, Italia, Ameye, Lieveke; KU Leuven, Department of Obstetrics and Gynecology Valentin, Lil; Lund University, Dept Obstetrics and Gynecology
Keywords:	ultrasonography, Doppler, ovarian neoplasms, three-dimensional ultrasound, vascular morphology
Manuscript Categories:	Gynecology

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

Supplementary video 1 corresponding to Fig 2a.mp4
Supplementary video 2 corresponding to Fig 2b.mp4
Supplementary video 3 corresponding to Fig 2c.mp4
Supplementary video 4 corresponding to Fig 2d.mp4
Supplementary video 5 corresponding to Fig 3a.mp4
Supplementary video 6 corresponding to Fig3b.mp4
Supplementary video 7 corresponding to Fig 3c.mp4

Vessel morphology depicted by three-dimensional power Doppler ultrasound as a second stage test in difficult adnexal tumors: a prospective diagnostic accuracy study

P. Sladkevicius[1,2], L. Jokubkiene[1,2], D. Timmerman[3,4], D. Fischerova[5], C. Van Holsbeke[6], D. Franchi[7], L. Savelli[8], E. Epstein[9], R. Fruscio[10], J. Kaijser[11], A. Czekierdowski[12], S. Guerriero[13], M. Angela Pascual[14], A. Testa[15], L. Ameye[3,16], L. Valentin[1,2]

[1]Department of Obstetrics and Gynecology, Skåne University Hospital, Malmö, Sweden

[2]Department of Clinical Sciences Malmö, Lund University, Sweden

[3] KU Leuven Department of Development and Regeneration, Herestraat 49, Box 7003, 3000 Leuven, Belgium

[4]Department of Obstetrics and Gynecology and Leuven Cancer Institute, University Hospitals Leuven, Herestraat 49, Box 7003, 3000 Leuven, Belgium

[5]Department of Obstetrics and Gynecology, First Faculty of Medicine, Charles University and First Faculty of Medicine, Prague, Czech Republic

[6]Department of Obstetrics and Gynecology, Ziekenhuis Oost Limburg, Schiepse Bos 6, 3600 Genk, Belgium

[7]Preventive Gynecology Unit, Division of Gynecology, European Institute of Oncology, Via Ripamonti 435**,** Milan 20141, Italy

[8]Gynecology and Reproductive Medicine Unit, S. Orsola-Malpighi Hospital, University of Bologna, Via Massarenti 13, Bologna 40138, Italy

[9]Department of Clinical Science and Education, Karolinska Institute, Södersjukhuset, Stockholm, Sweden

[10]Department of Obstetrics and Gynecology, San Gerardo Hospital, University of Milan-Bicocca, Via Pergolesi 33, 20052 Monza**,** Italy

[11]Department of Obstetrics and Gynecology, Ikazia Hospital Rotterdam, the Netherlands

[12]1st Department of Gynecological Oncology and Gynecology, Medical University of Lublin, Lublin, Poland

[13]Department of Obstetrics and Gynecology, University of Cagliari, Policlinico Universitario Duilio Casula, Monserrato, Cagliari, Italy

[14] Department of Obstetrics, Gynecology and Reproduction, Hospital Universitari Dexeus, Barcelona, Spain

[15]Department of Gynecological Oncology, Catholic University of Sacred Heart, Rome, Italy

[16]Jules Bordet Institute, Université Libre de Bruxelles, Brussels, Belgium

Corresponding author:

Lil Valentin

Department of Obstetrics and Gynecology

Skåne University Hospital Malmö

20502 Malmö, Sweden

e-mail: lil.valentin@med.lu.se

Short title: vascular morphology in difficult adnexal masses

Keywords: ultrasonography, Doppler; ovarian neoplasms; three-dimensional ultrasound; vascular morphology

Contribution

What are the novel findings of this work?

We show that vessel morphology depicted by three-dimensional power Doppler ultrasound differs between benign and malignant adnexal masses difficult to classify as benign or malignant by an experienced ultrasound examiner using subjective assessment of ultrasound images or by the IOTA logistic regression model LR1.

What are the clinical implications of this work?

Vessel morphology as depicted by three-dimensional power Doppler ultrasound may slightly improve discrimination between benign and malignant tumors judged difficult to classify by subjective assessment. For tumors in which the IOTA model LR1 yields an ambiguous result, subjective assessment is superior to vessel morphology as a second stage test.

Abstract

Objectives. The aim was to assess whether vessel morphology depicted by three-dimensional (3D) power Doppler ultrasound improves discrimination between benignity and malignancy if used as a second stage test in difficult adnexal masses.

Methods. This is a prospective observational international multicenter diagnostic accuracy study. 2403 consecutive patients with an adnexal mass underwent standardized transvaginal 2D gray scale and color or power Doppler ultrasound and 3D power Doppler ultrasound by an experienced examiner. We defined a difficult tumor as one in which the logistic regression model IOTA LR1 yielded an ambiguous result (risk of malignancy 8.3% to 25.5%), or as one in which the ultrasound examiner was uncertain whether the tumor was benign or malignant when using subjective assessment. Even when the ultrasound examiner was uncertain he/she was obliged to classify the tumor as most likely benign or most likely malignant. For each difficult tumor, one researcher created a 360° rotating 3D image of the vessel tree in the whole tumor and another of the vessels tree in a 5 cm^3 spherical volume selected from the most vascularized part of the tumor. Two other researchers, blinded to patient history, 2D ultrasound findings and histological diagnosis, independently described the vessel tree using predetermined vessel features. Their agreed classification was used. The reference standard was the histological diagnosis of the mass. We plotted the sensitivity of each test against 1 – specificity in a receiver operating characteristic diagram. The test with its symbol farthest from the reference line was considered to have the best discriminative ability.

Results There were 376/2403 (15.6%) difficult masses. Ultrasound volumes were available for 138 of these. In 79/138 masses the ultrasound examiner was uncertain about the diagnosis, in 87/138 IOTA LR1 yielded an ambiguous result, in 28/138 both methods gave an uncertain result. 38/138 (27%) masses were malignant. Among tumors difficult to classify by subjective assessment, the vessel feature 'densely packed vessels' had the best discriminative ability [sensitivity 67% (18/27), specificity 83% (43/52)] and was slightly superior to subjective assessment [sensitivity 74% (20/27), specificity 60% (31/52)]. In tumors in which IOTA LR1 yielded an ambiguous result, subjective assessment [sensitivity 82% (14/17), specificity 79% (55/70)] was superior to the best vascular feature, i.e. caliber changes of vessels in the whole tumor volume [sensitivity 71% (9/17), specificity 69% (48/70)].

Conclusion

Vessel morphology depicted by 3D power Doppler ultrasound may slightly improve discrimination between benign and malignant tumors difficult to classify by subjective assessment. For tumors in which the IOTA model LR1 yields an ambiguous result, subjective assessment is superior to vessel morphology as a second stage test.

Introduction

Subjective assessment of ultrasound findings (also called pattern recognition) in the hands of an experienced ultrasound examiner is the best ultrasound method to discriminate between benign and malignant adnexal masses[1,2]. However, even an experienced ultrasound examiner may find up to 10% of tumors impossible to confidently classify as benign or malignant using pattern recognition ("difficult tumors")[3-5]. The 10% risk cut-off of the International Ovarian Tumor Analysis (IOTA) logistic regression model 1 (LR1) has almost as good ability to discriminate between benign and malignant tumors as subjective assessment[6,7], but a risk of malignancy calculated by LR1 of 8.3 – 25.5% has been suggested to represent an ambiguous risk[8]. The tumor marker CA125 is clearly inferior to subjective assessment for discriminating between benign and malignant adnexal masses[9] and has no role for classifying difficult tumors[3-5,10]. Subjective assessment of ultrasound images is superior to computed tomography for discriminating between benign and malignant adnexal masses, while the role of magnetic resonance imaging is still unclear[11,12, 13]. A logistic regression model for calculating the risk of malignancy in tumors not classifiable as benign or malignant by an experienced ultrasound examiner using subjective assessment has been published, but its ability to discriminate between benign and malignant tumors was not superior to that of subjective assessment[4].

For tumors that are difficult to classify as benign or malignant using subjective assessment or using the IOTA model LR1, a second stage test capable of correctly classifying difficult tumors as benign or malignant would be valuable. A possible second stage test is three-dimensional (3D) power Doppler ultrasound examination of the vascular tree of tumors[14].

The aims of this study are to assess whether vessel morphology as depicted by 3D power Doppler ultrasound differs between benign and malignant difficult adnexal masses, and if vessel morphology improves discrimination between benign and malignant masses if used as a second stage test in difficult adnexal masses.

Patients and Methods

Study population

Our study population are those patients in the IOTA 3 study[15] with a difficult tumor (definition of difficult tumor below). The IOTA 3 study is a prospective observational international multicenter cross-sectional diagnostic accuracy study (study protocol available in Appendix 1 in supplementary material), that has been described in detail eslewhere[15]. Patients were recruited into IOTA 3 between October 2009 and May 2012 in 18 centers in six countries (Sweden, Belgium, Italy, Poland, Spain and Czech Republic). These centers were either oncology referral centers (i.e. tertiary referral centers with a specific gynecological oncology unit) or other hospitals or units with a special interest in gynecological ultrasound. The centers and type of center are listed after the main text. Ethics approval was obtained from the Ethics Committee of the University Hospitals Leuven (B32220095331/S51375) as well as from the local ethics committees of all contributing centers.

Patients referred to one of the participating centers for an ultrasound examination and found to have an adnexal mass were eligible for inclusion in IOTA 3. Consecutive patients with at least one adnexal mass judged not to be a functional cyst examined with transvaginal ultrasound by an experienced ultrasound examiner were included in IOTA 3, provided that they gave written and/or oral informed consent before the ultrasound scan. If more than one mass was detected, the mass with the most complex ultrasound morphology was used for statistical analysis. When masses with similar morphology were observed, the largest mass or the one most easily accessible with ultrasound was used. Criteria for excluding patients from IOTA 3 were pregnancy at the time of the ultrasound examination, surgical removal of the mass more than 120 days after the ultrasound examination, and data inconsistencies that persisted after final manual data checks.

Data collection.

A dedicated, secure electronic data-collection system was developed for the IOTA 3 study (IOTA 3 Study Screen; astraia Software, Munich, Germany). Patients automatically received a unique identifier. Data security was ensured by encrypting all data communication. Data integrity and completeness were ensured by client-side checks in the system supplied by astraia and by final data cleaning by a group of biostatisticians and expert ultrasound examiners. The study screen presented the risk of malignancy calculated using the IOTA logistic regression model LR1[6].

Ultrasound examination

All patients included in the IOTA 3 study[15] underwent a standardized transvaginal ultrasound examination by a gynecologist or radiologist very experienced in gynecologic ultrasound. High-end ultrasound systems were used. Gray scale and color or power Doppler ultrasound was used to obtain information on more than 40 ultrasound variables to characterize each adnexal mass. Details on the standardized ultrasound examination technique and the IOTA terminology used to describe the ultrasound images have been published elsewhere[16]. After completing the ultrasound examination, the ultrasound examiner classified each mass as benign or malignant on the basis of his/her subjective assessment of the gray scale and color or power Doppler ultrasound findings. In addition, the ultrasound examiner stated his/her level of confidence by classifying each mass as certainly benign, probably benign, uncertain, probably malignant or certainly malignant. This means that even when the ultrasound examiner was uncertain whether the tumor was benign or malignant he/she was obliged to classify it as most likely benign or most likely malignant. The ultrasound information was entered prospectively into the electronic data-collection system (see above),

was locked at the time of the examination and could not be changed thereafter. Decision regarding surgery for adnexal tumors was taken by the referring physician. It was based on clinical information (such as symptoms, age, operative risk, coexisting disease, etc.) and on the ultrasound report which was written using the results of subjective assessment.

In addition to performing the two-dimensional (2D) ultrasound examination as described above, ultrasound examiners in centers with access to a Voluson 730 Expert or GE E8 ultrasound system (GE Healthcare, Zipf, Austria) with a 5 – 9 MHz or a 6 – 12 MHz vaginal transducer were asked to acquire 3D power Doppler ultrasound volumes of all adnexal masses. The power Doppler and 3D settings are described in Appendix 2 in supplementary material. The ultrasound examiners were instructed to include the whole tumor in the volume, or if this was not possible, because the tumor was too large, to acquire several volumes to ensure that all parts of the tumor were captured. The patients were asked to lie still during the acquisition and if necessary to hold their breath. Volumes from difficult tumors, i.e. tumors in which the ultrasound examiner was uncertain whether the tumor was benign or malignant, or in which the IOTA model LR1 gave an ambiguous risk of malignancy (8.3-25.5%)[8] were sent to Skåne University Hospital, Malmö on compact discs, for analysis. The same methodology of analyzing the 3D volumes as previously described was used[14]. This method is briefly outlined below.

Analysis of 3D volumes and Audio Video Interleave (AVI) files

360° rotating 3D images (AVI-files) of the vessel tree of the tumors were prepared by the second author (LJ) using the virtual organ computer-aided analysis (VOCAL™) imaging program (4D-VIEW, version 7.0, GE Healthcare, Zipf, Austria) on a personal computer. For each tumor, a 360° rotating 3D image of the vessels tree in the whole tumor was created as well as a 360° rotating 3D image of the vessels tree in a 5 cm^3 spherical volume selected from the most vascularized part of the tumor. Subjective assessment was used to select the most vascularized part of the tumor[14]. Before creating the rotating image of the vessel tree, color transparency was adjusted to optimize the delineation of the vessels.

All AVI files created as described above were then analyzed independently by two members of the Malmö research team (LV, PS) who had no knowledge of patient history, 2D ultrasound findings, or histological diagnosis. The vessel tree in the whole tumor volume as well as in the 5 cm^3 sample was characterized using the same classification as previously described (Figure 1): branching, i.e. division of a vessel into two or more branches; caliber changes, i.e. changes in vessel width from narrow to wide and again from wide to narrow; 'splashes', i.e. areas of color in contrast to clearly separate vessels; tortuosity; areas with densely packed vessels; and "bridges", i.e. straight vessel connections between two nearby vessels; the presence of bridges was assessed only in the 5cm^3 samples and the presence of densely packed vessels was assessed only in the whole tumor volume[14]. In case of disagreement between the two observers, consensus was reached by discussion and we used the agreed classification for statistical analysis. We did this to decrease the risk of bias introduced by relying on one single observer. Ultrasound images of the vascular features are shown in Figures 2 and 3.

Reference standard.

The reference standard was the histologic classification of the excised mass as malignant or benign. Histological examination was carried out at the local centers. Central pathology review was not performed, because in a previous IOTA study no clinically significant differences between local and central pathology reports were observed[6]. Malignant tumors were classified according to the criteria recommended by the International Federation of Gynecology and Obstetrics[17]. Borderline ovarian tumors were classified as malignant. The pathologists were blinded to the ultrasound findings.

Statistical analysis

We used SAS version 9.4 (SAS Institute, Cary, NC, USA) for all statistical analyses.

The statistical significance of differences in proportions was determined using the Chi-square test or Fisher´s exact test and that of differences in continuous data using Student´s t-test or Mann-Whitney test as appropriate.

We evaluated the discriminative ability of each vascular morphology feature, of subjective assessment, and of the IOTA model LR1 when using the recommended 10% risk of malignancy cutoff[6] and expressed it as sensitivity, specificity, positive and negative likelihood ratio. We plotted the sensitivity of each test against 1 – specificity in a receiver operating characteristic diagram. The test with its symbol farthest from the reference line was considered to have the best discriminative ability.

We estimated inter-observer reliability (agreement beyond chance) in the assessment of the vascular tree by calculating Cohen´s Kappa[18]. Kappa values of 0.81-1.0 were taken to indicate excellent reliability, 0.61-0.80 good reliability, and 0.41-0.60 moderate reliability[19].

We defined a P-value < 0.05 as statistically significant and corrected for multiple testing using the permutation method[20].

Sample size calculation

We aimed to collect information on 300 difficult masses of which we expected a minimum of 30% (n = 90) to be malignant[3,8]. Ninety malignancies would give us a reasonable 95% confidence interval (CI) around the point estimate for sensitivity. Because about 7% of all adnexal masses are difficult to classify as benign or malignant using subjective assessment[4], and because the IOTA model LR1 may yield an ambiguous test result in 10% of all adnexal masses[8], we estimated that we needed to examine about 2000 women with an adnexal mass (see Study protocol in Appendix 1 in supplementary material).

We report the study using the Standards for Reporting Diagnostic Accuracy studies (STARD) guidelines[22].

Results

In total, 2541 women with an adnexal mass were enrolled for inclusion in IOTA 3, but 138 women were excluded. Reasons for exclusion were: >120 days between ultrasound examination and surgery (n=66), pregnancy (n=31), data errors that could not be solved by contacting principal investigators (n=28), and incomplete final histology (n=13). The final IOTA 3 dataset included 2403 patients[15] of whom 376 (15.6%) had a difficult tumor: subjective assessment gave an uncertain result in 168 (7%) tumors, LR1 in 259 (11%) tumors, and both methods yielded an uncertain result in 51 (2%) tumors. Serous and mucinous cystadenomas/cystadenofibromas, fibromas, and borderline tumors were substantially more common among the difficult tumors than in the others, while endometrioma, benign teratoma (dermoid cyst), primary ovarian cancer and metastases in the ovaries from another primary tumor were substantially less common with 2-fold to 3-fold differences in prevalence (Table S1 and S2). The distribution of histological diagnoses was similar in tumors in which the ultrasound examiner was uncertain about the diagnosis and in those in which LR1 yielded an ambiguous result, with the exception that benign teratomas and endometriomas were more common in the latter (Table S1). Unilocular solid tumors, multilocular solid tumors and papillary projections were substantially more common in the difficult tumors than in the others, while unilocular cysts, ground glass echogenicity of cyst fluid, and color score 1 and 4 were substantially less common with 2-fold to 3-fold differences in prevalence (Table S3 and S4). The ultrasound characteristics of the tumors that the ultrasound examiner found difficult to classify were similar to those in which LR1 yielded an ambiguous result, with the exception that ascites was more common and unilocular cysts were less common in the tumors that the ultrasound examiner found difficult to classify (Table S3).

3D ultrasound volumes were available from 138 of the 376 difficult tumors. Six centers provided no volumes at all (0/55). Four centers provided volumes for more than 80% of their difficult masses (73/87), four centers for between 40% and 57% (47/96), three centers for between 11% and 18% (16/107), and one center for 6% (2/31) of their difficult tumors. In three centers the reason for providing volumes for none or only a small proportion of the difficult tumors (6/78) was that the volumes stored in the ultrasound system had been deleted and there was no back-up, and in one center the GE ultrasound system required was not always available for research. One center (with 33 difficult tumors) did not have access to a Voluson 730 Expert or GE E8 ultrasound system and so could not provide any volumes. Seven centers reported forgetfulness or transient technical problems to be the explanation for not providing volumes of all their difficult masses, six centers gave no explanation. The number of patients and the proportion of difficult tumors contributed from each center are shown in Table S5. Patient flow is described in Figure 4.

The volumes from the 138 difficult tumors available and analyzed included 79 (57%) masses in which subjective assessment gave an uncertain result, 87 (63%) cases in which LR1 gave an ambiguous result, and 28 (20%) cases in which both methods gave an uncertain result. The histological diagnoses of the 138 difficult tumors with available volumes are shown in Table 1. One hundred tumors were benign, 38 (27.5%) were malignant. The tumor mix was similar to that in the 238 difficult tumors for which tumor volumes were not available, with the exception that the proportion of serous and mucinous cystadenomas was slightly higher among the tumors with volumes available (Figure 4, Table S6). The clinical background data and 2D ultrasound findings for the 138 difficult tumors are described in Table 2. They were similar to those in the 238 difficult tumors for which tumor volumes were not available with the following exceptions: unilocular cysts, incomplete septae, tender mass

at scan and color score 1 were less common in the 138 difficult tumors for which tumor volumes were available, while multilocular-solid tumors and tumors with color score 4 were more common (Figure 4, Table S7).

The ability of subjective assessment, the IOTA model LR1 and the vascular features to discriminate between benign and malignant difficult tumors in terms of sensitivity and specificity is shown in Table 3 (with 95% CIs shown in Table S8). All vessel features differed statistically significantly between the benign and malignant difficult tumors. Branching vessels, densely packed vessels, caliber changes, tortuous vessels, color splashes, and bridges between vessels were more common in the malignant than in the benign difficult tumors. However, none of the vessel features discriminated well between the benign and malignant difficult tumors. Figure 5 shows plots of sensitivity against 1 – specificity for subjective assessment, for IOTA model LR1 when using the 10% risk cutoff, and for the vessel features with the best discriminative ability. Subjective assessment was the best method for discriminating between benign and malignant masses in the total study population of 138 difficult masses, followed by densely packed vessels in the whole tumor volume and tortuous vessels in the tumor biopsy. Among the 79 tumors that were difficult to classify as benign or malignant using subjective assessment, densely packed vessels in the whole tumor volume and tortuous vessels in the tumor biopsy had the best discriminative ability. For those tumors in which the IOTA model LR1 yielded an ambiguous result, subjective assessment had the best discriminative ability and caliber changes in the tumor biopsy the second best.

Inter-observer agreement and reliability with regard to vessel morphology are shown in Table 4. Inter-observer reliability was moderate or good with Cohen´s Kappa values ranging from 0.55 to 0.77. It was best for densely packed vessels, branching vessels and tortuous vessels in the whole tumor volume (Cohen´s kappa 0.77, 0.71 and 0.69, respectively) and for branching vessels in the 5 cm^3 tumor biopsy (Cohen´s Kappa 0.70).

Discussion

We have shown that vessel morphology depicted by 3D power Doppler ultrasound differs between benign and malignant difficult tumors. Branching vessels, caliber changes, color splashes, tortuous vessels, densely packed vessels and bridges between vessels were more common in malignant than in benign difficult tumors. However, none of the vascular features discriminated well between benign and malignant difficult tumors. We have confirmed that inter-observer reliability with regard to vessel morphology depicted by 3D power Doppler is moderate to good[14].

To the best of our knowledge, there are no published studies exploring the ability of the morphology of tumor vessels depicted by 3D power Doppler ultrasound to discriminate between benign and malignant difficult tumors. It is a strength of our study that vessel morphology was assessed by observers that had no clinical information, no information on 2D gray-scale or color Doppler ultrasound findings, and no information on the histological diagnosis of the tumors. This means that our results of the evaluation of the vessel tree are unbiased and reflect the true discriminative capacity of vessel morphology. A limitation of our study is possible selections bias, because not all centers sent 3D volumes of all their difficult tumors. The histology and ultrasound features differed slightly between the difficult tumors included (i.e. those with tumor volumes available) and not included (i.e. tumor volumes not available) in that the proportion of serous and mucinous cystadenomas, multilocular-solid tumors and tumors with color score 3 or 4 was higher among the difficult tumors included than excluded, while the proportion of unilocular cysts was lower (Figure 4, Table S6 and S7). However, the histological diagnoses and the ultrasound features of the difficult tumors with available tumor volumes were quite similar to those of all difficult

tumors included in IOTA 3 (Table S1). Therefore, our study sample should be reasonably representative of all difficult tumors. The small number of difficult masses with available volumes is another limitation. It makes our estimates of sensitivity and specificity imprecise (Table S8). However, because this is an exploratory study, we think this is acceptable.

Our results confirm those of a previous study[4] that serous and mucinous cystadenomas, fibromas and borderline tumors are difficult to classify as benign or malignant, and that unilocular-solid and multilocular-solid tumors and tumors with papillary projections are overrepresented among difficult tumors[4]. The proportion of tumors that the ultrasound examiner found difficult to classify in the current study is identical to that in the published study[4] (168/2403 versus 244/3511, i.e. 7% versus 7%).

To the best of our knowledge, the ability of vessel morphology depicted by 3D ultrasound to discriminate between benign and malignant adnexal masses or to decrease diagnostic errors has been explored in only one published study[14]. That study included 104 adnexal masses reasonably representative of a general population of tumors scheduled for surgery. In that study, too, all vascular features differed between benign and malignant tumors, but the vessel features discriminated much better between benign and malignant tumors than in the current study[14]. This is not surprising, because in the study cited the tumors were more heterogeneous with less overlapping ultrasound features. In the published study[14], adding a vascular feature variable to a logistic regression model including only gray scale ultrasound variables improved the discriminative ability of the model only minimally (AUC increased from 0.98 to 0.99), because the gray-scale model itself performed extremely well. The authors concluded that in an ordinary population of ovarian tumors, 3D power Doppler ultrasound examination adds little to grayscale imaging. However, they hypothesized that 3D power Doppler ultrasound examination with assessment of the morphology of tumor vessels might be useful in difficult tumors.

Our results show that if the IOTA model LR1 gives an ambiguous risk estimate (8.3 – 25.5%), then subjective assessment by an experienced ultrasound examiner is superior to using vessel morphology depicted by 3D power Doppler ultrasound as a second stage test. However, if a mass cannot be confidently classified as benign or malignant by an experienced ultrasound examiner using subjective assessment, then assessing vessel morphology with 3D power Doppler ultrasound could be of some help. Both densely packed vessels in the whole tumor volume and tortuous vessels in a 5 cm^3 tumor biopsy can be used for discrimination (Figure 5). We recommend using densely packed vessels, because inter-observer reliability was better for this variable than for tortuous vessels in a tumor biopsy. On the other hand, it is more time consuming to analyze a whole tumor volume than a tumor biopsy[21]. In our experience, for a very experienced ultrasound examiner it takes a minimum of 2.5 min to create a rotating 3D image of the vascular tree of a whole tumor, while it takes a minimum of 1 min to create one of a 5 cm^3 biopsy selected from the most vascularized part of the tumor. When assessing vessel morphology, it is important to be aware of the pitfalls of Doppler ultrasound and to ensure that Doppler settings are correct. If the tumor is far away from the ultrasound probe, it might not be possible to detect Doppler signals from the whole or parts of the tumor. This limits the clinical usefulness of vessel morphology to classify tumors as benign or malignant. Another limiting factor is the subjectivity of the method. Evaluation of vessel morphology, including “densely packed vessels”, is based on pattern recognition and therefore difficult to standardize or define.

Correct classification of adnexal masses as benign or malignant is a requirement for optimal management, i.e. conservative management with follow-up examinations, surgery in a local hospital, or referral to a center specialized in gynecological oncology[22]. However, some tumors are difficult to confidently classify as benign or malignant. Vessel morphology depicted by 3D power Doppler ultrasound showed limited ability to discriminate between benign and malignant difficult tumors. It remains to be shown if new biomarkers, immune cells, proteins, or genetic information can improve classification of difficult tumors as benign or malignant.

Acknowledgements

This study was supported by the Swedish Medical Research Council (grants no. K2006-73X-11605-11-3); by the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement; Landstingsfinansierad regional forskning (a Swedish governmental grant from the region of Scania); funds administered by Skåne University Hospital; and Allmänna Sjukhusets i Malmö Stiftelse för bekämpande av cancer (the Malmö General Hospital Foundation for fighting against cancer).

Participating centers

Oncology referral centers

1. Ospedale S. Gerardo, Universita di Milano Bicocca, Monza, Italy (**OIT**)
2. General Faculty Hospital of Charles University, Prague, Czech Republic (**PCR**)
3. Skåne University Hospital Lund, Sweden (**LSW**)
4. Karolinska University Hospital, Stockholm, Sweden (**SSW**)
5. Istituto Europeo di Oncologia, Milan, Italy (**CIT**)
6. Medical University in Lublin, Poland (**LPO**)
7. University Hospitals Leuven, Leuven, Belgium (**LBE**)
8. Universita Cattolica del Sacro Cuore, Rome, Italy (**RIT**)
9. Universita degli Studi di Udine, Italy (**UDI**)
10. Instituto Nationale dei Tumori, Fondazione Pascale, Naples, Italy (**GIT**)
11. University of Bologna, Bologna, Italy (**BIT**)

Other centers

12. Skåne University Hospital Malmö, Sweden (**MSW**)
13. Departament d'Obstetrícia, Ginecologia i Reproducció, Institut Universitari Dexeus, Barcelona, Spain (**BSP**)
14. University of Cagliari, Ospedale San Giovanni di Dio, Cagliari, Italy (**SIT**)
15. Ziekenhuis Oost-Limburg, Genk, Belgium (**GBE**)
16. Childrens' Hospital Buzzi, Milan, Italy (**FIT**)
17. Universita degli Studi di Napoli, Naples, Italy (**NIT**)
18. DCS Sacco University of Milan, Milan, Italy (**MIT**)

References

1. 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.
2. Timmerman D. The use of mathematical models to evaluate pelvic masses; can they beat an expert operator? *Best Pract Res Clin Obstet Gynaecol* 2004;**18**:91-104.
3. Valentin L, Ameye L, Jurkovic D, Metzger U, Lécuru F, Van Huffel S, Timmerman D. Which extrauterine pelvic masses are difficult to correctly classify as benign or malignant on the basis of ultrasound findings and is there a way of making a correct diagnosis? *Ultrasound Obstet Gynecol* 2006; **27:** 438-444.
4. Valentin L, Ameye L, Savelli L, Fruscio R, Leone FP, Czekierdowski A, Lissoni AA, Fischerova D, Guerriero S, Van Holsbeke C, Van Huffel S, Timmerman D. 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* 2011; **38:** 456-465.
5. Valentin L, Ameye L, Savelli L, Fruscio R, Leone FP, Czekierdowski A, Lissoni AA, Fischerova D, Guerriero S, Van Holsbeke C, Van Huffel S, Timmerman D. Unilocular adnexal cysts with papillary projections but no other solid components: is there a diagnostic method that can classify them reliably as benign or malignant before surgery? *Ultrasound Obstet Gynecol* 2013; **41:** 570-581.
6. Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, Van Calster B, Collins WP, Vergote I, Van Huffel S, Valentin L; International Ovarian Tumor Analysis Group. 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.
7. Timmerman D, Van Calster B, Testa AC, Guerriero S, Fischerova D, Lissoni AA, Van Holsbeke C, Fruscio R, Czekierdowski A, Jurkovic D, Savelli L, Vergote I, Bourne T, Van Huffel S, Valentin L. Ovarian cancer prediction in adnexal masses using ultrasound-based logistic regression models: a temporal and external validation study by the IOTA group. *Ultrasound Obstet Gynecol* 2010; **36:** 226-234.

8. Daemen A, Valentin L, Fruscio R, Van Holsbeke C, Melis GB, Guerriero S, Czekierdowski A, Jurkovic D, Ombelet W, Rossi A, Vergote I, Bourne T, De Moor B, Timmerman D. Improving the preoperative classification of adnexal masses as benign or malignant by second-stage tests. *Ultrasound Obstet Gynecol* 2011; **37:** 100-106.
9. 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 CA125. *J Natl Cancer Inst* 2007; **99:** 1706-1714.
10. Valentin L, Jurkovic D, Van Calster B, Testa A, Van Holsbeke C, Bourne T, Vergote I, Van Huffel S, Timmerman D. Adding a single CA125 measurement to ultrasound imaging performed by an experienced examiner does not improve preoperative discrimination between benign and malignant adnexal masses. *Ultrasound Obstet Gynecol* 2009; **34:** 345-354.
11. Valentin L. Imaging in gynecology. *Best Pract Res Clin Obstet Gynaecol* 2006; **20:** 881-906.
12. Kaijser J, Vandecaveye V, Deroose CM, Rockall A, Thomassin-Naggara I, Bourne T, Timmerman D. Imaging techniques for the pre-surgical diagnosis of adnexal tumours. *Best Pract Res Clin Obstet Gynaecol* 2014; **28:** 683-695.
13. Pereira PN, Sarian LO, Yoshida A, Araújo KG, Silva ACB, de Oliveira Barros RH, Jales RM, Derchain S. Improving the performance of IOTA simple rules: sonographic assessment of adnexal masses with resource-effective use of a magnetic resonance scoring (ADNEX MR scoring system). Abdom Radiol (NY). 2019 Sep 3. doi: 10.1007/s00261-019-02207-9. Epub ahead of print. PMID: 31482379.
14. Sladkevicius P, Jokubkiene L, Valentin L. Contribution of morphological assessment of the vessel tree by three-dimensional ultrasound to a correct diagnosis of malignancy in ovarian masses. *Ultrasound Obstet Gynecol* 2007; **30:** 874-882.
15. Testa A, Kaijser J, Wynants L, Fischerova D, Van Holsbeke C, Franchi D, Savelli L, Epstein E, Czekierdowski A, Guerriero S, Fruscio R, Leone FP, Vergote I, Bourne T, Valentin L, Van Calster B, Timmerman D. Strategies to diagnose ovarian cancer: new evidence from phase 3 of the multicentre international IOTA study. *Br J Cancer* 2014; **111:** 680-688.

16. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I; International Ovarian Tumor Analysis (IOTA) Group. 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.
17. Heintz APM, Odicino F, Maisonneuve P, Beller U, Benedet JL, Creasman WT, Ngan HYS, Pecorelli S. Carcinoma of the Ovary, 25th Annual Report on the Results of Treatment in Gynecological Cancer. *Int J Gynecol Obstet* 2003; **83:** S135-S166 (suppl 1).
18. Cohen's kappa Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; **20**: 37-46.
19. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ 1992; **304**;1491-1494.
20. Westfall PH, Wolfinger RD. Multiple tests with discrete distributions. *Am Stat* 1997; 51: 3–8.
21. Jokubkiene L, Sladkevicius P, Valentin L. Does three-dimensional power Doppler ultrasound help in discrimination between benign and malignant ovarian masses? *Ultrasound Obstet Gynecol* 2007; **29**: 215-225.
22. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, Irwig L, Levine D, Reitsma JB, de Vet HC, Bossuyt PM. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;**6**:e012799
23. Froyman W, Landolfo C, De Cock B, Sladkevicius P, Testa AC, Van Holsbeke C, Domali E, Fruscio R, Epstein E, Dos Santos Bernardo MJ, Franchi D, Kudla MJ, Chiappa V, Alcazar JL, Leone FPG, Buonomo F, Hochberg L, Coccia ME, Guerriero S, Deo N, Jokubkiene L, Kaijser J, Coosemans A, Vergote I, Verbakel JY, Bourne T, van Calster B, Valentin L, Timmerman D. Risk of complications in patients with conservatively managed ovarian tumours (IOTA5): a 2-year interim analysis of a multicentre, prospective, cohort study. *Lancet Oncol* 2019; **20**: 448-458.

LEGENDS

Figure 1. Schematic drawing illustrating the vascular features evaluated in our study a) straight vessel with no branching b) branching vessel c) tortuous vessel d) vessel with caliber changes e) bridges (short, straight connections between two nearby vessels). Reproduced from Sladkevicius P et al. *Ultrasound Obstet Gynecol* 2007; **30**: 874-882.

Figure 2. Three-dimensional ultrasound images of the vessel tree of ovarian tumors illustrating a) dispersed (as opposed to densely packed), straight (as opposed to tortuous), branching vessels in a benign mucinous cystadenoma; b) dispersed, branching, tortuous vessels (thick arrow) with caliber changes (thin arrow) in a mucinous borderline tumor; c) densely packed, branching, tortuous vessels with caliber changes and color splashes (arrow shows splashes) in a functional cyst; d) densely packed, branching, tortuous vessels with caliber changes in clear cell ovarian cancer. The corresponding rotating images of the vessel tree are shown in Supplementary videos 1-4.

Figure 3. Three-dimensional ultrasound images of the vessel tree of 5 cm^3 samples from ovarian tumors showing a) bridges (arrows), i.e. straight connections between two nearby vessels in struma ovarii; b) branching, tortuous vessels with caliber changes and color splashes (arrow) in benign ovarian fibroma; c) branching, tortuous vessels with caliber changes in ovarian endometroid carcinoma (color splashes are seen in the rotating image of the same tumor in Supplementary file 7). The corresponding rotating images of the vessel tree are shown in Supplementary videos 5-7.

Figure 4. Flow chart showing recruitment of patients to the study. The prevalence of the histological diagnoses and the ultrasound features that differed most between difficult tumors and not difficult tumors are shown.

Figure 5. Sensitivity plotted against 1 - specificity in a) all 138 difficult tumors b) in those 79 tumors in which the ultrasound examiner was uncertain whether the tumor was benign or malignant when using subjective assessment c) in those 89 tumors in which the International Ovarian Tumor Analysis (IOTA) model LR1 yielded an ambiguous result (risk of malignancy 8.3% to 25.5%). bx, biopsy; whole, whole tumor volume

Supplementary material

Appendix 1. International Ovarian Tumor Analysis (IOTA) Phase 3 study protocol

Appendix 2. The power Doppler and three-dimensional (3D) ultrasound settings used

Supplementary video 1. Three-dimensional 360 degree rotating ultrasound image of the vessel tree of a benign mucinous cystadenoma showing a) dispersed (as opposed to densely packed), straight (as opposed to tortuous), branching vessels. A still image of the same tumor is shown in Figure 2a.

Supplementary video 2 Three-dimensional 360 degree rotating ultrasound image of the vessel tree in a mucinous borderline tumor showing dispersed, branching, tortuous vessels with caliber changes. A still image of the same tumor is shown in Figure 2b.

Supplementary video 3 Three-dimensional 360 degree rotating ultrasound image of the vessel tree in a functional cyst showing densely packed, branching, tortuous vessels with caliber changes and color splashes. A still image of the same tumor is shown in Figure 2c.

Supplementary video 4 Three-dimensional 360 degree rotating ultrasound image of the vessel tree in a clear cell ovarian cancer showing densely packed, branching, tortuous vessels with caliber changes. A still image of the same tumor is shown in Figure 2d.

Supplementary video 5. Three-dimensional 360 degree rotating ultrasound image of the vessel tree of a 5 cm^3 sample from struma ovarii showing bridges, i.e. straight connections between two nearby vessels. A still image of the same tumor is shown in Figure 3a.

Supplementary video 6 Three-dimensional 360 degree rotating ultrasound image of the vessel tree of a 5 cm^3 sample from ovarian fibroma showing branching, tortuous vessels with caliber changes. A still image of the same tumor is shown in Figure 3b.

Supplementary video 7 Three-dimensional 360 degree rotating ultrasound image of the vessel tree of a 5 cm^3 sample from ovarian endometroid carcinoma showing branching, tortuous vessels with caliber changes and color splashes. A still image of the same tumor is shown in Figure 3c.

Supplementary Table S1. Histological diagnoses

Supplementary Table S2. Histological diagnosis in difficult and not difficult tumors

Supplementary Table S3. Clinical characteristics of patients and ultrasound characteristics of tumors

Supplementary Table S4. Clinical and ultrasound characteristics of difficult and not difficult tumors

Supplementary Table S5. Number of patients and the proportion of difficult tumors contributed from each center

Supplementary Table S6. Histological diagnosis in difficult tumors with tumor volumes available
versus not available

Supplementary Table S7. Clinical background data and ultrasound characteristics in difficult tumors
with tumor volumes available versus not available

Supplementary Table S8. The ability of vessel morphology, subjective assessment and logistic regression model 1, to discriminate between benign and malignant difficult tumors

Table 1. Histological diagnoses of 138 difficult tumors with volumes available

Diagnosis	**n**	**(%)**
Benign tumors	100	(72)
Endometrioma	7	(5)
Teratoma	5	(4)
Simple cyst or parasalpingeal cyst	3	(2)
Functional cyst	3	(2)
Hydrosalpinx or salpingitis	3	(2)
Peritoneal pseudocyst	2	(1)
Abscess	2	(1)
Fibroma	18	(13)
Serous cystadenoma/cystadenofibroma	30	(22)
Mucinous cystadenoma/cystadenofibroma	23	(17)
Rare benign tumor*	4	(3)
Borderline tumors	15	(11)
Stage I	14	(10)
Stage II	1	(<1)
Stage III or IV	0	0
Primary invasive tumors	22	(16)
Stage I	7	(5)
Stage II	1	(<1)
Stage III	8	(6)
Stage IV	0	0
Rare malignant tumor†	6	(4)
Metastasis in ovary from another primary tumor	1	(<1)

*Rare benign tumors include one Brenner tumor and three cases of struma ovarii
†Rare malignant tumors include three granulosa cell tumors, one immature teratoma, one Sertoli cell tumor and one gastrointestinal stroma tumor

Table 2. Clinical background data and ultrasound characteristics for 138 difficult tumors with tumor volumes available

	Volume available and analyzed (n = 138)	
Clinical variables		
Age, years	54 ± 17	
Postmenopausal	73	(53)
Hysterectomy	9	(7)
Hormonal replacement therapy	13	(9)
Personal history ovarian cancer	4	(3)
Family history ovarian cancer	3	(2)
CA125, U/mL (n =117)	20 (4 - 1302)	
Gray scale ultrasound variables		
Largest diameter, mm	69 (10 - 310)	
Bilateral	22	(16)
Ascites	5	(4)
Type of mass		
Unilocular	1	(<1)
Unilocular solid	23	(17)
Multilocular	32	(23)
Multilocular solid	56	(41)
Solid	26	(19)
Number of locules in case of multilocular or multilocular solid tumor		
2	10	(11)
3	8	(9)
4	8	(9)
5-10	23	(26)
>10	39	(44)
Tender mass at ultrasound examination	9	(7)
Echogenicity of cyst fluid		
Anechoic	36	(26)
Low level	47	(34)
Ground glass	10	(7)
Hemorrhagic	3	(2)
Mixed	16	(12)
No cyst fluid	26	(19)
Papillary projections present	45	(33)
Flow in papillations, if papillations present	20	(44)
Number of papillations	2 (1 - ≥4)	
Height of papillations, mm	7 (3 - 45)	
Mass with solid components	105	(76)
Largest diameter of largest solid component, mm	24 (3 - 180)	
Incomplete septum	2	(1)
Irregular walls	77	(56)
Shadows	20	(14)
Doppler ultrasound variables		
Color Score		
Score 1	10	(7)
Score 2	44	(32)
Score 3	70	(51)
Score 4	14	(10)

Results are presented as n (%) or median (min-max) except for age (mean ± SD)

Table 3. The ability of vessel morphology, subjective assessment, and logistic regression model 1 to correctly discriminate between benign and malignant difficult tumors

Diagnostic method	Sensitivity		Specificity		LR+	LR-	P-value
Either US examiner or LR1 uncertain (n = 138)							
Whole tumor vessel morphology							
Branching vessels	89%	(34/38)	33%	(33/100)	1.34	0.32	0.009
Densely packed vessels	63%	(24/38)	83%	(83/100)	3.72	0.44	<0.001
Caliber changes in vessels	66%	(25/38)	68%	(68/100)	2.06	0.50	<0.001
Splashes	50%	(19/38)	78%	(78/100)	2.27	0.64	0.002
Tortuous vessels	66%	(25/38)	70%	(70/100)	2.19	0.49	<0.001
Biopsy vessel morphology							
Branching vessels	84%	(32/38)	34%	(34/100)	1.28	0.46	0.04
Caliber changes in vessels	79%	(30/38)	63%	(63/100)	2.13	0.33	<0.001
Splashes	53%	(20/38)	69%	(69/100)	1.70	0.69	0.02
Tortuous vessels	79%	(30/38)	64%	(64/100)	2.19	0.33	<0.001
Bridges between vessels	42%	(16/38)	81%	(81/100)	2.22	0.72	0.007
Subjective assessment	74%	(28/38)	74%	(74/100)	2.83	0.36	<0.001
LR1 (10% risk cutoff)	92%	(35/38)	23%	(23/100)	1.20	0.34	0.03
US examiner uncertain (n = 79)							
Whole tumor vessel morphology							
Branching vessels	89%	(24/27)	33%	(17/52)	1.32	0.34	0.06
Densely packed vessels	67%	(18/27)	83%	(43/52)	3.85	0.40	<0.001
Caliber changes in vessels	63%	(17/27)	67%	(35/52)	1.93	0.55	0.01
Splashes	52%	(14/27)	83%	(43/52)	3.00	0.58	0.003
Tortuous vessels	63%	(17/27)	63%	(33/52)	1.72	0.58	0.02

Cont

Diagnostic method	Sensitivity		Specificity		LR+	LR-	P-value
Biopsy vessel morphology							
Branching vessels	85%	(23/27)	35%	(18/52)	1.30	0.43	0.07
Caliber changes in vessels	78%	(21/27)	62%	(32/52)	2.02	0.36	0.001
Splashes	59%	(16/27)	71%	(37/52)	2.05	0.57	0.009
Tortuous vessels	81%	(22/27)	67%	(35/52)	2.49	0.28	<0.001
Bridges between vessels	44%	(12/27)	79%	(41/52)	2.10	0.71	0.03
Subjective assessment	74%	(20/27)	60%	(31/52)	1.83	0.44	0.004
LR1 (10% risk cutoff)	89%	(24/27)	19%	(10/52)	1.10	0.58	0.34
LR1 uncertain (n = 87)							
Whole tumor vessel morphology							
Branching vessels	94%	(16/17)	30%	(21/70)	1.35	0.20	0.06
Densely packed vessels	53%	(9/17)	83%	(58/70)	3.09	0.57	0.004
Caliber changes in vessels	71%	(12/17)	69%	(48/70)	2.25	0.43	0.005
Splashes	47%	(8/17)	77%	(54/70)	2.06	0.69	0.07
Tortuous vessels	65%	(11/17)	71%	(50/70)	2.27	0.49	0.01
Biopsy vessel morphology							
Branching vessels	76%	(13/17)	31%	(22/70)	1.12	0.75	0.77
Caliber changes in vessels	76%	(13/17)	64%	(45/70)	2.14	0.37	0.005
Splashes	47%	(8/17)	69%	(48/70)	1.50	0.77	0.26
Tortuous vessels	71%	(12/17)	60%	(42/70)	1.77	0.49	0.03
Bridges between vessels	35%	(6/17)	81%	(57/70)	1.90	0.80	0.19
Subjective assessment	82%	(14/17)	79%	(55/70)	3.84	0.23	<0.001
LR1 (10% risk cutoff)	100%	(17/17)	19%	(13/70)	1.23	not possible to calculate	0.06

LR+, positive likelihood ratio; LR-, negative likelihood ratio; LR1, logistic regression model 1 using the 10% risk cutoff to predict malignancy[6,7]
Cont.
Correction for multiple testing has not been done because this is an exploratory analysis

Table 4. Interobserver agreement with regard to describing vessel morphology in 138 difficult tumors

	% agreement (95% CI)	Kappa (95% CI)
Whole tumor		
Branching	88 (82 to 93)	0.71 (0.58 to 0.84)
Densely packed	90 (84 to 94)	0.77 (0.65 to 0.88)
Caliber changes	78 (71 to 84)	0.55 (0.41 to 0.69)
Splashes	81 (74 to 87)	0.57 (0.42 to 0.71)
Tortuous	85 (78 to 90)	0.69 (0.56 to 0.81)
Biopsy		
Branching	88 (81 to 92)	0.70 (0.57 to 0.83)
Caliber changes	78 (70 to 84)	0.55 (0.41 to 0.69)
Splashes	86 (79 to 90)	0.67 (0.54 to 0.80)
Tortuous	83 (75 to 88)	0.65 (0.52 to 0.78)
Bridges	86 (80 to 91)	0.63 (0.47 to 0.78)

Fig. 4

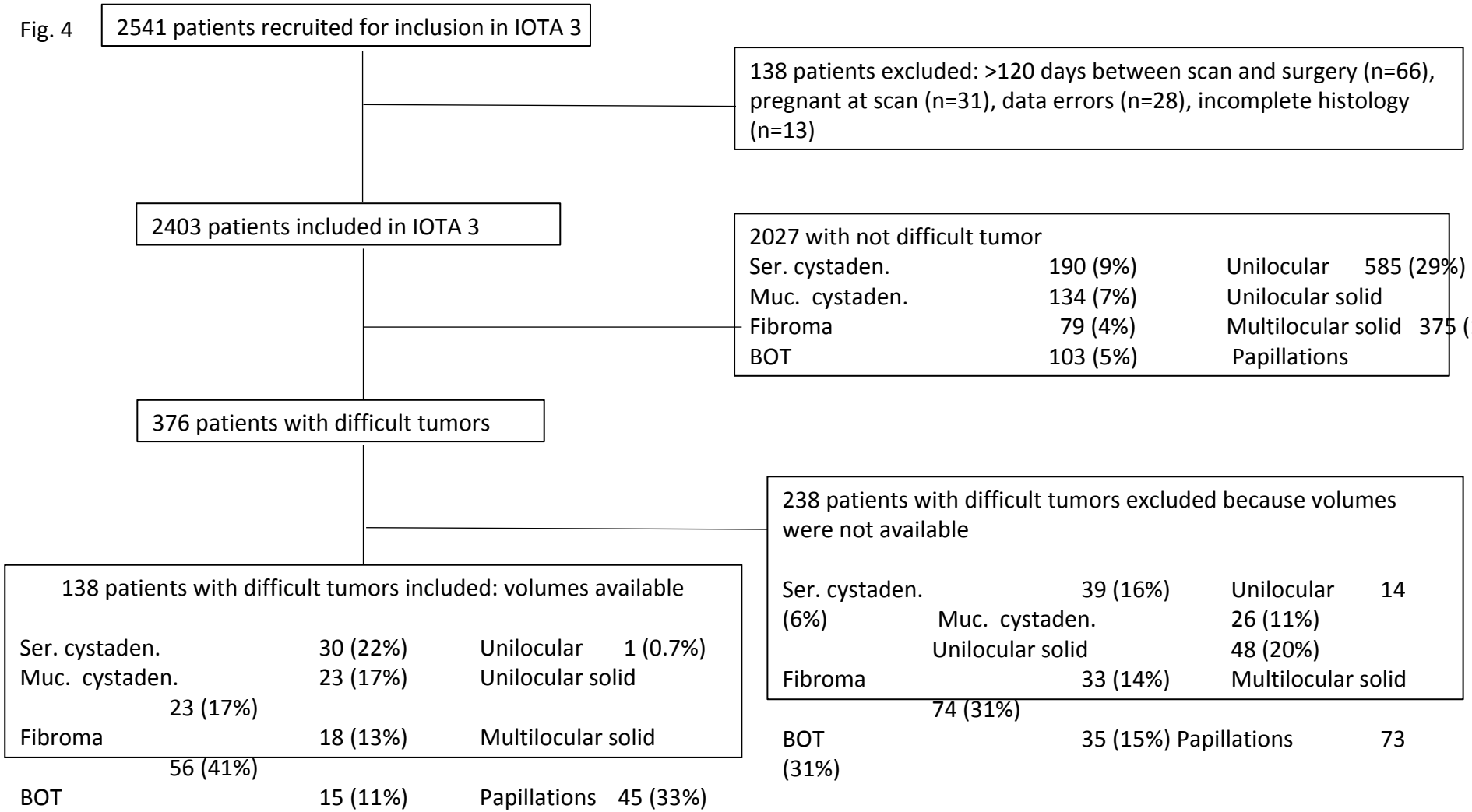

2541 patients recruited for inclusion in IOTA 3
138 patients excluded: >120 days between scan and surgery (n=66), pregnant at scan (n=31), data errors (n=28), incomplete histology (n=13)
2403 patients included in IOTA 3
2027 with not difficult tumor
Ser. cystaden. 190 (9%)
Muc. cystaden. 134 (7%)
Fibroma 79 (4%)
BOT 103 (5%)
Unilocular 585 (29%)
Unilocular solid
Multilocular solid 375 (1
Papillations
376 patients with difficult tumors
238 patients with difficult tumors excluded because volumes were not available
Ser. cystaden. 39 (16%)
Muc. cystaden. 26 (11%)
Fibroma 33 (14%)
BOT 35 (15%)
Unilocular 14 (6%)
Unilocular solid 48 (20%)
Multilocular solid 74 (31%)
Papillations 73 (31%)
138 patients with difficult tumors included: volumes available
Ser. cystaden. 30 (22%)
Muc. cystaden. 23 (17%)
Fibroma 18 (13%)
BOT 15 (11%)
Unilocular 1 (0.7%)
Unilocular solid 23 (17%)
Multilocular solid 56 (41%)
Papillations 45 (33%)

Table S1. Histological diagnoses

	Total		ALL										3D volume analyzed							
			Both US examiner and LR1 not uncertain		Either US examiner or LR1 uncertain		US examiner uncertain		LR1 uncertain		Both uncertain		Either US examiner or LR1 uncertain		US examiner uncertain		LR1 uncertain		Both uncertain	
	N=2403		N=2027		N=376		N=168		N=259		N=51		N=138		N=79		N=87		N=28	
Benign	**1423**	**(59%)**	**1169**	**(58%)**	**254**	**(68%)**	**111**	**(66%)**	**180**	**(70%)**	**37**	**(73%)**	**100**	**(72%)**	**52**	**(66%)**	**70**	**(80%)**	**22**	**(79%)**
Endometrioma	344	(14%)	324	(16%)	20	(5%)	5	(3%)	16	(6%)	1	(2%)	7	(5%)	2	(3%)	5	(6%)	-	-
Teratoma	231	(10%)	212	(10%)	19	(5%)	5	(3%)	15	(6%)	1	(2%)	5	(4%)	1	(1%)	4	(5%)	-	-
Simple cyst + parasalpingeal cyst	106	(4%)	96	(5%)	10	(3%)	5	(3%)	7	(3%)	2	(4%)	3	(2%)	1	(1%)	3	(3%)	1	(4%)
Functional cyst	40	(2%)	29	(1%)	11	(4%)	6	(4%)	7	(3%)	2	(4%)	3	(2%)	2	(3%)	2	(2%)	1	(4%)
Hydrosalpinx + salpingitis	47	(2%)	40	(2%)	7	(2%)	3	(2%)	5	(2%)	1	(2%)	3	(2%)	1	(1%)	3	(3%)	1	(4%)
Peritoneal pseudocyst	18	(<1%)	14	(<1%)	4	(1%)	2	(1%)	3	(1%)	1	(2%)	2	(1%)	1	(1%)	2	(2%)	1	(4%)
Abscess	17	(<1%)	14	(<1%)	3	(<1%)	1	(<1%)	2	(<1%)	-	-	2	(1%)	1	(1%)	1	(1%)	-	-
Fibroma	130	(5%)	79	(4%)	51	(14%)	22	(13%)	33	(13%)	4	(8%)	18	(13%)	9	(11%)	10	(11%)	1	(4%)
Serous cystadenoma	259	(11%)	190	(9%)	69	(18%)	35	(21%)	50	(19%)	16	(31%)	30	(22%)	21	(27%)	21	(24%)	12	(43%)
Mucinous cystadenoma	183	(8%)	134	(7%)	49	(13%)	21	(13%)	32	(12%)	4	(8%)	23	(17%)	10	(13%)	16	(18%)	3	(11%)
Rare benign	48	(2%)	37	(2%)	11	(3%)	6	(4%)	10	(4%)	5	(10%)	4	(3%)	3	(4%)	3	(3%)	2	(7%)
Borderline	**153**	**(6%)**	**103**	**(5%)**	**50**	**(13%)**	**24**	**(14%)**	**31**	**(12%)**	**5**	**(10%)**	**15**	**(11%)**	**7**	**(9%)**	**9**	**(10%)**	**1**	**(4%)**
Stage I	135	(6%)	86	(4%)	49	(13%)	24	(14%)	30	(12%)	5	(10%)	14	(10%)	7	(9%)	8	(9%)	1	(4%)
Stage II	6	(<1%)	5	(<1%)	1	(<1%)	-	-	1	(<1%)	-	-	1	(<1%)	-	-	1	(1%)	-	-
Stage III	12	(<1%)	12	(<1%)	-		-	-	-	-	-	-	-		-	-	-	-	-	-
Primary invasive	**701**	**(29%)**	**637**	**(31%)**	**64**	**(17%)**	**29**	**(17%)**	**44**	**(17%)**	**9**	**(18%)**	**22**	**(16%)**	**19**	**(24%)**	**8**	**(9%)**	**5**	**(18%)**
Stage I	128	(5%)	103	(5%)	25	(7%)	12	(7%)	17	(7%)	4	(8%)	7	(5%)	6	(8%)	3	(3%)	2	(7%)
Stage II	47	(2%)	44	(2%)	3	(<1%)	1	(<1%)	2	(<1%)	-	-	1	(<1%)	1	(1%)	-	-	-	-
Stage III	397	(17%)	378	(19%)	19	(5%)	8	(5%)	13	(5%)	2	(4%)	8	(6%)	6	(8%)	3	(3%)	1	(7%)
Stage IV	61	(3%)	60	(3%)	1	(<1%)	-	-	1	(<1%)	-	-	-	-	-	-	-	-	-	-
Rare	68	(3%)	52	(3%)	16	(4%)	8	(5%)	11	(4%)	3	(6%)	6	(4%)	6	(8%)	2	(2%)	2	(4%)
Metastatic	**126**	**(5%)**	**118**	**(6%)**	**8**	**(2%)**	**4**	**(2%)**	**4**	**(2%)**	**-**	**-**	**1**	**(<1%)**	**1**	**(1%)**	**-**	**-**	**-**	**-**

US, ultrasound; LR1, logistic regression model 1; 3D, three-dimensional

Table S2. Histological diagnosis in difficult and not difficult tumors

	Both US examiner and LR1 not uncertain N=2027		Either US examiner or LR1 uncertain N=376		P-value
Benign	**1169**	**(58%)**	**254**	**(68%)**	**<0.001**
Endometrioma	324	(16%)	20	(5%)	<0.001
Teratoma	212	(10%)	19	(5%)	0.01
Simple cyst + parasalpingeal cyst	96	(5%)	10	(3%)	0.72
Functional cyst	29	(1%)	11	(4%)	0.54
Hydrosalpinx + salpingitis	40	(2%)	7	(2%)	1
Peritoneal pseudocyst	14	(<1%)	4	(1%)	1
Abscess	14	(<1%)	3	(<1%)	1
Fibroma	79	(4%)	51	(14%)	<0.001
Serous cystadenoma	190	(9%)	69	(18%)	<0.001
Mucinous cystadenoma	134	(7%)	49	(13%)	<0.001
Rare benign	37	(2%)	11	(3%)	0.94
Borderline	**103**	**(5%)**	**50**	**(13%)**	**<0.001**
Stage I	86	(4%)	49	(13%)	<0.001
Stage II	5	(<1%)	1	(<1%)	1
Stage III	12	(<1%)	-		0.99
Primary invasive	**637**	**(31%)**	**64**	**(17%)**	**<0.001**
Stage I	103	(5%)	25	(7%)	0.98
Stage II	44	(2%)	3	(<1%)	0.82
Stage III	378	(19%)	19	(5%)	<0.001
Stage IV	60	(3%)	1	(<1%)	0.008
Rare	52	(3%)	16	(4%)	0.77
Metastatic	**118**	**(6%)**	**8**	**(2%)**	**0.008**

P-values are corrected for multiple testing with the permutation method (Westfall PH, Wolfinger RD. Multiple tests with discrete distributions. *Am Stat* 1997; **51**: 3-8)
US, ultrasound; LR1, logistic regression model 1

Table S3. Clinical characteristics of patients and ultrasound characteristics of tumors

		ALL					3D volume analyzed			
	Total	Both US examiner and LR1 not uncertain	Either US examiner or LR1 uncertain	US examiner uncertain	LR1 uncertain	Both uncertain	Either US examiner or LR1 uncertain	US examiner uncertain	LR1 uncertain	Both uncertain
	N=2403	N=2027	N=376	N=168	N=259	N=51	N=138	N=79	N=87	N=28
Clinical variables										
Age, years	50 ± 16	49 ± 16	53 ± 16	52 ± 16	53 ± 16	49 ± 14	54 ± 17	54 ± 18	51 ± 16	49 ± 13
Postmenopausal	1049 (44%)	861 (42%)	188 (50%)	79 (47%)	130 (50%)	21 (41%)	73 (53%)	41 (52%)	44 (51%)	12 (43%)
Hysterectomy	142 (6%)	115 (6%)	27 (7%)	15 (9%)	19 (7%)	7 (14%)	9 (7%)	6 (8%)	7 (8%)	4 (14%)
Hormonal replacement therapy	207 (9%)	174 (9%)	33 (9%)	26 (15%)	13 (5%)	6 (12%)	13 (9%)	12 (15%)	4 (5%)	3 (11%)
Personal history ovarian cancer	44 (2%)	36 (2%)	8 (2%)	2 (1%)	6 (2%)	- -	4 (3%)	1 (1%)	3 (3%)	- -
Family history ovarian cancer	74 (3%)	69 (3%)	5 (1%)	2 (1%)	3 (1%)	- -	3 (2%)	1 (1%)	2 (2%)	- -
CA125, N available	1451	1198	253	114	175	36	117	67	73	23
CA125	42 (1 – 14067)	52 (1 – 14067)	23 (3 – 1948)	24 (6 – 906)	21 (3 – 1948)	18 (6 – 313)	20 (4 – 1302)	21 (7 – 906)	18 (4 – 1302)	17 (7 – 313)
Gray scale ultrasound variables										
Largest diameter, mm	71 (10 – 550)	70 (10 – 550)	75 (10 – 322)	72 (10 – 322)	74 (10 – 300)	62 (10 – 169)	69 (10 – 310)	68 (10 – 310)	68 (10 – 300)	64 (10 – 122)
Bilateral	518 (22%)	465 (23%)	53 (14%)	16 (10%)	41 (16%)	4 (8%)	22 (16%)	9 (11%)	16 (18%)	3 (11%)
Ascites	340 (14%)	330 (16%)	10 (3%)	8 (5%)	2 (<1%)	- -	5 (4%)	5 (6%)	- -	- -
Type of mass										
Unilocular	600 (25%)	585 (29%)	15 (4%)	3 (2%)	12 (5%)	- -	1 (<1%)	- -	1 (1%)	- -
Unilocular solid	258 (11%)	187 (9%)	71 (19%)	40 (24%)	46 (18%)	15 (29%)	23 (17%)	17 (22%)	14 (16%)	8 (29%)
Multilocular	413 (17%)	331 (16%)	82 (22%)	30 (18%)	60 (23%)	8 (16%)	32 (23%)	12 (15%)	23 (26%)	3 (11%)
Multilocular solid	505 (21%)	375 (19%)	130 (35%)	59 (35%)	89 (34%)	18 (35%)	56 (41%)	34 (43%)	36 (41%)	14 (50%)
Solid	627 (26%)	549 (27%)	78 (21%)	36 (21%)	52 (20%)	10 (20%)	26 (19%)	16 (20%)	13 (15%)	3 (11%)
Nr locules	1 (0 to >10)	1 (0 to >10)	2 (0 to >10)	2 (0 to >10)	2 (0 to >10)	2 (0 to >10)	4 (0 to >10)	2 (0 to >10)	3 (0 to >10)	2 (0 to >10)
Pain at US examination	344 (14%)	294 (15%)	50 (13%)	20 (12%)	32 (12%)	2 (4%)	9 (7%)	4 (5%)	5 (6%)	- -
Echogenicity of cyst fluid										
Anechoic	589 (25%)	485 (24%)	104 (28%)	39 (23%)	77 (30%)	12 (24%)	36 (26%)	17 (22%)	27 (31%)	8 (29%)
Low level	516 (21%)	392 (19%)	124 (33%)	63 (38%)	82 (32%)	21 (41%)	47 (34%)	32 (41%)	28 (32%)	13 (46%)
Ground glass	350 (15%)	326 (16%)	24 (6%)	9 (5%)	16 (6%)	1 (2%)	10 (7%)	4 (5%)	6 (7%)	- -
Hemorrhagic	20 (<1%)	16 (<1%)	4 (1%)	2 (1%)	2 (<1%)	- -	3 (2%)	2 (3%)	1 (1%)	- -
Mixed	301 (13%)	259 (13%)	42 (11%)	19 (11%)	30 (12%)	7 (14%)	16 (12%)	8 (10%)	12 (14%)	4 (14%)
No cyst fluid	627 (26%)	549 (27%)	78 (21%)	36 (21%)	52 (20%)	10 (20%)	26 (19%)	16 (20%)	13 (15%)	3 (11%)
Papillary projections present	383 (16%)	265 (13%)	118 (31%)	69 (41%)	73 (28%)	24 (47%)	45 (33%)	35 (44%)	25 (29%)	15 (54%)
Flow in papillation	215 (56%)	169 (64%)	46 (39%)	32 (46%)	21 (29%)	7 (29%)	20 (44%)	16 (46%)	8 (32%)	4 (27%)
Number of papillations	2 (1 – ≥4)	2 (1 – ≥ 4)	2 (1 – ≥ 4)	2 (1 – ≥ 4)	1 (1 – ≥ 4)	2 (1 – ≥ 4)	2 (1 – ≥4)	2 (1 – ≥4)	1 (1 – ≥ 4)	2 (1 – ≥4)
Height of papillation, mm	10 (3 – 99)	11 (3 – 99)	7 (3 – 45)	8 (3 – 45)	6 (3 – 30)	7 (3 – 30)	7 (3 – 45)	7 (3 – 45)	5 (3 – 21)	6 (3 – 13)

Cont.

Table S3 continued

		ALL					3D volume analyzed			
	Total N=2403	Both US examiner and LR1 not uncertain N=2027	Either US examiner or LR1 uncertain N=376	US examiner uncertain N=168	LR1 uncertain N=259	Both uncertain N=51	Either US examiner or LR1 uncertain N=138	US examiner uncertain N=79	LR1 uncertain N=87	Both uncertain N=28
Mass with solid components	1390 (58%)	1111 (55%)	279 (74%)	135 (80%)	187 (72%)	43 (84%)	105 (76%)	67 (85%)	63 (72%)	25 (89%)
Largest diameter of largest solid component, mm	50 (3 – 300)	54 (3 – 300)	25 (3 – 200)	28 (3 – 200)	22 (3 – 196)	19 (5 – 112)	24 (3 – 180)	25 (3 – 162)	16 (3 – 180)	13 (5 – 112)
Incomplete septum	113 (5%)	88 (4%)	25 (7%)	9 (5%)	17 (7%)	1 (2%)	2 (1%)	1 (1%)	1 (1%)	- -
Irregular walls	957 (40%)	740 (37%)	217 (58%)	113 (67%)	141 (54%)	37 (73%)	77 (56%)	55 (70%)	44 (51%)	22 (79%)
Shadows	299 (12%)	239 (12%)	60 (16%)	20 (12%)	47 (18%)	7 (14%)	20 (14%)	10 (13%)	13 (15%)	3 (11%)
Doppler ultrasound variables										
Color Score										
Score 1	606 (25%)	553 (27%)	53 (14%)	21 (13%)	36 (14%)	4 (8%)	10 (7%)	5 (6%)	7 (8%)	2 (7%)
Score 2	762 (32%)	604 (30%)	158 (42%)	64 (38%)	115 (44%)	21 (41%)	44 (32%)	24 (30%)	31 (36%)	11 (39%)
Score 3	681 (28%)	539 (27%)	142 (38%)	71 (42%)	94 (36%)	23 (45%)	70 (51%)	40 (51%)	42 (48%)	12 (43%)
Score 4	354 (15%)	331 (16%)	23 (6%)	12 (7%)	14 (5%)	3 (6%)	14 (10%)	10 (13%)	7 (8%)	3 (11%)

US, ultrasound; LR1, logistic regression model 1 ; 3D, three-dimensional
Results are presented as n (%) or median (min-max) except for age (mean ± SD)

Table S4. Clinical and ultrasound characteristics of difficult and not difficult tumors

	Both US examiner and LR1 not uncertain N=2027		Either US examiner or LR1 uncertain N=376		P-value
Clinical variables					
Age, years	49 ± 16		53 ± 16		<0.001
Postmenopausal	861	(42%)	188	(50%)	0.007
Hysterectomy	115	(6%)	27	(7%)	0.27
Hormonal replacement therapy	174	(9%)	33	(9%)	0.90
Personal history ovarian cancer	36	(2%)	8	(2%)	0.67
Family history ovarian cancer	69	(3%)	5	(1%)	0.03
CA125, N available	1198		253		
CA125	52 (1 – 14067)		23 (3 – 1948)		<0.001
Gray scale ultrasound variables					
Largest diameter, mm	70 (10 – 550)		75 (10 – 322)		0.006
Bilateral	465	(23%)	53	(14%)	<0.001
Ascites	330	(16%)	10	(3%)	<0.001
Type of mass					<0.001
Unilocular	585	(29%)	15	(4%)	<0.001
Unilocular solid	187	(9%)	71	(19%)	<0.001
Multilocular	331	(16%)	82	(22%)	0.05
Multilocular solid	375	(19%)	130	(35%)	<0.001
Solid	549	(27%)	78	(21%)	0.04
Nr locules	1 (0 to >10)		2 (0 to >10)		<0.001
Pain at US examination	294	(15%)	50	(13%)	0.54
Echogenicity of cyst fluid					
Anechoic	485	(24%)	104	(28%)	0.52
Low level	392	(19%)	124	(33%)	<0.001
Ground glass	326	(16%)	24	(6%)	<0.001
Hemorrhagic	16	(<1%)	4	(1%)	0.98
Mixed	259	(13%)	42	(11%)	0.94
No cyst fluid	549	(27%)	78	(21%)	0.05
Papillary projections present	265	(13%)	118	(31%)	<0.001
Flow in papillation	169	(64%)	46	(39%)	<0.001
Number of papillations	2 (1 – ≥ 4)		2 (1 – ≥ 4)		0.006
Height of papillation, mm	11 (3 – 99)		7 (3 – 45)		<0.001
Mass with solid components	1111	(55%)	279	(74%)	<0.001
Largest diameter of largest solid component, mm	54 (3-300)		25 (3-200)		<0.001
Incomplete septum	88	(4%)	25	(7%)	0.06
Irregular walls	740	(37%)	217	(58%)	<0.001
Shadows	239	(12%)	60	(16%)	0.03
Doppler ultrasound variables					
Color Score					<0.001
Score 1	553	(27%)	53	(14%)	<0.001
Score 2	604	(30%)	158	(42%)	<0.001
Score 3	539	(27%)	142	(38%)	<0.001
Score 4	331	(16%)	23	(6%)	<0.001

US, ultrasound; LR, logistic regression model 1
Results are presented as n (%) or median (min-max) except for age (mean ± SD)
The P-values presented are corrected for multiple testing using the permutation method (Westfall PH, Wolfinger RD. Multiple tests with discrete distributions. *Am Stat* 1997; **51**: 3-8)

Table S5. Number of patients and the proportion of difficult tumors contributed from each center

	ALL										3D volume analyzed							
Center	**Total** n	(%)	Both US examiner and LR1 not uncertain		Either US examiner or LR1 uncertain		US examiner uncertain		LR1 uncertain		Both uncertain		Either US examiner or LR1 uncertain		US examiner uncertain		LR1 uncertain		Both uncertain	
BIT	213	(9%)	180	(9%)	33	(9%)	15	(9%)	21	(8%)	3	(6%)	-	-	-	-	-	-	-	-
BSP	37	(2%)	31	(2%)	6	(2%)	2	(1%)	4	(2%)	-	-	5	(4%)	2	(3%)	3	(3%)	-	-
CIT	218	(9%)	196	(10%)	22	(6%)	4	(2%)	21	(8%)	3	(6%)	11	(8%)	3	(4%)	10	(11%)	2	(7%)
FIT	21	(<1%)	20	(<1%)	1	(<1%)	1	(<1%)	-	-	-	-	-	-	-	-	-	-	-	-
GBE	228	(9%)	192	(9%)	36	(10%)	19	(11%)	24	(9%)	7	(14%)	4	(3%)	2	(3%)	3	(3%)	1	(4%)
GIT	6	(<1%)	5	(<1%)	1	(<1%)	1	(<1%)	-	-	-	-	-	-	-	-	-	-	-	-
LBE	129	(5%)	98	(5%)	31	(8%)	17	(10%)	16	(6%)	2	(4%)	2	(1%)	1	(1%)	1	(1%)	-	-
LPO	131	(5%)	117	(6%)	14	(4%)	4	(2%)	10	(4%)	-	-	2	(1%)	1	(1%)	1	(1%)	-	-
LSW	39	(2%)	30	(1%)	9	(2%)	6	(4%)	6	(2%)	3	(6%)	4	(3%)	3	(4%)	3	(3%)	2	(7%)
MIT	86	(4%)	75	(4%)	11	(3%)	-	-	11	(4%)	-	-	-	-	-	-	-	-	-	-
MSW	201	(8%)	141	(7%)	60	(16%)	32	(19%)	39	(15%)	11	(22%)	50	(36%)	31	(39%)	30	(34%)	11	(39%)
NIT	8	(<1%)	7	(<1%)	1	(<1%)	-	-	1	(<1%)	-	-	-	-	-	-	-	-	-	-
OIT	105	(4%)	91	(4%)	14	(4%)	7	(4%)	10	(4%)	3	(6%)	12	(9%)	7	(9%)	8	(9%)	3	(11%)
PCR	264	(11%)	234	(12%)	30	(8%)	3	(2%)	29	(11%)	2	(4%)	12	(9%)	3	(4%)	11	(13%)	2	(7%)
RIT	443	(18%)	386	(19%)	57	(15%)	26	(15%)	37	(14%)	6	(12%)	10	(7%)	9	(11%)	4	(5%)	3	(11%)
SIT	107	(4%)	100	(5%)	7	(2%)	1	(<1%)	6	(2%)	-	-	6	(4%)	1	(1%)	5	(6%)	-	-
SSW	120	(5%)	85	(4%)	35	(9%)	27	(16%)	16	(6%)	8	(16%)	20	(14%)	16	(20%)	8	(9%)	4	(14%)
UDI	47	(2%)	39	(2%)	8	(2%)	3	(2%)	8	(3%)	3	(6%)	-	-	-	-	-	-	-	-
All centers	**2403**		**2027**		**376**		**168**		**259**		**51**		**138**		**79**		**87**		**28**	

Percentages are calculated per column
US, ultrasound; LR1, logistic regression model 1; 3D, three-dimensional

BIT, Bologna, Italy; BSP, Barcelona, Spain; CIT, European Institute of Oncology, Milan, Italy; FIT, Children´s Hospital Buzzi, Milan Italy; GBE, Genk, Belgium; GIT, Instituto Nationale dei Tumori, Naples, Italy; LBE, Leuven, Belgium; LPO, Lublin, Poland; LSW, Lund, Sweden; MIT, Sacco University, Milan, Italy; MSW, Malmoe, Sweden; NIT, Universita degli Studi di Napoli, Naples, Italy; OIT, Monza, Italy; PCR, Prague, Czeck Republic; RIT, Rome, Italy; SIT, Cagliari, Italy; SSW, Stockholm, Sweden; UDI, Udine, Italy

Table S6. Histological diagnosis in difficult tumors with tumor volumes available versus not available

	Either ultrasound examiner or LR1 uncertain N=376				
	Volume not available N=238		Volume available and analyzed N=138		P-value
Benign	**154**	**(65%)**	**100**	**(72%)**	**0.32**
Endometrioma	13	(5%)	7	(5%)	
Teratoma	14	(6%)	5	(4%)	
Simple cyst or parasalpingeal cyst	7	(3%)	3	(2%)	
Functional cyst	8	(3%)	3	(2%)	
Hydrosalpinx or salpingitis	4	(2%)	3	(2%)	
Peritoneal pseudocyst	2	(<1%)	2	(1%)	
Abscess	1	(<1%)	2	(1%)	
Fibroma	33	(14%)	18	(13%)	
Serous cystadenoma	39	(16%)	30	(22%)	
Mucinous cystadenoma	26	(11%)	23	(17%)	
Rare benign	7	(3%)	4	(3%)	
Borderline	**35**	**(15%)**	**15**	**(11%)**	**0.69**
Stage I	35	(15%)	14	(10%)	
Stage II	0	0	1	(<1%)	
Stage III or IV	0	0	0	0	
Primary invasive	**42**	**(18%)**	**22**	**(16%)**	**0.98**
Stage I	18	(8%)	7	(5%)	
Stage II	2	(<1%)	1	(<1%)	
Stage III	11	(5%)	8	(6%)	
Stage IV	1	(<1%)	0	0	
Rare	10	(4%)	6	(4%)	
Metastatic	**7**	**(3%)**	**1**	**(<1%)**	**0.60**

The P-values presented have been corrected for multiple testing using the permutation method (Westfall PH, Wolfinger RD. Multiple tests with discrete distributions. *Am Stat* 1997; **51**: 3-8)

Table S7. Clinical background data and ultrasound characteristics in difficult tumors with tumor volumes available versus not available

	Either ultrasound examiner or LR1 uncertain		
	Volume not available N=238	Volume available and analyzed N=138	P-value
Clinical variables			
Age, years	52 ± 16	54 ± 17	0.50
Postmenopausal	115 (48%)	73 (53%)	0.39
Hysterectomy	18 (8%)	9 (7%)	0.84
Hormonal replacement therapy	20 (8%)	13 (9%)	0.74
Personal history ovarian cancer	4 (2%)	4 (3%)	0.47
Family history ovarian cancer	2 (<1%)	3 (2%)	0.36
CA125, number available	136	117	
CA125, U/mL	29 (3 - 1948)	20 (4 - 1302)	0.61
Gray scale ultrasound variables			
Largest diameter, mm	80 (14-322)	69 (10-310)	0.06
Bilateral	31 (13%)	22 (16%)	0.44
Ascites	5 (2%)	5 (4%)	0.51
Type of mass			0.06
Unilocular	14 (6%)	1 (<1%)	
Unilocular solid	48 (20%)	23 (17%)	
Multilocular	50 (21%)	32 (23%)	
Multilocular solid	74 (31%)	56 (41%)	
Solid	52 (22%)	26 (19%)	
Number of locules if multilocular or multilocular solid			0.17
2	19 (15%)	10 (11%)	
3	16 (13%)	8 (9%)	
4	9 (7%)	8 (9%)	
5-10	37 (30%)	23 (26%)	
>10	43 (35%)	39 (44%)	
Tender mass at ultrasound examination	41 (17%)	9 (7%)	0.003
Echogenicity of cyst fluid			0.65
Anechoic	68 (29%)	36 (26%)	
Low level	77 (32%)	47 (34%)	
Ground glass	14 (6%)	10 (7%)	
Hemorrhagic	1 (<1%)	3 (2%)	
Mixed	26 (11%)	16 (12%)	
No cyst fluid	52 (22%)	26 (19%)	
Papillary projections present	73 (31%)	45 (33%)	0.70
Flow in papillation, if papillation present	26 (36%)	20 (44%)	0.34
Number of papillations	2 (1 – ≥4)	2 (1 – ≥4)	0.92
Height of papillation, mm	7 (3 – 30)	7 (3 – 45)	0.27
Mass with solid components	174 (73%)	105 (76%)	0.52
Largest diameter of largest solid component, mm	26 (4-200)	24 (3-180)	0.45
Incomplete septum	23 (10%)	2 (1%)	0.002
Irregular walls	140 (59%)	77 (56%)	0.57
Shadows	40 (17%)	20 (14%)	0.55
Doppler ultrasound variables			
Color Score			<0.001
Score 1	43 (18%)	10 (7%)	
Score 2	114 (48%)	44 (32%)	
Score 3	72 (30%)	70 (51%)	
Score 4	9 (4%)	14 (10%)	

LR1, logistic regression model 1
Results are shown as n (%) or median (min – max) except for age (mean ± SD)
No correction for multiple testing because no further testing was done for the subcategories

Table S8. The ability of vessel morphology, subjective assessment and logistic regression model 1, to discriminate between benign and malignant difficult tumors

Diagnostic method	Sensitivity [95% CI]	Specificity [95% CI]	LR+ [95% CI]	LR- [95% CI]	P-value
Either US examiner or LR1 uncertain (N=138)					
Whole tumor vessel morphology					
Branching vessels	89% (34/38) [76% to 96%]	33% (33/100) [25% to 43%]	1.34 [1.12 to 1.59]	0.32 [0.12 to 0.84]	0.009
Densely packed vessels	63% (24/38) [47% to 77%]	83% (83/100) [74% to 89%]	3.72 [2.26 to 6.10]	0.44 [0.29 to 0.68]	<0.001
Caliber changes in vessels	66% (25/38) [50% to 79%]	68% (68/100) [58% to 76%]	2.06 [1.43 to 2.97]	0.50 [0.32 to 0.80]	<0.001
Splashes	50% (19/38) [35% to 65%]	78% (78/100) [69% to 85%]	2.27 [1.40 to 3.70]	0.64 [0.46 to 0.90]	0.002
Tortuous vessels	66% (25/38) [50% to 79%]	70% (70/100) [60% to 78%]	2.19 [1.50 to 3.20]	0.49 [0.31 to 0.77]	<0.001
Biopsy vessel morphology					
Branching vessels	84% (32/38) [70% to 93%]	34% (34/100) [25% to 44%]	1.28 [1.05 to 1.55]	0.46 [0.21 to 1.02]	0.04
Caliber changes in vessels	79% (30/38) [64% to 89%]	63% (63/100) [53% to 72%]	2.13 [1.58 to 2.89]	0.33 [0.18 to 0.63]	<0.001
Splashes	53% (20/38) [37% to 68%]	69% (69/100) [59% to 77%]	1.70 [1.12 to 2.58]	0.69 [0.48 to 0.98]	0.02
Tortuous vessels	79% (30/38) [64% to 89%]	64% (64/100) [54% to 73%]	2.19 [1.61 to 2.99]	0.33 [0.18 to 0.62]	<0.001
Bridges between vessels	42% (16/38) [28% to 58%]	81% (81/100) [72% to 87%]	2.22 [1.28 to 3.84]	0.72 [0.54 to 0.95]	0.007
Subjective assessment	74% (28/38) [58% to 85%]	74% (74/100) [65% to 82%]	2.83 [1.94 to 4.15]	0.36 [0.21 to 0.61]	<0.001
LR1 (10% risk cutoff)	92% (35/38) [79% to 97%]	23% (23/100) [16% to 32%]	1.20 [1.04 to 1.38]	0.34 [0.11 to 1.08]	0.03
US examiner uncertain (N=79)					
Whole tumor vessel morphology					
Branching vessels	89% (24/27) [72% to 96%]	33% (17/52) [22% to 46%]	1.32 [1.05 to 1.67]	0.34 [0.11 to 1.06]	0.06
Densely packed vessels	67% (18/27) [48% to 81%]	83% (43/52) [70% to 91%]	3.85 [2.01 to 7.39]	0.40 [0.23 to 0.70]	<0.001
Caliber changes in vessels	63% (17/27) [44% to 78%]	67% (35/52) [54% to 78%]	1.93 [1.19 to 3.13]	0.55 [0.33 to 0.93]	0.01
Splashes	52% (14/27) [34% to 69%]	83% (43/52) [70% to 91%]	3.00 [1.49 to 6.01]	0.58 [0.39 to 0.88]	0.003
Tortuous vessels	63% (17/27) [44% to 78%]	63% (33/52) [50% to 75%]	1.72 [1.09 to 2.73]	0.58 [0.34 to 0.99]	0.02
Biopsy vessel morphology					
Branching vessels	85% (23/27) [68% to 94%]	35% (18/52) [23% to 48%]	1.30 [1.01 to 1.68]	0.43 [0.16 to 1.14]	0.07
Caliber changes in vessels	78% (21/27) [59% to 89%]	62% (32/52) [48% to 74%]	2.02 [1.36 to 3.01]	0.36 [0.17 to 0.76]	0.001
Splashes	59% (16/27) [41% to 75%]	71% (37/52) [58% to 82%]	2.05 [1.21 to 3.49]	0.57 [0.35 to 0.93]	0.009
Tortuous vessels	81% (22/27) [63% to 92%]	67% (35/52) [54% to 78%]	2.49 [1.62 to 3.83]	0.28 [0.12 to 0.62]	<0.001
Bridges between vessels	44% (12/27) [28% to 63%]	79% (41/52) [66% to 88%]	2.10 [1.07 to 4.11]	0.71 [0.49 to 1.02]	0.03

Cont.

Table S8 continued

Diagnostic method	Sensitivity [95% CI]	Specificity [95% CI]	LR+ [95% CI]	LR- [95% CI]	P-value
Subjective assessment	74% (20/27) [55% to 87%]	60% (31/52) [46% to 72%]	1.83 [1.23 to 2.73]	0.44 [0.22 to 0.86]	0.004
LR1 (10% risk cutoff)	89% (24/27) [72% to 96%]	19% (10/52) [11% to 32%]	1.10 [0.91 to 1.33]	0.58 [0.17 to 1.93]	0.34
LR1 uncertain (N=87)					
Whole tumor vessel morphology					
Branching vessels	94% (16/17) [73% to 99%]	30% (21/70) [21% to 42%]	1.35 [1.11 to 1.63]	0.20 [0.03 to 1.36]	0.06
Densely packed vessels	53% (9/17) [31% to 74%]	83% (58/70) [72% to 90%]	3.09 [1.56 to 6.11]	0.57 [0.34 to 0.95]	0.004
Caliber changes in vessels	71% (12/17) [47% to 87%]	69% (48/70) [57% to 78%]	2.25 [1.41 to 3.57]	0.43 [0.20 to 0.91]	0.005
Splashes	47% (8/17) [26% to 69%]	77% (54/70) [66% to 85%]	2.06 [1.06 to 4.00]	0.69 [0.43 to 1.09]	0.07
Tortuous vessels	65% (11/17) [41% to 83%]	71% (50/70) [60% to 81%]	2.27 [1.36 to 3.77]	0.49 [0.26 to 0.96]	0.01
Biopsy vessel morphology					
Branching vessels	76% (13/17) [53% to 90%]	31% (22/70) [22% to 43%]	1.12 [0.82 to 1.52]	0.75 [0.30 to 1.89]	0.77
Caliber changes in vessels	76% (13/17) [53% to 90%]	64% (45/70) [53% to 75%]	2.14 [1.42 to 3.23]	0.37 [0.15 to 0.88]	0.005
Splashes	47% (8/17) [26% to 69%]	69% (48/70) [57% to 78%]	1.50 [0.81 to 2.76]	0.77 [0.48 to 1.24]	0.26
Tortuous vessels	71% (12/17) [47% to 87%]	60% (42/70) [48% to 71%]	1.77 [1.16 to 2.69]	0.49 [0.23 to 1.05]	0.03
Bridges between vessels	35% (6/17) [17% to 59%]	81% (57/70) [71% to 89%]	1.90 [0.85 to 4.27]	0.80 [0.55 to 1.15]	0.19
Subjective assessment	82% (14/17) [59% to 94%]	79% (55/70) [68% to 87%]	3.84 [2.33 to 6.33]	0.23 [0.08 to 0.63]	<0.001
LR1 (10% risk cutoff)	100% (17/17) [82% to 100%]	19% (13/70) [11% to 29%]	1.23 [1.10 to 1.37]	Not possible to calculate	0.06

LR+, positive likelihood ratio; LR-, negative likelihood ratio; CI, confidence interval; LR1, logistic regression model 1 using the 10% risk cutoff to predict malignancy suggested in *J Clin Oncol* 2005; **23:** 8794-8801 and *Ultrasound Obstet Gynecol* 2010; **36:** 226-234
No corrections have been made for multiple testing because this is an exploratory study

Vessel morphology depicted by three-dimensional power Doppler ultrasound as second-stage test in difficult adnexal tumors: prospective diagnostic accuracy study

P. Sladkevicius[1,2], L. Jokubkiene[1,2], D. Timmerman[3,4], D. Fischerova[5], C. Van Holsbeke[6], D. Franchi[7], L. Savelli[8], E. Epstein[9], R. Fruscio[10], J. Kaijser[11], A. Czekierdowski[12], S. Guerriero[13], M. Angela Pascual[14], A. C. Testa[15], L. Ameye[3,16], L. Valentin[1,2]

[1]Department of Obstetrics and Gynecology, Skåne University Hospital, Malmö, Sweden

[2]Department of Clinical Sciences Malmö, Lund University, Sweden

[3] KU Leuven Department of Development and Regeneration, Herestraat 49, Box 7003, 3000 Leuven, Belgium

[4]Department of Obstetrics and Gynecology and Leuven Cancer Institute, University Hospitals Leuven, Herestraat 49, Box 7003, 3000 Leuven, Belgium

[5]Department of Obstetrics and Gynecology, First Faculty of Medicine, Charles University and First Faculty of Medicine, Prague, Czech Republic

[6]Department of Obstetrics and Gynecology, Ziekenhuis Oost Limburg, Schiepse Bos 6, 3600 Genk, Belgium

[7]Preventive Gynecology Unit, Division of Gynecology, European Institute of Oncology, Via Ripamonti 435**,** Milan 20141, Italy

[8]Gynecology and Reproductive Medicine Unit, S. Orsola-Malpighi Hospital, University of Bologna, Via Massarenti 13, Bologna 40138, Italy

[9]Department of Clinical Science and Education, Karolinska Institute, Södersjukhuset, Stockholm, Sweden

[10]Department of Obstetrics and Gynecology, San Gerardo Hospital, University of Milan-Bicocca, Via Pergolesi 33, 20052 Monza**,** Italy

[11]Department of Obstetrics and Gynecology, Ikazia Hospital Rotterdam, the Netherlands

[12]1st Department of Gynecological Oncology and Gynecology, Medical University of Lublin, Lublin, Poland

[13]Department of Obstetrics and Gynecology, University of Cagliari, Policlinico Universitario Duilio Casula, Monserrato, Cagliari, Italy

[14] Department of Obstetrics, Gynecology and Reproduction, Hospital Universitari Dexeus, Barcelona, Spain

[15]Department of Gynecological Oncology, Catholic University of Sacred Heart, Rome, Italy

[16]Jules Bordet Institute, Université Libre de Bruxelles, Brussels, Belgium

Corresponding author:

Lil Valentin

Department of Obstetrics and Gynecology

Skåne University Hospital Malmö

20502 Malmö, Sweden

e-mail: lil.valentin@med.lu.se

Short title: Vascular morphology in difficult adnexal masses

Keywords: ultrasonography, Doppler; ovarian neoplasms; three-dimensional ultrasound; vascular morphology

Contribution

What are the novel findings of this work?

Vessel morphology, depicted by three-dimensional power Doppler ultrasound, differs between benign and malignant adnexal masses that are difficult to classify as benign or malignant by an experienced ultrasound examiner using subjective assessment of ultrasound images or by the IOTA logistic regression model 1 (LR1).

What are the clinical implications of this work?

Vessel morphology, as depicted by three-dimensional power Doppler ultrasound, may slightly improve discrimination between benign and malignant adnexal tumors judged to be difficult to classify by subjective assessment. For tumors in which the IOTA LR1 model yields an ambiguous result, subjective assessment is superior to vessel morphology as a second stage test.

Abstract

Objectives To assess whether vessel morphology depicted by three-dimensional (3D) power Doppler ultrasound improves discrimination between benignity and malignancy if used as a second stage test in adnexal masses that are difficult to classify.

Methods This was a prospective observational international multicenter diagnostic accuracy study. Consecutive patients with an adnexal mass underwent standardized transvaginal 2D gray scale and color or power Doppler and 3D power Doppler ultrasound examination by an experienced examiner, and those with a difficult tumor were included in the current analysis. A difficult tumor was defined as one in which the IOTA logistic regression model 1 (LR1) yielded an ambiguous result (risk of malignancy, 8.3% to 25.5%), or as one in which the ultrasound examiner was uncertain regarding classification as benign or malignant when using subjective assessment. Even when the ultrasound examiner was uncertain, they were obliged to classify the tumor as most likely benign or most likely malignant. For each difficult tumor, one researcher created a 360° rotating 3D power Doppler image of the vessel tree in the whole tumor and another of the vessel tree in a 5 cm^3 spherical volume selected from the most vascularized part of the tumor. Two other researchers, blinded to patient history, 2D ultrasound findings and histological diagnosis, independently described the vessel tree using predetermined vessel features. Their agreed classification was used. The reference standard was the histological diagnosis of the mass. Sensitivity of each test for discriminating between benign and malignant difficult tumors was plotted against 1 – specificity on a receiver operating characteristics diagram, and the test with the point farthest from the reference line was considered to have the best ability.

Commented [KP1]: AU: I have added 'and those with a difficult tumor were included in the current analysis'. Do you agree? (KP)

Commented [KP2]: AU: 'power Doppler' added here. OK? (KP)

Results Of 2403 women with an adnexal mass, 376/2403 (15.6%) had a difficult mass. Ultrasound volumes were available for 138 of these cases. In 79/138 masses, the ultrasound examiner was uncertain about the diagnosis based on subjective assessment, in 87/138 IOTA LR1 yielded an ambiguous result and in 28/138 both methods gave an uncertain result. Of the masses, 38/138 (27%) were malignant. Among tumors that were difficult to classify as benign or malignant by subjective assessment, the vessel feature 'densely packed vessels' had the best discriminative ability [sensitivity 67% (18/27), specificity 83% (43/52)] and was slightly superior to subjective assessment [sensitivity 74% (20/27), specificity 60% (31/52)]. In tumors in which IOTA LR1 yielded an ambiguous result, subjective assessment [sensitivity 82% (14/17), specificity 79% (55/70)] was superior to the best vascular feature, i.e. caliber changes of vessels in the whole tumor volume [sensitivity 71% (12/17), specificity 69% (48/70)].

Conclusion

Vessel morphology depicted by 3D power Doppler ultrasound may slightly improve discrimination between benign and malignant adnexal tumors that are difficult to classify by subjective ultrasound assessment. For tumors in which the IOTA LR1 model yields an ambiguous result, subjective assessment is superior to vessel morphology as a second stage test.

Commented [KP3]: AU: According to style, the number of overall women has been moved to the Abstract Results from the Abstract Methods. OK? (KP)

Commented [KP4]: AU: '9/17' changed here to '12/17' for sensitivity, based on the reported percentage and Table 3. OK? (KP)

Introduction

Subjective assessment of ultrasound findings (also called pattern recognition) by an experienced ultrasound examiner is the best ultrasound method to discriminate between benign and malignant adnexal masses[1,2]. However, even an experienced ultrasound examiner may find up to 10% of tumors impossible to classify confidently as benign or malignant using pattern recognition (termed "difficult tumors")[3-5]. The 10% risk cut-off of the International Ovarian Tumor Analysis (IOTA) logistic regression model 1 (LR1) has almost as good ability to discriminate between benign and malignant tumors as subjective assessment[6,7], but a risk of malignancy calculated by LR1 of 8.3–25.5% has been suggested to represent an ambiguous risk[8]. The tumor marker CA125 is clearly inferior to subjective assessment for discriminating between benign and malignant adnexal masses[9] and has no role for classifying difficult tumors[3-5,10]. Subjective assessment of ultrasound images is superior to computed tomography for discriminating between benign and malignant adnexal masses, while the role of magnetic resonance imaging is still unclear[11,12, 13]. A logistic regression model for calculating the risk of malignancy in tumors not classifiable as benign or malignant by an experienced ultrasound examiner using subjective assessment has been published, but its ability to discriminate between benign and malignant tumors was not superior to that of subjective assessment[4].

For tumors that are difficult to classify as benign or malignant using subjective assessment or using the IOTA LR1 model, a second stage test capable of classifying correctly difficult tumors as benign or malignant would be valuable. A possible second stage test is three-dimensional (3D) power Doppler ultrasound examination of the vascular tree of tumors[14].

The aims of this study were to assess whether vessel morphology, as depicted by 3D power Doppler ultrasound, differs between benign and malignant difficult adnexal masses, and whether vessel morphology improves discrimination between benign and malignant masses when used as a second stage test in difficult adnexal masses.

Patients and Methods

Study population

Our study population comprised those patients in the IOTA 3 study[15] who had a difficult adnexal tumor (defined below). The IOTA 3 study is a prospective observational international multicenter cross-sectional diagnostic accuracy study (study protocol available in Appendix S1), that has been described in detail elsewhere[15]. Patients were recruited into IOTA 3 between October 2009 and May 2012 in 18 centers in six countries (Sweden, Belgium, Italy, Poland, Spain and Czech Republic). These centers were either oncology referral centers (i.e. tertiary referral centers with a specific gynecological oncology unit) or other hospitals or units with a special interest in gynecological ultrasound. The centers and type of center are listed after the main text. Ethics approval was obtained from the Ethics Committee of the University Hospitals Leuven (B32220095331/S51375) as well as from the local ethics committees of all contributing centers.

Patients referred to one of the participating centers for an ultrasound examination and found to have an adnexal mass were eligible for inclusion in IOTA 3. Consecutive patients with at least one adnexal mass judged not to be a functional cyst, examined with transvaginal ultrasound by an experienced ultrasound examiner, were included in IOTA 3, provided that they gave written and/or oral informed consent before the ultrasound scan. If more than one mass was detected, the mass with the most complex ultrasound morphology was used for statistical analysis. When masses with similar morphology were observed, the largest mass or the one most easily accessible with ultrasound was used. Criteria for excluding patients from IOTA 3 were pregnancy at the time of the ultrasound examination, surgical removal of the mass more than 120 days after the ultrasound examination and data inconsistencies that persisted after final manual data checks.

Data collection

A dedicated, secure electronic data-collection system was developed for the IOTA 3 study (IOTA 3 Study Screen; astraia Software, Munich, Germany). Patients automatically received a unique identifier. Data security was ensured by encrypting all data communication. Data integrity and completeness were ensured by client-side checks in the system supplied by astraia and by final data cleaning by a group of biostatisticians and expert ultrasound examiners. The study screen presented the risk of malignancy calculated using the IOTA LR1 model[6].

Ultrasound examination

All patients included in the IOTA 3 study[15] underwent a standardized transvaginal ultrasound examination by a gynecologist or radiologist very experienced in gynecologic ultrasound. High-end ultrasound systems were used. Gray scale and color or power Doppler ultrasound was used to obtain information on more than 40 ultrasound variables to characterize each adnexal mass. Details on the standardized ultrasound examination technique and the IOTA terminology used to describe the ultrasound images have been described elsewhere[16]. After completing the ultrasound examination, the ultrasound examiner classified each mass as benign or malignant on the basis of their subjective assessment of the gray scale and color or power Doppler ultrasound findings. In addition, the ultrasound examiner stated their level of confidence by classifying each mass as certainly benign, probably benign, uncertain, probably malignant or certainly malignant. This means that, even when the ultrasound examiner was uncertain whether the tumor was benign or malignant, they were obliged to classify it as most likely benign or most likely malignant. The ultrasound information was entered prospectively into the electronic data-collection system (see above),

and was locked at the time of the examination and could not be changed thereafter. The decision regarding surgery for adnexal tumors was made by the referring physician, based on clinical information, such as symptoms, age, operative risk and coexisting disease, and on the ultrasound report which was written using the results of subjective assessment.

In addition to performing the two-dimensional (2D) ultrasound examination as described above, ultrasound examiners in centers with access to a Voluson 730 Expert or GE E8 ultrasound system (GE Healthcare, Zipf, Austria) with a 5–9 MHz or a 6–12 MHz vaginal transducer were asked to acquire 3D power Doppler ultrasound volumes of all adnexal masses. The power Doppler and 3D settings are described in Appendix S2. The ultrasound examiners were instructed to include the whole tumor in the volume or, if this was not possible because the tumor was too large, to acquire several volumes to ensure that all parts of the tumor were captured. The patients were asked to lie still during the acquisition and, if necessary, to hold their breath. Volumes from difficult tumors, i.e. tumors in which the ultrasound examiner was uncertain whether the tumor was benign or malignant based on subjective assessment, or those in which the IOTA LR1 model gave an ambiguous risk of malignancy (8.3-25.5%)[8], were sent to Skåne University Hospital, Malmö on compact discs, for analysis. The same methodology of analyzing the 3D volumes as described previously was used[14]. This method is outlined briefly below.

Analysis of 3D volumes and Audio Video Interleave (AVI) files

For each tumor, 360° rotating 3D power Doppler images (AVI-files) of the vessel tree in the whole tumor as well as in a 5 cm^3 spherical volume selected from the most vascularized part of the tumor were prepared by the second author (LJ) using the virtual organ computer-aided analysis (VOCAL™) imaging program (4D-VIEW, version 7.0, GE Healthcare, Zipf,

Austria) on a personal computer. Subjective assessment was used to select the most vascularized part of the tumor[14]. Before creating the rotating image of the vessel tree, color transparency was adjusted to optimize the delineation of the vessels.

All AVI files were then analyzed independently by two members of the Malmö research team (LV, PS) who had no knowledge of patient history, 2D ultrasound findings or histological diagnosis. The vessel tree in the whole tumor volume, as well as in the 5 cm^3 sample, was characterized using the same classification as described previously (Figure 1): branching, i.e. division of a vessel into two or more branches; caliber changes, i.e. changes in vessel width from narrow to wide and again from wide to narrow; 'splashes', i.e. areas of color in contrast to clearly separate vessels; tortuosity; areas with densely packed vessels; and "bridges", i.e. straight vessel connections between two nearby vessels. The presence of bridges was assessed only in the 5cm^3 samples and the presence of densely packed vessels was assessed only in the whole tumor volume[14]. In cases of disagreement between the two observers, consensus was reached by discussion, and the agreed classification was used for statistical analysis. We did this to decrease the risk of bias introduced by relying on one single observer. Ultrasound images of the vascular features are shown in Figures 2 and 3.

Reference standard

The reference standard was the histologic classification of the excised mass as malignant or benign. Histological examination was carried out at the local centers. Central pathology review was not performed because, in a previous IOTA study, no clinically significant differences between local and central pathology reports were observed[6]. Malignant tumors were classified according to the criteria recommended by the International Federation of Gynecology and Obstetrics[17]. Borderline ovarian tumors were classified as malignant. The pathologists were blinded to the ultrasound findings.

Statistical analysis

We used SAS version 9.4 (SAS Institute, Cary, NC, USA) for all statistical analyses.

The statistical significance of differences in proportions was determined using the Chi-square test or Fisher's exact test and that of differences in continuous data using Student's t-test or Mann-Whitney U-test, as appropriate.

We evaluated the discriminative ability of each vascular morphology feature, of subjective assessment and of the IOTA LR1 model when using the recommended 10% risk cut-off for malignancy[6], and expressed it as sensitivity, specificity and positive and negative likelihood ratios. The sensitivity of each test was plotted against 1 – specificity in a receiver operating characteristic diagram. The test with the point farthest from the reference line was considered to have the best discriminative ability.

We estimated inter-observer reliability (agreement beyond chance) in the assessment of the vascular tree of the tumors by calculating Cohen's Kappa[18]. Kappa values of 0.81-1.0 were taken to indicate excellent reliability, 0.61-0.80 indicated good reliability and 0.41-0.60 indicated moderate reliability[19].

We defined a P-value < 0.05 as statistically significant and corrected for multiple testing using the permutation method[20].

Sample size calculation

We aimed to collect information on 300 difficult masses, of which we expected a minimum of 30% (n = 90) to be malignant[3,8]. Ninety malignancies would give us a reasonable 95% confidence interval (CI) around the point estimate for sensitivity. Because about 7% of all adnexal masses are difficult to classify as benign or malignant using subjective assessment[4], and because the IOTA LR1 model may yield an ambiguous test result in 10% of

all adnexal masses[8], we estimated that we needed to examine about 2000 women with an adnexal mass (Appendix S1).

The study is reported using the Standards for Reporting Diagnostic Accuracy studies (STARD) guidelines[22].

Results

In total, 2541 women with an adnexal mass were enrolled for inclusion in IOTA 3, of whom 138 were excluded. Reasons for exclusion were >120 days between ultrasound examination and surgery (n=66), pregnancy (n=31), data errors that could not be solved by contacting principal investigators (n=28) and incomplete final histology (n=13). The final IOTA 3 dataset included 2403 patients[15], of whom 376 (15.6%) had a difficult tumor; an uncertain result regarding malignancy was obtained for 168 (7%) tumors on subjective assessment, for 259 (11%) tumors using the LR1 model and for 51 (2%) tumors using both methods. Serous and mucinous cystadenomas/cystadenofibromas, fibromas and borderline tumors were substantially more common among the difficult tumors than in the others, while endometriomas, benign teratomas (dermoid cysts), primary ovarian cancers and metastases in the ovaries from another primary tumor were substantially less common, with 2-fold to 3-fold differences in prevalence (Tables S1 and S2). The distribution of histological diagnoses was similar in tumors in which the ultrasound examiner was uncertain about the diagnosis on subjective assessment and in those in which LR1 yielded an ambiguous result, with the exception that benign teratomas and endometriomas were more common in the latter (Table S1). Unilocular solid tumors, multilocular solid tumors and papillary projections were substantially more common in difficult tumors than in the others, while unilocular cysts, ground glass echogenicity of cyst fluid, and color scores of 1 and 4 were substantially less common, with 2-fold to 3-fold differences in prevalence (Tables S3 and S4). The ultrasound characteristics of tumors that the ultrasound examiner found difficult to classify on subjective assessment were similar to those in which LR1 yielded an ambiguous result, with the exception that ascites was more common and unilocular cysts were less common in tumors that the ultrasound examiner found difficult to classify on subjective assessment (Table S3).

3D ultrasound volumes were available for 138 of the 376 difficult tumors. Six centers did not provide volumes for any of their difficult masses (0/55), four centers provided volumes for more than 80% (73/87), four centers for between 40% and 57% (47/96), three centers for between 11% and 18% (16/107) and one center for 6% (2/31). In three centers, the reason for providing volumes for none or only a small proportion of the difficult tumors (6/78) was that the volumes stored in the ultrasound system had been deleted and there was no back-up, and in one center the GE ultrasound system required was not always available for research. One center (with 33 difficult tumors) did not have access to a Voluson 730 Expert or GE E8 ultrasound system and so could not provide any volumes. Seven centers reported forgetfulness or transient technical problems to be the explanation for not providing volumes for all their difficult masses, and six centers gave no explanation. The number of patients and the proportion of difficult tumors contributed by each center are shown in Table S5. Patient flow is described in Figure 4.

The available volumes from the 138 difficult tumors included 79 (57%) masses in which subjective assessment gave an uncertain result, 87 (63%) cases in which LR1 gave an ambiguous result and 28 (20%) cases in which both methods gave an uncertain result. The histological diagnoses of the 138 difficult tumors with available volumes are shown in Table 1. One hundred (72.5%) tumors were benign and 38 (27.5%) were malignant. The distribution of histological diagnoses was similar to that in the 238 difficult tumors for which tumor volumes were not available, with the exception that the proportion of serous and mucinous cystadenomas was slightly higher among the tumors with available volumes (Figure 4, Table S6). The clinical background data and 2D ultrasound findings for the 138 difficult tumors with available volumes are described in Table 2. They were similar to those in the 238 difficult tumors for which volumes were not available, with the following exceptions:

Commented [KP5]: AU: 'tumor mix' changed to 'distribution of histological diagnoses'. Do you agree? (KP)

Commented [KP6]: AU: May I please just check whether instances of 'cystadenomas' should instead be 'cystadenomas/cystadenofibromas' (as you have in the first paragraph of the Results, for example)? (KP)

unilocular cysts, incomplete septae, tender mass at scan and color score of 1 were less common in the 138 difficult tumors for which tumor volumes were available, while multilocular-solid tumors and tumors with color score of 3 or 4 were more common (Figure 4, Table S7).

Commented [KP7]: AU: '4' changed here to '3 or 4', based on the wording in the Discussion and the results in Table S7. Is this correct? (KP)

The ability of subjective assessment, the IOTA LR1 model and the vascular features to discriminate between benign and malignant difficult tumors, in terms of sensitivity and specificity, is shown in Table 3 (with the 95% CIs shown in Table S8). All vessel features differed significantly between benign and malignant difficult tumors. Branching vessels, densely packed vessels, caliber changes, tortuous vessels, color splashes and bridges between vessels were more common in the malignant than in the benign difficult tumors. However, none of the vessel features discriminated well between benign and malignant difficult tumors. Figure 5 shows plots of sensitivity against 1 – specificity for subjective assessment, for the IOTA LR1 model when using a 10% risk cut-off and for the vessel features with the best discriminative ability. Subjective assessment was the best method for discriminating between benign and malignant masses in the total study population of 138 difficult masses, followed by densely packed vessels in the whole tumor volume and tortuous vessels in the tumor biopsy. Among the 79 tumors that were difficult to classify as benign or malignant using subjective assessment, densely packed vessels in the whole tumor volume and tortuous vessels in the tumor biopsy had the best discriminative ability. For those tumors in which the IOTA LR1 model yielded an ambiguous result, subjective assessment had the best discriminative ability, while caliber changes in the tumor biopsy had the second best.

Inter-observer agreement and reliability with regard to vessel morphology are shown in Table 4. Inter-observer reliability was moderate or good, with Cohen's Kappa values ranging from 0.55 to 0.77. Agreement was best for assessment of densely packed vessels, branching vessels and tortuous vessels in the whole tumor volume (Cohen's kappa, 0.77, 0.71 and 0.69,

respectively) and for branching vessels in the 5 cm^3 tumor biopsy (Cohen's Kappa, 0.70).

Discussion

We have shown that vessel morphology, depicted by 3D power Doppler ultrasound, differs between benign and malignant difficult adnexal tumors. Branching vessels, caliber changes, color splashes, tortuous vessels, densely packed vessels and bridges between vessels were more common in malignant than in benign difficult tumors. However, none of the vascular features discriminated well between benign and malignant difficult tumors. Our findings confirm that inter-observer reliability with regard to vessel morphology depicted by 3D power Doppler is moderate to good[14].

To the best of our knowledge, there are no published studies exploring the ability of the morphology of tumor vessels depicted by 3D power Doppler ultrasound to discriminate between benign and malignant difficult tumors. A strength of our study is that vessel morphology was assessed by observers who were blinded to clinical information, 2D gray-scale or color Doppler ultrasound findings and histological diagnosis of the tumors. This means that our results regarding the evaluation of the vessel tree are unbiased and reflect the true discriminative capacity of vessel morphology. A limitation of our study is possible selection bias, because not all centers provided 3D volumes of all their difficult tumors. The histology and ultrasound features differed slightly between the difficult tumors included (i.e. those with tumor volumes available) and those not included (i.e. those with tumor volumes not available) in that the proportion of serous and mucinous cystadenomas, multilocular-solid tumors and tumors with a color score of 3 or 4 was higher among the difficult tumors that were included, while the proportion of unilocular cysts was lower (Figure 4, Table S6 and S7). However, the histological diagnoses and ultrasound features of the difficult tumors with available volumes were relatively similar to those of all difficult tumors included in IOTA 3 (Table S1). Therefore, our study sample should be reasonably representative of all difficult adnexal tumors. The small number of difficult masses with available volumes is another

limitation, making our estimates of sensitivity and specificity imprecise (Table S8). However, because this is an exploratory study, we believe that this is acceptable.

Our results confirm those of a previous study[4] demonstrating that serous and mucinous cystadenomas, fibromas and borderline tumors are difficult to classify as benign or malignant, and that unilocular-solid and multilocular-solid tumors and tumors with papillary projections are overrepresented among difficult tumors[4]. The proportion of tumors that the ultrasound examiner found difficult to classify based on subjective assessment in the current study is the same as that in the previously published study[4] (168/2403 (7%) versus 244/3511 (7%)).

To the best of our knowledge, the ability of vessel morphology depicted by 3D ultrasound to discriminate between benign and malignant adnexal masses or to decrease diagnostic errors has been explored in only one published study[14]. That study included 104 adnexal masses reasonably representative of a general population of tumors scheduled for surgery. Similar to our findings, all vascular features differed between benign and malignant tumors, but the vessel features discriminated much better between benign and malignant tumors than in the current study[14]. This is not surprising because, in the cited study, the tumors were more heterogeneous with fewer overlapping ultrasound features. They found that adding a vascular feature variable to a logistic regression model including only gray scale ultrasound variables improved the discriminative ability of the model only minimally (increase in AUC from 0.98 to 0.99) because the gray-scale model itself performed extremely well. The authors concluded that, in an ordinary population of ovarian tumors, 3D power Doppler ultrasound examination adds little to grayscale imaging. However, they hypothesized that 3D power Doppler ultrasound examination with assessment of the morphology of tumor vessels might be useful in difficult tumors.

Our results show that, if the IOTA LR1 model gives an ambiguous risk estimate (8.3–25.5%), then subjective assessment by an experienced ultrasound examiner is superior to using vessel morphology depicted by 3D power Doppler ultrasound as a second stage test. However, if a mass cannot be confidently classified as benign or malignant by an experienced ultrasound examiner using subjective assessment, then assessing vessel morphology with 3D power Doppler ultrasound could be useful. Both densely packed vessels in the whole tumor volume and tortuous vessels in a 5-cm^3 tumor biopsy can be used for discrimination (Figure 5). We recommend using densely packed vessels because inter-observer reliability was better for this variable than for tortuous vessels in a tumor biopsy. On the other hand, it is more time consuming to analyze a whole tumor volume than a tumor biopsy[21]. In our experience, for a very experienced ultrasound examiner, it takes a minimum of 2.5 min to create a rotating 3D image of the vascular tree of a whole tumor, while it takes a minimum of 1 min to create one for a 5 cm^3 biopsy selected from the most vascularized part of the tumor. When assessing vessel morphology, it is important to be aware of the pitfalls of Doppler ultrasound and to ensure that Doppler settings are correct. If the tumor is far away from the ultrasound probe, it might not be possible to detect Doppler signals from the whole or parts of the tumor. This limits the clinical usefulness of vessel morphology to classify tumors as benign or malignant. Another limiting factor is the subjectivity of the method. Evaluation of vessel morphology, including “densely packed vessels”, is based on pattern recognition and is therefore difficult to standardize or define.

Correct classification of adnexal masses as benign or malignant is a requirement for optimal management, i.e. conservative management with follow-up examinations, surgery in a local hospital or referral to a center specialized in gynecological oncology[22]. However, some tumors are difficult to classify confidently as benign or malignant. Vessel morphology depicted by 3D power Doppler ultrasound showed limited ability to discriminate between benign and malignant difficult tumors. It remains to be shown if new biomarkers, immune cells, proteins or genetic information can improve classification of difficult tumors as benign or malignant.

Acknowledgements

This study was supported by the Swedish Medical Research Council (grants no. K2006-73X-11605-11-3); by the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement; Landstingsfinansierad regional forskning (a Swedish governmental grant from the region of Scania); funds administered by Skåne University Hospital; and Allmänna Sjukhusets i Malmö Stiftelse för bekämpande av cancer (the Malmö General Hospital Foundation for fighting against cancer).

Participating centers

Oncology referral centers

1. Ospedale S. Gerardo, Universita di Milano Bicocca, Monza, Italy (**OIT**)
2. General Faculty Hospital of Charles University, Prague, Czech Republic (**PCR**)
3. Skåne University Hospital Lund, Sweden (**LSW**)
4. Karolinska University Hospital, Stockholm, Sweden (**SSW**)
5. Istituto Europeo di Oncologia, Milan, Italy (**CIT**)
6. Medical University in Lublin, Poland (**LPO**)
7. University Hospitals Leuven, Leuven, Belgium (**LBE**)
8. Universita Cattolica del Sacro Cuore, Rome, Italy (**RIT**)
9. Universita degli Studi di Udine, Italy (**UDI**)
10. Instituto Nationale dei Tumori, Fondazione Pascale, Naples, Italy (**GIT**)
11. University of Bologna, Bologna, Italy (**BIT**)

Other centers

12. Skåne University Hospital Malmö, Sweden (**MSW**)
13. Departament d'Obstetrícia, Ginecologia i Reproducció, Institut Universitari Dexeus, Barcelona, Spain (**BSP**)
14. University of Cagliari, Ospedale San Giovanni di Dio, Cagliari, Italy (**SIT**)
15. Ziekenhuis Oost-Limburg, Genk, Belgium (**GBE**)
16. Childrens' Hospital Buzzi, Milan, Italy (**FIT**)
17. Universita degli Studi di Napoli, Naples, Italy (**NIT**)
18. DCS Sacco University of Milan, Milan, Italy (**MIT**)

References

1. 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.
2. Timmerman D. The use of mathematical models to evaluate pelvic masses; can they beat an expert operator? *Best Pract Res Clin Obstet Gynaecol* 2004;**18**:91-104.
3. Valentin L, Ameye L, Jurkovic D, Metzger U, Lécuru F, Van Huffel S, Timmerman D. Which extrauterine pelvic masses are difficult to correctly classify as benign or malignant on the basis of ultrasound findings and is there a way of making a correct diagnosis? *Ultrasound Obstet Gynecol* 2006; **27:** 438-444.
4. Valentin L, Ameye L, Savelli L, Fruscio R, Leone FP, Czekierdowski A, Lissoni AA, Fischerova D, Guerriero S, Van Holsbeke C, Van Huffel S, Timmerman D. 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* 2011; **38:** 456-465.
5. Valentin L, Ameye L, Savelli L, Fruscio R, Leone FP, Czekierdowski A, Lissoni AA, Fischerova D, Guerriero S, Van Holsbeke C, Van Huffel S, Timmerman D. Unilocular adnexal cysts with papillary projections but no other solid components: is there a diagnostic method that can classify them reliably as benign or malignant before surgery? *Ultrasound Obstet Gynecol* 2013; **41:** 570-581.
6. Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, Van Calster B, Collins WP, Vergote I, Van Huffel S, Valentin L; International Ovarian Tumor Analysis Group. 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.
7. Timmerman D, Van Calster B, Testa AC, Guerriero S, Fischerova D, Lissoni AA, Van Holsbeke C, Fruscio R, Czekierdowski A, Jurkovic D, Savelli L, Vergote I, Bourne T, Van Huffel S, Valentin L. Ovarian cancer prediction in adnexal masses using ultrasound-based logistic regression models: a temporal and external validation study by the IOTA group. *Ultrasound Obstet Gynecol* 2010; **36:** 226-234.

8. Daemen A, Valentin L, Fruscio R, Van Holsbeke C, Melis GB, Guerriero S, Czekierdowski A, Jurkovic D, Ombelet W, Rossi A, Vergote I, Bourne T, De Moor B, Timmerman D. Improving the preoperative classification of adnexal masses as benign or malignant by second-stage tests. *Ultrasound Obstet Gynecol* 2011; **37:** 100-106.
9. 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 CA125. *J Natl Cancer Inst* 2007; **99:** 1706-1714.
10. Valentin L, Jurkovic D, Van Calster B, Testa A, Van Holsbeke C, Bourne T, Vergote I, Van Huffel S, Timmerman D. Adding a single CA125 measurement to ultrasound imaging performed by an experienced examiner does not improve preoperative discrimination between benign and malignant adnexal masses. *Ultrasound Obstet Gynecol* 2009; **34:** 345-354.
11. Valentin L. Imaging in gynecology. *Best Pract Res Clin Obstet Gynaecol* 2006; **20:** 881-906.
12. Kaijser J, Vandecaveye V, Deroose CM, Rockall A, Thomassin-Naggara I, Bourne T, Timmerman D. Imaging techniques for the pre-surgical diagnosis of adnexal tumours. *Best Pract Res Clin Obstet Gynaecol* 2014; **28:** 683-695.
13. Pereira PN, Sarian LO, Yoshida A, Araújo KG, Silva ACB, de Oliveira Barros RH, Jales RM, Derchain S. Improving the performance of IOTA simple rules: sonographic assessment of adnexal masses with resource-effective use of a magnetic resonance scoring (ADNEX MR scoring system). Abdom Radiol (NY). 2019 Sep 3. doi: 10.1007/s00261-019-02207-9. Epub ahead of print. PMID: 31482379.
14. Sladkevicius P, Jokubkiene L, Valentin L. Contribution of morphological assessment of the vessel tree by three-dimensional ultrasound to a correct diagnosis of malignancy in ovarian masses. *Ultrasound Obstet Gynecol* 2007; **30:** 874-882.
15. Testa A, Kaijser J, Wynants L, Fischerova D, Van Holsbeke C, Franchi D, Savelli L, Epstein E, Czekierdowski A, Guerriero S, Fruscio R, Leone FP, Vergote I, Bourne T, Valentin L, Van Calster B, Timmerman D. Strategies to diagnose ovarian cancer: new evidence from phase 3 of the multicentre international IOTA study. *Br J Cancer* 2014; **111:** 680-688.

16. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I; International Ovarian Tumor Analysis (IOTA) Group. 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.
17. Heintz APM, Odicino F, Maisonneuve P, Beller U, Benedet JL, Creasman WT, Ngan HYS, Pecorelli S. Carcinoma of the Ovary, 25th Annual Report on the Results of Treatment in Gynecological Cancer. *Int J Gynecol Obstet* 2003; **83:** S135-S166 (suppl 1).
18. Cohen's kappa Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; **20**: 37-46.
19. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ 1992; **304**;1491-1494.
20. Westfall PH, Wolfinger RD. Multiple tests with discrete distributions. *Am Stat* 1997; 51: 3–8.
21. Jokubkiene L, Sladkevicius P, Valentin L. Does three-dimensional power Doppler ultrasound help in discrimination between benign and malignant ovarian masses? *Ultrasound Obstet Gynecol* 2007; **29**: 215-225.
22. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, Irwig L, Levine D, Reitsma JB, de Vet HC, Bossuyt PM. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;**6**:e012799
23. Froyman W, Landolfo C, De Cock B, Sladkevicius P, Testa AC, Van Holsbeke C, Domali E, Fruscio R, Epstein E, Dos Santos Bernardo MJ, Franchi D, Kudla MJ, Chiappa V, Alcazar JL, Leone FPG, Buonomo F, Hochberg L, Coccia ME, Guerriero S, Deo N, Jokubkiene L, Kaijser J, Coosemans A, Vergote I, Verbakel JY, Bourne T, van Calster B, Valentin L, Timmerman D. Risk of complications in patients with conservatively managed ovarian tumours (IOTA5): a 2-year interim analysis of a multicentre, prospective, cohort study. *Lancet Oncol* 2019; **20**: 448-458.

LEGENDS

Figure 1 Schematic diagrams illustrating vascular features evaluated in adnexal masses in our study: a) straight vessel with no branching; b) branching vessel; c) tortuous vessel; d) vessel with caliber changes; and e) bridges (short, straight connections between two nearby vessels). Reproduced from Sladkevicius P et al. *Ultrasound Obstet Gynecol* 2007; **30**: 874-882.

Figure 2 Three-dimensional power Doppler ultrasound images of vessel tree in whole volume of ovarian tumors, showing: a) dispersed (as opposed to densely packed), straight (as opposed to tortuous), branching vessels in benign mucinous cystadenoma; b) dispersed, branching, tortuous vessels (thick arrow) with caliber changes (thin arrow) in mucinous borderline tumor; c) densely packed, branching, tortuous vessels with caliber changes and color splashes (arrow shows splashes) in functional cyst; and d) densely packed, branching, tortuous vessels with caliber changes in ovarian clear cell cancer. Corresponding rotating images of vessel trees are shown in Videoclips S1-S4.

Figure 3 Three-dimensional power Doppler ultrasound images of vessel tree in 5 cm^3 samples of ovarian tumors, showing: a) bridges (arrows), i.e. straight connections between two nearby vessels, in struma ovarii; b) branching, tortuous vessels with caliber changes and color splashes (arrow) in benign ovarian fibroma; c) branching, tortuous vessels with caliber changes in ovarian endometroid carcinoma (color splashes are seen in rotating image of same tumor in Videoclip S7). Corresponding rotating images of vessel trees are shown in Videoclips S5-S7.

Figure 4 Flowchart summarizing recruitment of patients with adnexal tumor that was difficult to classify as benign or malignant. Prevalence of histological diagnoses and ultrasound features that differed most between tumors that were and those that were not difficult to classify are shown.

Figure 5 Ability of subjective ultrasound assessment, IOTA LR1 model and best performing 3D power Doppler vessel morphology features, to discriminate between benign and malignant difficult adnexal tumors, in: a) all 138 difficult tumors; b) 79 difficult tumors in which ultrasound examiner was uncertain whether tumor was benign or malignant when using subjective assessment; and c) 89 difficult tumors in which the International Ovarian Tumor Analysis (IOTA) LR1 model yielded ambiguous result (risk of malignancy, 8.3% to 25.5%). bx, biopsy; whole, whole tumor volume

Commented [KP8]: AU: 'power Doppler' added here and in the legends of Figure 3 and the videoclips. OK? (KP)

Commented [KP9]: AU: 'in whole volume' added here (to differentiate them from biopsy samples in Figure 3). Do you agree? (KP)

Supplementary material

Appendix S1 International Ovarian Tumor Analysis (IOTA) Phase 3 study protocol

Appendix S2 Power Doppler and three-dimensional (3D) ultrasound settings used in study

Videoclip S1 Three-dimensional 360 degree rotating power Doppler ultrasound image of vessel tree of benign mucinous cystadenoma, showing dispersed (as opposed to densely packed), straight (as opposed to tortuous), branching vessels. Still image of same tumor is shown in Figure 2a.

Videoclip S2 Three-dimensional 360 degree rotating power Doppler ultrasound image of vessel tree of mucinous borderline tumor, showing dispersed, branching, tortuous vessels with caliber changes. Still image of same tumor is shown in Figure 2b.

Videoclip S3 Three-dimensional 360 degree rotating power Doppler ultrasound image of vessel tree in functional cyst, showing densely packed, branching, tortuous vessels with caliber changes and color splashes. Still image of same tumor is shown in Figure 2c.

Videoclip S4 Three-dimensional 360 degree rotating power Doppler ultrasound image of vessel tree in ovarian clear cell cancer, showing densely packed, branching, tortuous vessels with caliber changes. Still image of same tumor is shown in Figure 2d.

Videoclip S5 Three-dimensional 360 degree rotating power Doppler ultrasound image of vessel tree in 5 cm^3 sample of struma ovarii showing bridges, i.e. straight connections between two nearby vessels. Still image of same tumor is shown in Figure 3a.

Videoclip S6 Three-dimensional 360 degree rotating power Doppler ultrasound image of vessel tree in 5 cm^3 sample of ovarian fibroma, showing branching, tortuous vessels with caliber changes and color splashes. Still image of same tumor is shown in Figure 3b.

Commented [KP10]: AU: 'and color splashes' added to Videoclip S6 legend based on the Figure 3 legend. Is this correct? (KP)

Videoclip S7 Three-dimensional 360 degree rotating power Doppler ultrasound image of vessel tree in 5 cm^3 sample of ovarian endometroid carcinoma, showing branching, tortuous vessels with caliber changes and color splashes. Still image of same tumor is shown in Figure 3c.

Table 1 Histological diagnoses of 138 adnexal tumors that were difficult to classify as benign or malignant and for which ultrasound volumes were available

Diagnosis	n	(%)
Benign tumor	100	(72)
Endometrioma	7	(5)
Teratoma	5	(4)
Simple cyst or parasalpingeal cyst	3	(2)
Functional cyst	3	(2)
Hydrosalpinx or salpingitis	3	(2)
Peritoneal pseudocyst	2	(1)
Abscess	2	(1)
Fibroma	18	(13)
Serous cystadenoma/cystadenofibroma	30	(22)
Mucinous cystadenoma/cystadenofibroma	23	(17)
Rare benign tumor*	4	(3)
Borderline tumor	15	(11)
Stage I	14	(10)
Stage II	1	(<1)
Stage III or IV	0	0
Primary invasive tumor	22	(16)
Stage I	7	(5)
Stage II	1	(<1)
Stage III	8	(6)
Stage IV	0	0
Rare malignant tumor†	6	(4)
Metastasis in ovary from another primary tumor	1	(<1)

*Rare benign tumors included one Brenner tumor and three cases of struma ovarii. †Rare malignant tumors included three granulosa cell tumors, one immature teratoma, one Sertoli cell tumor and one gastrointestinal stroma' tumor

Table 2 Clinical background and ultrasound characteristics of 138 adnexal tumors that were difficult to classify as benign or malignant and for which ultrasound volumes were available

Characteristic	Value	
Clinical variables		
Age, years	54 ± 17	
Postmenopausal	73	(53)
Hysterectomy	9	(7)
Hormonal replacement therapy	13	(9)
Personal history of ovarian cancer	4	(3)
Family history of ovarian cancer	3	(2)
CA125, U/mL	20 (4 - 1302)*	
Gray scale ultrasound variables		
Largest diameter, mm	69 (10 - 310)	
Bilateral tumors	22	(16)
Ascites	5	(4)
Type of mass		
Unilocular	1	(<1)
Unilocular solid	23	(17)
Multilocular	32	(23)
Multilocular solid	56	(41)
Solid	26	(19)
Number of locules in cases of multilocular or multilocular solid tumor		
2	10/88	(11)
3	8/88	(9)
4	8/88	(9)
5-10	23/88	(26)
>10	39/88	(44)
Tender mass at ultrasound examination	9	(7)
Echogenicity of cyst fluid		
Anechoic	36	(26)
Low level	47	(34)
Ground glass	10	(7)
Hemorrhagic	3	(2)
Mixed	16	(12)
No cyst fluid	26	(19)
Papillary projections	45	(33)
Flow in papillations	20/45	(44)
Number of papillations	2 (1 - ≥4)	
Height of papillations, mm	7 (3 - 45)	
Mass with solid component	105	(76)
Largest diameter of largest solid component, mm	24 (3 - 180)	
Incomplete septum	2	(1)
Irregular walls	77	(56)
Shadows	20	(14)
Doppler ultrasound variables		
Color Score		
1	10	(7)
2	44	(32)
3	70	(51)
4	14	(10)

Data are given as mean ± SD, n (%) or median (min-max). *Data available for 117 cases.

Commented [KP11]: AU: According to style, the number of cases with available data for CA125 has been moved to the footnote from the body of the table. OK? (KP)

Table 3 Ability of vessel morphology on 3D power Doppler, subjective ultrasound assessment and IOTA logistic regression model 1 (LR1) to discriminate correctly between benign and malignant difficult adnexal tumors

Diagnostic method	**Sensitivity**		**Specificity**		**LR+**	**LR-**	**P**
US examiner uncertain or LR1 result ambiguous (n = 138)							
Whole tumor vessel morphology							
Branching vessels	89%	(34/38)	33%	(33/100)	1.34	0.32	0.009
Densely packed vessels	63%	(24/38)	83%	(83/100)	3.72	0.44	<0.001
Caliber changes in vessels	66%	(25/38)	68%	(68/100)	2.06	0.50	<0.001
Color splashes	50%	(19/38)	78%	(78/100)	2.27	0.64	0.002
Tortuous vessels	66%	(25/38)	70%	(70/100)	2.19	0.49	<0.001
Biopsy vessel morphology							
Branching vessels	84%	(32/38)	34%	(34/100)	1.28	0.46	0.04
Caliber changes in vessels	79%	(30/38)	63%	(63/100)	2.13	0.33	<0.001
Color splashes	53%	(20/38)	69%	(69/100)	1.70	0.69	0.02
Tortuous vessels	79%	(30/38)	64%	(64/100)	2.19	0.33	<0.001
Bridges between vessels	42%	(16/38)	81%	(81/100)	2.22	0.72	0.007
Subjective assessment	74%	(28/38)	74%	(74/100)	2.83	0.36	<0.001
LR1 (10% risk cut-off)	92%	(35/38)	23%	(23/100)	1.20	0.34	0.03
US examiner uncertain (n = 79)							
Whole tumor vessel morphology							
Branching vessels	89%	(24/27)	33%	(17/52)	1.32	0.34	0.06
Densely packed vessels	67%	(18/27)	83%	(43/52)	3.85	0.40	<0.001
Caliber changes in vessels	63%	(17/27)	67%	(35/52)	1.93	0.55	0.01
Color splashes	52%	(14/27)	83%	(43/52)	3.00	0.58	0.003
Tortuous vessels	63%	(17/27)	63%	(33/52)	1.72	0.58	0.02

Commented [KP12]: AU: What specifically are these P-values in Table 3 for? Should this be made clear? (KP)

Biopsy vessel morphology							
Branching vessels	85%	(23/27)	35%	(18/52)	1.30	0.43	0.07
Caliber changes in vessels	78%	(21/27)	62%	(32/52)	2.02	0.36	0.001
Color splashes	59%	(16/27)	71%	(37/52)	2.05	0.57	0.009
Tortuous vessels	81%	(22/27)	67%	(35/52)	2.49	0.28	<0.001
Bridges between vessels	44%	(12/27)	79%	(41/52)	2.10	0.71	0.03
Subjective assessment	74%	(20/27)	60%	(31/52)	1.83	0.44	0.004
LR1 (10% risk cut-off)	89%	(24/27)	19%	(10/52)	1.10	0.58	0.34
LR1 result ambiguous (n = 87)							
Whole tumor vessel morphology							
Branching vessels	94%	(16/17)	30%	(21/70)	1.35	0.20	0.06
Densely packed vessels	53%	(9/17)	83%	(58/70)	3.09	0.57	0.004
Caliber changes in vessels	71%	(12/17)	69%	(48/70)	2.25	0.43	0.005
Color splashes	47%	(8/17)	77%	(54/70)	2.06	0.69	0.07
Tortuous vessels	65%	(11/17)	71%	(50/70)	2.27	0.49	0.01
Biopsy vessel morphology							
Branching vessels	76%	(13/17)	31%	(22/70)	1.12	0.75	0.77
Caliber changes in vessels	76%	(13/17)	64%	(45/70)	2.14	0.37	0.005
Color splashes	47%	(8/17)	69%	(48/70)	1.50	0.77	0.26
Tortuous vessels	71%	(12/17)	60%	(42/70)	1.77	0.49	0.03
Bridges between vessels	35%	(6/17)	81%	(57/70)	1.90	0.80	0.19
Subjective assessment	82%	(14/17)	79%	(55/70)	3.84	0.23	<0.001
LR1 (10% risk cut-off)	100%	(17/17)	19%	(13/70)	1.23	Not possible to calculate	0.06

LR1 model uses 10% risk cut-off to predict malignancy[6,7]. LR+, positive likelihood ratio; LR-, negative likelihood ratio.
Correction for multiple testing was not done because this is exploratory analysis

Table 4 Interobserver agreement for assessment of vessel morphology by 3D power Doppler ultrasound in 138 difficult adnexal tumors

Variable	% agreement (95% CI)	Kappa (95% CI)
Whole tumor		
Branching	88 (82 to 93)	0.71 (0.58 to 0.84)
Densely packed	90 (84 to 94)	0.77 (0.65 to 0.88)
Caliber changes	78 (71 to 84)	0.55 (0.41 to 0.69)
Color splashes	81 (74 to 87)	0.57 (0.42 to 0.71)
Tortuous	85 (78 to 90)	0.69 (0.56 to 0.81)
Biopsy		
Branching	88 (81 to 92)	0.70 (0.57 to 0.83)
Caliber changes	78 (70 to 84)	0.55 (0.41 to 0.69)
Color splashes	86 (79 to 90)	0.67 (0.54 to 0.80)
Tortuous	83 (75 to 88)	0.65 (0.52 to 0.78)
Bridges	86 (80 to 91)	0.63 (0.47 to 0.78)

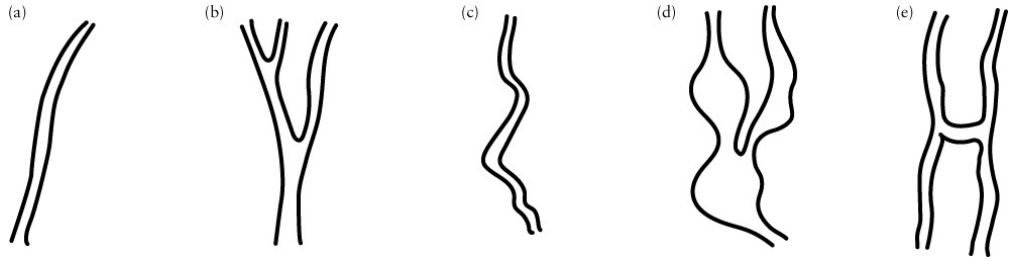


Figure 1

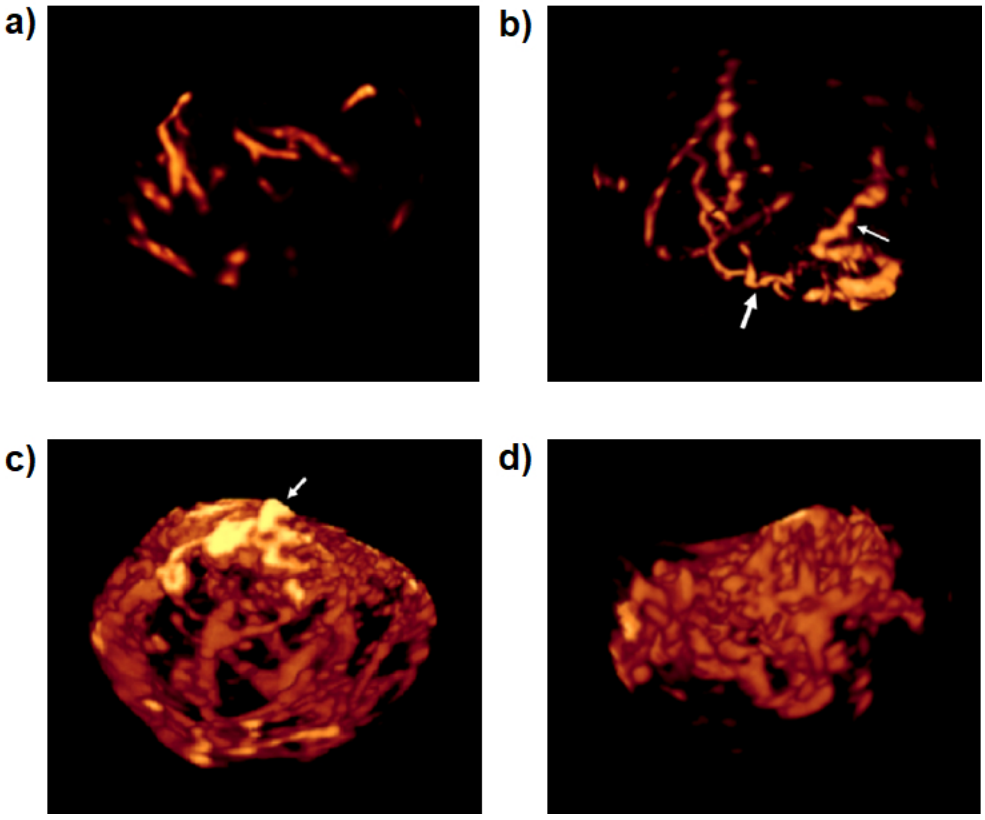


Figure 2

198x164mm (96 x 96 DPI)

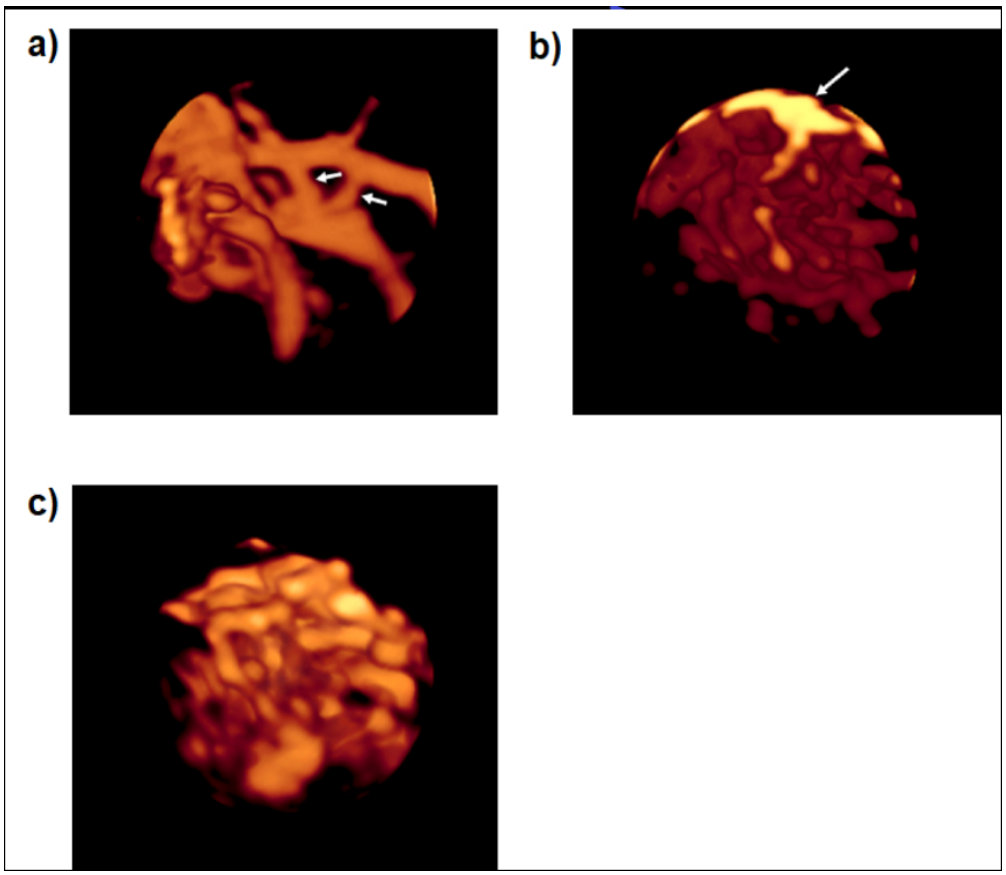


Figure 3

228x197mm (96 x 96 DPI)

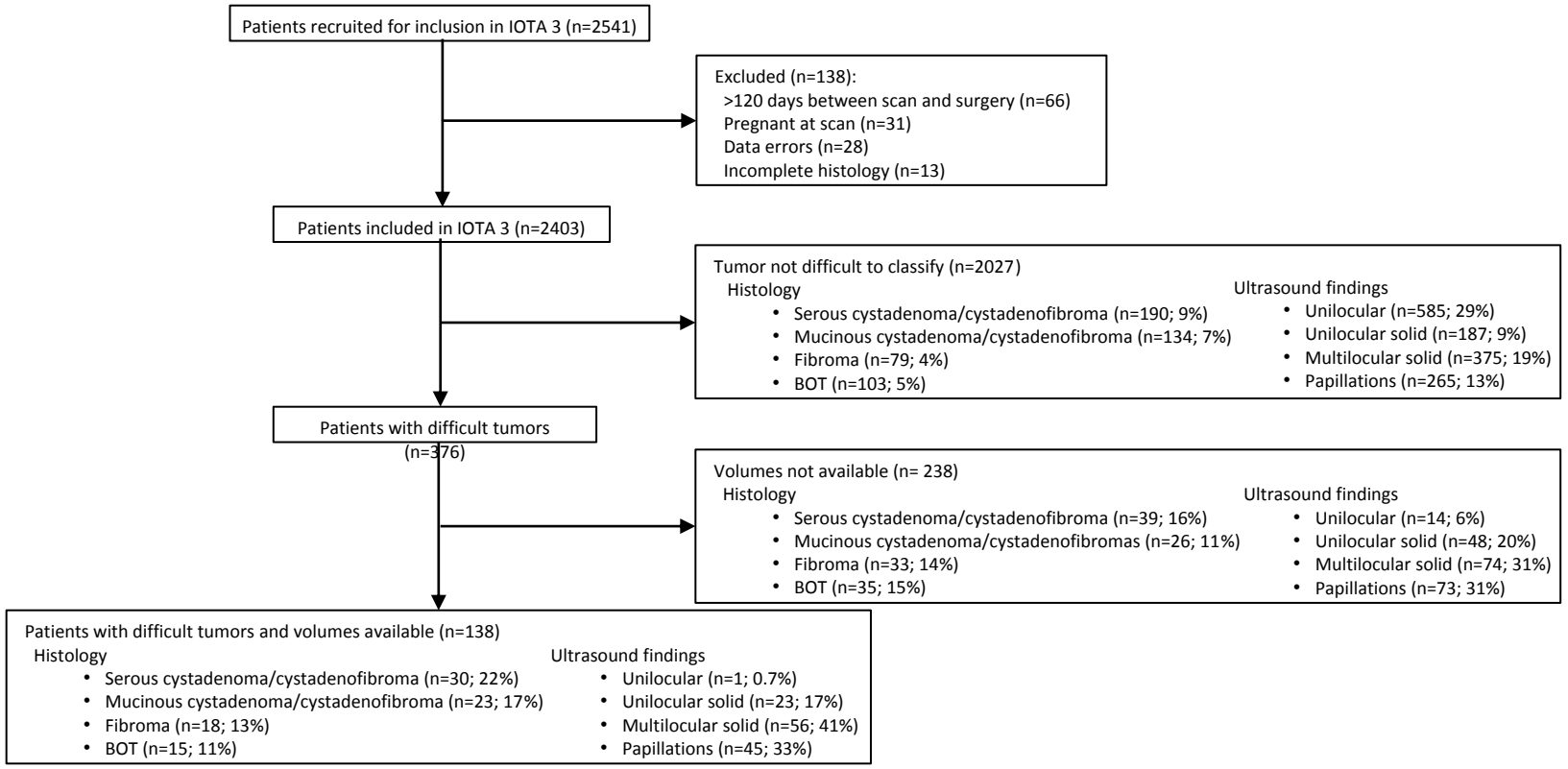
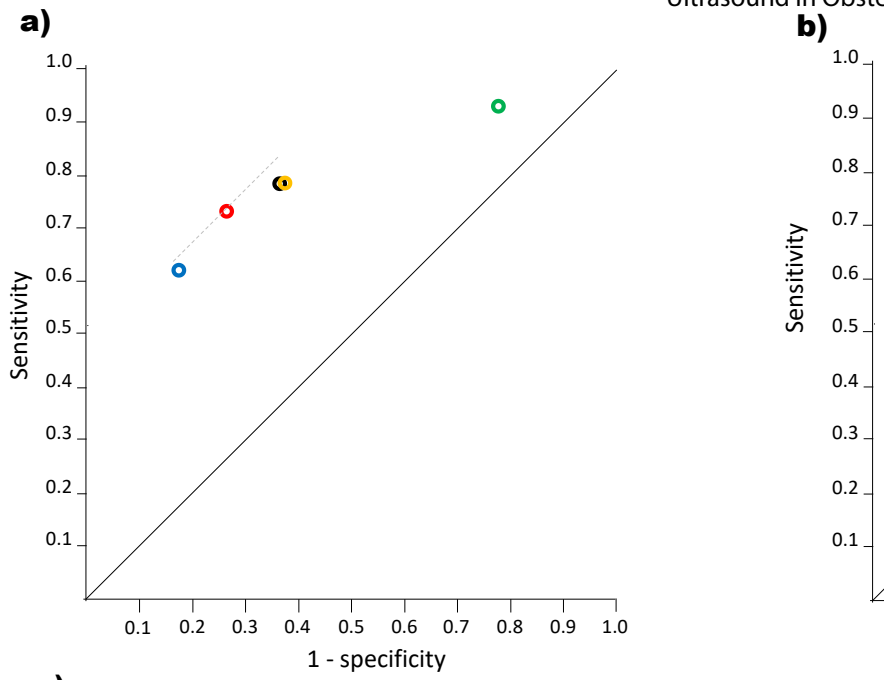
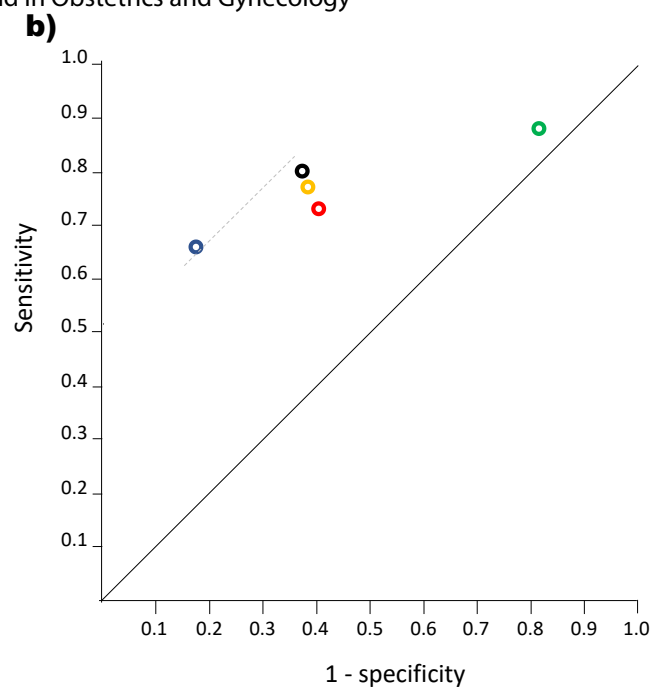
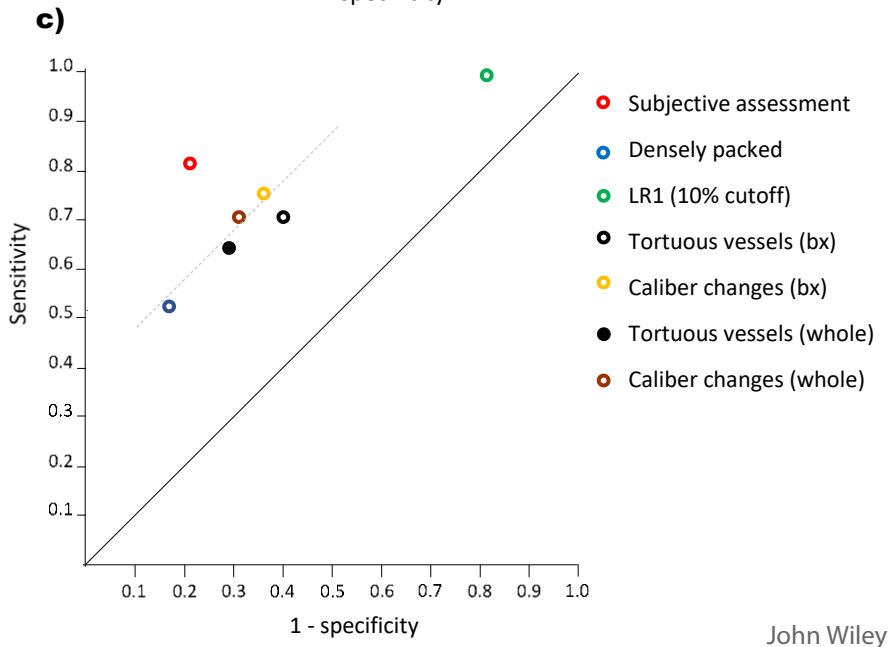

Patients recruited for inclusion in IOTA 3 (n=2541)
Excluded (n=138):
>120 days between scan and surgery (n=66)
Pregnant at scan (n=31)
Data errors (n=28)
Incomplete histology (n=13)
Patients included in IOTA 3 (n=2403)
Tumor not difficult to classify (n=2027)
Histology
Serous cystadenoma/cystadenofibroma (n=190; 9%)
Mucinous cystadenoma/cystadenofibroma (n=134; 7%)
Fibroma (n=79; 4%)
BOT (n=103; 5%)
Ultrasound findings
Unilocular (n=585; 29%)
Unilocular solid (n=187; 9%)
Multilocular solid (n=375; 19%)
Papillations (n=265; 13%)
Patients with difficult tumors (n=376)
Volumes not available (n= 238)
Histology
Serous cystadenoma/cystadenofibroma (n=39; 16%)
Mucinous cystadenoma/cystadenofibromas (n=26; 11%)
Fibroma (n=33; 14%)
BOT (n=35; 15%)
Ultrasound findings
Unilocular (n=14; 6%)
Unilocular solid (n=48; 20%)
Multilocular solid (n=74; 31%)
Papillations (n=73; 31%)
Patients with difficult tumors and volumes available (n=138)
Histology
Serous cystadenoma/cystadenofibroma (n=30; 22%)
Mucinous cystadenoma/cystadenofibroma (n=23; 17%)
Fibroma (n=18; 13%)
BOT (n=15; 11%)
Ultrasound findings
Unilocular (n=1; 0.7%)
Unilocular solid (n=23; 17%)
Multilocular solid (n=56; 41%)
Papillations (n=45; 33%)

Fig. 5







International Ovarian Tumour Analysis (IOTA) Phase 3

A multicentre study on the pre-operative characterisation of ovarian tumours based on artificial intelligence models.

STUDY CO-ORDINATOR

Dirk Timmerman, MD, PhD
Department of Obstetrics and Gynaecology, University Hospitals Leuven,
Herestraat 49, B-3000 Leuven, BELGIUM.
Telephone: + 32 16 344201 (office) Fax: + 32 16 344205
+ 32 16 344215 (secretary) Mobile + 32 497 896025
E-mail: dirk.timmerman@uzleuven.be

STEERING COMMITTEE

Tom Bourne (Imperial College, University College London)
Dirk Timmerman (University Hospitals, Leuven)
Antonia C. Testa (Università Cattolica di Sacro Cuore, Roma)
Lil Valentin (University of Lund / Malmö)
Sabine Van Huffel (ESAT-SISTA, K.U. Leuven)
Ignace Vergote (University Hospitals, Leuven)

LIST OF PRINCIPAL INVESTIGATORS AND PARTICIPATING CENTRES (AND CENTRE CODE)

LPO (i.e. centre code)
Artur Czekierdowski, MD, PhD
Professor, 1st Department of Gynecology, Medical University in Lublin
16 Staszica Str., 20-081 Lublin, POLAND
Telephone: +4881 532 78 47 Fax: +4881 532 06 08
E-mail: ginonkol@am.lublin.pl
E-mail: a.czekierdowski@am.lublin.pl

LSW
Elisabeth Epstein, MD, PhD
Department of Obstetrics and Gynaecology, Lund University,
Lund, SWEDEN
Telephone: +46-46-171000
E-mail: Elisabeth.Epstein@med.lu.se

PCR
Daniela Fischerová, MD
Oncogynecological Center Department of Obstetrics and Gynecology
General Faculty Hospital of Charles University
Apolinarska 18, Prague 2, CZECH REPUBLIC
Telephone: +42 0603285097 (mobile)

E-mail: daniela.fischerova@seznam.cz

CIT
Dorella Franchi, MD
Dt.ssa Ginecologic Oncology Unit, Division of Gynecology,
IEO, Milano, ITALY
E-mail: dorella.franchi@ieo.it

GIT
Stefano Greggi, MD, PhD
Director Gynecologic Oncology
Instituto Nat del Tumori, Fondazione Pascale
Via Mariano Semmola, 1
80121 Napoli, ITALY
Telephone: +39 81 5903417 Fax: +39 81 5903851
E-mail: s.greggi@tin.it

SIT
Stefano Guerriero, MD, PhD
Department of Obstetrics and Gynaecology, University of Cagliari,
Ospedale San Giovanni di Dio
Via Ospedale 46, 09124, Cagliari, Sardinia, ITALY
Telephone: +39 070 6092467 Fax: +39 070 668575
E-mail: gineca.sguerriero@tiscali.it

BCH
Jingzhang, MD
Professor, Ultrasound Department, Chinase PLA General Hospital
No.28 Fuxing Road, Beijing, P.R. of CHINA 100853
Telephone: +86 010 88626007 Fax: +86 010 88626007
E-mail: zjbch@vip.sina.com
E-mail: zjbch@hotmail.com

KUK
Davor Jurkovic, MD, PhD, MRCOG
Consultant, Department of Obstetrics and Gynaecology,
University College Hospital, London, UK
e-mail davor.jurkovic@kcl.ac.uk

MIT
Francesco P.G. Leone, MD, PhD
DSC L. Sacco, Università di Milano, Milano, ITALY.
Telephone: +39 02 39042264 Fax: +39 02 3565061
E-mail: f.leone@hsacco.it

OIT
Robert Fruscio, MD
Andrea A. Lissoni, MD
Clinica Ostetrica e Ginecologica, Ospedale S. Gerardo, Università di Milano Bicocca
via Solferino 16, Monza I-20052, ITALY
E-mail: andreaalberto.lissoni@unimib.it

OCA
Henry Muggah, MD, FRCSC
Professor, Department of Obstetrics and Gynaecology, McMaster University, St. Joseph's Hospital
301 James Street South, 2nd Floor, Fontbonne Building, Hamilton, Ontario L8N 4A6, CANADA
Telephone: +1 905 521 6041 Fax: +1 905 521 6089
E-mail: muggah@mcmaster.ca

NIT
Dario Paladini, MD.
Professor, Obstetrics and Gynecology
Università degli Studi di Napoli "Federico II" (University Federico II of Naples)
Via Petrarca, 72, 81022, Napoli, ITALY
Telephone: +39 081 7462951 Fax: +39 081 7463255
Telephone: +39 339 2461688 (mobile)
E-mail: paladini@unina.it

UDI
Alberto Rossi, MD
Clinica Ostetrica Ginecologica
Università degli Studi di Udine
Piazza Misericordia, 33100 – UDINE – Italy
Telephone +39 349 5606530
E-mail: roalbert@tiscali.it

BIT
Luca Savelli, MD
Reproductive Medicine Unit, Department of Obstetrics and Gynecology
University of Bologna, Via Massarenti, 13, 40138 Bologna, ITALY
Telephone: +39 051 6364424 (work) Fax: +39 051 6360892
Via Mengoli 31/4, 40138 Bologna, ITALY (home) Telephone: +39 347 4248767 (mobile)
E-mail: savelliluca@libero.it E-mail: savelli@aosp.bo.it

RIT
Antonia Carla Testa, MD
Unità Operativa di Ginecologia Oncologica

Salvatore Mancuso, MD, PhD
Head of the Department of Obstetrics and Gynaecology
Istituto di Clinica Ostetrica e Ginecologica, Università Cattolica di Sacro Cuore
Largo Agostino Gemelli 8, Roma, ITALY.
Telephone: +39 06 30154979 Mobile: +39 339 7044256 Fax: +39 06 35510031
E-mail: atesta@rm.unicatt.it, foc.ovest@libero.it

LBE
Dirk Timmerman, MD, PhD
Professor and clinical head, Department of Obstetrics and Gynaecology,
University Hospitals KU Leuven,
Herestraat 49, B-3000 Leuven, BELGIUM.
Telephone: + 32 16 344215 (office) Fax: + 32 16 344205 (office)
E-mail: dirk.timmerman@uz.kuleuven.ac.be

Caroline Van Holsbeke, MD,
Department of Obstetrics and Gynaecology, University Hospitals KU Leuven
Herestraat 49, B-3000 Leuven.
Telephone: + 32 16 343642 (office) Fax: + 32 16 344205 (office)
E-mail: caroline.vanholsbeke@uz.kuleuven.ac.be

VIT
Diego Trio, MD
Dottor, Department of Obstetrics and Gynecology
Mauro Busacca, MD
Professor, Department of Obstetrics and Gynecology
Macedonio Melloni Hospital, University of Milan,
Via Melloni, 52, 20122 Milano, ITALY
Fax: +39 02 710645
E-mail: diego.trio@libero.it

MSW
Lil Valentin, MD, PhD
Professor, Obstetrics and Gynecology
Malmö University Hospital, SE20502 Malmö, SWEDEN. (Lövviksg 7B, S21374 Malmö)
Telephone: +46 40 949726 (home) Fax: +46 40 949726 (home)
Telephone: +46 40 332149 or 332094 (office) Fax: +46 40 962600 (office)
E-mail: lil.valentin@obst.mas.lu.se
E-mail: lil.valentin@med.lu.se

GBE
Caroline Van Holsbeke, MD
Willem Ombelet, MD, PhD, Head of Department
Department of Obstetrics and Gynaecology,
Ziekenhuis Oost-Limburg, Genk (ZOL)
Schiepse Bos 6, 3600 Genk, BELGIUM

Tel: +32 89 327524 Fax: +32 89 327920
Email: caroline.van.holsbeke@skynet.be

OTHER CONTRIBUTORS

Lieveke Ameye, M.Sc., PhD
Departement of electrical engineering (ESAT SCD-SISTA), K.U.Leuven
KasteelparkArenberg 10, B-3001 Heverlee-Leuven, BELGIUM.
Telephone: +32 477 794332 (mobile) Fax: + 32 16 321970 (office)
Email: lieveke.ameye@hotmail.com
Email: lieveke.ameye@esat.kuleuven.be

Anneleen Daemen, M.Sc.
Department of electrical engineering (ESAT SCD-SISTA), K.U. Leuven
Kasteelpark Arenberg 10, B-3001 Heverlee-Leuven, BELGIUM.
Telephone: +32 473 529251 (mobile) Fax: + 32 16 321970 (office)

Lieven De Clercq, M.Sc.
Department of electrical engineering (ESAT SCD-SISTA), K.U. Leuven
Kasteelpark Arenberg 10, B-3001 Heverlee-Leuven, BELGIUM.
Telephone: + 32 16 321143 (office) Fax: + 32 16 321970 (office)
E-mail: lieven.declercq@esat.kuleuven.be

Bart De Moor, M.Sc., PhD
Professor, Head of BIOI, Department of electrical engineering (ESAT SCD-SISTA), K.U. Leuven
Kasteelpark Arenberg 10, B-3001 Heverlee-Leuven, BELGIUM.
Telephone: + 32 16 321715 (office) Fax: + 32 16 321970 (office)
E-mail: bart.demoor@esat.kuleuven.be

Ben Van Calster, M.Sc., PhD
Department of electrical engineering (ESAT SCD-SISTA), K.U. Leuven
Kasteelpark Arenberg 10, B-3001 Heverlee-Leuven, BELGIUM.
Telephone: + 32 498 857512 (mobile) 32 16 321970 (office)
E-mail: ben.vancalster@esat.kuleuven.be

Sabine Van Huffel, M.Sc., PhD
Professor, Head of the BIOMED research group
Department of electrical engineering (ESAT SCD-SISTA), K.U. Leuven
Kasteelpark Arenberg 10, B-3001 Heverlee-Leuven, BELGIUM.
Telephone: + 32 16 321703 (office) Fax: + 32 16 321970 (office)
E-mail: sabine.vanhuffel@esat.kuleuven.be

Joan Veldman, MD
Erika Werbrouck, MD
Gynaecologists, Research fellows,
Department of Obstetrics and Gynaecology, University Hospitals KU Leuven,
Herestraat 49, B-3000 Leuven, BELGIUM.
Telephone: + 32 16 343642 (office) Fax: + 32 16 344205 (office)

E-mail: joan.veldman@uzleuven.be and erika.werbrouck@uzleuven.be

Ignace Vergote, MD, PhD
Professor, Head of Department of Obstetrics and Gynaecology, University Hospitals KU Leuven, Herestraat 49, B-3000 Leuven, BELGIUM.
Telephone: + 32 16 344636 (office) Fax: + 32 16 344629 (office)
E-mail: ignace.vergote@uzleuven.be

IOTA website : http://www.iota-group.org

Summary

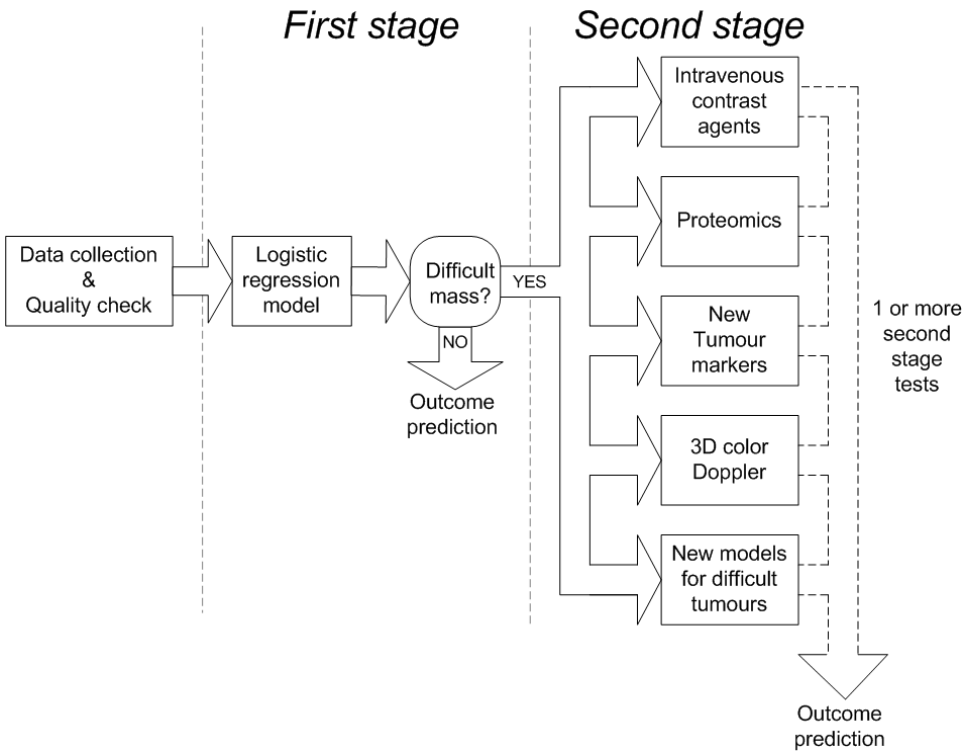
The aim is to prospectively test and implement previously developed mathematical algorithms for the pre-operative classification of adnexal masses. IOTA phase 3 is a prospective multicentre study. The information obtained will be used to define the optimal management of patients presenting with adnexal tumours. 2,000 patients, with at least one adnexal mass, will be recruited and studied within 3 months before investigative surgery. Medical and family histories will be recorded. Transvaginal ultrasonography with colour Doppler imaging will be used to derive indices of tumour form and blood flow. A sample of peripheral venous blood will be taken for the analyses of serum CA-125 and other tumour markers and in certain centres additional blood will be taken and stored appropriately to test proteomic pattern analysis. Findings at surgery and the histological classification of excised tissues as malignant or benign (and by cell type) will be used as outcome measures. Conventional and novel algorithms (including proteomic patterns), which can be used to effectively classify difficult adnexal masses will be tested prospectively in centres throughout the world.

1. Study objectives

The main objective of this project is to improve preoperative diagnosis and subsequent management of patients with adnexal tumours by advanced algorithms in order to decrease morbidity and costs and in order to improve survival of patients with ovarian cancer. In particular, we aim to validate the added value of mathematical models as new diagnostic tool in the prediction of ovarian cancer in clinical practice. First of all, we aim to prove their enhanced diagnostic performance and generalized applicability as a first stage examination. In cases where prediction is unreliable, we aim to further improve the predictive performance of this diagnostic tool with second stage tests. As a side effect, a consistent large database of prospectively collected clinical, biochemical and ultrasound data will be generated for research into disease processes, as well as for the development and validation of other advanced algorithms.
To achieve these objectives the project is divided in work packages each having its own sub-objective.

1. Prospective external validation of predictive mathematical models and pattern recognition to distinguish between malignant and benign adnexal masses (Responsible person: D Timmerman; all centres)
2. Intravenous ultrasound contrast agents: quantitative analysis of contrast uptake and washout in tumour (Responsible person: A Testa; RIT, LBE, MSW, LSW,GBE, MIT, UDI). Not in UK centres.
3. Proteomic analysis (D Timmerman; only in LBE, MSW, GBE)
4. New set of tumour markers (D Timmerman; all centres with ethical approval and informed consent)
5. Validation of 3D power Doppler (L Valentin; all centres with GE Voluson equipment: LBE, GBE, SIT, LPO, MSW, RIT, CIT, KUK/UCH, NIT, MIT, LSW)
6. Validation of a new model, based on grey scale ultrasound and colour Doppler information, specifically designed for "difficult" tumours (all centres)

Schematic representation of the diagnostic algorithm applied in IOTA Phase 3:



2. Background

At present malignant ovarian tumours are diagnosed at an advanced stage in 75 % of the cases and they result in the highest mortality figures of all gynaecological cancers. An estimated 22,430 new cases of ovarian cancer are expected to be diagnosed in 2007 in the United States, according to the American Cancer Society, with about 15,280 deaths. Worldwide there are more than 190,000 new cases of ovarian cancer each year, according to the International Agency for Research on Cancer. These figures underscore the need for effective tests to diagnose ovarian cancer at early stages and to improve management of patients with adnexal tumours.

The **IOTA study (International Ovarian Tumour Analysis)** is a multicentre collaborative project for the pre-operative characterisation of ovarian tumours based on predictive computer models.

The **first phase** of IOTA was conducted between 1999 and 2002. Several new mathematical models were developed based on the prospectively collected data of 1066 patients with a persisting adnexal tumour in 9 European Centres. Between 2002 and 2005 three centres continued the prospective collection in order to be able to perform an internal validation of mathematical models developed in IOTA phase 1. In this so-called IOTA **phase 1b** study a new dataset of 507 new patients was prospectively collected in 3 out of the 9 original IOTA centres. All models proved to perform excellent with area under the ROC curves of more than 0.94.

The **second phase** of IOTA consisted of an external validation of the models and this was conducted between 2005 and 2007. The diagnostic algorithms were prospectively validated on 2,093 patients with adnexal tumours in 19 centres in Belgium, Italy, UK, Sweden, Poland, Czech Republic, Canada, and China. A first analysis showed that overall performance of the logistic regression models was excellent (area under the ROC curve 0.94). A subgroup of "uncertain" tumours needs a reliable second stage test in order to help even experienced ultrasound examiners.

The **third phase** of the IOTA study is planned for 2009.

Rationale

Diagnostic and therapeutic relevance of predicting ovarian cancer preoperatively:

The pre-operative assessment of adnexal tumours remains a major challenge for clinicians. Advances in surgery have provided more treatment options, but their potential usefulness depends upon a prior assessment of the mass using non-invasive procedures. It is often difficult to determine pre-operatively the nature (benign or malignant) of ovarian tumours. However, this knowledge is essential for obtaining a reliable diagnosis and for an appropriate management, both of which will influence the outcome for the patient and the medical costs. In case of benign functional cysts a laparoscopy or any other surgical intervention should be avoided in order not to create unnecessary morbidity and impaired fertility. On the contrary, most benign non-functional tumours can be treated laparoscopically or with a low transverse abdominal incision. Patients with a suspect adnexal mass require a series of expensive and unpleasant staging examinations and in most cases also an exploratory laparotomy through a median incision is indicated, because the rupture of a stage 1 ovarian cancer during the operation may worsen the prognosis. This procedure should only be performed by a surgeon with proper skills and experience in debulking surgery, since the amount of residual malignant tissue after primary surgery is one of the most important prognostic factors in ovarian cancer. Furthermore, appropriate pre-, per-, and postoperative measures should be taken.

During recent years several techniques have been introduced to differentiate between benign and malignant ovarian lesions. These techniques include serum tumour markers, transvaginal ultrasonography with a variety of morphological scoring systems, or sometimes combined with colour Doppler sonography, computerised tomography (CT scan), and magnetic resonance imaging (MRI). Despite numerous publications on each of these techniques, an insufficient amount of data is available to provide for a rational clinical management.

Previous studies to distinguish between benign and malignant ovarian tumours were small and single-centre studies. The IOTA group first published a consensus statement on terms, definitions and measurements, which is now widely used. The IOTA phase 1 and phase 2 studies were by far the largest multicentre studies ever conducted in this area and they received international scientific attention and wide media coverage. The third phase is now needed to validate the use of second stage tests in order to provide reliable classifications in cases where the presently developed mathematical models result in uncertain diagnoses.

Official approval by the Ethical Committee

The multicentre project IOTA phase 3 will be submitted to the Ethical Committee of the University Hospitals Leuven as main investigating centre as well as in each participating centre.
Second stage tests are not uniformly performed in all centres and therefore each participating centre will provide appropriate patient information leaflets and request specific ethical approval to their Ethical Committee.

Insurance policy

This multicentre international study is initiated by the University Hospitals Leuven, Belgium. Each participating centre outside Belgium is fully responsible for optimal patient care and its own patient management in agreement with local laws. Each centre is also responsible for all legal aspects and for its own insurance of all matters related to this study.

Financial Support

The IOTA phase 3 project is supported by an Applied Biomedical Research grant (Toegepast Biomedisch Onderzoek, TBM) from the Flanders Institute for Scientific and Technological Research: IWT Flanders, Belgium (IWT-TBM 070706). This grant covers costs of central data collection, proteomic analysis, analysis of new tumour markers and statistical analyses.

There is no financial recompensation for principal investigators nor patients.

3. Design

Number of patients / tumours

2,000 patients with at least one histologically examined adnexal mass will be recruited for the third part of the IOTA trial (i.e. the prospective testing of new mathematical models and second stage tests).

Outcome measures

The histological classification of removed tissue and the findings at surgery.

Duration

In **year 1**, the IOTA phase 3 study protocol will be implemented in the different participating centres. Data collection with minor built-in quality checks will occur similarly as in IOTA phase 2. At the end of year 1, complete data of 1,000 patients will be available.
In **year 2**, data collection will continue until the target sample size of 2,000 patients is reached. At the end of data registration, a detailed data quality control will be applied. All mathematical models developed in IOTA phase 1 and 2 will be prospectively validated as 1st stage test. At the end of year 2, the predictive performance of two 2nd stage tests in case of difficult tumours will also be validated: 3D power Doppler and a logistic regression model using grey scale ultrasound that was specifically constructed for difficult tumours.
In **year 3**, the validation of the three remaining 2nd stage tests will be investigated: intravenous ultrasound contrast agents, proteomic analysis and new tumour markers. During the third year different publications each highlighting different second stage tests will be prepared as well as overall study reports in high impact medical journals (e.g. JCO, J Natl Cancer Inst…)
No problems are expected with the data collection, data quality control and data storage, and data analysis since these will be similar to the successful IOTA phase 2 study and because of the existing expertise in the IOTA group.

4. Patient entry

Inclusion criteria

All patients assessed with transvaginal ultrasound by the principal investigators (or their appropriately trained medical staff) and found to have an apparent persistent extrauterine pelvic mass* (from this point termed: mass), provided that

1. the patient is fit for surgery
2. the patient gives informed consent
3. the patient has at least one remaining ovary

- *: a persistent extrauterine pelvic mass is defined as a mass judged by ultrasonography to be of adnexal origin, and not consistent with normal physiology
- Patients with bilateral tumours will be included in the study. Both tumours are examined. The worst case mass based on morphological characteristics (according to the sonographer) is recorded first (a list of examples is attached). If both masses are morphologically equal, the largest one is recorded first.

Exclusion criteria

- Patients with more than one lesion in the same ovary or in two different ovaries will be included in the data collection, but in case of two different non-physiological conditions (based on the pathology report) the results will be excluded from the main analysis of the data. The reason for this approach is to avoid several difficulties: e.g. to assess the Doppler signals of different masses separately, and to know the contribution of each mass in particular elevation of serum CA 125 levels).
- Pregnancy is not an exclusion criterium, but the data from pregnant patients are analysed separately, because the physiology is different (e.g. different colour Doppler findings and higher serum CA 125 levels).
- Preferably surgery is scheduled within one week after ultrasonography. If more than 90 days have passed before the patient has been operated upon, the patient is excluded for analysis, except if a new scan has been performed prior to surgery.

Consent / information leaflet

Information leaflets are at the discretion of the participating centres.
Approval of the local Ethical Committee for clinical studies is necessary.
Written informed consent is necessary for storing blood samples for later analysis, and for the intravenous contrast study.

Collection of clinical data

Family history:	Number of first degree relatives with ovarian cancer (0-...)
	Number of first degree relatives with breast cancer (0-...)
Medical history:	Personal history of ovarian or breast cancer
	Age (years); Parity (number of deliveries)
	Hysterectomy (yes/no)
	Menopausal status (pre-(1) or postmenopausal (3))
	Years after menopause;
	Day of cycle. Hormonal therapy (yes, no).
	Pelvic pain during the scan: “is the mass painful?” (yes/no)

5. Diagnostic methods

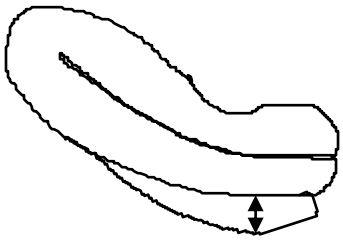
Ultrasonography

Ultrasound variables / definitions

All ultrasound variables are included in the dedicated software, which shows the requested parameters. In the database 0 always means NO and 1 always means YES.
The adnexal lesion is that part of an ovary or of an adnexal mass that is judged by ultrasonography to be not consistent with normal physiology. This can be a persistent unilocular cyst, surrounded by normal looking ovarian stroma with some follicles. In this case the whole ovary containing the cyst is the ‘ovary’, whereas the unilocular cyst is the ‘lesion’. Both are measured and the cyst is described as being ‘unilocular’ and not ‘unilocular-solid’. In other cases the lesion is separate from the ovary (e.g. hydrosalpinx). Again, both ovary and lesion are measured separately. In other cases no normal ovarian stroma is seen. In these cases the lesion and the ovary are undistinguishable and the measurement of lesion and ovary will be the same.

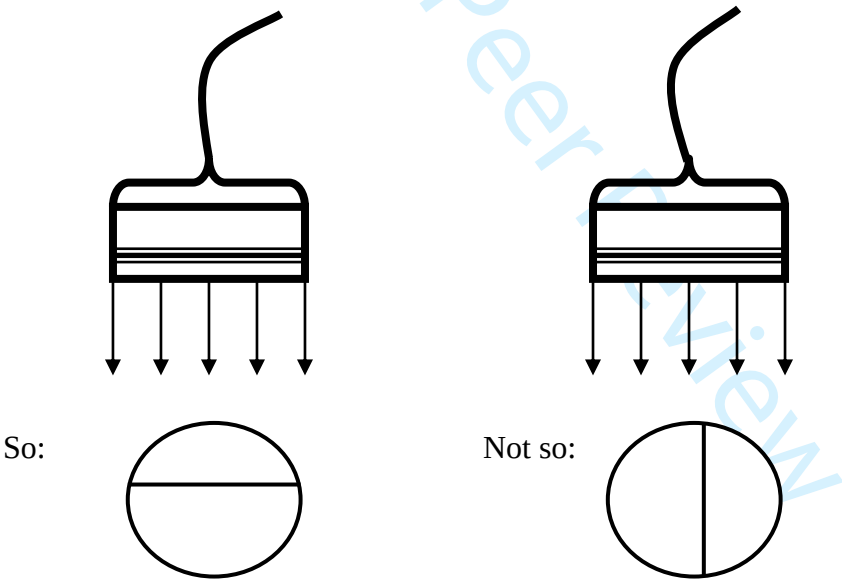
Measurements (in mm): The ovary in two perpendicular planes
The lesion in two perpendicular planes
The volume of the tumor is calculated from the three diameters in two perpendicular planes

- The presence of ascites (*i.e.fluid outside* the pouch of Douglas) is noted (yes/no).
- Fluid in the pouch of Douglas is measured in a sagittal plane (the largest anteroposterior diameter is given).
(see Figure)

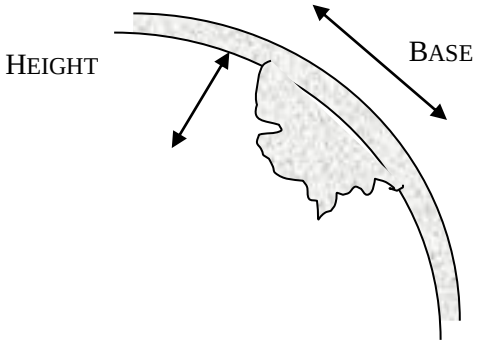


- A septum is defined as a thin strand of tissue running across the cyst cavity from one internal surface to the contralateral side. The thickness of the thickest septum is measured where it appears to be at its widest (other than at its interface at the internal surface of the cyst wall)

It is preferable to measure a septum which is perpendicular to the ultrasound beam



- An incomplete septum (as seen in hydrosalpinges) is defined as a thin strand of tissue running across the cyst cavity from one internal surface to the contralateral side, but which is not complete in some scanning planes. Its presence is noted. If a cyst only has incomplete septa and no real septa, it is unilocular, despite the fact that in certain sections the cyst appears to be multilocular.
- Solid means echogenicity suggesting the presence of tissue (e.g. the myometrium, the ovarian stroma, myomas, fibromas). Methods to distinguish between blood clots and the presence of solid tissue are the use of colour Doppler and to look for internal movement when gently pushing to the structure with the transducer. The presence of flow (with the appropriate settings) is diagnostic for solid tissue. The absence of flow is not informative. In cases of doubt whether it is a blood clot or a solid area, call it solid.
- Solid papillary projections are defined as any solid projections into the cyst cavity from the cyst wall greater than or equal to 3 mm in height

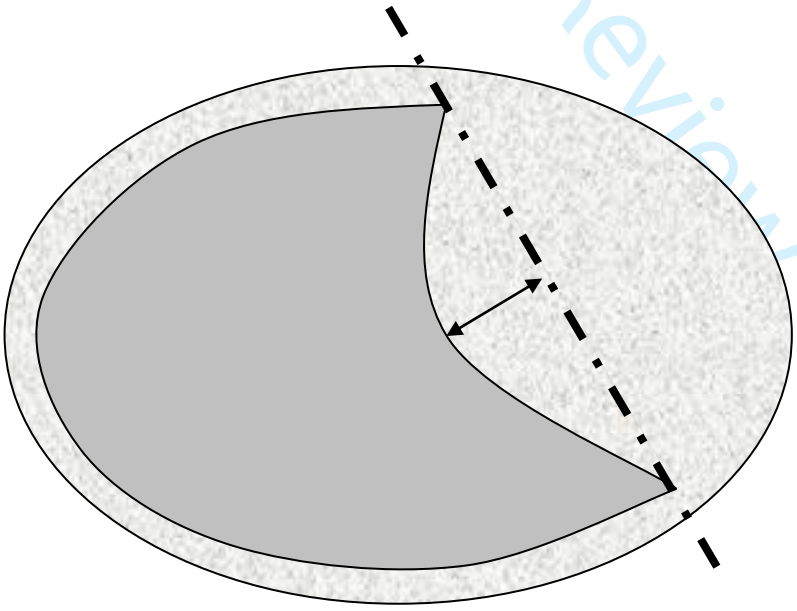


If it is unsure whether solid papillary projections or an incomplete septum are present, the 'worse case scenario' is used. E.g. 'cogwheel excrescences' and 'beads-on-a-string' (as seen in hydrosalpinges) should be classified as papillary excrescences if their height is greater than or equal to 3 mm. The 'white ball' in a dermoid, however, should not be classified as a solid papillary projection.

The 'sludge' on the internal walls of endometriotic cysts is not regarded as a papillary projection. In these cases the internal walls are usually 'irregular'.

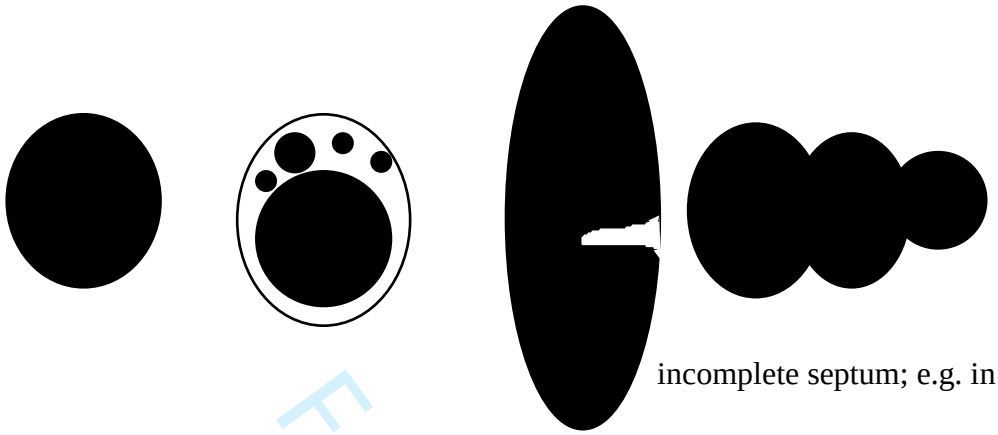
- The number of separate papillary projections is noted (1/2/3/more).
- The presence of flow within some of these projections is noted (yes/no).
- Solid papillary projections are described as being 'smooth' or 'irregular' (e.g. cauliflower-like).

In some cases it is difficult to judge whether it is a papillary projection and from which point to measure the projection. In these cases it may be helpful to use an imaginary line as shown in the following schematic drawing:



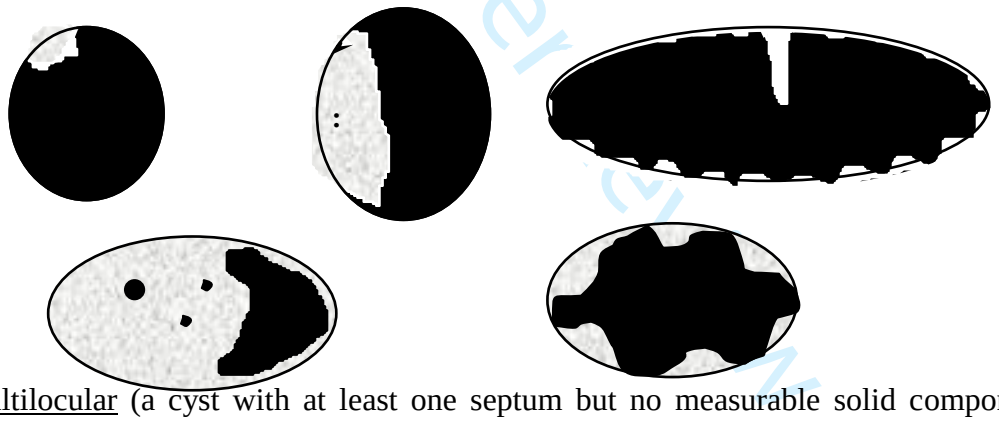
All lesions are qualitatively classified into one of 5 categories:

1. unilocular (a unilocular cyst without septa and without solid parts or papillary structures). Normal ovarian stroma is not regarded as 'solid' (e.g. a peritoneal cyst, containing a normal ovary, is unilocular and not unilocular-solid).

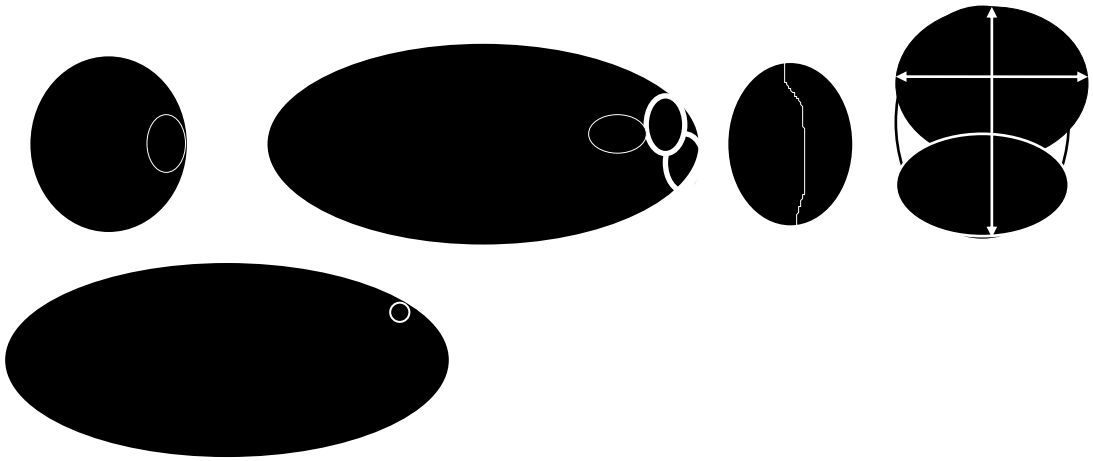


incomplete septum; e.g. in hydrosalpinx)

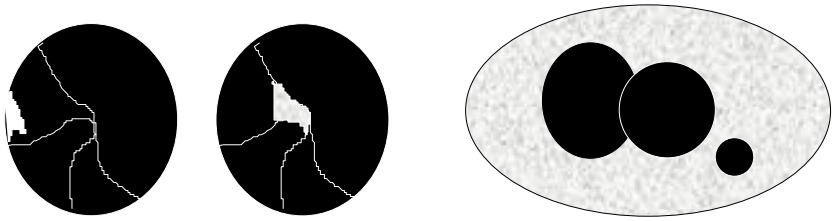
2. unilocular cyst with solid component (a unilocular cyst with a measurable solid component or at least one papillary structure). This category may include pyo- or hydrosalpinges with the so-called 'beads-on-a-string' or 'cogwheel' appearance if ≥ 3 mm. If the solid part contains very small cysts the mass might be unilocular-solid (see below).



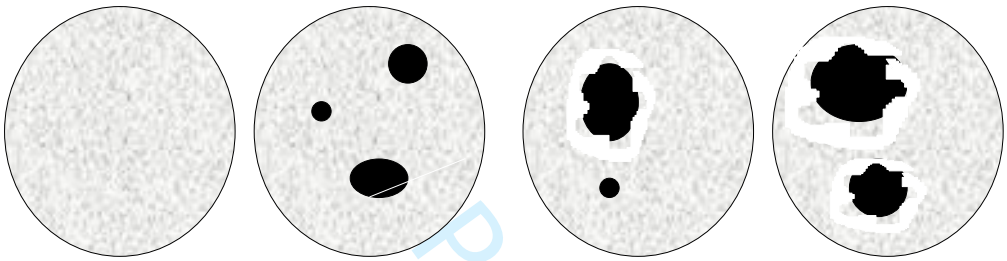
3. multilocular (a cyst with at least one septum but no measurable solid components or papillary projections). The 'lesion' is measured as indicated by the arrows.



4. multilocular with solid component (a multilocular cyst with a measurable solid component or at least one papillary structure)



5. solid (a tumour where the solid components comprise 80% or more of the tumour when assessed in a two-dimensional section).



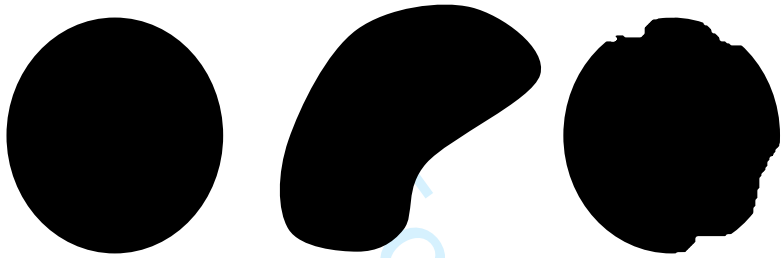
(solid tumour with an irregular cyst wall)

A solid tumour may contain papillary projections protruding into the small cysts.

6. not classifiable because of poor visualization (e.g. strong acoustic shadowing due to calcifications or as seen in certain dermoids *('tip of the ice-berg' sign)*

QUANTITATIVE ASSESSMENT OF MORPHOLOGY

- In cystic-solid tumours the largest solid component is measured separately (in three perpendicular planes). The solid component is noted as being smooth or irregular (e.g. cauliflower-like). In some cases a solid papillary projection is the largest solid component and thus the papillary projection is recorded both as papillary projection and as solid component.
- The internal wall is also noted as being smooth or irregular.

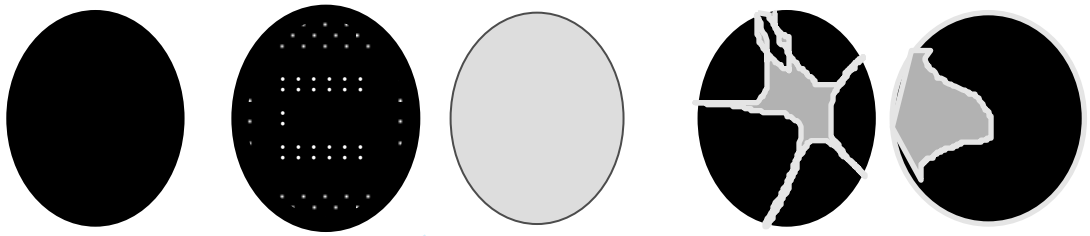


Smooth Smooth Irregular

If there is a solid papillary projection, then the wall is irregular by definition.

- The external wall of cyts is not looked at.
- In cases of solid tumours the description of the internal wall being smooth or irregular is usually not applicable but the outline of the tumour is described as smooth or irregular.
- If there is any irregularity in either the inner wall of any cyst or in the outer wall of a solid tumour or on the surface of a solid component, the lesion is described as 'irregular'.

- The dominant feature of the cystic contents is described as anechoic (black), low-level echogenic (homogeneous low level echogenic as seen in mucinous tumours), 'ground glass' appearance (homogeneously dispersed echogenic cystic contents, as often seen in endometriotic cysts), hemorrhagic (with internal thread-like structures, representing fibrin strands; it is possible to describe the echogenicity as star-shaped, cobweb-like or jelly-like) or mixed echogenic (as often seen in teratomas) (see images attached).



Anechoic Low level Ground glass Hemorrhagic Hemorrhagic

Mixed Mixed (old blood-fluid level or fat-fluid level) Mixed (e.g. abscess)

- The presence of acoustic shadows, defined as loss of acoustic echo behind a sound-absorbing structure, is noted as well. Solid tumours are identified by the appearance of the internal texture, by the absence of internal movement when moving the transducer or by colour Doppler imaging (presence of central flow).
- In solid tumours the dominant feature of any cystic contents is described only if it can be assessed.
- Acoustic streaming, defined as the bulk movement of fluid due to a sound field. It results in movement of fluid particles in the direction of the ultrasound beam away from the transducer ("absent" or "present", mandatory new variable for phase II)
- 'Ovarian crescent sign', defined as the presence of normal ovarian tissue adjacent to an adnexal tumour. ("absent" or "present", mandatory new variable for phase 3)
- Ultrasound evidence of metastases (e.g. "omental cake" or peritoneal tumoural implants). ("absent" or "present", mandatory new variable for phase 3)

Colour Doppler imaging and blood flow indices

Subsequently, the entire tumor is surveyed by CDI. The power, gain and pulse repetition frequency are initially adjusted for maximum sensitivity of low blood flow states. The lowest velocity signals are filtered out by gradually increasing the pulse repetition frequency and flow analysis is concentrated on the highest velocity signals. A subjective semiquantitative assessment of the amount of blood flow (area and colour scale) within the septa, cyst walls, or solid tumor areas is made: a score of 1 is given when no blood flow can be found in the lesion; a score of 2 is given when only minimal flow can be detected; 3 is given when moderate flow is present and 4 is given when the adnexal mass appears highly vascular with marked blood flow using colour Doppler. This colour score refers only to colour Doppler image and not to Doppler shift spectrum. It is only given once (for the tumour as a whole). Multiple photographic prints are made of relevant structures and Doppler signals.

Quality control

Several informative images of all adnexal masses should be made. Preferably, these are stored digitally. Photographs or video are acceptable as well.

6. Subjective assessment

After ultrasonographic examination of the mass the investigator gives his subjective assessment of the mass:

A: Malignant (1) or benign (0)?
B: Probability of malignancy: (=level of certainty)
1 = benign
2 = probably benign
3 = uncertain
4 = probably malignant
5 = malignant
C: Self impression: presumed histological diagnosis (e.g. dermoid, serous cystadenoma, endometrioma…)

7. Serum tumour markers

Sample collection / treatment / storage

Ideally, CA 125 measurements are centralised, using the CA 125 II immunoradiometric assay (Centocor, Malvern, PA) on frozen samples, but we accept local measurements of serum CA 125 levels if it is difficult to store serum samples.
In some centres additional blood samples will be taken for use in proteomics pattern analyses. Written informed consent will be obtained for this purpose.

8. Histopathology and staging

Surgical investigation

Surgery is indicated in case of a persistent mass after 6-12 weeks. In cases of symptomatic adnexal masses, suspected malignancy or at patient's request immediate surgery may be performed. Premenopausal patients with a persistent unilocular smooth cyst < 5 cm and postmenopausal patients with a persistent unilocular smooth cyst < 3 cm may be suitable for follow-up.
Surgery is performed either laparoscopically or transabdominally, depending on the surgeon's clinical judgement. Cytology of cyst fluid or fine needle biopsies are not sufficient for histological classification. If cyst fluid has been obtained, the cytological classification is noted.

Tissue collection

Preferably the whole tumour should be removed. However, representative biopsies may be sufficient (e.g. in advanced ovarian cancer or endometrioma).

Sampling

All surgically removed tumours should be extensively sampled for histological examination. The number of prepared blocks is dependent of the size and the nature of the tumour.

Tumour classification

Tumours are classified according to the criteria recommended by the International Federation of Gynecology and Obstetrics (FIGO). In malignant tumours the degree of differentiation is included.

9. Data collection

Clinical report forms: not necessary, because the dedicated software in Astraia (www.astraia.com) will be used for direct retrieval of the collected data.

10. Statistical analysis

New mathematical models have been constructed in IOTA phase 1 and 2. In phase 3 these models will prospectively be evaluated in terms of sensitivity and specificity and overall accuracy (i.e. the proportion of the sum of true positives and true negatives over the total number of cases). Independent risk factors for malignancy will be expressed as odds. The models will be compared by constructing receiver operating characteristic (ROC) curves.
In addition new models based on new variables will be constructed. The data will be thoroughly inspected using univariate analyses (contingency tables and basic descriptive statistics) and multivariate analyses (principal component analysis, biplots, scattermatrices, canonical correlation analysis and logistic regression) to explore the data and to detect multicollinearity and outliers. Manual checks of any outliers will be performed to eliminate mistakes that had occurred during submission of the data.

11. Study supervision

Central supervision: the Steering Committee is responsible for the protocol, quality control, interim analyses of the data and final analysis and reporting of the study.

Local supervision: the Principal Investigators are responsible for the data collection in their centres.

Dirk Timmerman is responsible for the co-ordination of the IOTA project and the contact between the centres.

Sabine Van Huffel and Bart De Moor are responsible for the data management and the development of new algorithms, in collaboration with B. Van Calster, L. Ameye, A. Daemen O. Gevaert, V. Van Belle, L. De Clercq, P. Antal, and J. Vandewalle.

12. Publication policy

The steering committee is responsible for publication of the data in scientific journals. As such the members are co-authors in all resulting clinically relevant papers, to which they made significant contributions. By the time of the final analysis the principal investigators have to have contributed at least 50 cases to the study. They are co-authors, according to the number of patients they contributed to the study (depending on the journal's restriction of the number of co-authors).

Purely mathematical papers without clinical relevance related to the study data are published by S. Van Huffel, B De Moor and co-workers at ESAT with reference to the IOTA group and the inclusion of as many as possible of the clinical contributors.

The Katholieke Universiteit Leuven represented by its department K.U.LEUVEN RESEARCH & DEVELOPMENT, having its office in 3000 Leuven, Minderbroedersstraat 8A – box 5105, Belgium, VAT number BE 419.052.173 holds intellectual property rights that might result from the IOTA 3 project.

Detailed Research Description

State of the art in the literature

It is of the utmost importance to have preoperatively a correct idea of the benign or malignant character of an adnexal mass because this will influence not only the whole treatment strategy but also the prognosis of the patient [1]. If an adnexal mass is most likely to be benign we could suffice with minimal invasive surgery or even conservative treatment with no need for referral to a gynaecologic oncologist. In this way the duration of hospitalisation and revalidation will be limited and morbidity will be minimal in comparison with surgery after midline laparotomy [2,3]. On the other hand, whenever a mass is considered to be malignant, laparoscopic surgery is contra-indicated because spilling of the cyst content during surgery may worsen the prognosis of the patient [1] and the patient should be referred to a gynaecologic oncologist for proper staging and debulking of the tumour [4].

O1: **Prospective external validation of predictive mathematical models and pattern recognition to distinguish between malignant and benign adnexal masses**

In the past several scoring systems and mathematical models using ultrasound variables have been developed for the preoperative prediction of probability of malignancy of an adnexal mass. However, before these models can be used in daily practice, they should be tested prospectively on a new population. Some models have been tested prospectively on small data sets, but results were disappointing [5-9]. We recently published the results of prospective testing of 17 scoring systems and mathematical models for the calculation of the risk of malignancy in adnexal masses on a large dataset collected by the International Ovarian Tumour Analysis (IOTA) collaborative group, the IOTA phase 1 dataset [10]. Unfortunately, most models and scoring systems performed worse than was originally reported. Of the 17 models tested, only the previously published neural networks, the vector machine models and one of the logistic regression models had good performance on prospective testing [10].
The primary aim of the IOTA group was to prospectively collect clinical information and standardised data from ultrasound examinations in a large number of patients with adnexal masses in order to create new scoring systems and mathematical models to distinguish between benign and malignant adnexal tumours. The multicentre approach was chosen as the most likely method to achieve a model which might be more effectively applicable to prospective studies in different clinic populations.

O2: **Validation of intravenous contrast agents: quantitative analysis of contrast uptake and washout in tumour**

Accurate preoperative diagnosis of ovarian tumours is essential for planning appropriate patient management and improves patient outcome if a malignancy is present. Transvaginal ultrasonography is at present the most effective method for early diagnosis of ovarian tumours, with a sensitivity of 80-85%. A characteristic of malignant ovarian tumours is the presence of neovascularisation which allows the tumour to grow. Changes in vessels may be visualised before tumour detection. An ultrasound technique that can be used to describe ovarian vascularisation is colour Doppler sonography [5,11-14]. In recent years, the depiction of intratumoural vessels with colour and power Doppler sonography has improved, but visualisation of vessels smaller than 0.1 mm in diameter remains impossible. The use of intravascular contrast agents increases the signal-to-noise ratio, resulting in an improved detection of low-volume blood flow. In this way signals from vessels less than 200 μm in diameter can be depicted. Orden et. al [15] showed a difference in degree, onset and duration of Doppler US enhancement between malignant and benign adnexal lesions after the injection of microbubble contrast agents.

O3: **Validation of proteomic analysis**

Since ovarian cancer is clinically quiet, it is often called the “silent killer”. Therefore ovarian cancer diagnosis would greatly benefit from more advanced technologies that can be applied for diagnostic purposes. In 2002 Petricoin et al. [16] applied Surface-Enhanced Laser Desorption and Ionization (SELDI) mass spectrometry to assess differences in the spectra of benign and malignant ovarian masses. They used 50 serum samples with ovarian cancer and 50 serum samples without ovarian cancer and used an iterative search algorithm to develop a

pattern that distinguishes benign from malignant. Next, this model was tested on 116 new serum samples, 50 from women with ovarian cancer and 66 from unaffected women or non-malignant ovarian masses. Using SELDI mass spectrometry this resulted in a pattern based on the amplitudes at key mass-to-charge (m/z) values of 534, 989, 2111, 2251 and 2465. This pattern had a sensitivity of 100% and specificity of 95% with a positive predictive value of 94%. However, no identification of the key m/z values was performed, therefore this pattern constitutes a black box model (i.e. no biological interpretation is possible). The authors also claimed that this model can be used for ovarian cancer screening.

The results of this study lead to a significant increase in research on mass spectrometry-based proteomics applied for cancer diagnosis and prognosis. However, the approach of Petricoin et al. [16] is highly questionable [17] and initiated a debate in the literature [18-24]. These concerns show that the approach by Petricoin and colleagues has several disadvantages and may not be ready for clinical use. However, mass spectrometry-based technology should definitely not be dismissed as an interesting concept [25] and has several advantages when more effort is invested in the use of high quality instrumentation and processing. The use of more advanced technologies such as high pressure liquid chromatography and MALDI-based mass spectrometry can deliver substantially better sensitivity and specificity, with the added possibility of obtaining more biological insight via protein identification.

O4: **Validation of a new set of tumour markers**

Serum CA 125 is a glycoprotein and at present it is the tumour marker with the highest sensitivity for ovarian cancer. Recently different other tumour markers for diagnosis of ovarian cancer have been discovered. Some of the most promising include CA 15-3, CA 72-4, CEA, macrophage colony-stimulating factor (M-CSF), HE4, mesothelin, osteopontin, kallikrein(s), and soluble EGF receptor. However, external validaton is needed to test the value of these and other markers in patients with adnexal tumours [26].

O5:**Validation of 3D power Doppler**

Several strategies can be followed for the preoperative sonographic assessment of adnexal masses. In general, the first step is the ultrasonographic evaluation using subjective evaluation or pattern recognition. In the past, some studies have shown that expert sonologists can reach a sensitivity of 98% and a specificity of 90% solely by using their subjective evaluation [27-28]. Unfortunately, this is not the performance of the less experienced sonographer [27]. A number of different scoring systems and mathematical models have been developed to help the less experienced sonographer in the discrimination between the benign or malignant character of an adnexal mass [29-33]. The first models only used grey scale ultrasound variables. Later on 2D colour Doppler became available and several studies investigated the benefit of this new technique. A disadvantage of the technique is that it is highly dependent on the ultrasound equipment that is used and the blood vessel that is interrogated. The first reports on velocity and resistance indices were promising but later on it became clear that there was too much overlap between benign and malignant masses [34-35].

In a study from Valentin et al. where 173 cases were examined using grey-scale and colour Doppler examination, only 5 extra cases were classified correctly after the use of colour Doppler [36]. Probably this is also one of the reasons that mathematical models that included these Doppler parameters did not perform well when tested prospectively [10,37].

Advantage of 3D Power Doppler examination in comparison with 2D power Doppler examination. Introduction of three-dimensional (3D) power Doppler ultrasound has made it possible to assess vascularisation in a whole organ or tumour in a more objective way.

The vascularisation index (VI) is the ratio of colour voxels to all voxels in the region of interest expressed as a percentage. It reflects the density of vessels in the volume analyzed. The flow index (FI) is the sum of weighted colour voxels divided by the number of all colour voxels in the region of interest, and it reflects the number of blood corpuscles in the vessels of the volume. The vascularisation-flow index (VFI) is the sum of weighted colour Doppler voxels divided by all voxels in the region of interest. It reflects both the density of vessels and the number of blood corpuscles flowing in the vessels of the volume.

Another opportunity is to be able to look at the vascular tree of a tumour. Some authors showed that the morphology of vessels is different in benign vs. malignant tumours. Vessels in a malignant tumour tend to be more branched and tortuous [38]. Although several reports have showed that 3D power Doppler can quite reliably

distinguish between benign and malignant masses [13,39-41], Jokubkiene et al. and Guerriero et al. demonstrated that 3D power Doppler added little to the correct diagnosis of malignancy in an unselected ordinary population and was not superior to grey-scale imaging [42-43]. Jokubkiene et al. compared the performance of logistic regression models using grey-scale variables only, subjective evaluation and logistic regression models that added 3D power Doppler variables as well. These data showed that 3D power Doppler added little benefit to the use of grey scale only [42].

O6: **Validation of a new model, based on grey scale ultrasound and colour Doppler information, specifically designed for "difficult" tumours**

An experienced ultrasound examiner using a good ultrasound system can be expected to correctly discriminate between benign and malignant adnexal masses in 9 of the 10 cases [5,27,28]. The reported sensitivity and specificity with regard to malignancy in the studies cited were around 90%. For less experienced operators, the sensitivity and specificity were 86% and 80% [27]. The main IOTA logistic regression model applicable to all tumours [44] could be of considerable help to these less experienced clinicians, but does not outperform experienced ultrasonographers. But in approximately 1 in 10 cases even an experienced ultrasound examiner is likely to fail to make a confident and correct diagnosis [45]. Until now, no successful model has been constructed in these "uncertain" cases. Before a method capable of distinguishing between benignity and malignancy in such difficult pelvic masses is found (the new method might next to grey scale ultrasound or colour Doppler also involve ultrasound contrast (O2), proteomics (O3), new tumour markers (O4) or 3D power Doppler (O5)), we must accept that some women will need to undergo an unnecessary operation – or perhaps an unnecessarily extensive operation – because of our inability to reliably exclude malignancy before surgery.

State of the art in the consortium and comparative advantage

O1: **Prospective external validation of predictive mathematical models and pattern recognition to distinguish between malignant and benign adnexal masses**

IOTA phase 1. During IOTA phase 1 we collected more than 50 ultrasound, demographic and clinical variables of 1,066 patients with an adnexal mass. A histopathologic diagnosis was available for all patients. Patients were collected in 9 different international centres where the ultrasound examination was performed by one of the IOTA expert sonologists following a standardised protocol. In this way we obtained a large and multicentre database that is more likely to resemble the general population [44].
Based on this dataset 12 mathematical models were developed (2 logistic regression models [44], 1 scoring system (unpublished data), 3 neural networks [46] and 6 kernel based models [47] (3 support vector machine models (SVM) and 3 relevance vector machine models (RVM)). On the test set of IOTA phase 1 all models obtained an area under the ROC curve (AUC) of more than 0.92.
IOTA phase 1b. In IOTA phase 1b a new dataset of 507 new patients was prospectively collected in 3 of the 9 IOTA centres. All models proved to perform excellent with again AUCs of more than 0.94 (unpublished data).
IOTA phase 2. The aim of IOTA phase 2 was to test the IOTA models in new centres with different population characteristics and different levels of ultrasound experience. More than 2,000 patients have been included in 20 centres throughout the world. In a preliminary analysis the main logistic regression model (1) was tested and in every centre the AUC was above 0.87. When tested on the whole dataset the AUC was above 0.93. In IOTA phase 1 and 2 we found that both for the expert sonologist using pattern recognition (subjective impression) as for the mathematical models, approximately 8% of the adnexal masses were difficult to classify as benign or malignant and lied close to the decision boundary. Models that were built especially for the classification of these difficult masses showed disappointing results [45].
IOTA phase 3 . During IOTA phase 3, the same centres that participated in IOTA phase 2 will prospectively include a new dataset of more than 2,000 patients but will now be able to evaluate the prediction of the main logistic regression model immediately after performing the ultrasound scan.

O2: **Intravenous contrast agents: quantitative analysis of contrast uptake and washout in tumour**

We evaluated the efficacy of contrast-dedicated ultrasound technology, contrast-tuned (CnTI) imaging and using the second-generation contrast agent SonoVue (Bracco International BV, Amsterdam, the Netherlands), in

comparison with the standard ultrasound examination in ovarian tumours [48-49]. Eighty-nine patients were enrolled in the study in 4 different clinical centres. The study included 40 uncertain pervic adnexal masses, 10 pelvic masses indicative of recurrence of gyneacologic tumours, 26 uterine pathologic features and 13 cervical lesions. Pictures of the intralesional microvascularization after SonoVue injection differed dramatically from those obtained during colour Doppler examination. By use of the CnTI technology, it was possible to improve the ability of the operator to distinguish benign from malignant lesions.
Preliminary results indicate an area under the ROC curve (AUC) of 0.90 for peak intensity of contrast enhanced signals as opposed to 0.79 for subjective assessment to discriminate between malignant and benign or borderline tumours. An AUC of 0.85 and 0.82 is obtained to separate benign from malignant and borderline tumours for the area under the intensity curve and subjective assessment, respectively. Within this project the preliminary results can be checked on a larger sample set, permitting to fine-tune the decision thresholds and to create more complex models.

O3: **Validation of proteomic analysis**

Pilot study. We adopted the MALDI approach to benefit from the potential diagnostic power of mass spectrometry-based proteomics. We have started a pilot study consisting of 39 serum samples from 20 patients with benign ovarian masses and 19 patients with malignant ovarian masses. The patients in the malignant group consist of 3 stage I tumours, 15 stage III tumours and 1 stage IV tumour. In this first phase the goal is to detect interesting m/z ranges that display differential patterns in benign and malignant ovarian cancer (PILOTMODEL). Subsequently, a second phase consisting of identification of proteins and/or peptides within these mass ranges will be initiated. The use of MALDI-TOF offers a wide range of advantages compared to the SELDI platform described earlier. It provides a better resolution and wider mass range compared to the SELDI platform. Secondly MALDI offers the possibility of identifying interesting peaks at key m/z values by subsequent fragmentation which is impossible with SELDI. The resulting mass spectral patterns can be submitted to mass spectrum interpretation engines such as MASCOT for comparison with known proteins and peptides in databases, returning the most probable biomolecule responsible for the mass spectral measurement.
Proteomics as second stage test: innovative aspects. In this project we aim to expand the pilot study: firstly, by increasing the number of patients analysed with mass spectrometry as a second stage test and secondly, by validating the PILOTMODEL on a larger set of samples. This entails the following innovative aspects: Firstly, by increasing the number of patients analysed more robust analysis can be performed. In both benign and malignant ovarian tumours different subgroups exist. By expanding the number of samples analysed, these effects can be taken into account to look for proteins and/or peptides which are present in all benign samples and absent in all malignant samples and vice versa. Secondly, the MALDI-TOF/TOF work-flow that was applied in the pilot study and that will be used in this project, has not been previously used in this context. Finally, SCD/BIOI has gained much experience in the analysis of high dimensional data, both in pre-processing as in the actual modelling. Therefore, the most recent advances in mathematical modelling in medical informatics and bioinformatics will be used to develop mathematical models that can discriminate between benign and malignant ovarian cancer. For this purpose SCD/BIOI can rely on previously developed models and techniques on clinical and microarray data.

O4: **Validation of a new set of tumour markers**

We prospectively assessed the value of serum CA-125 [50-51]. These studies challenged the current clinical practice, where CA-125 is an integral part of the preoperative work up of patients with adnexal masses. These studies proved for the first time that a single measurement of serum CA-125 does not add to the diagnostic confidence of experienced ultrasound examiners nor to the diagnostic performance of logistic regression models to distinguish between benign and malignant tumours. Therefore, we need to test combinations of new tumour markers that might really improve the classification of difficult tumours. A pilot study on proteomics has been performed in our consortium and this was first presented at the ESGO 2007 meeting (European Society of Gynaecologic Oncology) [52]. External validaton is now needed to test the value of combinations of novel tumour markers with clinical and ultrasound variables in other patients with adnexal tumours.

O5: **Validation of 3D power Doppler**

2D pulsed Doppler information did not improve mathematical models developed to distinguish between benign and malignant adnexal masses. In the already validated IOTA models none of the velocity or resistance Doppler parameters were selected as independent variables [10,31,44,46,47]. Only the colour score showed to be an independent variable. This is a semiquantitative score between 1 and 4 that gives the score of 1 when the tumour shows no vascularisation and 4 if the tumour is highly vascularised [12]. The drawback is the fact that the 2D power Doppler ultrasound result is related to the blood vessel that is examined. It does not offer the investigator an overall image of the vascularization in the tumour, this in contrast to 3D power Doppler. Several IOTA group members published papers on the use of 3D power Doppler to distinguish between benign and malignant adenexal tumours [42,43]. Therefore, 3D power Doppler ultrasound will be offered as a second stage test to the investigators in order to assess the added value of this examination to the clinician's preoperative diagnosis.

O6: **Validation of a new model, based on grey scale ultrasound and colour Doppler information, specifically designed for "difficult" tumours**

Using subjective evaluation of grey scale and Doppler ultrasound findings, an experienced ultrasound examiner using a good ultrasound system can correctly classify ovarian tumours as benign or malignant in most cases. The experts reached in IOTA 1 and IOTA 1b a sensitivity of 88% (234/266) and 90% (129/143), respectively, and a specificity of 95% (762/800) and 93% (338/364) ([44], unpublished data). In 8% of the cases, the ultrasound examiner found it difficult to discriminate between benign and malignant tumours (90/1066 in IOTA 1, 39/507 in IOTA 1b) [45]. Borderline tumours were over-represented among these "uncertain" masses, being three times more common among the "uncertain" masses than among the other ones. Of those borderline tumours, 60% (33/55) were correctly classified (i.e., classified as malignant) by the expert, compared to 95% (762/800) of the benign tumours, 96% (162/169) of the primary invasive tumours and 93% (39/42) of the metastatic tumours [45]. A logistic regression model has been constructed using data of the 54 "uncertain" cases with papillary projections in IOTA 1, but its performance was not stable: the area under the ROC curve was 0.88 on the development set, but dropped to 0.73 on a independent test set. Combining all data available in IOTA 1, 1b and 2 (IOTA 2 data collection just recently closed), we now have full preoperative information of 3,500 tumours including 250 cases which were difficult to peoperatively predict according to the investigators. Moreover, tumours for which the output of the main logistic regression is close to the decision boundary can also be regarded as difficult to classifiy. Using all available grey scale ultrasound and colour Doppler data of difficult tumours in IOTA 1, 1b and 2, a logistic regresison model has specifically been developed for these difficult masses, with an area under the ROC curve of 0.85 on both development and test set (unpublished data).
In IOTA Phase 3, the investigators will have the preoperative prediction based on this new logistic regression model. Besides this, they can also apply ultrasound contrast (O2), proteomics (O3), new tumour markers (O4) or 3D power Doppler (O5)) depending on their preference. This will make it possible to define an optimal algorithm to preoperatively identify maligancy among the small but yet important proportion of ovarian masses for which until now malignancy could not be ruled out, not even by a highly experienced ultrasonographer.

4. Detailed research description & Work Plan

Introduction: IOTA Phase 3

The work Plan is entirely focused to IOTA Phase 3 data collection and validation of the complete diagnostic algorithm. The aim of IOTA Phase 3 is to incorporate the IOTA mathematical models that predict the character of an adnexal mass in daily clinical practice using centre-specific cut-offs and to evaluate second stage tests in patients with adnexal masses that are difficult to classify in order to obtain more reliable results. For clarity of exposition we repeat here the objectives.

The objective of IOTA Phase 3

The main objective of this project is to improve preoperative diagnosis and subsequent management of patients with adnexal tumours by advanced algorithms in order to decrease morbidity and costs and in order to improve survival of patients with ovarian cancer. In particular, we aim to validate the added value of mathematical models as a new diagnostic tool in the prediction of ovarian cancer in clinical practice. First of all, we aim to prove their enhanced diagnostic performance and generalized applicability as a first stage examination. In cases where

prediction is unreliable, we aim to further improve the predictive performance of this diagnostic tool with second stage tests.
To achieve these objectives the project is divided in work packages each having its own sub-objective.

WP1: Prospective external validation of predictive mathematical models and pattern recognition to distinguish between malignant and benign adnexal masses
WP2: Validation of intravenous ultrasound contrast agents: quantitative analysis of contrast uptake and washout in tumor
WP3: Validation of proteomic analysis
WP4: Validation of a new set of tumour markers
WP5: Validation of 3D power Doppler
WP6: Validation of a new model, based on grey scale ultrasound and colour Doppler information, specifically designed for "difficult" tumours

The Work Plan now describes each of these Work Packages and details how each objective is achieved. The link between the different work packages is made clear in the schematic representation of the diagnostic algorithm. Realisation of the work plan should result in optimising all parameters of this diagnostic scheme with respect to diagnostic performance.

Work package 1:

Objective 1: **Prospective external validation of predictive mathematical models and pattern recognition to distinguish between malignant and benign adnexal masses**

The aim of this work package is to validate the IOTA mathematical models that predict the character of an adnexal mass in daily clinical practice using centre-specific cut-offs and to select those patients that are difficult to classify or to diagnose (i.e. with uncertain subjective assessment or uncertain predictive value) for a second stage examination. This work package is further subdivided in 6 tasks.

Task 1.1: Study protocol
The detailed study protocol for IOTA Phase 3 will be approved by the IOTA Steering Committee and then submitted to all relevant Ethical Committees for clinical studies.
Patients will give informed consent prior to participate in this study. Blood samples are only taken after obtaining written informed consent from each patient.

Task 1.2: Data collection and management
For the data collection of IOTA phase 1 and IOTA phase 2, a dedicated, secure data collection system was developed. A unique identifier was generated automatically for each patient's record based on the identifier for the centre, the patient's birthday and the date of the ultrasound scan. Clinicians at each centre could only view or update patients records from their own centre. In IOTA phase 1, the data collection was performed web-based, the investigators needed to login and to complete online a case report from. Data security was ensured by not recording the patient's name and by encrypting all data communication using a 54-bit SSL (Secure Socket Layer) certificate. For IOTA phase 2, an Astraia software application has been developed (www.astraia.com) which enables the researchers to complete the data offline in an electronical case report form. And afterwards, they transferred it to a central database located in Belgium. Since this second method was approved much more user-friendly by the clinicians, the data collection of IOTA phase 3 will be similarly executed.
Each participating centre inserts the clinical and ultrasound data in a specifically designed Astraia study screen. No missings are allowed during data submission, except for the CA-125 marker. A subjective assessment whether the tumour seems benign or malignant as well as the degree of certainty is given by the ultrasound examiner and this assessment is frozen in the local Astraia database before the results of the mathematical model become available to the examiner.

VARIABLES needed to test mathematical models in IOTA PHASE 3

Variable	Acronym	LR	BMLP	BPER	LS-SVM/ RVM	MC1	MC2
Personal history of ov. cancer	Pershistovca	X	X	X	X	X	X
Current use of hormonal therapy	Hormtherapy	X	X		X		
Age	Age	X	X	X	X	X	X
Diameters of the lesion	LesD1-3	X	X	X		X	X
Pain during examination	Pain	X	X				
Ascites	Ascites	X	X	X	X	X	X
Blood flow within papillary proj.	Papflow	X	X	X	X	X	X
Locularity	Locularity	X	X	X	X	X	X
Diameters of solid component	SolidD1-3	X	X	X	X	X	X
Irregular internal cyst walls	Wallregularity	X	X	X	X	X	X
Acoustic shadows	Shadows	X	X	X	X		X
Color Doppler score	Colscore	X	X	X	X		
Number of papillary projections	Papnr		X				X
Diameters of the ovary	OvD1-3				X		
Bilateral tumors	Bilateral					X	X
Suspected origin (ovary vs other)	Origin				X		

LR (logistic regression):
- LR1
- LR2
- objLR
- regLR

BMLP (Bayesian multi-layer perceptron):
- BMLP11-2a
- BMLP11-2b

BPER (Bayesian perceptron):
- BPER11

LS-SVM/RVM (Bayesian least squares support vector machines/relevance vector machines):
- BLSSVMlin (using linear kernel)
- BLSSVMrbf (using radial basis function kernel)
- BLSSVMaddrbf (using additive RBF kernel)
- RVMlin
- RVMrbf
- RVMaddrbf

MC1 and MC2 (Multi-class models):
- MLR (multi-category logistic regression)
- LR-PC (binary LR models combined using pairwise coupling)
- LR-PC2 (similar, but different inputs)
- BLSSVM-PC (binary Bayesian LS-SVMs + PC)
- KLR-PC (binary kernel logistic regression models + PC)
- MKLR (multi-class KLR)

The data are stored locally and by clicking “send data from FTP” all data are automatically and anonymously transferred to the central server in Leuven, where the datamanager checks them.

Task 1.3: Data quality checks and storage
In Leuven a full protocol has been developed for quality checks: 1) cross checks, e.g. in multilocular cysts you may not find papillary structures and 2) manual verification of the outliers. This protocol has already succesfully been applied to the data of IOTA phase 1 and 2.
Once the data has passed all quality checks, it is stored in the central database.

Task 1.4: Prospective external model validation
Once the data have been collected, all models of IOTA phase 1 and 2 (i.e. logistic regression models, scoring systems, neural networks, support and relevance vector machines) will be validated prospect-ively. The AUC, sensitivity, specificity, positive and negative predictive values will be computed.

Task 1.5: Patient selection for Second stage tests
The IOTA 2 dataset enables us to describe the population characteristics of each centre in correlation with the performance of the models in that centre. This gives us the opportunity to define centre-specific cut-offs above which an adnexal mass is classified as malignant. Whenever the probability of malignancy is amongst the 15% that lies close to the decision boundary of LR1, the examiner will be obliged to perform one of the second stage tests. In this way, we expect that the adnexal masses of approximately 300 patients will be classified as "difficult" and will need a second stage test.

Work package 2:

Objective 2: **Validation of intravenous ultrasound contrast agents: quantitative analysis of contrast uptake and washout in tumour**

The aim of this work package is to validate the preliminary study. Patients enrolled in task 2.1 will undergo a baseline (unenhanced) and contrast-enhanced examination. Correlations among results from B-mode, colour Doppler and CnTI-SonoVue examinations and the histopathologic diagnosis will be investigated. This work package is further subdivided in 2 tasks:

Task 2.1: Data collection
To validate the existing model on its discriminating power new patients will be enrolled. Each patient will undergo an unenhanced (baseline) and a contrast-enhanced ultrasound evaluation. Ultrasound examinations will be performed with a high-resolution (9.0- to 5.0-MHz) Technos MPX endovaginal probe (Esaote SpA, Genoa, Italy). The contrast-enhanced examination will be performed with the CnTI technology applied to the transvaginal probe (Esaote SpA, Genoa, Italy) and with SonoVue (Bracco International BV, Amsterdam, the Netherlands). The CnTI technology works at very low acoustic pressure (derated pressure =126 kPa) or a very low mechanical index (<0.1). Each patient will receive SonoVue in a bolus dose of 4.8 ml. The examination starts from the injection of the bolus of SonoVue for 3 minutes or until the end of the contrast effect.

Task 2.2: Validation of the model
The presence, the amount of vascularisation (colour score) and the pattern (regular or chaotic) of blood flow detected with colour Doppler and contrast-enhanced images will be investigated to assess whether the use of CnTI technology provides advantages with respect to the unenhanced ultrasound examination in the assessment of gynaecologic diseases, as seen in the preliminary results. Results will be validated using a case-control design. The above tumours for which intravenous contrast agents have been applied are considered as the "cases". In IOTA 1, 1b and 2 a "historical control group" of difficult masses will be composed with two matching factors: 1) the same tumour type and 2) age (<50 versus ≥50). The difference in sensitivity and specificity of the investigator's diagnosis between the cases with intravenous contrast and the control group will be assessed. The value of quantitative assessement of contrast ultrasound as second stage test in difficult tumours will also be compared with other second stage tests that were applied in the same patients.

Work package 3:

Objective 3: ***Validation of proteomic analysis***

The proteomics work package aims to develop a second stage test for distinguishing benign and malignant adnexal masses using liquid chromatography (LC) and MALDI-based mass spectrometry. It will continue seamlessly the work that is currently being done in our pilot study. The pilot study consists of a set of 39 serum samples, 20 benign samples and 19 malignant samples which are being analysed using MALDI-TOF based proteomic profiling. After proper processing of the data, mathematical and probabilistic methods will be used to develop models based on a panel of m/z peaks. The combination of m/z peaks with the highest diagnostic performance will constitute the PILOTMODEL. This workpackage aims to (1) expand the number of samples analysed and (2) validate the PILOTMODEL. This will be accomplished by making use of the state-of-the-art mass spectral analysis capabilities that have recently become available to us at the Katholieke Universiteit Leuven through the foundation of the Interfaculty Centre for Proteomics and Metabolomics (ProMeta).
The number of samples will be increased by 100 and this new set of patients will be called the prospective data set. The additional proteomics experiments will be carried out on set of patients which reflect the distribution of benign and malignant samples. Then, after proper processing of the data, this set of patients will be used to assess the predictive performance of the PILOTMODEL. If unsatisfactory this set of patients can be used to further improve and refine the PILOTMODEL. This entails the identification and verification of the proteins and/or peptides that are part of the diagnostic model with the best discriminatory performance. The identification allows to replace the m/z values by their corresponding biomolecules and allows to decouple the model from mass spectrometry.

Hereby we will make use of mathematical and probabilistic methods that model the data and predict clinically relevant classes. First, pre-processing techniques will be developed and used to allow comparisons between spectra. Next, both univariate and multivariate techniques will be used to develop models that are able to distinguish benign from malignant neoplasms. This will be accomplished by using both standard and advanced statistical and mathematical data analysis techniques (e.g. logistic regression, LS-SVMs, Bayesian networks). This work package is further subdivided in 3 tasks:

Task 3.1: Sample collection for proteomics second stage test
The aim is to test the clinical application of proteomics for the pre-operative classification of ovarian tumours. The following protocol is being used to collect samples where approximately 100 patients with an adnexal mass will be recruited and studied within 35 days before investigative surgery. Medical and family histories will be recorded. A sample of peripheral venous blood will be taken for the proteomic analysis. Findings at surgery and the histological classification of excised tissues as malignant or benign (and by cell type) will be used as outcome measures. Informed consent must be signed before collection of blood sample. Standard blood clotting tubes without any anti-coagulantia are used for the collection of serum samples. Venous blood is collected in an 8-10 ml clotting tube, this will yield ± 3-4 ml serum. Samples are collected by adequately trained medical personnel only (nurse or doctor). Once the blood is taken, the tube is mixed gently end over end. The tubes are placed in a plastic bag (per donor) and placed at 4°C before and during transport to the laboratory. The maximum allowed time between collection and preparation of the serum sample is 3 hours.

Task 3.2: LC-MALDI Mass spectrometry
Mass spectrometric measurements are carried out in the gas phase on ionized analytes. By definition, a mass spectrometer consists of an ion source, a mass analyser that measures the mass-to-charge ratio (m/z) of the ionized analytes, and a detector that registers the number of ions at each m/z value (Aebersold and Mann, 2003). The ion source is known as Matrix Assisted Laser Desorption/Ionization (MALDI). Its function is to volatise and ionise the proteins and peptides for further analysis. It does this by using laser pulses to desorp and ionise the molecules under study out of a dry, crystalline chemical matrix. The mass analyser is based on a Time-Of-Flight tube design (TOF) and allows for single MS measurements to be performed. For identification of molecules two TOF analysers are used in a tandem MS setting which allows for the fragmentation of the proteins under study. For biomacromolecules such as proteins, fragmentation into smaller peptide fragments is a prerequisite for reliable identification.

Task 3.3: Pre-processing and mathematical modeling (*in silico* analysis)
The analysis of mass-spectrometry data is far from straightforward because of its high dimensional nature. It is virtually impossible to manually analyse results and search for interesting diagnostic proteins and/or peptides.

Therefore advanced methods that draw from statistics, machine learning, probability theory and mathematics are necessary to analyse the data that results from MALDI-TOF/TOF. Furthermore, the analysis is subdivided in two steps corresponding to two objectives in this work package: pre-processing and model development.

Pre-processing**:** The goal of pre-processing is to account for time and mass disturbances between spectra of different samples. Peaks in the mass spectrum correspond to individual proteins or peptides (or fractions thereof) and their peak heights are related to their concentration. When analysing mass spectrometry data for biomarker discovery only a subset of peaks that result from ionisation of biomolecules such as peptides or proteins are biologically significant and of use in applications. In order to detect and locate these peaks, the raw data is subjected to several pre-processing steps: base line correction, smoothing, peak detection and peak alignment. Base line correction and smoothing can be handled using filtering methods and curve fitting techniques. Peak detection and alignment will be handled using wavelet analysis. Wavelets have been widely used to de-noise signals in several number of contexts (e.g. magnetic resonance, ultrasound blood flow and computed tomography) and can be used to extract desirable features from the data. This results in the location of peaks and their quantification by combining elements from signal processing with wavelet analysis.

Model development: The pilot data set will be used to develop a model (i.e. PILOTMODEL) which will subsequently be validated using the prospectively collected data form this project. Both univariate and multivariate data analysis will be used for model development on the pre-processed data. Univariate data analysis will be used to look for single diagnostic peaks. For this purpose, non-parametric statistics will be used (i.e. Wilcoxon rank sum tests). Moreover the predictive performance of each single peak will be estimated using sensitivity, specificity, Area Under the ROC curve (AUC) and likelihood ratios. We expect that a single biomarker will be unsatisfactory to accurately discriminate benign and malignant ovarian cancer. Therefore multivariate analysis will be performed. This will allow to develop a panel of peaks which may have poor predictive performance in a univariate approach but reach significance when combined with other peaks. For this purpose we will use both standard and advanced mathematical techniques. Currently we plan to use stepwise logistic regression, Least-Squares Support Vector Machines and Bayesian network analysis. These three methods have already proven their usefulness in medical and biological data analysis and offer a complementary view on the data.

Work package 4:

Objective 4: ***Validation of new set of tumour markers***

External validaton is needed to test the added value of combinations of novel tumour markers to the standard clinical and ultrasound variables in patients with adnexal tumours. Therefore, two tasks have to be executed:

Task 4.1: Data collection

In tumours needing a 2nd stage test in order to make a reliable preoperative diagnosis, a blood sample is taken and later on a combination of the following markers will be assessed:

1. CA 125 II
2. CA 15-3
3. CA 72-4
4. CEA
5. macrophage colony-stimulating factor (M-CSF)
6. HE4
7. Mesothelin
8. Osteopontin
9. kallikrein(s)

However, if in the meantime more promising markers are proposed, some of these markers may be replaced by others.

Task 4.2: Validation of the model

Similarly as to the validation of 2nd stage tests in WP2, each new tumour marker will be assessed using a case-control study design with tumours where a new marker was present considered as "cases" and difficult masses present in the data of IOTA 1, 1b and 2 that match the "cases" in type of tumour and patient's age (<50 versus ≥50) considered as "controls". Each new tumour marker will be validated in terms of possible increased sensitivity or specificity.

Work package 5:

Objective 5: **Validation of 3D power Doppler**

The aim of this study is to evaluate the potential benefit of 3D power Doppler as a second stage examination in order to be able to classify the difficult masses more correctly as benign or malignant.

Task 5.1: Data collection
In centres where a 3D ultrasound equipment is available, 3D power Doppler volumes of the adnexal mass will be collected. The vascularisation index will be calculated and the vascular tree described. This will be performed in the centre where the volumes are collected. Afterwards a copy of all of the stored 3D volumes will be sent to Prof. Lil Valentin (University Hospital, Malmö, Sweden) where she will independently recalculate all of the vascularisation indices. In this way, we will be able to correct for interobserver variability.

Task 5.2: Validation of the model
A case-control study will be performed to validate the use of 3D ultrasound as a second stage test. Masses in IOTA Phase 3 which have been investigated with 3D ultrasound are denoted as the "cases". A historical control group is constructed using data of difficult masses available in IOTA Phase 1, 1b and 2. Two matching factors will be applied: 1) same tumour type and 2) age (<50 versus ≥50). This will enable us to assess whether 3D ultrasound results in a higher detection rate of malignant masses (higher sensitivity) or a decrease in the number of unnecessary exploratory laparotomies (higher specificity).

Work package 6:

Objective 6: **Validation of a new model, based on grey scale ultrasound and colour Doppler information, specifically designed for difficult tumours.**

The aim is to validate a logistic regression model, based on grey scale ultrasound and colour Doppler, specifically designed for difficult tumours. The main question is: "Can unnecessary exploratory laporotomies be avoided in a number of cases?" Study participants in the prospective set will be offered in a randomised way the standard hospital approaches or standard hospital approaches + the model result. This work package is further subdivided in 2 tasks:

Task 6.1: Data collection
At least 2,000 new patients with persisting adnexal tumours will be examined. This results in a collection of roughly 300 difficult masses. The hypothesis is that adding the prediction by the model, developed for this specific subgroup of tumours, to the decision tree allows a significant reduction in the number of exploratory laparotomies for benign tumours when compared to the standard hospital procedure. Therefore, cases where the clinicians state to be uncertain about the diagnosis, will be randomised in two arms in 1:2 ratios:

- arm (A): standard hospital procedure approach
- arm (B): standard hospital procedure approach + result of difficult tumours model.

So, we will have roughly 100 cases in the control arm (arm A) and 200 cases in the model arm (arm B).

Task 6.2: Validation of the model

In both arms, the sensitivity and specificity will be calculated and compared. An increase of 10% specificity in the model arm (arm B), at an equivalent sensitivity in arm A and B, will be regarded as clinically relevant.

This analysis will be performed twice: 1) overall and 2) stratified by the degree of expertise of the ultrasound examiner. The 2nd analysis is planned due to the fact that less experienced ultrasound examiners might benefit more from the use of such model for difficult tumours than highly experienced investigators.

Expertise in supporting centres

Division SCD, Department of Electrical Engineering (ESAT) of the Katholieke Universiteit Leuven.

SCD's major research objective is to design and build advanced methods for crucial problems in information processing. It builds on the enormous growth in computer power, communication bandwidth and available data and the various needs in society for effective use of these opportunities. The strength of the group is the use of mathematical engineering/engineering mathematics from mathematical fields such as linear and multi-linear algebra, statistics, discrete mathematics, differential geometry, and optimization. In this way, SCD has built up world recognised expertise in bioinformatics, biomedical data processing, signal (audio, communications) processing, cryptography, embedded systems, data mining, neural networks, identification, control, Two research groups with complementary expertise within the Division of SCD, are participating in the project, i.e.

1. The **SCD/BIOMED** team consists of 1 staff member, 3 postdocs, 15 PhD students. Research – fundamental/theoretical as well as application oriented- is performed in the domain of (multi)linear algebra, (non)linear signal analysis, classification and system identification with special focus to the development of numerically reliable and robust algorithms for improving medical diagnostics. In this domain the group has built up an international reputation (more than 150 publications, look http://www.esat.kuleuven.ac.be/sista/). Applications under study are: quantification of metabolite concentrations using in-vivo Magnetic Resonance Spectroscopic (MRS) data and images, quantification of brain oxygenation in neonates using (functional) Near-Infrared Spectroscopy, quantification of cardiovascular dynamics and auto-regulation, heart-rate variability, detection of somato-sensory evoked potentials in EEG, preoperative classification of (brain, ovarian, prostate) tumours and prediction/detection of epileptic seizures based on scalp-EEG monitoring. At present, research focuses on the integrated multimodal and multichannel data processing, analysis and decision support for simultaneously acquired biomedical data such as EEG, ECG, EMG, (functional) MRI, ultrasound, PET and SPECT.
2. The **SCD/BIOI** group is one of the largest bioinformatics groups in Belgium with extensive internationally renowned experience in the integration of mathematical, statistical, and computational methods to analyze biological, biochemical, and biophysical data. The group is specialised in bioinformatics and more particularly in the analysis of large, complex data sets to identify biological and/or clinical relationships between different parameters. A high level of expertise in bioinformatics is a key element to successfully accomplish the goals of the current project proposal. Within the ESAT-SCD group, expertise in all necessary domains is present as well as the necessary hardware and software background. Our research focus is on gene prioritisation, gene network inference, motif detection and disease management, see www.kuleuven.be/bioinformatics for more detailed information. SCD/Bioi operates internationally at the cutting edge of its frontline enabling discipline. The scope of its activities are regional, national and international.

Regarding **proteomics** the research in this project is performed in the Interfaculty Centre for Proteomics and Metabolomics (ProMeta) at the KULeuven. ProMeta has cutting-edge wet-lab and in silico technologies.

UZ Leuven

The Department of Women & Child has a large experience in gynaecologic ultrasound and ovarian cancer treatment. It is the largest referral centre for gynaecologic oncology in Flanders.

A prospective collection of the data on ovarian tumours started in 1994 in UZ Leuven. Since 1996 there is a close cooperation between D. Timmerman and I. Vergote of the University Hospitals of Leuven and the SCD division of ESAT which has resulted in many publications and projects. The Department also initiated the IOTA consortium. The IOTA study is a multicentric cooperation with renowned universities in Lund/Malmö, Leuven, Rome, London, Milan, Monza, Napoli, Lublin, Cagliari, Beijing, Udine and Prague. This international collaboration is technically supported by ESAT-SCD, K.U. Leuven, with software and internet-applications.

Phase 1 of the IOTA study started in 1999 and was concluded in 2002. In this phase data from 1,066 patients with an adnexal tumour were used to construct a large database with potentially important medical parameters. Several new mathematical models were developed. Phase 2 started in October 2005 till October 2007 and further expanded this database with 1,938 patient data from 19 centres.

References

1. Vergote I, De Brabanter J, Fyles A, Bertelsen K, Einhorn N, Sevelda P. Prognostic importance of degree of differentation and cyst rupture in stage I invasive epithelial ovarian carcinoma. Lancet 2001; 357: 176-182
2. Medeiros LR, Fachel JM, Garry R, Stein AT, Furness S. Laparoscopy versus laparotomy for benign ovarian tumours. Cochrane Database Syst Rev. 2005; 20: CD004751
3. Carley ME, Klingele CJ, Gebhart JB, Webb MJ, Wilson TO. Laparoscopy versus laparotomy in the management of benign unilateral adnexal masses. J Am Assoc Gynecol Laparosc 2002; 9: 321-6
4. Hacker NF, Berek JS, Lagasse LD, Primary cytoreductive surgery for epithelial ovarian cancer. Obstet Gynecol 1983; 61:431-420
5. 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: 273-83
6. Ferrazzi E, Zanetta G, Dordoni D, Berlanda N, Mezzopane R, Lissoni G. Transvaginal ultrasonographic characterization of ovarian masses : a comparison of five scoring systems in a multicenter study. Ultrasound Obstet Gynecol 1997; 10: 192-197
7. Aslam N, Banerjee S, Carr J, Savvas M, Hooper R, Jurkovic D. Prospective evaluation of logistic regression models for the diagnosis of ovarian cancer. Obstet Gynecol 2000; 96: 75-80
8. Mol BW, Boll D, De Kanter M, Heintz P, Sijmons E, Oei G. Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. Gyn Oncol 2001; 80: 162-167
9. Valentin L. Comparison of Lerner score, Doppler ultrasound examination, and their combination for discrimination between benign and adnexal masses. Ultrasound Obstet Gynecol 2000; 15: 143-147
10. Van Holsbeke C, Van Calster B, Valentin L, Testa AC, Ferrazzi E, Dimou I, Lu C, Moerman P, Van Huffel S, Vergote I, Timmerman D. External validation of mathematical models to distinguish between benign and malignant adnexal tumors: a multicenter study by the international ovarian tumor analysis group. Clin Cancer Res 2007; 13: 4440-7
11. Tekay A, Jouppila P. Validity of pulsatility and resistance indices in classification of adnexal tumors with transvaginal color Doppler ultrasound. Ultrasound Obstet Gynecol 1992; 2:338–344.
12. 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;16:500-5. Review.
13. Marret H, Ecochard R, Giraudeau B, Golfier F, Raudrant D, Lansac J. Color Doppler energy prediction of malignancy in adnexal masses using logistic regression models. Ultrasound Obstet Gynecol. 2002; 20: 597-604.
14. Marret H, Sauget S, Giraudeau B, Body G, Tranquart F. Power Doppler vascularity index for predicting malignancy of adnexal masses. Ultrasound Obstet Gynecol. 2005; 25: 508-13
15. Orden M.R., Jurvelin J.S. and Kirkinen P.P.. Kinetics of a US contrast agent in benign and malignant adnexal tumors, Radiology 226 (2003), pp. 405–410.
16. Petricoin EF *et al.* Use of proteomic patterns in serum to identify ovarian cancer. The Lancet 2002, vol 359, 572-577
17. Check E. Proteomics and cancer: Running before we can walk? 2004 Nature 429, 496-497
18. Diamandis EP. OvaCheck: doubts voiced soon after publication 2004 Nature 430, 611
19. Diamandis EP. Proteomic patterns in serum and identification of ovarian cancer Lancet. 2002 Jul 13;360(9327):170
20. Baggerly K, Morris J, Coombes K. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics 2004 20(5),777-785
21. Sorace J, Zhan M. A data review and re-assessment of ovarian cancer serum proteomic profiling. BMC Bioinformatics 2003,4:24
22. Rockhill B. Proteomic patterns in serum and identification of ovarian cancer. Lancet. 2002 Jul 13;360(9327):169
23. Elwood M. Proteomic patterns in serum and identification of ovarian cancer. Lancet. 2002 Jul 13;360(9327):170
24. Pearl DC, Proteomic patterns in serum and identification of ovarian cancer. Lancet. 2002 Jul 13;360(9327):169-70
25. Villanueva J, Tempst P. OvaCheck: let's not dismiss the concept. 2004 Nature 430, 611
26. Bast RC Jr, Badgwell D, Lu Z, Marquez R, Rosen D, Liu J, Baggerly KA, Atkinson EN, Skates S, Zhang Z, Lokshin A, Menon U, Jacobs I, Lu K. New tumor markers: CA125 and beyond. Int J Gynecol Cancer. 2005 Nov-Dec;15 Suppl 3:274-81.
27. Timmerman D, Schwärzler P, Collins WP, Claerhout F, Coenen M, Amant F. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. Ultrasound Obstet Gynecol 1999; 13: 11-16
28. Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. Ultrasound Obstet Gynecol 1999; 14: 338-347
29. Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas J. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. BJOG 1990; 97: 922-929
30. Lu C., Suykens J.A.K., Timmerman D., Vergote I., Van Huffel S. Linear and nonlinear preoperative classification of ovarian tumors. Chapter 11 of Knowledge Based Intelligent System for Health Care, (Ichimura T. and Yoshida K. , ed.), vol. 7 of International Series on Advanced Intelligence, Advanced Knowledge International (Magill, Australia), 2004, 343-382
31. Timmerman D, Bourne T, Tailor A, Collins WP, Verrelst H, Vandenberghe K. A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: The development of a new logistic regression model. Am J Obstet Gynecol 1999; 181: 57-65
32. Timmerman D, Verrelst H, Bourne TH, De Moor B, Collins WP, Vergote I. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound Obstet Gynecol 1999; 13: 17-25

33. Tailor A, Jurkovic D, Bourne T, Collins WP, Campbell S. Sonographic prediction of malignancy in adnexal masses using multivariate logistic regression analysis. Ultrasound Obstet Gynecol 1997; 10: 41-47
34. Valentin L, Sladkevicius P, Marsál K. Limited contribution of Doppler velocimetry to the differential diagnosis of extrauterine pelvic tumors. Obstet Gynecol 1994; 83: 425-433
35. Tekay A, Jouppila P. Controversies in assessment of ovarian tumors with transvaginal color Doppler ultrasound. Acta Obstet Gynecol Scand 1996; 75: 316-329
36. Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. UOG 1999; 14: 338-347
37. Mol BW, Boll D, De Kanter M, Heintz P, Sijmons E, Oei G. Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. Gyn Oncol 2001; 80: 162-167
38. Sladkevicius P, Jokubkiene L, Valentin L. Contribution of morphological assessment of the vessel tree by three-dimensional ultrasound to a correct diagnosis of malignancy in ovarian masses. UOG 2007; 30: 874-82
39. Oral communication on the 15th ISUOG world congress in Vancouver. Sladkevicius P, Jokubkiene L, Valentin L. Assessment of the vascular trees in ovarian tumors using 3D power Doppler ultrasound
40. Cohen L, Escobar P, Scharm C, Glimco B, Fishman D. Three-dimensional power Doppler ultrasound improves the diagnostic accuracy for ovarian cancer prediction. Gynecol Oncol 2002; 84: 352-3
41. Kurjak A, Kupesic S, Sparac V, Kosuta D. Three-dimensional ultrasonographic and power Doppler characterization of ovarian lesions. Ultrasound Obstet Gynecol 2000; 16: 365-71
42. Jokubkiene L, Sladkevicius P, Valentin L. Does three-dimensional power Doppler ultrasound help in discrimination between benign and malignant ovarian masses? Ultrasound Obstet Gynecol 2007; 29: 215-225
43. Guerriero S, Ajossa S, Piras S, Gerada M, Floris S, Garau N, Minerba L, Paoletti AM, Melis GB. Three-dimensional quantification of tumor vascularity as a tertiary test after B-mode and power Doppler evaluation for detection of ovarian cancer. J Ultrasound Med 2007; 26: 1271-8
44. Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML., Van Calster B., Collins W.P., 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: 8794-801
45. Valentin L, Ameye L, Jurkovic D, Metzger F, Lécuru F, Van Huffel S, Timmerman D. Which extrauterine pelvic masses are difficult to correctly classify as benign or malignant on the basis of ultrasound findings and is there a way of making a correct diagnosis? UOG 2006; 27: 438-444
46. Van Calster B, Timmerman D, Nabney I, Valentin L, Testa AC5, Van Holsbeke C, Vergote I, Van Huffel S. Using Bayesian Neural Networks with ARD input selection to detect malignant adnexal masses prior to surgery 2007; Neural computing and applications. Accepted for publication.
47. Van Calster B, Timmerman D, Lu C, Suykens J, Valentin L, Van Holsbeke C, Amant F, Vergote I, Van Huffel S. Preoperative diagnosis of ovarian tumors using Bayesian kernel-based methods. UOG 2007; 29: 496-504
48. Testa A, Ferrandina G, Fruscella E, Van Holsbeke C, Ferrazzi E, Leone F, Arduini D, Exacoustos C, Bokor D, Scambia G, Timmermean D. The use of contrasted transvaginal sonography in the diagnosis of gynecologic diseases. J Ultrasound Med 24 (2005), pp. 1267-1278.
49. Testa A.C., Timmerman D., Exacoustos C., Fruscella E., Van Holsbeke C., Bokor D., Arduini D., Scambia G.and Ferrandina G. The role of CnTI-SonoVue in the diagnosis of ovarian masses with papillary projections: a preliminary study. *Ultrasound Obstet Gynecol*, 26: 644-650, 2005.
50. Timmerman D., Van Calster B., Jurkovic D., Valentin L., Testa A.C., Bernard J.P. et al. The inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and malignant adnexal tumors. J Clin Oncol 2007, 25, 4194-4200
51. Van Calster B., Timmerman D., Bourne T., Testa A., 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
52. Van Gorp T., Gevaert O., Van de Plas R., Waelkens E., De Moor B., Timmerman D., Vergote I. Biomarker discovery workflow with MALDI-TOF to distinguish between benign and malignant adnexal masses. Abstracts ESGO meeting, Berlin, 2007.

Supplementary material

Appendix 2.

The power Doppler and three-dimensional (3D) ultrasound settings used.

Settings recommended when acquiring 3D power Doppler volumes using a Voluson 730 expert with a vaginal transducer 5 – 9 MHz

Power 100%

Gain 0.8

Frequency mid

Quality normal

Wall motion filter (WMF) low 1

Pulse repetition Frequency (PRF) 0.6KHz

Submenu

Smooth 5/6

Ensemble 16

Flow res set high

Line dens 7

PD map 5

Balance G > 170

Artefact on

L filter 3

It is important to get a good image of the vascular tree, and in some patients other settings might yield better results. It is important that the settings that detect most small vessels without artefacts (most sensitive settings possible) are used.

Settings recommended when acquiring 3D power Doppler volumes using a GE E8 version 7.03 or 7.05, with a vaginal transducer 5 – 9 MHz

Power 100%

Gain –0.0

Frequency mid

Quality high

Wall motion filter (WMF) mid 1

Pulse repetition Frequency (PRF) 0.6 KHz

Submenu

Smooth rise 7

Smooth fall 7

PD map 5

Flow res high

Line dens 7

Ensemble 21

L filter 2

Artefact on

Balance 205

It is important to get a good image of the vascular tree, and in some patients other settings might yield better results. It is important that the settings that detect most small vessels without artefacts (most sensitive settings possible) are used.

Settings recommended when acquiring 3D power Doppler volumes using a GE E8 version 7.03 or 7.05, with a vaginal transducer 6 – 12 MHz

Power 100%

Gain –8.0

Frequency low

Quality normal

Wall motion filter (WMF) mid 1

Pulse repetition Frequency (PRF) 0.6 KHz

Submenu

Smooth rise 7

Smooth fall 7

PD map 5

Flow res mid 2

Line dens 7

Ensemble 13

L filter 2

Artefact on

Balance 205

It is important to get a good image of the vascular tree, and in some patients other settings might yield better results. It is important that the settings that detect most small vessels without artefacts (most sensitive settings possible) are used.

Table S1 Histological diagnoses of 2403 adnexal tumors, according to whether tumor was difficult to classify on each assessment and availability of ultrasound volumes

			ALL										**3D volume analyzed**							
	Total		Both US examiner and LR1 not uncertain		Either US examiner or LR1 uncertain		US examiner uncertain		LR1 uncertain		Both uncertain		Either US examiner or LR1 uncertain		US examiner uncertain		LR1 uncertain		Both uncertain	
	N=2403		N=2027		N=376		N=168		N=259		N=51		N=138		N=79		N=87		N=28	
Benign	**1423**	**(59%)**	**1169**	**(58%)**	**254**	**(68%)**	**111**	**(66%)**	**180**	**(70%)**	**37**	**(73%)**	**100**	**(72%)**	**52**	**(66%)**	**70**	**(80%)**	**22**	**(79%)**
Endometrioma	344	(14%)	324	(16%)	20	(5%)	5	(3%)	16	(6%)	1	(2%)	7	(5%)	2	(3%)	5	(6%)	-	-
Teratoma	231	(10%)	212	(10%)	19	(5%)	5	(3%)	15	(6%)	1	(2%)	5	(4%)	1	(1%)	4	(5%)	-	-
Simple cyst + parasalpingeal cyst	106	(4%)	96	(5%)	10	(3%)	5	(3%)	7	(3%)	2	(4%)	3	(2%)	1	(1%)	3	(3%)	1	(4%)
Functional cyst	40	(2%)	29	(1%)	11	(4%)	6	(4%)	7	(3%)	2	(4%)	3	(2%)	2	(3%)	2	(2%)	1	(4%)
Hydrosalpinx + salpingitis	47	(2%)	40	(2%)	7	(2%)	3	(2%)	5	(2%)	1	(2%)	3	(2%)	1	(1%)	3	(3%)	1	(4%)
Peritoneal pseudocyst	18	(<1%)	14	(<1%)	4	(1%)	2	(1%)	3	(1%)	1	(2%)	2	(1%)	1	(1%)	2	(2%)	1	(4%)
Abscess	17	(<1%)	14	(<1%)	3	(<1%)	1	(<1%)	2	(<1%)	-	-	2	(1%)	1	(1%)	1	(1%)	-	-
Fibroma	130	(5%)	79	(4%)	51	(14%)	22	(13%)	33	(13%)	4	(8%)	18	(13%)	9	(11%)	10	(11%)	1	(4%)
Serous cystadenoma	259	(11%)	190	(9%)	69	(18%)	35	(21%)	50	(19%)	16	(31%)	30	(22%)	21	(27%)	21	(24%)	12	(43%)
Mucinous cystadenoma	183	(8%)	134	(7%)	49	(13%)	21	(13%)	32	(12%)	4	(8%)	23	(17%)	10	(13%)	16	(18%)	3	(11%)
Rare benign	48	(2%)	37	(2%)	11	(3%)	6	(4%)	10	(4%)	5	(10%)	4	(3%)	3	(4%)	3	(3%)	2	(7%)
Borderline	**153**	**(6%)**	**103**	**(5%)**	**50**	**(13%)**	**24**	**(14%)**	**31**	**(12%)**	**5**	**(10%)**	**15**	**(11%)**	**7**	**(9%)**	**9**	**(10%)**	**1**	**(4%)**
Stage I	135	(6%)	86	(4%)	49	(13%)	24	(14%)	30	(12%)	5	(10%)	14	(10%)	7	(9%)	8	(9%)	1	(4%)
Stage II	6	(<1%)	5	(<1%)	1	(<1%)	-	-	1	(<1%)	-	-	1	(<1%)	-	-	1	(1%)	-	-
Stage III	12	(<1%)	12	(<1%)	-		-	-	-	-	-	-	-		-	-	-	-	-	-
Primary invasive	**701**	**(29%)**	**637**	**(31%)**	**64**	**(17%)**	**29**	**(17%)**	**44**	**(17%)**	**9**	**(18%)**	**22**	**(16%)**	**19**	**(24%)**	**8**	**(9%)**	**5**	**(18%)**
Stage I	128	(5%)	103	(5%)	25	(7%)	12	(7%)	17	(7%)	4	(8%)	7	(5%)	6	(8%)	3	(3%)	2	(7%)
Stage II	47	(2%)	44	(2%)	3	(<1%)	1	(<1%)	2	(<1%)	-	-	1	(<1%)	1	(1%)	-	-	-	-
Stage III	397	(17%)	378	(19%)	19	(5%)	8	(5%)	13	(5%)	2	(4%)	8	(6%)	6	(8%)	3	(3%)	1	(7%)
Stage IV	61	(3%)	60	(3%)	1	(<1%)	-	-	1	(<1%)	-	-	-	-	-	-	-	-	-	-
Rare	68	(3%)	52	(3%)	16	(4%)	8	(5%)	11	(4%)	3	(6%)	6	(4%)	6	(8%)	2	(2%)	2	(4%)
Metastatic	**126**	**(5%)**	**118**	**(6%)**	**8**	**(2%)**	**4**	**(2%)**	**4**	**(2%)**	**-**	**-**	**1**	**(<1%)**	**1**	**(1%)**	**-**	**-**	**-**	**-**

US, ultrasound; LR1, logistic regression model 1; 3D, three-dimensional

Table S2 Histological diagnoses of 2403 adnexal tumors, according to whether tumor was difficult to classify as benign or malignant

	Both US examiner and LR1 not uncertain N=2027		Either US examiner or LR1 uncertain N=376		P-value
Benign	**1169**	**(58%)**	**254**	**(68%)**	**<0.001**
Endometrioma	324	(16%)	20	(5%)	<0.001
Teratoma	212	(10%)	19	(5%)	0.01
Simple cyst + parasalpingeal cyst	96	(5%)	10	(3%)	0.72
Functional cyst	29	(1%)	11	(4%)	0.54
Hydrosalpinx + salpingitis	40	(2%)	7	(2%)	1
Peritoneal pseudocyst	14	(<1%)	4	(1%)	1
Abscess	14	(<1%)	3	(<1%)	1
Fibroma	79	(4%)	51	(14%)	<0.001
Serous cystadenoma	190	(9%)	69	(18%)	<0.001
Mucinous cystadenoma	134	(7%)	49	(13%)	<0.001
Rare benign	37	(2%)	11	(3%)	0.94
Borderline	**103**	**(5%)**	**50**	**(13%)**	**<0.001**
Stage I	86	(4%)	49	(13%)	<0.001
Stage II	5	(<1%)	1	(<1%)	1
Stage III	12	(<1%)	-		0.99
Primary invasive	**637**	**(31%)**	**64**	**(17%)**	**<0.001**
Stage I	103	(5%)	25	(7%)	0.98
Stage II	44	(2%)	3	(<1%)	0.82
Stage III	378	(19%)	19	(5%)	<0.001
Stage IV	60	(3%)	1	(<1%)	0.008
Rare	52	(3%)	16	(4%)	0.77
Metastatic	**118**	**(6%)**	**8**	**(2%)**	**0.008**

P-values are corrected for multiple testing with the permutation method (Westfall PH, Wolfinger RD. Multiple tests with discrete distributions. *Am Stat* 1997; **51**: 3-8)
US, ultrasound; LR1, logistic regression model 1

Table S3 Clinical and ultrasound characteristics of 2403 adnexal tumors, according to whether tumor was difficult to classify on each assessment and availability of ultrasound volumes

		ALL					**3D volume analyzed**			
	Total N=2403	Both US examiner and LR1 not uncertain N=2027	Either US examiner or LR1 uncertain N=376	US examiner uncertain N=168	LR1 uncertain N=259	Both uncertain N=51	Either US examiner or LR1 uncertain N=138	US examiner uncertain N=79	LR1 uncertain N=87	Both uncertain N=28
Clinical variables										
Age, years	50 ± 16	49 ± 16	53 ± 16	52 ± 16	53 ± 16	49 ± 14	54 ± 17	54 ± 18	51 ± 16	49 ± 13
Postmenopausal	1049 (44%)	861 (42%)	188 (50%)	79 (47%)	130 (50%)	21 (41%)	73 (53%)	41 (52%)	44 (51%)	12 (43%)
Hysterectomy	142 (6%)	115 (6%)	27 (7%)	15 (9%)	19 (7%)	7 (14%)	9 (7%)	6 (8%)	7 (8%)	4 (14%)
Hormonal replacement therapy	207 (9%)	174 (9%)	33 (9%)	26 (15%)	13 (5%)	6 (12%)	13 (9%)	12 (15%)	4 (5%)	3 (11%)
Personal history ovarian cancer	44 (2%)	36 (2%)	8 (2%)	2 (1%)	6 (2%)	- -	4 (3%)	1 (1%)	3 (3%)	- -
Family history ovarian cancer	74 (3%)	69 (3%)	5 (1%)	2 (1%)	3 (1%)	- -	3 (2%)	1 (1%)	2 (2%)	- -
CA125, N available	1451	1198	253	114	175	36	117	67	73	23
CA125	42 (1 – 14067)	52 (1 – 14067)	23 (3 – 1948)	24 (6 – 906)	21 (3 – 1948)	18 (6 – 313)	20 (4 – 1302)	21 (7 – 906)	18 (4 – 1302)	17 (7 – 313)
Gray scale ultrasound variables										
Largest diameter, mm	71 (10 – 550)	70 (10 – 550)	75 (10 – 322)	72 (10 – 322)	74 (10 – 300)	62 (10 – 169)	69 (10 – 310)	68 (10 – 310)	68 (10 – 300)	64 (10 – 122)
Bilateral	518 (22%)	465 (23%)	53 (14%)	16 (10%)	41 (16%)	4 (8%)	22 (16%)	9 (11%)	16 (18%)	3 (11%)
Ascites	340 (14%)	330 (16%)	10 (3%)	8 (5%)	2 (<1%)	- -	5 (4%)	5 (6%)	- -	- -
Type of mass										
Unilocular	600 (25%)	585 (29%)	15 (4%)	3 (2%)	12 (5%)	- -	1 (<1%)	- -	1 (1%)	- -
Unilocular solid	258 (11%)	187 (9%)	71 (19%)	40 (24%)	46 (18%)	15 (29%)	23 (17%)	17 (22%)	14 (16%)	8 (29%)
Multilocular	413 (17%)	331 (16%)	82 (22%)	30 (18%)	60 (23%)	8 (16%)	32 (23%)	12 (15%)	23 (26%)	3 (11%)
Multilocular solid	505 (21%)	375 (19%)	130 (35%)	59 (35%)	89 (34%)	18 (35%)	56 (41%)	34 (43%)	36 (41%)	14 (50%)
Solid	627 (26%)	549 (27%)	78 (21%)	36 (21%)	52 (20%)	10 (20%)	26 (19%)	16 (20%)	13 (15%)	3 (11%)
Nr locules	1 (0 to >10)	1 (0 to >10)	2 (0 to >10)	2 (0 to >10)	2 (0 to >10)	2 (0 to >10)	4 (0 to >10)	2 (0 to >10)	3 (0 to >10)	2 (0 to >10)
Pain at US examination	344 (14%)	294 (15%)	50 (13%)	20 (12%)	32 (12%)	2 (4%)	9 (7%)	4 (5%)	5 (6%)	- -
Echogenicity of cyst fluid										
Anechoic	589 (25%)	485 (24%)	104 (28%)	39 (23%)	77 (30%)	12 (24%)	36 (26%)	17 (22%)	27 (31%)	8 (29%)
Low level	516 (21%)	392 (19%)	124 (33%)	63 (38%)	82 (32%)	21 (41%)	47 (34%)	32 (41%)	28 (32%)	13 (46%)
Ground glass	350 (15%)	326 (16%)	24 (6%)	9 (5%)	16 (6%)	1 (2%)	10 (7%)	4 (5%)	6 (7%)	- -
Hemorrhagic	20 (<1%)	16 (<1%)	4 (1%)	2 (1%)	2 (<1%)	- -	3 (2%)	2 (3%)	1 (1%)	- -
Mixed	301 (13%)	259 (13%)	42 (11%)	19 (11%)	30 (12%)	7 (14%)	16 (12%)	8 (10%)	12 (14%)	4 (14%)
No cyst fluid	627 (26%)	549 (27%)	78 (21%)	36 (21%)	52 (20%)	10 (20%)	26 (19%)	16 (20%)	13 (15%)	3 (11%)
Papillary projections present	383 (16%)	265 (13%)	118 (31%)	69 (41%)	73 (28%)	24 (47%)	45 (33%)	35 (44%)	25 (29%)	15 (54%)
Flow in papillation	215 (56%)	169 (64%)	46 (39%)	32 (46%)	21 (29%)	7 (29%)	20 (44%)	16 (46%)	8 (32%)	4 (27%)
Number of papillations	2 (1 – ≥4)	2 (1 – ≥ 4)	2 (1 – ≥ 4)	2 (1 – ≥ 4)	1 (1 – ≥ 4)	2 (1 – ≥ 4)	2 (1 – ≥4)	2 (1 – ≥4)	1 (1 – ≥ 4)	2 (1 – ≥4)

Cont.

Table S3 continued

	Total	ALL					3D volume analyzed			
		Both US examiner and LR1 not uncertain	Either US examiner or LR1 uncertain	US examiner uncertain	LR1 uncertain	Both uncertain	Either US examiner or LR1 uncertain	US examiner uncertain	LR1 uncertain	Both uncertain
	N=2403	N=2027	N=376	N=168	N=259	N=51	N=138	N=79	N=87	N=28
Height of papillation, mm	10 (3 – 99)	11 (3 – 99)	7 (3 – 45)	8 (3 – 45)	6 (3 – 30)	7 (3 – 30)	7 (3 – 45)	7 (3 – 45)	5 (3 – 21)	6 (3 – 13)
Mass with solid components	1390 (58%)	1111 (55%)	279 (74%)	135 (80%)	187 (72%)	43 (84%)	105 (76%)	67 (85%)	63 (72%)	25 (89%)
Largest diameter of largest solid component, mm	50 (3 – 300)	54 (3 – 300)	25 (3 – 200)	28 (3 – 200)	22 (3 – 196)	19 (5 – 112)	24 (3 – 180)	25 (3 – 162)	16 (3 – 180)	13 (5 – 112)
Incomplete septum	113 (5%)	88 (4%)	25 (7%)	9 (5%)	17 (7%)	1 (2%)	2 (1%)	1 (1%)	1 (1%)	- -
Irregular walls	957 (40%)	740 (37%)	217 (58%)	113 (67%)	141 (54%)	37 (73%)	77 (56%)	55 (70%)	44 (51%)	22 (79%)
Shadows	299 (12%)	239 (12%)	60 (16%)	20 (12%)	47 (18%)	7 (14%)	20 (14%)	10 (13%)	13 (15%)	3 (11%)
Doppler ultrasound variables										
Color Score										
Score 1	606 (25%)	553 (27%)	53 (14%)	21 (13%)	36 (14%)	4 (8%)	10 (7%)	5 (6%)	7 (8%)	2 (7%)
Score 2	762 (32%)	604 (30%)	158 (42%)	64 (38%)	115 (44%)	21 (41%)	44 (32%)	24 (30%)	31 (36%)	11 (39%)
Score 3	681 (28%)	539 (27%)	142 (38%)	71 (42%)	94 (36%)	23 (45%)	70 (51%)	40 (51%)	42 (48%)	12 (43%)
Score 4	354 (15%)	331 (16%)	23 (6%)	12 (7%)	14 (5%)	3 (6%)	14 (10%)	10 (13%)	7 (8%)	3 (11%)

US, ultrasound; LR1, logistic regression model 1 ; 3D, three-dimensional
Results are presented as n (%) or median (min-max) except for age (mean ± SD)

Table S4 Clinical and ultrasound characteristics of 2403 adnexal tumors, according to whether tumor was difficult to classify as benign or malignant

	Both US examiner and LR1 not uncertain N=2027		Either US examiner or LR1 uncertain N=376		P-value
Clinical variables					
Age, years	49 ± 16		53 ± 16		<0.001
Postmenopausal	861	(42%)	188	(50%)	0.007
Hysterectomy	115	(6%)	27	(7%)	0.27
Hormonal replacement therapy	174	(9%)	33	(9%)	0.90
Personal history ovarian cancer	36	(2%)	8	(2%)	0.67
Family history ovarian cancer	69	(3%)	5	(1%)	0.03
CA125, N available	1198		253		
CA125	52 (1 – 14067)		23 (3 – 1948)		<0.001
Gray scale ultrasound variables					
Largest diameter, mm	70 (10 – 550)		75 (10 – 322)		0.006
Bilateral	465	(23%)	53	(14%)	<0.001
Ascites	330	(16%)	10	(3%)	<0.001
Type of mass					<0.001
Unilocular	585	(29%)	15	(4%)	<0.001
Unilocular solid	187	(9%)	71	(19%)	<0.001
Multilocular	331	(16%)	82	(22%)	0.05
Multilocular solid	375	(19%)	130	(35%)	<0.001
Solid	549	(27%)	78	(21%)	0.04
Nr locules	1 (0 to >10)		2 (0 to >10)		<0.001
Pain at US examination	294	(15%)	50	(13%)	0.54
Echogenicity of cyst fluid					
Anechoic	485	(24%)	104	(28%)	0.52
Low level	392	(19%)	124	(33%)	<0.001
Ground glass	326	(16%)	24	(6%)	<0.001
Hemorrhagic	16	(<1%)	4	(1%)	0.98
Mixed	259	(13%)	42	(11%)	0.94
No cyst fluid	549	(27%)	78	(21%)	0.05
Papillary projections present	265	(13%)	118	(31%)	<0.001
Flow in papillation	169	(64%)	46	(39%)	<0.001
Number of papillations	2 (1 – ≥ 4)		2 (1 – ≥ 4)		0.006
Height of papillation, mm	11 (3 – 99)		7 (3 – 45)		<0.001
Mass with solid components	1111	(55%)	279	(74%)	<0.001
Largest diameter of largest solid component, mm	54 (3-300)		25 (3-200)		<0.001
Incomplete septum	88	(4%)	25	(7%)	0.06
Irregular walls	740	(37%)	217	(58%)	<0.001
Shadows	239	(12%)	60	(16%)	0.03
Doppler ultrasound variables					
Color Score					<0.001
Score 1	553	(27%)	53	(14%)	<0.001
Score 2	604	(30%)	158	(42%)	<0.001
Score 3	539	(27%)	142	(38%)	<0.001
Score 4	331	(16%)	23	(6%)	<0.001

US, ultrasound; LR, logistic regression model 1
Results are presented as n (%) or median (min-max) except for age (mean ± SD)
The P-values presented are corrected for multiple testing using the permutation method (Westfall PH, Wolfinger RD. Multiple tests with discrete distributions. *Am Stat* 1997; **51**: 3-8)

Table S5 Number of patients and proportion of difficult adnexal tumors contributed by each center

	ALL										3D volume analyzed							
Center	**Total** n	(%)	Both US examiner and LR1 not uncertain		Either US examiner or LR1 uncertain		US examiner uncertain		LR1 uncertain		Both uncertain		Either US examiner or LR1 uncertain		US examiner uncertain		LR1 uncertain		Both uncertain	
BIT	213	(9%)	180	(9%)	33	(9%)	15	(9%)	21	(8%)	3	(6%)	-	-	-	-	-	-	-	-
BSP	37	(2%)	31	(2%)	6	(2%)	2	(1%)	4	(2%)	-	-	5	(4%)	2	(3%)	3	(3%)	-	-
CIT	218	(9%)	196	(10%)	22	(6%)	4	(2%)	21	(8%)	3	(6%)	11	(8%)	3	(4%)	10	(11%)	2	(7%)
FIT	21	(<1%)	20	(<1%)	1	(<1%)	1	(<1%)	-	-	-	-	-	-	-	-	-	-	-	-
GBE	228	(9%)	192	(9%)	36	(10%)	19	(11%)	24	(9%)	7	(14%)	4	(3%)	2	(3%)	3	(3%)	1	(4%)
GIT	6	(<1%)	5	(<1%)	1	(<1%)	1	(<1%)	-	-	-	-	-	-	-	-	-	-	-	-
LBE	129	(5%)	98	(5%)	31	(8%)	17	(10%)	16	(6%)	2	(4%)	2	(1%)	1	(1%)	1	(1%)	-	-
LPO	131	(5%)	117	(6%)	14	(4%)	4	(2%)	10	(4%)	-	-	2	(1%)	1	(1%)	1	(1%)	-	-
LSW	39	(2%)	30	(1%)	9	(2%)	6	(4%)	6	(2%)	3	(6%)	4	(3%)	3	(4%)	3	(3%)	2	(7%)
MIT	86	(4%)	75	(4%)	11	(3%)	-	-	11	(4%)	-	-	-	-	-	-	-	-	-	-
MSW	201	(8%)	141	(7%)	60	(16%)	32	(19%)	39	(15%)	11	(22%)	50	(36%)	31	(39%)	30	(34%)	11	(39%)
NIT	8	(<1%)	7	(<1%)	1	(<1%)	-	-	1	(<1%)	-	-	-	-	-	-	-	-	-	-
OIT	105	(4%)	91	(4%)	14	(4%)	7	(4%)	10	(4%)	3	(6%)	12	(9%)	7	(9%)	8	(9%)	3	(11%)
PCR	264	(11%)	234	(12%)	30	(8%)	3	(2%)	29	(11%)	2	(4%)	12	(9%)	3	(4%)	11	(13%)	2	(7%)
RIT	443	(18%)	386	(19%)	57	(15%)	26	(15%)	37	(14%)	6	(12%)	10	(7%)	9	(11%)	4	(5%)	3	(11%)
SIT	107	(4%)	100	(5%)	7	(2%)	1	(<1%)	6	(2%)	-	-	6	(4%)	1	(1%)	5	(6%)	-	-
SSW	120	(5%)	85	(4%)	35	(9%)	27	(16%)	16	(6%)	8	(16%)	20	(14%)	16	(20%)	8	(9%)	4	(14%)
UDI	47	(2%)	39	(2%)	8	(2%)	3	(2%)	8	(3%)	3	(6%)	-	-	-	-	-	-	-	-
All centers	**2403**		**2027**		**376**		**168**		**259**		**51**		**138**		**79**		**87**		**28**	

Percentages are calculated per column
US, ultrasound; LR1, logistic regression model 1; 3D, three-dimensional

BIT, Bologna, Italy; BSP, Barcelona, Spain; CIT, European Institute of Oncology, Milan, Italy; FIT, Children´s Hospital Buzzi, Milan Italy; GBE, Genk, Belgium; GIT, Instituto Nationale dei Tumori, Naples, Italy; LBE, Leuven, Belgium; LPO, Lublin, Poland; LSW, Lund, Sweden; MIT, Sacco University, Milan, Italy; MSW, Malmoe, Sweden; NIT, Universita degli Studi di Napoli, Naples, Italy; OIT, Monza, Italy; PCR, Prague, Czeck Republic; RIT, Rome, Italy; SIT, Cagliari, Italy; SSW, Stockholm, Sweden; UDI, Udine, Italy

Table S6 Histological diagnoses of 376 difficult adnexal tumors, according to whether ultrasound volumes were available

	Either ultrasound examiner or LR1 uncertain N=376				
	Volume not available N=238		Volume available and analyzed N=138		P-value
Benign	**154**	**(65%)**	**100**	**(72%)**	**0.32**
Endometrioma	13	(5%)	7	(5%)	
Teratoma	14	(6%)	5	(4%)	
Simple cyst or parasalpingeal cyst	7	(3%)	3	(2%)	
Functional cyst	8	(3%)	3	(2%)	
Hydrosalpinx or salpingitis	4	(2%)	3	(2%)	
Peritoneal pseudocyst	2	(<1%)	2	(1%)	
Abscess	1	(<1%)	2	(1%)	
Fibroma	33	(14%)	18	(13%)	
Serous cystadenoma	39	(16%)	30	(22%)	
Mucinous cystadenoma	26	(11%)	23	(17%)	
Rare benign	7	(3%)	4	(3%)	
Borderline	**35**	**(15%)**	**15**	**(11%)**	**0.69**
Stage I	35	(15%)	14	(10%)	
Stage II	0	0	1	(<1%)	
Stage III or IV	0	0	0	0	
Primary invasive	**42**	**(18%)**	**22**	**(16%)**	**0.98**
Stage I	18	(8%)	7	(5%)	
Stage II	2	(<1%)	1	(<1%)	
Stage III	11	(5%)	8	(6%)	
Stage IV	1	(<1%)	0	0	
Rare	10	(4%)	6	(4%)	
Metastatic	**7**	**(3%)**	**1**	**(<1%)**	**0.60**

The P-values presented have been corrected for multiple testing using the permutation method (Westfall PH, Wolfinger RD. Multiple tests with discrete distributions. *Am Stat* 1997; **51**: 3-8)

Table S7 Clinical and ultrasound characteristics of 376 difficult adnexal tumors, according to whether ultrasound volumes were available

	Either ultrasound examiner or LR1 uncertain		
	Volume not available N=238	Volume available and analyzed N=138	P-value
Clinical variables			
Age, years	52 ± 16	54 ± 17	0.50
Postmenopausal	115 (48%)	73 (53%)	0.39
Hysterectomy	18 (8%)	9 (7%)	0.84
Hormonal replacement therapy	20 (8%)	13 (9%)	0.74
Personal history ovarian cancer	4 (2%)	4 (3%)	0.47
Family history ovarian cancer	2 (<1%)	3 (2%)	0.36
CA125, number available	136	117	
CA125, U/mL	29 (3 - 1948)	20 (4 - 1302)	0.61
Gray scale ultrasound variables			
Largest diameter, mm	80 (14-322)	69 (10-310)	0.06
Bilateral	31 (13%)	22 (16%)	0.44
Ascites	5 (2%)	5 (4%)	0.51
Type of mass			0.06
Unilocular	14 (6%)	1 (<1%)	
Unilocular solid	48 (20%)	23 (17%)	
Multilocular	50 (21%)	32 (23%)	
Multilocular solid	74 (31%)	56 (41%)	
Solid	52 (22%)	26 (19%)	
Number of locules if multilocular or multilocular solid			0.17
2	19 (15%)	10 (11%)	
3	16 (13%)	8 (9%)	
4	9 (7%)	8 (9%)	
5-10	37 (30%)	23 (26%)	
>10	43 (35%)	39 (44%)	
Tender mass at ultrasound examination	41 (17%)	9 (7%)	0.003
Echogenicity of cyst fluid			0.65
Anechoic	68 (29%)	36 (26%)	
Low level	77 (32%)	47 (34%)	
Ground glass	14 (6%)	10 (7%)	
Hemorrhagic	1 (<1%)	3 (2%)	
Mixed	26 (11%)	16 (12%)	
No cyst fluid	52 (22%)	26 (19%)	
Papillary projections present	73 (31%)	45 (33%)	0.70
Flow in papillation, if papillation present	26 (36%)	20 (44%)	0.34
Number of papillations	2 (1 – ≥4)	2 (1 – ≥4)	0.92
Height of papillation, mm	7 (3 – 30)	7 (3 – 45)	0.27
Mass with solid components	174 (73%)	105 (76%)	0.52
Largest diameter of largest solid component, mm	26 (4-200)	24 (3-180)	0.45
Incomplete septum	23 (10%)	2 (1%)	0.002
Irregular walls	140 (59%)	77 (56%)	0.57
Shadows	40 (17%)	20 (14%)	0.55
Doppler ultrasound variables			
Color Score			<0.001
Score 1	43 (18%)	10 (7%)	
Score 2	114 (48%)	44 (32%)	
Score 3	72 (30%)	70 (51%)	
Score 4	9 (4%)	14 (10%)	

LR1, logistic regression model 1
Results are shown as n (%) or median (min – max) except for age (mean ± SD)
No correction for multiple testing because no further testing was done for the subcategories

Table S8 Ability (with 95% CI) of vessel morphology on 3D power Doppler, subjective ultrasound assessment and IOTA logistic regression model 1 to discriminate correctly between benign and malignant difficult adnexal tumors

Diagnostic method	Sensitivity [95% CI]	Specificity [95% CI]	LR+ [95% CI]	LR- [95% CI]	P-value
Either US examiner or LR1 uncertain (N=138)					
Whole tumor vessel morphology					
Branching vessels	89% (34/38) [76% to 96%]	33% (33/100) [25% to 43%]	1.34 [1.12 to 1.59]	0.32 [0.12 to 0.84]	0.009
Densely packed vessels	63% (24/38) [47% to 77%]	83% (83/100) [74% to 89%]	3.72 [2.26 to 6.10]	0.44 [0.29 to 0.68]	<0.001
Caliber changes in vessels	66% (25/38) [50% to 79%]	68% (68/100) [58% to 76%]	2.06 [1.43 to 2.97]	0.50 [0.32 to 0.80]	<0.001
Splashes	50% (19/38) [35% to 65%]	78% (78/100) [69% to 85%]	2.27 [1.40 to 3.70]	0.64 [0.46 to 0.90]	0.002
Tortuous vessels	66% (25/38) [50% to 79%]	70% (70/100) [60% to 78%]	2.19 [1.50 to 3.20]	0.49 [0.31 to 0.77]	<0.001
Biopsy vessel morphology					
Branching vessels	84% (32/38) [70% to 93%]	34% (34/100) [25% to 44%]	1.28 [1.05 to 1.55]	0.46 [0.21 to 1.02]	0.04
Caliber changes in vessels	79% (30/38) [64% to 89%]	63% (63/100) [53% to 72%]	2.13 [1.58 to 2.89]	0.33 [0.18 to 0.63]	<0.001
Splashes	53% (20/38) [37% to 68%]	69% (69/100) [59% to 77%]	1.70 [1.12 to 2.58]	0.69 [0.48 to 0.98]	0.02
Tortuous vessels	79% (30/38) [64% to 89%]	64% (64/100) [54% to 73%]	2.19 [1.61 to 2.99]	0.33 [0.18 to 0.62]	<0.001
Bridges between vessels	42% (16/38) [28% to 58%]	81% (81/100) [72% to 87%]	2.22 [1.28 to 3.84]	0.72 [0.54 to 0.95]	0.007
Subjective assessment	74% (28/38) [58% to 85%]	74% (74/100) [65% to 82%]	2.83 [1.94 to 4.15]	0.36 [0.21 to 0.61]	<0.001
LR1 (10% risk cutoff)	92% (35/38) [79% to 97%]	23% (23/100) [16% to 32%]	1.20 [1.04 to 1.38]	0.34 [0.11 to 1.08]	0.03
US examiner uncertain (N=79)					
Whole tumor vessel morphology					
Branching vessels	89% (24/27) [72% to 96%]	33% (17/52) [22% to 46%]	1.32 [1.05 to 1.67]	0.34 [0.11 to 1.06]	0.06
Densely packed vessels	67% (18/27) [48% to 81%]	83% (43/52) [70% to 91%]	3.85 [2.01 to 7.39]	0.40 [0.23 to 0.70]	<0.001
Caliber changes in vessels	63% (17/27) [44% to 78%]	67% (35/52) [54% to 78%]	1.93 [1.19 to 3.13]	0.55 [0.33 to 0.93]	0.01
Splashes	52% (14/27) [34% to 69%]	83% (43/52) [70% to 91%]	3.00 [1.49 to 6.01]	0.58 [0.39 to 0.88]	0.003
Tortuous vessels	63% (17/27) [44% to 78%]	63% (33/52) [50% to 75%]	1.72 [1.09 to 2.73]	0.58 [0.34 to 0.99]	0.02
Biopsy vessel morphology					
Branching vessels	85% (23/27) [68% to 94%]	35% (18/52) [23% to 48%]	1.30 [1.01 to 1.68]	0.43 [0.16 to 1.14]	0.07
Caliber changes in vessels	78% (21/27) [59% to 89%]	62% (32/52) [48% to 74%]	2.02 [1.36 to 3.01]	0.36 [0.17 to 0.76]	0.001
Splashes	59% (16/27) [41% to 75%]	71% (37/52) [58% to 82%]	2.05 [1.21 to 3.49]	0.57 [0.35 to 0.93]	0.009
Tortuous vessels	81% (22/27) [63% to 92%]	67% (35/52) [54% to 78%]	2.49 [1.62 to 3.83]	0.28 [0.12 to 0.62]	<0.001
Bridges between vessels	44% (12/27) [28% to 63%]	79% (41/52) [66% to 88%]	2.10 [1.07 to 4.11]	0.71 [0.49 to 1.02]	0.03

Cont.

Table S8 continued

Diagnostic method	Sensitivity [95% CI]	Specificity [95% CI]	LR+ [95% CI]	LR- [95% CI]	P-value
Subjective assessment	74% (20/27) [55% to 87%]	60% (31/52) [46% to 72%]	1.83 [1.23 to 2.73]	0.44 [0.22 to 0.86]	0.004
LR1 (10% risk cutoff)	89% (24/27) [72% to 96%]	19% (10/52) [11% to 32%]	1.10 [0.91 to 1.33]	0.58 [0.17 to 1.93]	0.34
LR1 uncertain (N=87)					
Whole tumor vessel morphology					
Branching vessels	94% (16/17) [73% to 99%]	30% (21/70) [21% to 42%]	1.35 [1.11 to 1.63]	0.20 [0.03 to 1.36]	0.06
Densely packed vessels	53% (9/17) [31% to 74%]	83% (58/70) [72% to 90%]	3.09 [1.56 to 6.11]	0.57 [0.34 to 0.95]	0.004
Caliber changes in vessels	71% (12/17) [47% to 87%]	69% (48/70) [57% to 78%]	2.25 [1.41 to 3.57]	0.43 [0.20 to 0.91]	0.005
Splashes	47% (8/17) [26% to 69%]	77% (54/70) [66% to 85%]	2.06 [1.06 to 4.00]	0.69 [0.43 to 1.09]	0.07
Tortuous vessels	65% (11/17) [41% to 83%]	71% (50/70) [60% to 81%]	2.27 [1.36 to 3.77]	0.49 [0.26 to 0.96]	0.01
Biopsy vessel morphology					
Branching vessels	76% (13/17) [53% to 90%]	31% (22/70) [22% to 43%]	1.12 [0.82 to 1.52]	0.75 [0.30 to 1.89]	0.77
Caliber changes in vessels	76% (13/17) [53% to 90%]	64% (45/70) [53% to 75%]	2.14 [1.42 to 3.23]	0.37 [0.15 to 0.88]	0.005
Splashes	47% (8/17) [26% to 69%]	69% (48/70) [57% to 78%]	1.50 [0.81 to 2.76]	0.77 [0.48 to 1.24]	0.26
Tortuous vessels	71% (12/17) [47% to 87%]	60% (42/70) [48% to 71%]	1.77 [1.16 to 2.69]	0.49 [0.23 to 1.05]	0.03
Bridges between vessels	35% (6/17) [17% to 59%]	81% (57/70) [71% to 89%]	1.90 [0.85 to 4.27]	0.80 [0.55 to 1.15]	0.19
Subjective assessment	82% (14/17) [59% to 94%]	79% (55/70) [68% to 87%]	3.84 [2.33 to 6.33]	0.23 [0.08 to 0.63]	<0.001
LR1 (10% risk cutoff)	100% (17/17) [82% to 100%]	19% (13/70) [11% to 29%]	1.23 [1.10 to 1.37]	Not possible to calculate	0.06

LR+, positive likelihood ratio; LR-, negative likelihood ratio; CI, confidence interval; LR1, logistic regression model 1 using the 10% risk cutoff to predict malignancy suggested in *J Clin Oncol* 2005; **23:** 8794-8801 and *Ultrasound Obstet Gynecol* 2010; **36:** 226-234
No corrections have been made for multiple testing because this is an exploratory study