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?

Lil Valentin[1], MD, PhD, Lieveke Ameye[2], MS, Davor Jurkovic[3], MD, PhD, Ulrike Metzger[4], MD, Fabrice Lécuru[4], MD, PhD, Sabine Van Huffel[2], PhD, Dirk Timmerman[5], MD, PhD.

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

[2] Department of Electrical Engineering, ESAT-SCD, Katholieke Universiteit Leuven, Leuven, Belgium

[3] Department of Obstetrics and Gynaecology, King's College Hospital, Denmark Hill, London, UK

[4] Service de Chirurgie Gynécologique, Hôpital Européen Georges Pompidou, Paris, France

[5] Department of Obstetrics and Gynecology, University Hospitals KU Leuven, Leuven, Belgium.

Short title: Difficult tumors

Key words: ultrasonography, regression analysis, ovarian neoplasms, pelvic neoplasms

Address for correspondence:

Professor Lil Valentin

Department of Obstetrics and Gynecology

Malmö University Hospital

SE – 205 02 Malmö, Sweden

Phone: Int +46 40 33 2094 Telefax: Int +46 40 96 26 00

E-mail: lil.valentin@obst.mas.lu.se

Abstract

Objective. To determine which extrauterine pelvic masses are difficult to correctly classify as benign or malignant on the basis of ultrasound findings, and to determine if the use of logistic regression models for calculation of individual risk of malignancy would improve the diagnostic accuracy in difficult tumors.

Material and methods. In a prospective international European multi-center study involving nine centers, 1066 women with a pelvic mass judged to be of extrauterine origin underwent transvaginal ultrasound examination by an experienced ultrasound examiner before surgery. A standardized examination technique and predefined definitions of ultrasound characteristics were used. On the basis of subjective evaluation of ultrasound findings, the examiner classified each mass as being certainly benign, probably benign, unclassifiable, probably malignant, or certainly malignant. Even when the examiner found the mass unclassifiable (i.e., difficult mass), he or she was obliged to state whether the mass was more likely to be benign or malignant. **Borderline tumors were classified as malignant.**

Results. There were 90 (8%) unclassifiable masses. Multiple logistic regression analysis showed papillary projections, >10 locules in a cyst without solid components, low level echogenicity of cyst fluid, and moderate vascularization as assessed subjectively at color Doppler examination to be ultrasound variables independently associated with unclassifiable mass. Borderline malignant tumors (n=55) proved to be most difficult to assess with only 47% being correctly classified (i.e., classified as malignant), 29% being incorrectly classified (i.e., classified as benign), and 24% being unclassifiable vs. 89% of non-borderline tumors being correctly classified, 3% being incorrectly classified, and 8% being unclassifiable ($P <$ 0.0001). **Papillary cystadeno(fibr)omas, myomas and cases of struma ovarii were also more common among the unclassifiable masses than among the classifiable ones (5.6% vs. 1.1%, p = 0.008; 4.4% vs. 0.9%, p = 0.02; 4.4% vs. 0.2%, p = 0.0006).** No ultrasound

variable or clinical variable (including CA 125) entered a logistic regression model to predict malignancy in difficult masses. A model could be constructed for difficult masses containing papillary projections, but this model performed no better than subjective evaluation of the ultrasound image. **Sensitivity and specificity of subjective evaluation with regard to malignancy in the group of unclassifiable masses were 56% (14/25) and 77% (50/65) vs. 91% (220/241) and 97% (712/735) in the classifiable masses.**

  Conclusions Borderline tumors cause great diagnostic difficulties, **but so do papillary cystadeno(fibro)mas, struma ovarii and some myomas.** Logistic regression models do not solve the diagnostic problem in difficult pelvic masses.

Introduction

Using subjective evaluation of gray scale and Doppler ultrasound findings, an experienced ultrasound examiner using a good ultrasound system can correctly classify extrauterine pelvic masses as benign or malignant in most cases [1, 2], and suggest a correct specific diagnosis (e.g., endometrioma, dermoid cyst, or hydrosalpinx) in many cases[3]. **The reported sensitivity and specificity with regard to malignancy in the studies cited were 88% and 96%[1], and 96% and 90%[2], respectively, and the reported sensitivity and specificity with regard to endometrioma were 92% and 97%, with regard to dermoid cyst 90% and 98%, and with regard to hydrosalpinx 100% and 100%[3].** However, in a small proportion of cases even a very experienced ultrasound examiner will find it difficult to discriminate between benignity and malignancy[1].

The aim of this study was to determine which extrauterine pelvic masses are difficult to correctly classify as benign or malignant when using subjective evaluation of ultrasound findings as the diagnostic method, and to determine if the use of a logistic regression model for calculation of individual risk of malignancy would improve diagnostic accuracy in 'difficult' masses.

Subjects and methods

This is a prospective international multicenter study (International Ovarian Tumor Analysis study, IOTA) including the following nine centers: Malmö University Hospital, Lund University (Sweden), University Hospitals, Leuven (Belgium), Universita del Sacro Cuore, Rome (Italy), ISBM L.Sacco University of Milan, Milano (Italy), Hôpital Boucicaut, Paris (France), Hôpital Européen Georges Pompidou, Paris (France), Centre Medical des Pyramides, Maurepas (France), King's College Hospital, London (UK), ISBM Ospedale, San Gerado Universita di Milano, Monza (Italy), Universita degli Studi di Napoli, Naples (Italy). Recruitment was from June 1999 to June 2002.

Patients presenting with at least one overt pelvic mass judged to be of extrauterine origin and who were examined by a principal investigator at one of the participating centers were eligible for inclusion. **In case of bilateral masses, data from the mass with the most complex ultrasound morphology** were used. Our exclusion criteria were: pregnant patient, inability to tolerate transvaginal sonography, surgery >120 days after sonographic assessment, or incomplete submission of data. **An expert external pathologist reviewed 10% of all pathological specimens. Cases were excluded if there was disagreement between the original histopathological diagnosis and that of the external reviewer.**

A history, including the number of first degree relatives with ovarian cancer or breast cancer, **and use of hormone replacement therapy**, was taken from each patient following a standardized protocol. **A woman was considered to be postmenopausal, if she reported a period of at least 12 months of amenorrhea after the age of 40 years, provided that medication or disease did not explain the amenorrhea. Women 50 years or older who had undergone hysterectomy were also defined as postmenopausal.**

A transvaginal gray scale and color Doppler ultrasound examination using a high-end ultrasound system equipped with a transvaginal transducer with a frequency of 4 – 8 MHz was performed in all cases. Transabdominal sonography was added, if a large mass could not be seen in its entirety using a transvaginal probe. A standardized examination technique and standardized definitions of ultrasound terms were used in all centers. **These have been published[4]. A papillary projection was defined as any solid protrusion into a cyst cavity from the cyst wall with a height greater than or equal to 3 mm[4].** When intra-tumoral blood flow velocity waveforms were not detected, the peak systolic velocity, time averaged maximum velocity, pulsatility index and resistance index were coded as 2.0 cm/sec, 1 cm/sec, 3.0 and 1.0, respectively, for use in mathematical modeling. The presence or absence of pain during the examination was noted. Finally, on the basis of subjective evaluation of the

ultrasound findings, the ultrasound examiner classified each mass as being certainly benign, probably benign, unclassifiable, probably malignant, or certainly malignant. Even when the examiner found the mass unclassifiable (= difficult mass), he or she was obliged to state whether the mass was more likely to be benign or malignant.

Blood samples were drawn for analysis of CA 125, but the availability of this biochemical end-point was not an essential requirement for recruitment into the study. The immunoradiometric assay CA 125 II (Centocor, Malvern, PA or Cis-Bio, Gif-sur-Yvette, Cedex, France) or Abbott Axsym system, REF 3B41-22 (Abbott Laboratories Diagnostic Division, Abbott Park, IL 60064 USA) was used. CA 125 results were unavailable to the ultrasound examiner at the time of the ultrasound examination.

Data were submitted via the internet to a central database using a dedicated, secure data collection system developed for the study[5].

Statistical analysis was carried out using the SAS System release 8.02. Student´s t-test and Mann-Whitney´s test were used to test the statistical significance of differences in continuous data; the Chi-squared test and Fisher´s exact test were used to test the statistical significance of differences in discrete data. We used logistic regression with stepwise selection of variables to determine which ultrasound variables were independently associated with unclassifiable mass and for building a model to predict malignancy in difficult masses. Two-tailed P-values < 0.05 were considered statistically significant.

Results

A total of 1149 patients were recruited. Data from 83 patients (7%) were excluded (eight because of pregnancy, 31 because surgery was undertaken more than 120 days from the sonographic assessment, 42 because of incomplete submission of data, two because of

disagreement between pathologists over the histological diagnosis). Data from 1066 patients

(93%) were available for statistical analysis and model development.

   Clinical information including CA 125 values for the women included are shown in Table

1. Women with unclassifiable masses (n = 90, i.e., 8% of all masses; 95% CI 7 – 10%) were

older and of higher parity than those with classifiable masses (n = 976). Histological

diagnoses are presented in Table 2. Endometriomas and primary invasive malignancies Stage

II-IV were less common among the unclassifiable masses than among the classifiable ones.

Borderline tumors, papillary cystadenomas and (cyst)adenofibromas, myomas, and struma

ovarii were over-represented among the unclassifiable masses, these diagnoses being three to

twenty times more common among the unclassifiable masses than among the others. Of the

borderline tumors, 47% (26/55) were correctly classified (i.e., classified as malignant) by the

ultrasound examiner, 29% (16/55) were incorrectly classified (i.e., classified as benign), and

24% (13/55) were unclassifiable. Of the non-borderline tumors 90% (906/1011) were

correctly classified, 3% (28/1011) were incorrectly classified, and 8% (77/1011) were

unclassifiable. This difference between borderline and non-borderline tumors was statistically

significant (P <0.0001).

   Ultrasound findings are shown in Table 3. Multiple logistic regression analysis showed

papillary projections, >10 locules in a cyst without solid components, low level echogenicity

of cyst fluid, and color score 3 (indicating moderate vascularization[4]) to be the only

ultrasound variables independently associated with unclassifiable mass with odds ratio

estimates of 5.1 (P < 0.0001), 3.7 (P = 0.0175), 2.5 (P = 0.0002), and 1.8 (P = 0.0109). In our

study, 64% (35/55) of the borderline tumors contained papillary projections vs. 24%

(242/1011) of the non-borderline tumors (P = < 0.0001), 11% (6/55) vs. 2% (19/1011) were

multilocular cysts with >10 locules (P = 0.012), and 33% (18/55) vs.19% (191/1011)

contained cyst fluid with low-level echogenicity (P = 0.023).

Examples of unclassifiable masses are shown in Figure 1.

In unclassifiable masses the sensitivity and specificity of subjective evaluation of ultrasound findings with regard to malignancy were 56% (14/25) and 77% (50/65), the positive and negative likelihood ratios being 2.43 and 0.57. For the classifiable masses the corresponding figures were 91% (220/241), 97% (712/735), 28.5 and 0.09 (P < 0.0001 for the difference in sensitivity and P < 0.0001 for the difference in specificity).

Substituting missing values for information about papillary projections and solid components (papillary flow, papillary volume, volume of solid component, etc) in tumors without papillary projections and solid components with zeros, no ultrasound variable or clinical variable entered a logistic regression model to predict malignancy in difficult masses. However, a model could be constructed for masses with papillary projections (39 masses in the training set, and 15 in the test set), see Table 4. CA125 did not add any significant information to the model. The interpretation of the model is that for each one-unit increase in height of the largest papillary projection the odds of malignancy increased 1.23 times, for each one-unit increase in the number of papillary projections the odds increased 2.67 times, and for each one-unit increase in thickness of the thickest septum the odds decreased 0.54 times. **Areas under ROC curves, sensitivity and specificity with regard to malignancy, and positive and negative likelihood ratios of the logistic regression model and of subjective evaluation of the ultrasound image in the training set and test set of difficult tumors with papillary projections are shown in Table 5.**

Discussion

Our multicenter study has shown that experienced ultrasound examiners using high-end ultrasound systems find slightly less than 10% of pelvic masses judged to be of adnexal origin difficult to classify as benign or malignant on the basis of gray scale and color Doppler

ultrasound findings. Masses with papillary projections, multilocular cysts with > 10 locules, cyst with low level echogenicity of cyst fluid and masses moderately vascularized at color Doppler ultrasound examination seem to be more difficult to classify than other types of tumor. The histological diagnoses that present the greatest diagnostic difficulties are borderline tumors, struma ovarii, papillary (cyst)adeno(fibro)mas, and myomas. We found no logistic regression model that was useful to distinguish between benignity and malignancy in difficult masses. However, a logistic regression model built on unclassifiable masses with papillary projections suggested that the more papillary projections and the larger the papillary projections the greater the risk of malignancy **including borderline malignancy**.

It is a strength of our study that it is large and involves many centers. Therefore, our results are likely to be generalizable to other experienced ultrasound examiners using good ultrasound systems provided that they are exposed to a population similar to our study population. We believe that our sample of masses is fairly representative of the types of extrauterine pelvic mass currently considered appropriate to remove surgically.

It was an unexpected finding that myomas were common among difficult masses. This is likely to be explained by some of the myomas in this series not being ordinary myomas. They were all suspected to be an adnexal mass both clinically and at ultrasound examination, and their ultrasound morphology was unclear enough to justify surgical removal.

Masses with papillary projections, multilocular cysts with >10 locules, and masses with low level echogenicity of cyst fluid were clearly overrepresented among the difficult masses. This is not surprising, because we found that these ultrasound features are characteristic of borderline tumors, and borderline tumors were overrepresented among the difficult tumors. Others, too, found papillary projections and multilocularity to be characteristic of borderline tumors[6, 7].

It is important to be able to reliably discriminate between benign and malignant adnexal masses in order to be able to correctly evaluate the need for surgery and to choose appropriate time and mode of operation. We have shown in this study that an experienced ultrasound examiner using a good ultrasound system can be expected to be able to correctly discriminate between benign and malignant adnexal masses in > 90% of all cases. However, in approximately 10 % of cases even an experienced ultrasound examiner using a good ultrasound system is likely to fail to make a confident and correct diagnosis. Before we have found a method capable of distinguishing between benignity and malignancy in such difficult pelvic masses (these new methods might prove to be use of intravenous ultrasound contrast, or proteomic pattern recognition), 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 operation.

IOTA Steering Committee

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

IOTA principal investigators (alphabetical order)

Jean-Pierre Bernard, Maurepas, FRANCE.

Thomas H. Bourne, London, U.K.

Enrico Ferrazzi, Milan, ITALY.

Davor Jurkovic, London, UK

Fabrice Lécuru, Paris, FRANCE.

Andrea Lissoni, Monza, ITALY.

Ulrike Metzger, Paris, FRANCE.

Dario Paladini, Napels, ITALY.

Antonia Testa, Roma, ITALY.

Dirk Timmerman, Leuven, BELGIUM.

Lil Valentin, Malmö, SWEDEN.

Sabine Van Huffel, Leuven, BELGIUM.

Caroline Van Holsbeke, Leuven, BELGIUM.

Ignace Vergote, Leuven, BELGIUM.

Gerardo Zanetta†, Monza, ITALY.

References

1. Valentin L. Prospective cross-validation of Doppler ultrasound examination and gray-scale ulrasound imaging for discrimination of benign and malignant pelvic masses *Ultrasound Obstet Gynecol* 1999; 14 : 273 – 283.

2. Timmerman D, Schwarzler P, Collins WP, Claerhout F, Coenen M, Amant F, Vergote I, Bourne TH. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. Ultrasound Obstet Gynecol. 1999;13:11 – 16

3. Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. *Ultrasound Obstet Gynecol* 1999;14:338 – 347

4. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H and Vergote I. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) group. *Ultrasound Obstet Gynecol* , 2000; 16:500 – 505.

5. Aerts S, Antal P, Timmerman D, De Moor B, Moreau Y. Web based data collection for ovarian cancer: a case study. In: P Kokol (Hrsg): *Proceedings of the 15th IEEE Symposium on Computer Based Medical Systems (CBMS)*, 4 – 7 June 2002, Maribor, Slovenia. Los Alamitos, Calif: IEEE Computer Society 2002, pp. 213 – 218

6. Exacoustos C, Romanini ME, Rinaldo D, Amoroso C, Szabolcs B, Zupi E, Arduini D. Preoperative sonographic features of borderline ovarian tumors. Ultrasound Obstet Gynecol 2005; 25: 50 – 59

7. Pascual MA, Tresserra F, Grases PJ, Labastida R, Dexeus S. Borderline cystic tumors of the ovary: gray scale and color Doppler sonographic findings. J Clin Ultrasound 2002; 30. 76 – 82

Legends

Figure 1. Ultrasound images of unclassifiable masses a) unilocular cyst with anechoic cyst fluid and one papillary projection with irregular surface; this is a benign seropapillary paraovarian cyst b) multilocular solid cyst; this is a benign papillary cystadenofibroma **c) color Doppler image of myoma judged to be 'unclassifiable' d) gray scale image of struma ovarii judged to be 'unclassifiable' e) color Doppler image of the same case of struma ovarii judged to be 'unclassifiable'**

Table1. Clinical information on the women included

		All n = 1066	Unclassifiable mass n = 90	Classifiable mass n = 976	P-value*
Age, years; median; range		46 (17-94)	53 (18-87)	46 (17-94)	0.0013
Parity; median; range		1 (0-10)	2 (0-9)	1 (0-10)	0.0044
Postmenopausal; n (%)		432 (40.5%)	51 (56.7%)	381 (39.0%)	0.0011
Hysterectomy; n (%)		78 (7.3%)	9 (10.0%)	69 (7.1%)	0.3070
Hormonal replacement therapy; n (%)	yes vs. no	235 (22.1%)	16 (17.8%)	219 (22.4%)	0.3075
Personal history of ovarian cancer; n (%)	yes vs. no	14 (1.3%)	3 (3.3%)	11 (1.1%)	0.1073
Family history of ovarian cancer; n (%)	yes vs. no	33 (3.1%)	2 (2.2%)	31 (3.2%)	> 0.999
Personal history of breast cancer; n (%)	yes vs. no	38 (3.6%)	5 (5.6%)	33 (3.4%)	0.8614
Family history of breast cancer; n (%)	yes vs. no	119 (11.2%)	9 (10.0%)	110 (11.3%)	0.7142
CA 125, U/mL**; median		23	21.5	23	0.6777
range		1 – 31610	1 – 31610	4 – 2080	

Cont.

Table1. Continued

	All n = 1066	Unclassifiable mass n = 90	Classifiable mass n = 976	P-value*

*The P-values refer to comparison between unclassifiable and classifiable masses

** There are 809 CA125 measurements, 66 for unclassifiable masses and 743 for classifiable masses

Table 2. Histological diagnoses

	All		Unclassifiable mass		Classifiable mass		P-value*
	N=1066		N=90		N=976		
Benign lesions	800	(75.0%)	65	(72.2%)	735	(75.3%)	0.5175
Endometrioma; n (%)	199	(18.7%)	6	(6.7%)	193	(19.8%)	0.0014
Serous cystadenoma; n (%)	138	(12.9%)	9	(10.0%)	129	(13.2%)	0.5107
Teratoma; n (%)	118	(11.1%)	5	(5.6%)	113	(11.6%)	0.1113
Mucinous cystadenoma; n (%)	85	(8.0%)	9	(10.0%)	76	(7.8%)	0.4189
Non-papillary adenofibroma; n (%)	36	(3.4%)	6	(6.7%)	30	(3.1%)	0.1149
Para-ovarian/para tubal/peritoneal cyst; n (%)	27	(2.3%)	3	(3.3%)	24	(2.5%)	0.4920
Functional cyst; n (%)	24	(2.3%)	2	(2.2%)	22	(2.3%)	>0.9999
Hydrosalpinx; n (%)	22	(2.1%)	0		22	(2.3%)	0.2481
Fibroma, fibrothecoma; n (%)	20	(1.9%)	3	(3.3%)	17	(1.7%)	0.2348
Pelvic inflammatory disease; n (%)	20	(1.9%)	2	(2.2%)	18	(1.8%)	0.6831
Simple cyst; n (%)	18	(1.8%)	0	(0%)	18	(1.8%)	0.3909
Papillary (cyst)adeno(fibro)ma; n (%)	16	(1.5%)	5	(5.6%)	11	(1.1%)	0.0079
Myoma, adenomyoma; n (%)	13	(1.2%)	4	(4.4%)	9	(0.9%)	0.0188

Cont.

Table 2. Histological diagnoses cont.

	All		Unclassifiable mass		Classifiable mass		P-value*
	N=1066		N=90		N=976		
Struma ovari; n (%)	6	(0.6%)	4	(4.4%)	2	(0.2%)	0.0006
Rare benign tumor**; n (%)	9	(0.8%)	1	(1.1%)	8	(0.8%)	0.5493
Other***; n (%)	49	(4.6%)	6	(6.7%)	43	(4.4%)	0.2952
Malignancies; n (%)	211	(19.8%)	12	(13.3%)	199	(20.4%)	0.1079
Primary invasive epithelial cancer; n (%)	144	(13.5%)	6	(6.7%)	13.8	(14.1%)	0.0520
stage I; n (%)	42	(3.9%)	3	(3.3%)	39	(4.0%)	> 0.9999
stage II-IV; n (%)	102	(9.6%)	3	(3.3%)	99	(10.1%)	0.03560
Primary invasive rare type[#], all stages; n (%)	25	(2.3%)	2	(2.2%)	23	(2.4%)	> 0.9999
Metastatic	42	(3.9%)	4	(4.4%)	38	(3.9%)	0.7751
Borderline tumor[##]; n (%)	55	(5.2%)	13	(14.4%)	42	(4.3%)	< 0.0001

* P-values refer to comparison between unclassifiable and classifiable masses

** Brenner tumor (n=3), granulosa cell tumor (n=1), lymphangioma (n=1), Schwanoma (n=1), Sertoli cell tumor (n=1), Stromal tumor (n=2)

*** The group "other" contains 36 cases with more than one histological diagnosis in the same adnexa, e.g., mucinous cystadenoma and endometrioma in the same ovary, or chronic salpingitis and endomtriosis in the same adnexa, and 13 other cases, e.g., tuberculous granuloma in the tube, mucinous histiocystoma, etcetera

[#] Granulosa cell tumor (n=3), Brenner tumor (n = 2), carcinosarcoma (n=4), immature teratoma (n=2), invasive Leydig cell tumor (n = 2), complex germinal tumor (n=1), dysgerminoma and yolk sac tumor (n=1), endodermal sinus tumor (n=1), gynandroblastoma (n=1), hindgut carcinoid of uncertain origin (n=1), leiomyosarcoma (n=1), lymphoma(n=1), mixed mullerian tumor(n=1), pure choriocarcinoma (n=1), sex-cord stromal ovarian tumor (n=1), small bowel gastrointestinal stroma cell tumor (n=1), small cell cancer of pulmunary type (n=1).

[##] Stage I (n = 50), stage II (n = 5)