Ultrasound Obstet Gynecol 2008; **32**: 828–831
Published online 16 October 2008 in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/uog.6214

Real-time ultrasound vs. evaluation of static images in the preoperative assessment of adnexal masses

C. VAN HOLSBEKE*†, J. YAZBEK‡, T. K. HOLLAND‡, A. DAEMEN§, B. DE MOOR§, A. C. TESTA¶, L. VALENTIN** , D. JURKOVIC‡ and D. TIMMERMAN*

**Department of Obstetrics and Gynaecology, University Hospitals Leuven, †Department of Obstetrics and Gynaecology, Ziekenhuis Oost-Limburg, Genk and §Department of Electrical Engineering, ESAT-SCD, Katholieke Universiteit Leuven, Belgium, ‡Early Pregnancy and Gynaecology Assessment Unit, Kings' College Hospital, London, UK, ¶Istituto di Clinica Ostetrica e Ginecologica, Università Cattolica del Sacro Cuore, Roma, Italy and **Department of Obstetrics and Gynaecology, Malmö University Hospital, Lund University, Sweden*

KEYWORDS: adnexal tumors; ovarian cancer; pattern recognition; risk of malignancy; subjective impression; ultrasound

ABSTRACT

Objective *To determine if the prediction of the malignancy of an adnexal mass using pattern recognition, i.e. subjective evaluation of gray-scale and Doppler ultrasound findings, is as accurate when based on static images as it is when based on a real-time ultrasound examination.*

Methods *The static images of 166 non-consecutive patients with 'difficult' adnexal masses, who all underwent surgery and for whom a histopathological diagnosis was available, were evaluated by three experts in gynecological ultrasound ('image experts'). All ultrasound examinations had been performed and the static images saved by a fourth expert sonologist ('real-time' sonologist). All four sonologists classified the adnexal masses as benign or malignant based on their subjective impression and stated with what degree of confidence their diagnosis was made. The diagnostic performance of the real-time sonologist was compared with that of each of the three image experts and with that of the 'consensus opinion' of the image experts (i.e. the diagnosis suggested by at least two of the latter).*

Results *The real-time sonologist correctly predicted the diagnosis with an accuracy of 89% (148/166) vs. 85% (141/166) for the consensus opinion of static images (*$P = 0.0707$*). Equivalent values for sensitivity and specificity were 80% (56/70) vs. 83% (58/70) (*$P = 0.4142$*) and 96% (92/96) vs. 86% (83/96) (*$P = 0.0027$*), respectively.*

Conclusions *The preoperative diagnosis of an adnexal mass made on the basis of a real-time ultrasound examination is more precise than a diagnosis made on the basis of saved static ultrasound images. Evaluation of static images is associated with lower diagnostic specificity. Copyright © 2008 ISUOG. Published by John Wiley & Sons, Ltd.*

INTRODUCTION

Trying to make an accurate prediction of the benign or malignant nature of an adnexal mass during a preoperative ultrasound examination is an important task because the preoperative diagnosis influences the treatment strategy and, as a consequence, the prognosis of the patient[1–5]. Several reports have shown that subjective evaluation of ultrasound findings (pattern recognition) by an expert sonologist is one of the best methods for discriminating between benign and malignant adnexal masses before surgery[6–9]. If the evaluation is performed by a less experienced sonologist, the diagnostic performance is less good[6]. In training hospitals, a patient with an adnexal mass is often first scanned by a junior or senior registrar. Subsequently, the images are discussed with a supervising sonologist, who will often rescan the patient because he or she believes that this will help in gaining the information required to make a correct diagnosis. However, to the best of our knowledge, it has never been confirmed in a scientific study that real-time scanning is superior to evaluating saved static images of an adnexal mass.

The aim of this study was to determine if the prediction of the malignancy of an adnexal mass using pattern recognition, i.e. subjective evaluation of gray-scale and Doppler ultrasound findings, is as accurate when based on static images as it is when based on a real-time ultrasound examination.

Correspondence to: Dr C. Van Holsbeke, Department of Obstetrics and Gynaecology, Ziekenhuis Oost-Limburg, Schiepse Bos 6, B-3600 Genk, Belgium (e-mail: caroline.van.holsbeke@skynet.be)

Accepted: 15 April 2008

METHODS

For this prospective observational study, a dataset of non-consecutive patients with an adnexal mass was created in the Early Pregnancy and Gynaecology unit of King's College Hospital, London, UK. Patients were included if they had complex or difficult-to-classify masses when using pattern recognition. A large number of patients were referred to this tertiary referral center after they had undergone an ultrasound examination in their referring hospital. In this way we collected a dataset of 171 selected patients with adnexal masses. The study was approved by the local ethics committee.

All the patients included in this study were preoperatively scanned by one of the expert sonologists at the tertiary referral center, the 'real-time' sonologist (D. J.). The ultrasound examination was performed transvaginally and transabdominally using the gray-scale and color/power Doppler modes of an Aloka SSD-5000 ultrasound machine (Aloka Co., Tokyo, Japan), and the images were saved electronically. The anonymized electronic images of all 171 adnexal masses were later independently evaluated by three expert sonologists, the 'image experts' (D.T., A.C.T. and L.V.). Patients were excluded if one of the image experts found the quality of the images insufficient to make a reliable diagnosis. The image experts received relevant clinical information and information on the color score if the color content of the tumor scan was not clearly demonstrated in the static ultrasound images. The color score is a subjective score ranging from 1 to 4 indicating the degree of vascularization of the examined mass as determined by color Doppler examination (a color score of 1 means no vascularization and a color score of 4 means that the mass is highly vascularized)[10]. Both the real-time sonologist and the image experts classified each mass as benign or malignant using pattern recognition. Borderline tumors were regarded as malignant. They also stated with what degree of confidence they made their diagnosis (certainly benign or malignant, probably benign or malignant, uncertain). All four ultrasound experts involved in this study had more than 10 years of experience in gynecological ultrasound at the start of the study, and they are all senior clinicians in tertiary referral centers.

The primary outcome was the histopathological classification of the adnexal mass following the WHO guidelines[11]. In case of malignancy, patients were staged according to the International Federation of Gynecology and Obstetrics (FIGO) criteria[12].

The accuracy, sensitivity and specificity of the prediction of malignancy were calculated for the real-time sonologist and for each of the image experts. The performance of the real-time expert was compared with the performance of each image expert and also with the 'consensus opinion' of the three image experts. The consensus opinion was defined as the diagnosis predicted by at least two of the three image experts. McNemar's test was used to determine the statistical significance of a difference in nominal variables between matched samples. Statistical analyses were performed using SAS Version 9.1.3 for Windows (SAS Institute Inc., Cary, NC, USA).

RESULTS

Five of the 171 masses were excluded because of poor image quality, leaving 166 adnexal masses for evaluation. Seventy masses (42%) were malignant, of which 34 (49%) were borderline tumors. Ninety-six masses (58%) were benign. Table 1 shows the histopathological diagnoses.

Table 2 shows the performance of the real-time sonologist, that of each image expert and that of the consensus opinion of the image experts. The accuracy and specificity of the real-time sonologist were superior to those of two of the image experts and also to the 'consensus opinion' of the three image experts, although the latter was only significant with respect to specificity. The consensus opinion had nine more false positive cases in comparison with the real-time sonologist, reducing the specificity significantly. These false positive cases were one dermoid, two fibromas and six cystadenomas. The fibromas were presumed to be rare malignant tumors by the image experts and the cystadenomas and dermoid were presumed to be borderline tumors. In only one of these nine cases, two of the three image experts were very confident about their diagnosis of malignancy; and in three cases one image expert was certain. In the remaining five false positive cases the image experts were either completely unsure or not really confident (probably benign or probably malignant). The real-time expert stated that he was very confident of his benign diagnosis in four of the nine cases and stated that the diagnosis was probably benign in the remaining five. The four cases that were misclassified by both the consensus opinion and the real-time expert included three cystadenofibromas (one with Brenner tumor component), all of which were misclassified as borderline tumors, and one thecoma that was misclassified as a rare malignant tumor by all the experts.

Table 1 Histopathological diagnoses of the adnexal masses included in the study

Histopathological diagnosis	n *(%)*
Benign	96 (57.8)
Dermoid cyst	35 (21.1)
Cystadenoma/fibroma	35 (21.1)
Endometrioma	16 (9.6)
Fibroma	6 (3.6)
Simple cyst/functional cyst	2 (1.2)
Abscess	1 (0.6)
Rare benign tumor	1 (0.6)
Malignant	70 (42.2)
Mucinous borderline tumor	16 (9.6)
Serous borderline tumor	18 (10.8)
Primary invasive carcinoma	25 (15.1)
Rare malignant tumor	11 (6.6)

Table 2 Accuracy, sensitivity and specificity with regard to malignancy of subjective evaluation of gray-scale and Doppler ultrasound findings in an adnexal mass during scanning ('real-time' sonologist) and by three 'image experts' (A, B and C) who evaluated the static images saved by the real-time sonologist, and of the consensus opinion of the three image experts (i.e. the diagnosis suggested by at least two of the three experts)

Parameter	*Real-time sonologist*	*Image expert A*	*Image expert B*	*Image expert C*	*Consensus opinion of experts A, B and C*
Accuracy (% (*n*))	89 (148/166)	89 (147/166)	82 (136/166)	83 (138/166)	85 (141/166)
		$P = 0.8084$	$P = 0.0186$	$P = 0.0184$	$P = 0.0707$
Sensitivity (% (*n*))	80 (56/70)	86 (60/70)	86 (60/70)	80 (56/70)	83 (58/70)
		$P = 0.1573$	$P = 0.2059$	$P = 1.0000$	$P = 0.4142$
Specificity (% (*n*))	96 (92/96)	91 (87/96)	79 (76/96)	85 (82/96)	86 (83/96)
		$P = 0.0956$	$P < 0.0001$	$P = 0.0016$	$P = 0.0027$

The *P*-values show the statistical significance of the difference between the real-time sonologist and the others.

DISCUSSION

Our results indicate that predicting the character of an adnexal mass by evaluating static ultrasound images is less accurate than by evaluating ultrasound findings during a real-time scan. The difference in accuracy is explained by a difference in specificity, and it seems that ultrasound examiners tend to be more confident in excluding malignancy during a real-time scan than when looking at saved static ultrasound images of an adnexal mass. Because a second opinion is more frequently requested in difficult cases than in obviously benign or malignant cases, it was deliberate that our study population comprised mainly 'difficult' and complex masses, i.e. an unrealistically high number of borderline tumors, which are known to be difficult to classify as benign or malignant using pattern recognition[13]. Masses that can be obviously identified as being either benign or malignant on real-time ultrasound examination can probably also be more easily and accurately classified (in comparison to difficult cases) by the assessment of static ultrasound images, and in such cases the difference in accuracy between real-time and static image assessment that was found in this study may not exist. It is likely that the performance of the image experts in this study would have been poorer if the real-time examination had been performed, and the static ultrasound images saved, by a less experienced sonologist. An expert sonologist is more likely to be able to create representative ultrasound images of an adnexal mass demonstrating relevant ultrasound features. On the other hand, it is possible that the performance of the image experts would have been better if they had had the opportunity to evaluate a volume of static images collected by three-dimensional ultrasound instead of only a few selected two-dimensional (2D) ultrasound images. By analyzing volumes, it is possible to scroll through the volume and look for important ultrasound features that might not have been demonstrated in the selected electronic 2D images. However, when looking at static images, the dynamic aspect of the real-time ultrasound examination is lost. On static images a liquefying blood clot in a hemorrhagic cyst or recent bleeding into an endometrioma may be misinterpreted as an irregular papillary projection of solid tissue, but during a real-time ultrasound examination clots may be seen sliding against the cyst wall when the cyst is pushed using the ultrasound probe. Blood clots mimicking a multilocular cyst on a static image may demonstrate the typical 'jelly' movement when pushed during a real-time scan. On a static image adhesions may give the impression of thick septa in a cystic tumor, while gently pushing on them with the probe will show the 'flapping sail' sign[14]. Thus, the performance of the image experts might have been better if they had been provided with representative video clips or with the use of four-dimensional cine loops of volumes instead of with static images.

In conclusion, a preoperative diagnosis of an adnexal mass made on the basis of a real-time ultrasound examination is more accurate than a diagnosis made on the basis of saved static ultrasound images. Evaluation of static images is associated with lower specificity. The dynamic aspect of the real-time scan gives the examiner supplementary information and makes it possible to scrutinize every part of the mass.

REFERENCES

1. Buchweitz O, Matthias S, Muller-Steinhardt M, Malik E. Laparoscopy in patients over 60 years old: a prospective randomized evaluation of laparoscopic versus open adnexectomy. *Am J Obstet Gynecol* 2005; **193**: 1364–1368.
2. Medeiros LR, Fachel JM, Garry R, Stein AT, Furness S. Laparoscopy versus laparotomy for benign ovarian tumours. *Cochrane Database Syst Rev* 2005; **B**: 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**: 321–326.
4. 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.
5. Hacker NF, Berek JS, Lagasse LD, Nieberg RK, Elashoff RM. Primary cytoreductive surgery for epithelial ovarian cancer. *Obstet Gynecol* 1983; **61**: 413–420.
6. Timmerman D, Schwärzler P, Collins WP, Claerhout F, Coenen M, Amant F, Vergote I, Bourne TH. Subjective assessment of adnexal masses using ultrasonography: an analysis of interobserver variability and experience. *Ultrasound Obstet Gynecol* 1999; **13**: 11–16.

7. Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. *Ultrasound Obstet Gynecol* 1999; **14**: 338–347.
8. 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.
9. Valentin L. Use of morphology to characterize and manage common adnexal masses. *Best Pract Res Clin Obstet Gynaecol* 2004; **18**: 105–123.
10. 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.
11. Senoy SF, Scully RE, Sobin LH. *The World Health Organization international histological classificiation of ovarian tumours*. World Health Organization: Geneva, 1973.
12. International Federation of Gynecology and Obstetrics. Classification and staging of malignant tumours in the female pelvis. *Acta Obstet Gynecol Scand* 1971; **50**: 1–7.
13. Valentin L, Ameye L, Jurkovic D, Metzger U, Lécuru F, Van Hufel S, Timmerman D. Which extrauterine pelvic masses are difficult to 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.
14. Savelli L, de Iaco P, Ghi T, Bovicelli L, Rosati F, Cacciatore B. Transvaginal sonographic appearance of peritoneal pseudocysts. *Ultrasound Obstet Gynecol* 2004; **23**: 284–288.

