Jeroen Kaijser[1]
MD MSc

Tom Bourne[1,2,5]
MD PhD FRCOG

Sylvie De Rijdt[1]
MD

Caroline Van Holsbeke[1, 3]
MD PhD

Ahmad Sayasneh[2]
MD

Lil Valentin[4]
MD PhD

Ben Van Calster[5]
MSc PhD

Dirk Timmerman[1,5]
MD PhD

[1]Department of Obstetrics and Gynaecology
University Hospital Leuven-Leuven
Belgium

[2]Department of Obstetrics and Gynaecology
Queen Charlotte's & Chelsea Hospital
Imperial College
London
United Kingdom

[3]Department of Obstetrics and Gynaecology
Ziekenhuis Oost-Limburg
Genk
Belgium

[4]Department of Obstetrics and Gynaecology
Skåne University Hospital
Lund University
Malmö
Sweden

[5]Academic Department of Development and Regeneration
KU Leuven
Belgium

Correspondence to email
jeroen.kaijser@uzleuven.be

Key findings from the International Ovarian Tumor Analysis (IOTA) study: an approach to the optimal ultrasound based characterisation of adnexal pathology

Abstract

The principal aim of the IOTA project has been to develop approaches to the evaluation of adnexal pathology using ultrasound that can be transferred to all examiners. Creating models that use simple, easily reproducible ultrasound characteristics is one approach.

Keywords: ovarian neoplasms, predictive model, risk assessment, scoring system, ultrasonography.

Introduction

Subjective assessment using grey scale and colour Doppler ultrasonography in the hands of expert examiners is the optimal way to characterise the nature of an adnexal mass. Optimising the diagnostic performance of transvaginal ultrasonography by less experienced examiners has been attempted many times in the past using multimodal scoring systems such as the risk of malignancy index (RMI), morphology scores or models based on logistic regression analysis, neural networks, and support vector machines. However, none of these have retained their performance when externally validated on different patient populations.[1–3,4,5,6] Historically, no biomarker test or algorithm has reached the same accuracy as an expert sonographer.[6,7] An important explanation for this discrepancy is that most models were developed on small populations derived from single centres, with heterogenous tumour populations, and variations in the definitions of ultrasound terms used. The IOTA collaborative group was set up in 1999 to create predictive models that are able to perform as well as expert examiners by studying a large number of adnexal masses, recruited in different centres, using a clearly defined standardised ultrasound protocol.[8] Agreement on a unified approach to describe all possible ultrasound variables is a key element to future successful implementation of any ultrasound based protocol for ovarian pathology.

Main IOTA logistic regression models

In phase 1 of the IOTA study (1999–2002), data from 1066 non-pregnant women with at least one persistent adnexal mass were collected in nine clinical centres, and split into a training set for model development, and a test set for internal validation of the models.[9] In total, 11 new models were created using different techniques (scoring systems, logistic regression models, artificial neural networks (ANN), and vector machine models). We used the logistic regression models LR 1 and LR 2 as our main models, since they were more straightforward and easy to use, compared to our other models. Table 1 shows which ultrasound variables were selected for use in both models. On internal validation, both LR1 and LR2 performed very well with an area under the ROC curve (AUC) for the prediction of malignancy on the test set of 0. 94 and 0.92, respectively.[9] Using a risk threshold of 10% to predict malignancy, the sensitivity and specificity were 93 and 76% for LR1 and 89 and 73% for LR2.[9]

In IOTA phase 2 (2005–2007) a further 1938 patients from 19 centres were included in a validation study.[10] On temporal validation, using the 941 patients recruited in the original development centres LR1 had an AUC of 0.95, 93% sensitivity and 81% specificity. LR2 had an AUC of 0.92, 89% sensitivity and 80% specificity.[10]

On external validation, using the 997 patients recruited by the 12 new centres, LR1 had an AUC of 0.96, 92% sensitivity and 86% specificity. LR2 had an AUC of 0.95, 92% sensitivity and 86% specificity.[10] An important finding was that the performance of the logistic regression models LR1 and LR2 on external validation was equivalent to subjective assessment by gynaecologists and radiologists with a special interest in the ultrasound examination of adnexal tumours (AUC 0.96, 93% sensitivity, 93%

Table 1: Main IOTA models.

Reference	Type of model	Variables used	Cutoff
LR 1 (18)	Logistic regression	(1) personal history of ovarian cancer, (2) current use of hormonal therapy, (3) age, (4) maximal diameter of the lesion, (5) pain, (6) ascites, (7) blood flow within papillary projection, (8) solid tumor, (9) maximal diameter of the largest solid component (bounded at 50 mm), (10) irregular internal cyst walls, (11) acoustic shadows, and (12) color score of intratumoral blood flow	10%
LR 2 (18)	Logistic regression	(1) Age, (2) ascites, (3) blood flow within a solid papillary projection, (4) maximal diameter of the largest solid component (bounded at 50 mm), (5) irregular internal cyst walls, (6) acoustic shadows	10%
Reprinted with permission from Van Holsbeke C, Van Calster B, Testa AC, *et al. Clin Cancer Res* 2009;15: 684–691.			

specificity).[10] However, it remains to be shown if the excellent performance of the models is retained when they are used by ultrasound examiners with a different training background or more variable levels of expertise. This is the subject of a further arm of the IOTA study.

When LR1 and LR 2 were compared to the principal non IOTA models including RMI, the IOTA models clearly outperformed all other models.[11,12] The difference in performance between LR1or LR2 and RMI was more pronounced in premenopausal than postmenopausal patients.[12]

Notwithstanding the evidence that exists to the contrary, many clinicians confronted with an adnexal mass diagnosed using ultrasonography will arrange for laboratory studies to be carried out including biomarker analysis (CA-125, human epididymis secretory protein-4 (HE-4), OVA-1, and the risk of ovarian malignancy algorithm (ROMA)).[13]

However, the IOTA studies have shown that ultrasonography by experienced examiners performs better than serum CA-125 measurements[14] and that incorporating measurements of serum CA-125 into logistic regression models[15] or adding them to subjective assessment by experienced examiners in cases of difficult to characterise ovarian masses does not increase test performance.[16] Furthermore, in a validation study in Leuven, both HE-4 nor ROMA were unable to improve on the diagnostic performance of measurements of serum CA-125 alone.[17]

Even in the hands of experienced ultrasound examiners almost 10% of all adnexal masses remain difficult or impossible to classify.[18] When applied in this subgroup of difficult tumours, the IOTA logistic regression models do not offer any improvement over subjective assessment.[19] Future IOTA studies will evaluate whether second-stage tests such as three-dimensional power Doppler assessment of the vascular tree, the use of ultrasound contrast, or proteomics are useful in characterising this cohort of difficult tumours.

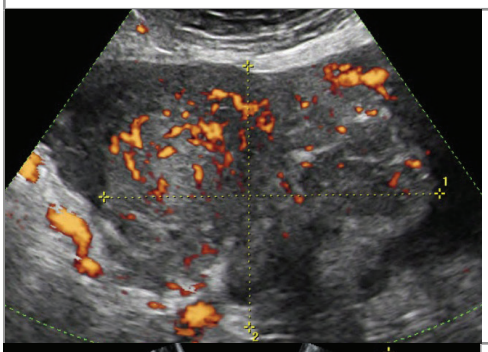
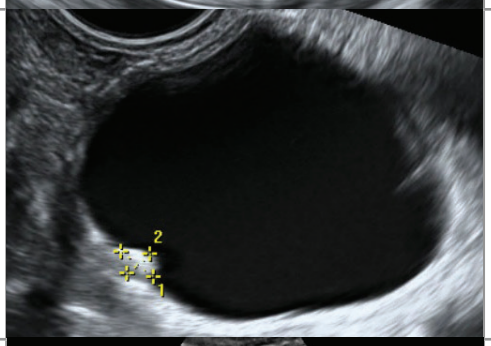
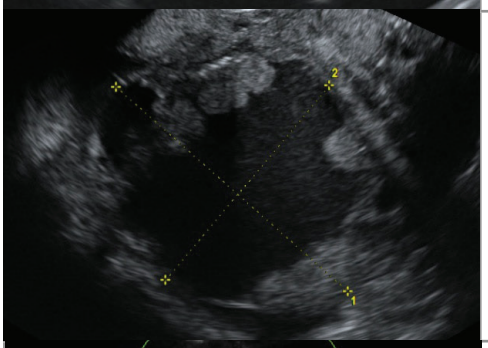
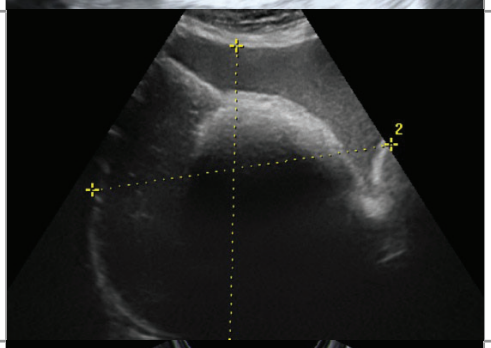
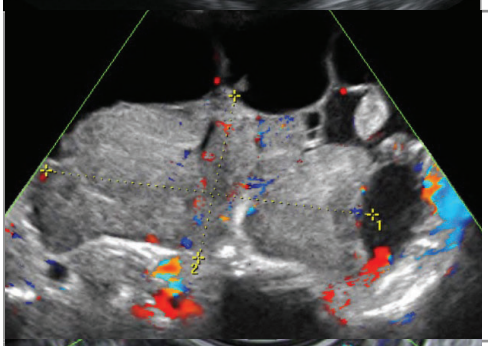
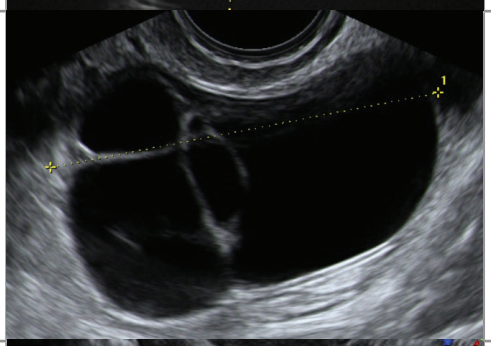
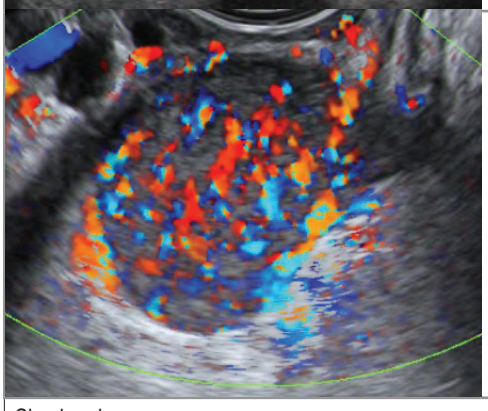
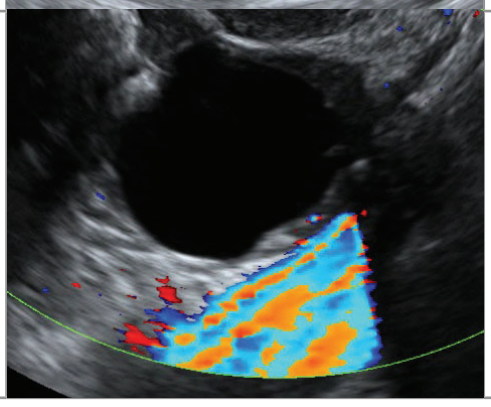
Simple ultrasound-based rules

The principal aim of the IOTA project has been to develop approaches to the evaluation of adnexal pathology using ultrasound that can be transferred to all examiners. Creating models that use simple, easily reproducible ultrasound characteristics is one approach. An analysis of the IOTA study phase 1 data, led to the development of a set of simple rules based on the ultrasound features of a mass.[20] Five ultrasound features to predict malignancy (M-features) and five to predict a benign tumour (B-features) were identified on the basis of the highest positive predictive value (PPV) for malignancy for M-rules and the lowest PPV for malignancy for B-rules.[20] Using this approach, a mass is classified as malignant if one or more M-features apply in the absence of a B-feature. A mass is classified as benign if one or more B-features apply in the absence of an M-feature. A mass cannot be classified if either both M-features and B-features apply, or if no B- or M-feature is present. On temporal and external validation using the IOTA 2 dataset, the simple rules could be applied in 77% of masses.[21] The sensitivity and specificity for ovarian cancer were 92% and 96%, respectively, which was similar to the performance of subjective assessment by an experienced operator (91% and 96%, respectively). If we used simple rules as a triage test and subjective assessment by an experienced examiner to evaluate masses when the rules did not apply, an overall sensitivity of 91% and specificity of 93% were obtained.[21] We believe that any sonographer with adequate training should be able to identify ultrasound M- and B-features in order to use the "simple rules" and so should be able to characterise most ovarian pathology. If there is any doubt about the diagnosis, then review by an examiner with a special interest in gynaecological ultrasound seems the optimal approach to take. This protocol incorporating the initial use of IOTA simple rules to classify pathology has now been included in the Royal College of Obstetricians and Gynaecologists (RCOG) guidance for the evaluation of ovarian pathology in pre-menopausal women.[22]

Key points

- Transvaginal ultrasonography by expert examiners is currently the optimal approach to evaluating an ovarian mass.
- Measurement of serum CA-125 does not enhance the diagnostic performance of ultrasonography in the hands of experienced examiners.
- Neither HE-4 nor ROMA improve on the diagnostic accuracy of measurements of serum CA-125 alone.
- The logistic regression models LR1 and LR2 can be applied to all adnexal masses and will provide a reliable classification in most cases.
- Both LR 1 and LR 2 perform significantly better than the RMI. Furthermore, they provide the clinician with a percentage risk of cancer instead of a simple positive or negative diagnosis.

Table 2: Simple rules for identifying a benign or malignant tumour.

Features for predicting a malignant tumor (M-features)		Features for predicting a benign tumor (B-features)	
	**M1:** Irregular solid tumour		**B1:** Unilocular
	**M2:** Presence of ascites		**B2:** Presence of solid components where the largest solid component has a largest diameter < 7 mm
	**M3:** At least four papillary structures		**B3:** Presence of acoustic shadows
	**M4:** Irregular multilocular solid tumour with largest diameter ≥100 mm		**B4:** Smooth multilocular tumour with largest diameter < 100 mm
	**M5:** Very strong blood flow (colour score 4)		**B5:** No blood flow (colour score 1) (the external iliac vein is visualised next to the ovary)

Simple rules:
- If one or more M-features apply in the absence of a B-feature, the mass is classified as malignant.
- If one or more B-features apply in the absence of an M-feature, the mass is classified as benign.
- If both M-features and B-features apply, the mass cannot be classified. If no feature applies, the mass cannot be classified.

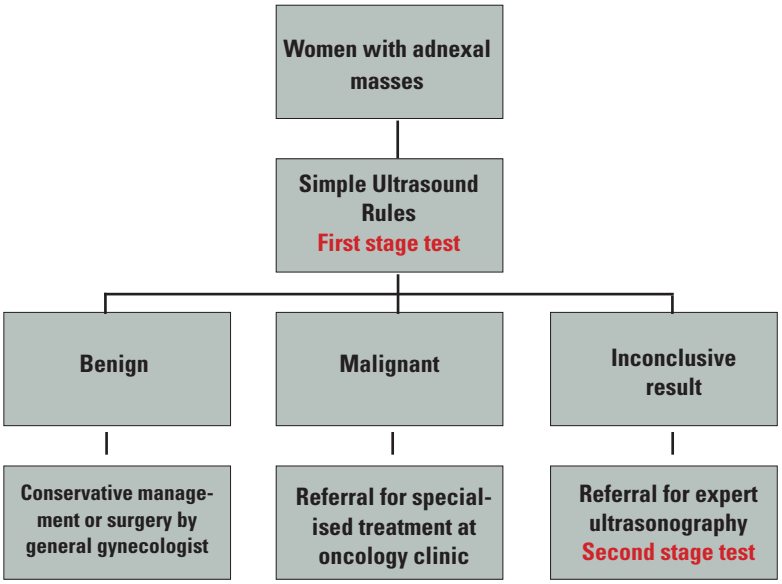


Figure 1: An algorithm for evaluating women with adnexal pathology based on simple ultrasound based rules.

- Ultrasound based "simple rules" do not require a computer for their use and have been shown to have the same test performance as subjective assessment by expert examiners in masses where the rules can be applied. They have been externally validated in 1983 women from 19 ultrasound centres in eight countries.
- Current evidence suggests that using the IOTA simple rules as a triage test, gives the best diagnostic performance and most straightforward approach to characterising most ovarian pathology.
- In cases where simple rules cannot be applied patients should be referred to a specialist in gynaecological ultrasound. An algorithm, based on this approach, which can be used in clinical centres to evaluate women with adnexal pathology is shown in Figure 1.

Acknowledgements

The authors thank all participating centres, the principal investigators, and the study participants for their contribution to the IOTA study

Recruitment Centres: The recruitment centres include University Hospitals Leuven (Belgium); Ospedale S. Gerardo, Università di Milano Bicocca, Monza (Italy); Ziekenhuis Oost-Limburg (ZOL), Genk (Belgium); Medical University in Lublin (Poland); University of Cagliari, Ospedale San Giovanni di Dio, Cagliari (Italy); Malmö University Hospital, Lund University (Sweden), University of Bologna (Italy); Università Cattolica del Sacro Cuore Rome (Italy); DCS Sacco University of Milan (Milan A, Italy); General Faculty Hospital of Charles University, Prague (Czech Republic); Chinese PLA General Hospital, Beijing (PR China); King's College Hospital London (UK); Universita degli Studi di Napoli, Napoli (Naples A, Italy); IEO, Milano (Milan B, Italy); Lund University Hospital, Lund (Sweden); Macedonio Melloni Hospital, University of Milan (Milan C, Italy); Università degli Studi di Udine (Italy); McMaster University, St. Joseph's Hospital, Hamilton, Ontario (Canada); and Instituto Nationale dei Tumori, Fondazione Pascale, Napoli (Naples B, Italy).

IOTA Steering Committee: The members of IOTA steering committee are D. Timmerman, Leuven, Belgium; L. Valentin, Malmö, Sweden; T. Bourne, London, UK; A.C. Testa, Rome, Italy; S. Van Huffel, Leuven, Belgium; Ignace Vergote, Leuven, Belgium; and B. Van Calster, Leuven, Belgium.

IOTA Principal Investigators: IOTA principal investigators (in alphabetical order) are A. Czekierdowski, Lublin, Poland; Elisabeth Epstein, Lund, Sweden; Daniela Fischerová, Prague, Czech Republic; Dorella Franchi, Milano, Italy; Robert Fruscio, Monza, Italy; Stefano Greggi, Napoli, Italy; S. Guerriero, Cagliari, Italy; Jingzhang, Beijing, PR China; Davor Jurkovic, London, UK; Francesco P.G. Leone, Milano, Italy; A.A. Lissoni, Monza, Italy; Henry Muggah, Hamilton, Ontario, Canada; Dario Paladini, Napoli, Italy; Alberto Rossi, Udine, Italy; L. Savelli, Bologna, Italy; A.C. Testa, Roma, Italy; D. Timmerman, Leuven, Belgium; Diego Trio, Milano, Italy; L. Valentin, Malmö, Sweden; and C. Van Holsbeke, Genk, Belgium.

Disclosure of interests

The authors have no conflict of interest to disclose.

Contribution to authorship

All coauthors helped in writing the manuscript and they all approved the final version.

Details of ethics approval

The study protocol was approved by the central ethics committee for clinical studies at the University Hospitals Leuven, Belgium, and by the local ethics committee at each recruitment centre.

Funding

This research was supported by the Research Council KUL: GOA MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC); Research Foundation–Flanders (FWO): projects G.0302.07 (SVM), G.0341.07 (Data fusion); IWT: TBM070706-IOTA3; Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, "Dynamical systems, control and optimization", 2007–2011); IBBT (Flemish Government); Swedish Medical Research

Council: grants nos. K2001-72X 11605-06A, K2002-72X-11605-07B, K2004-73X-11605-09A and K2006-73X-11605-11-3; funds administered by Malmö University Hospital; and two Swedish governmental grants (ALF-medel and Landstingsfinansierad Regional Forskning). Ben Van Calster is a postdoctoral fellow of the Research Foundation - Flanders (FWO) For IOTA 5 a project grant from the Research Foundation-Flanders (FWO Vlaanderen; grant G049312N) was received. There is no financial compensation for principal investigators or patients.

References

1 Geomini P, Kruitwagen R, Bremer GL, Cnossen J, Mol BW. The accuracy of risk scores in predicting ovarian malignancy: a systematic review. *Obstet Gynecol* 2009; 113: 384–94.

2 Mol BW, Boll D, De Kanter M, Heintz AP, Sijmons EA, Oei SG, *et al.* Distinguishing the benign and malignant adnexal mass: An external validation of prognostic models. *Gynecol Oncol* 2001; 80: 162–67.

3 Ferrazzi E, Zanetta G, Dordoni D, Berlanda N, Mezzopane R, Lissoni AA. Transvaginal ultrasonographic characterization of ovarian masses: comparison of five scoring systems in a multicenter study. *Ultrasound Obstet Gynecol* 1997; 10: 192–97.

4 Aslam N, Banerjee S, Carr JV, Savvas M, Hooper R, Jurkovic D. Prospective evaluation of logistic regression models for the diagnosis of ovarian cancer. *Obstet Gynecol* 2000; 96: 75–80.

5 Van Holsbeke C, Van Calster B, Valentin L, Testa AC, Ferrazzi E, Dimou I, *et al.* 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: 4440–47.

6 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: 357–65.

7 Van Calster B, Timmerman D, Bourne T, Testa A, Van Holsbeke C, Domali E, *et al.* Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125. *J Natl Cancer Inst* 2007; 99: 1706–14.

8 Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group. *Ultrasound Obstet Gynecol* 2000; 16: 500–05.

9 Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, *et al.* 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.

10 Timmerman D, Van Calster B, Testa AC, Guerriero S, Fischerova D, Lissoni AA, *et al.* 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: 226–34.

11 Van Calster B, Timmerman D, Valentin L, McIndoe A, Ghaem-Maghami S, Testa A, *et al.* Triaging women with ovarian masses for surgery: observational diagnostic study to compare RCOG guidelines with an International Ovarian Tumour Analysis (IOTA) group protocol. *BJOG* 2012 Mar 6. Epub.

12 Van Holsbeke C, Van Calster B, Bourne T, Ajossa S, Testa AC, Guerriero S, *et al.* External Validation of Diagnostic Models to Estimate the Risk of Malignancy in Adnexal Masses. *Clin Cancer Res* 2012; 18 (3): 815–25.

13 Myers ER, Bastian LA, Havrilesky L, Kulasingam SL, Terplan MS, Cline KE, *et al.* Management of Adnexal Mass. Evidence Report/Technology Assessment No130(Prepared by the Duke Evidence-based Practice Center under Contract No 290-02-0025) AHRQ Publication No 06-E004. Rockville, MD: Agency for Healthcare Research and Quality; 2006.

14 Van Calster B, Timmerman D, Borune T, Testa AC, Van Holsbeke C, Domali E, *et al.* Discrimination between benign and malignant adnexal masses by specialized ultrasound examination versus serum CA-125. *J Natl Cancer Inst* 2007; 99: 1706–14.

15 Timmerman D, Van Calster B, Jurkovic D, Valentin L, Testa A, Bernard J, *et al.* Inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and malignant adnexal tumors. *J Clin Oncol* 2007; 25: 4194–200.

16 Valentin L, Jurkovic D, Van Calster B, Testa AC, Van Holsbeke C, Bourne T, *et al.* Adding a single CA-125 measurement to ultrasound performed by an experienced examiner does not improve preoperative discrimination between benign and malignant adnexal masses. *Ultrasound Obstet Gynecol* 2009; 34: 345–54.

17 Van Gorp T, Cadron I, Despierre E, Daemen A, Leunen K, Amant F, *et al.* HE4 and CA125 as a diagnostic test in ovarian cancer: prospective validation of the Risk of Ovarian Malignancy Algorithm. *Br J Cancer* 2011; 104 (5): 863–70.

18 Valentin L, Ameye L, Jurkovic D, Metzger U, Lecuru F, Van Huffel S, *et al.* 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: 438–44.

19 Valentin L, Ameye L, Savelli L, Fruscio R, Leone FP, Czekierdowski A, *et al.* 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 (4): 456–65.

20 Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, *et al.* Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol* 2008; 31: 681–90.

21 Timmerman D, Ameye L, Fischerova D, Epstein E, Melis GB, Guerriero S, *et al.* Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. *BMJ* 2010; 341: c6839.

22 Guidance RC. Management of ovarian cysts in premenopausal women. www.rcog.org.uk.