



Improving strategies for diagnosing ovarian cancer: a summary of the International Ovarian Tumor Analysis (IOTA) studies

J. KAIJSER*, T. BOURNE*†‡, L. VALENTIN§, A. SAYASNEH†, C. VAN HOLSBEKE*¶, I. VERGOTE*, A. C. TESTA**, D. FRANCHI††, B. VAN CALSTER*‡ and D. TIMMERMAN*‡

*Department of Obstetrics and Gynecology, University Hospitals KU Leuven, Leuven, Belgium; †Department of Obstetrics and Gynecology, Queen Charlotte's & Chelsea Hospital, Imperial College, London, UK; ‡Department of Development and Regeneration, KU Leuven, Leuven, Belgium; §Department of Obstetrics and Gynecology, Skåne University Hospital Malmö, Lund University, Malmö, Sweden; ¶Department of Obstetrics and Gynecology, Ziekenhuis Oost-Limburg, Genk, Belgium; **Department of Obstetrics and Gynecology, Università Cattolica del Sacro Cuore, Rome, Italy; ††Preventive Gynecology Unit, Division of Gynecology, European Institute of Oncology, Milan, Italy

KEYWORDS: biomarkers; decision support techniques; logistic models; ovarian neoplasms; ultrasonography

ABSTRACT

In order to ensure that ovarian cancer patients access appropriate treatment to improve the outcome of this disease, accurate characterization before any surgery on ovarian pathology is essential. The International Ovarian Tumor Analysis (IOTA) collaboration has standardized the approach to the ultrasound description of adnexal pathology. A prospectively collected large database enabled previously developed prediction models like the risk of malignancy index (RMI) to be tested and novel prediction models to be developed and externally validated in order to determine the optimal approach to characterize adnexal pathology preoperatively. The main IOTA prediction models (logistic regression model 1 (LR1) and logistic regression model 2 (LR2)) have both shown excellent diagnostic performance (area under the curve (AUC) values of 0.96 and 0.95, respectively) and outperform previous diagnostic algorithms. Their test performance almost matches subjective assessment by experienced examiners, which is accepted to be the best way to classify adnexal masses before surgery. A two-step strategy using the IOTA simple rules supplemented with subjective assessment of ultrasound findings when the rules do not apply, also reached excellent diagnostic performance (sensitivity 90%, specificity 93%) and misclassified fewer malignancies than did the RMI. An evidence-based approach to the preoperative characterization of ovarian and other adnexal masses should include the use of LR1, LR2 or IOTA simple rules and subjective assessment by an experienced examiner. Copyright © 2013 ISUOG. Published by John Wiley & Sons, Ltd.

BACKGROUND

Correctly discriminating between benign or malignant adnexal masses is the essential starting point for optimal management. Most women with an adnexal mass do not have cancer¹. Identifying women with benign pathology is important in order to avoid unnecessary morbidity as well as unnecessary costs². Conversely, recognizing cancer means that treatment is not delayed and appropriate staging can be carried out in specialized surgical centers^{3–6}.

To characterize ovarian pathology as benign or malignant, biomarkers or various prediction models have been used to try to optimize diagnostic accuracy. These include simple scores based on the morphological appearance of a mass using ultrasonography^{7–11}; an index including information on serum CA 125 levels, menopausal status and ultrasound findings (the risk of malignancy index (RMI))¹²; and more advanced mathematical models using logistic regression¹³, neural networks¹⁴ and other complex computational approaches¹⁵.

The RMI¹² remains the most widely used prediction model for characterizing ovarian pathology in many countries. Although the RMI is based on several ultrasound markers, the serum CA 125 level heavily influences the predictions. A systematic review in 2009 concluded that the RMI was the best available test to triage patients with ovarian tumors for referral to tertiary oncology units¹⁶.

In the USA, the American College of Obstetricians and Gynecologists (ACOG) guidelines for selecting patients for referral to a gynecologic oncology center rely on the use of biomarkers. However, although useful for

Correspondence to: Prof. D. Timmerman, Department of Obstetrics and Gynecology, University Hospitals KU Leuven, Herestraat 49, 3000 Leuven, Belgium (e-mail: dirk.timmerman@uzleuven.be)

Accepted: 3 October 2012

predicting advanced-stage ovarian cancer, the ACOG protocol performs poorly for the detection of early-stage disease and in the subgroup of premenopausal women¹⁷. When the multivariate biomarker assay, OVA-1, was incorporated instead of serum CA 125 into these referral guidelines¹⁸, the false-positive rate reached unacceptably high levels¹⁹.

Several prediction models, other than those discussed above, have been developed with the aim of improving preoperative diagnostic tests for ovarian cancer. Most did not retain their original accuracy when subjected to external validation^{16,20–24}. This can be explained by the relatively small sample size of most studies in which models were developed, the use of single-center populations for model development, the heterogeneity of the tumors studied, variations in the definitions of ultrasound terms used and a lack of consistency regarding the reporting of histological findings. To minimize these shortcomings and to develop robust rules and prediction models that can be used by different examiners in various clinical settings, the International Ovarian Tumor Analysis (IOTA) study was established.

AIM OF THE IOTA STUDIES

The principal IOTA investigators set out to study a large cohort of patients with a persistent adnexal mass in different clinical centers, using a standardized ultrasound protocol²⁵. The primary aim of the study was to develop rules and models to characterize ovarian pathology and subsequently to demonstrate their utility by both temporal and external validation in the hands of examiners with different levels of ultrasound expertise. Other aims related to the different IOTA projects included establishing the role of measurements of CA 125 and other serum tumor markers for diagnosis, and developing a better understanding of the characteristics of ovarian tumors difficult to classify as benign or malignant using ultrasound. Both a strength and a limitation of the early IOTA phases is that final histology is required for inclusion. This is common to the majority of studies seeking to characterize ovarian pathology. An important ongoing phase for IOTA is studying the long-term behavior of adnexal masses characterized as benign in order to validate IOTA models or rules also in patients in whom surgery is not performed (Figure 1).

DEVELOPMENT AND PERFORMANCE OF PREDICTIVE MODELS

The first step was to agree on standardized terms and definitions that could be used to describe adnexal pathology. These were published in 2000²⁵. Subsequently, in Phase 1 of the study (1999–2002), ultrasound data from 1066 non-pregnant women with at least one persistent adnexal mass were collected from nine clinical centers in five countries. A training set of 754 patients (70.7%) was used for model development, and a test

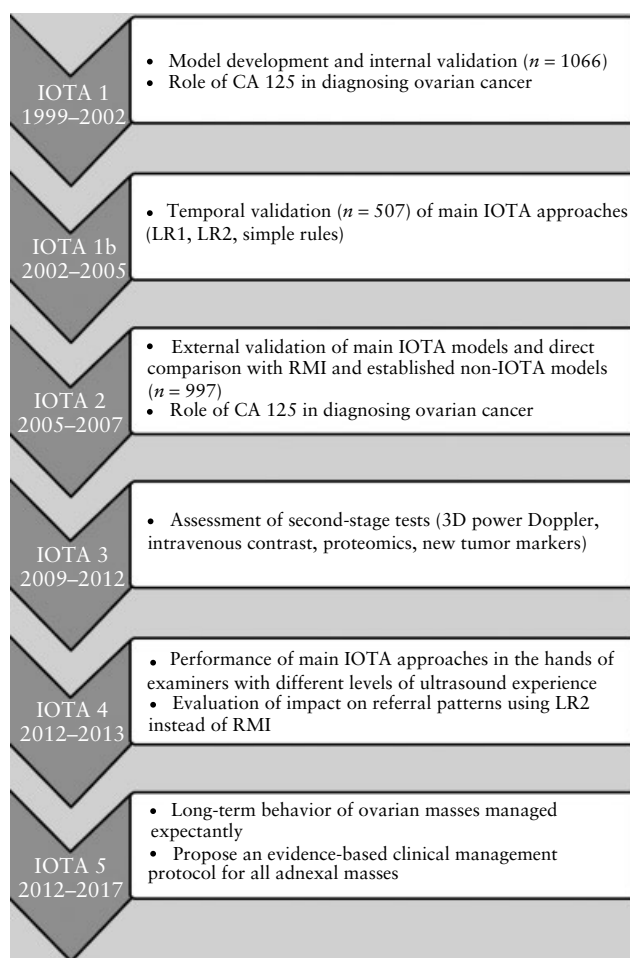


Figure 1 Objectives of the different phases of the International Ovarian Tumor Analysis (IOTA) study. 3D, three dimensional; LR1, logistic regression model 1; LR2, logistic regression model 2; RMI, risk of malignancy index.

set containing the remaining 312 patients was used for internal validation of the models²⁶.

Between 2002 and 2005 (IOTA Phase 1b), we recruited 507 new consecutive patients, at three centers participating in Phase 1, for prospective temporal validation of the models that seemed to perform best on internal validation in Phase 1²⁷. The aim of IOTA Phase 2 (2005–2007) was to externally validate the models. This involved the recruitment of a further 997 patients in 12 new centers that did not take part in Phase 1, and of 941 patients in seven centers from Phase 1 for further temporal validation (Figure 1)^{28,29}.

Initially, 11 prediction models were derived from the IOTA 1 dataset. Scoring systems, simple ultrasound rules, logistic regression analysis, artificial neural networks (ANN) and kernel methods, such as support vector machine models, were developed^{26,30–33}. We found that more complex statistical modeling did not improve test performance appreciably in comparison with more simple statistical approaches, such as logistic regression²⁷. Accordingly we designated two relatively simple logistic regression models (logistic regression model 1 (LR1) and logistic regression model 2 (LR2)) as our principal models.

Table 1 Diagnostic performance of the main predictive models and rules for discrimination between benign and malignant adnexal masses derived by the International Ovarian Tumor Analysis (IOTA) study and of the risk of malignancy index (RMI)

Model or rules	IOTA phase	Type of validation	n	Sensitivity (%)	Specificity (%)	LR+	LR–	DOR	AUC
LR1 (cut-off 10%)	1	Development data ²⁶	754	93	77	4.01	0.10	42.1	0.95
	1	Internal (test set) ²⁶	312	93	76	3.81	0.09	45.6	0.94
	1b	Temporal ²⁷	507	95	74	3.68	0.07	55.8	0.95
	2	Temporal ^{28,29}	941	93	81	4.77	0.09	52.8	0.94
	2	External ^{28,29}	997	92	87	6.84	0.09	75.7	0.96
LR2 (cut-off 10%)	1	Development data ²⁶	754	92	75	3.71	0.10	35.5	0.93
	1	Internal (test set) ²⁶	312	89	73	3.36	0.15	23.1	0.92
	1b	Temporal ²⁷	507	95	74	3.64	0.07	55.0	0.95
	2	Temporal ^{28,29}	941	89	80	4.42	0.14	32.7	0.92
	2	External ^{28,29}	997	92	86	6.36	0.10	66.1	0.95
Simple rules with subjective expert assessment*	1	Development data ³²	1066	91	90	8.84	0.10	84.4	N/A
	1b	Temporal ³²	507	92	90	9.08	0.09	106	N/A
	2	Temporal ³⁷	941	92	93	12.28	0.09	142	N/A
	2	External ³⁷	997	90	93	12.63	0.11	120	N/A
Subjective expert assessment	1	N/A	1066	88	95	18.52	0.13	147	N/A
	1b	N/A	507	90	93	12.63	0.11	120	N/A
	2	N/A	941	93	93	14.15	0.07	190	N/A
	2	N/A	997	87	92	11.00	0.14	80.7	N/A
RMI† (cut-off 200)	2	External ²⁸	997	67	95	12.7	0.34	36.8	0.91

*Results are shown for simple rules supplemented with subjective assessment of ultrasound findings when the rules did not apply. †Missing values for CA 125 were handled using multiple imputation, and missing values for metastases were handled as explained in our external validation study²⁸. AUC, area under the receiver–operating characteristics curve; DOR, diagnostic odds ratio; LR1, IOTA logistic regression model 1; LR2, IOTA logistic regression model 2; LR+, positive likelihood ratio; LR–, negative likelihood ratio; N/A, not applicable.

LR1 included 12, and LR2 included six, demographic and ultrasound variables. The 12 variables used in LR1 were: (1) personal history of ovarian cancer; (2) current hormonal therapy; (3) age of the patient; (4) maximum diameter of the lesion; (5) pain during examination; (6) ascites; (7) blood flow within a solid papillary projection; (8) a purely solid tumor; (9) the maximum diameter of the solid component; (10) irregular internal cyst walls; (11) acoustic shadows; and (12) color score. The six variables in LR2 were: (1) age; (2) ascites; (3) blood flow within a solid papillary projection; (4) maximal diameter of the solid component; (5) irregular internal cyst walls; and (6) acoustic shadows. Any qualified ultrasound examiner scanning women with adnexal masses should be able to retrieve information on the variables required for both models.

Both models had excellent diagnostic performance on both the training and test data²⁶ and retained their accuracy at prospective temporal validation in three clinical centers using the IOTA 1b dataset (Table 1)²⁷. In the IOTA study we have emphasized that good sensitivity is more important than specificity. However, interpreting indices of diagnostic performance is dependent upon the prevalence of pathology in the studied population. In the IOTA study, the overall prevalence of cancer was 28%, which implies that at a fixed specificity level of 75% with a sensitivity of 90%, for every five patients who undergo surgery for a presumed malignant mass only two will have a benign histology.

A problem with diagnostic models is that they are prone to produce good results on the populations on which they were developed. Therefore, a crucial step before incorporating any diagnostic test or prediction model into clinical practice is establishing whether they work

in different patient populations and in different clinical settings. This involves external validation in centers unrelated to those in which the tests were developed^{34,35}.

Phase 2 of the IOTA project (2005–2007) was designed to externally validate LR1 and LR2 and to compare their test performance with the RMI and other previously published non-IOTA models²⁸. A useful tool to evaluate test performance is to construct receiver–operating characteristics (ROC) curves and to compare the area under the curve (AUC) for different diagnostic tests. The advantage of this approach is that the AUC is independent of the cut-off value applied and is therefore a more useful description of how a test performs. Using a risk threshold of 10%, LR1 outperformed other tests, such as RMI, for estimating the risk of malignancy in an ovarian mass, with an AUC of 0.96 (95% CI, 0.94–0.97) and sensitivity and specificity of 92% and 87%, respectively. LR2 achieved an AUC of 0.95, sensitivity of 92% and specificity of 86% using the same 10% risk threshold. In contrast, the RMI achieved an AUC of 0.91, sensitivity of 67% and specificity of 95% (Table 1). The adopted risk-threshold of 10% means that a tumor predicted by the model as having a risk of 10% or more should be classified as malignant. Altering the risk threshold according to personal preference affects test performance. If a 5% risk of malignancy is considered more appropriate, the sensitivity for cancer would increase but at the expense of increasing the false-positive rate.

The ROC curves for LR1, LR2, RMI and serum CA 125 for both premenopausal and postmenopausal women are shown in Figure 2. Figure 2a demonstrates an important diagnostic advantage for both LR1 and LR2 for characterizing adnexal tumors in premenopausal patients compared with the current reference test RMI or using CA 125 alone.

The difference in AUC between LR1 and LR2 was small, irrespective of menopausal status, and this small difference is unlikely to be of clinical importance. This implies that a model with only six variables (LR2) has diagnostic performance very similar to one using 12 variables (LR1). The lower number of variables needed for LR2 and its excellent test performance may lead to clinicians favoring the use of LR2 over LR1 in clinical practice.

In postmenopausal patients (Figure 2b), there is little difference in diagnostic performance between the main IOTA models and RMI. However, the ultrasound-based models have the advantage of offering an instant diagnosis.

The performance of any test for the detection of primary Stage I disease is of particular interest, because treatment for early-stage disease is associated with high survival rates. For Stage I tumors, we found that the logistic regression models had a higher detection rate than the RMI²⁸. The LR2 missed fewer malignancies of any kind than did the RMI or CA 125 alone when applying their cut-off points most often used clinically. Table 2 shows the number and types of malignancies that were missed by LR2, RMI and serum CA 125. Of course, the number of false negatives depends on the cut-off adopted.

We have retrospectively compared the RMI-based triage system advocated by the Royal College of Obstetricians and Gynaecologists (RCOG) with an IOTA-based alternative protocol using LR2. The IOTA protocol classified women as being at high risk if the estimated probability of malignancy according to LR2 was at least 25%, at low risk if the estimated probability was below 5% and at intermediate risk if the estimated probability was at least 5% but less than 25%³⁶. This analysis suggests that if implemented, the IOTA alternative is likely to be better at avoiding unnecessary surgery or unnecessarily extensive surgery in benign disease whilst selecting more patients with cancer for appropriate referral to an oncological surgeon than the RCOG system³⁶. This result held true, irrespective of the menopausal status of the patients and the type of unit in which the patient was examined³⁶.

DEVELOPMENT, VALIDATION AND ROLE OF THE SIMPLE ULTRASOUND-BASED RULES

Subjective assessment of ultrasound images by experienced clinicians is the best way of characterizing ovarian pathology²⁴. Many adnexal masses have a typical ultrasound appearance and will therefore be instantly correctly classified even by relatively inexperienced ultrasound examiners. To take this idea forward we established simple rules based on a number of clearly defined ultrasound features that can guide examiners without the need for a computer³². Using these simple rules no risk estimates are produced, but tumors are classified as benign, malignant or unclassifiable.

The simple rules consist of five ultrasound features of malignancy (M-features) and five ultrasound features

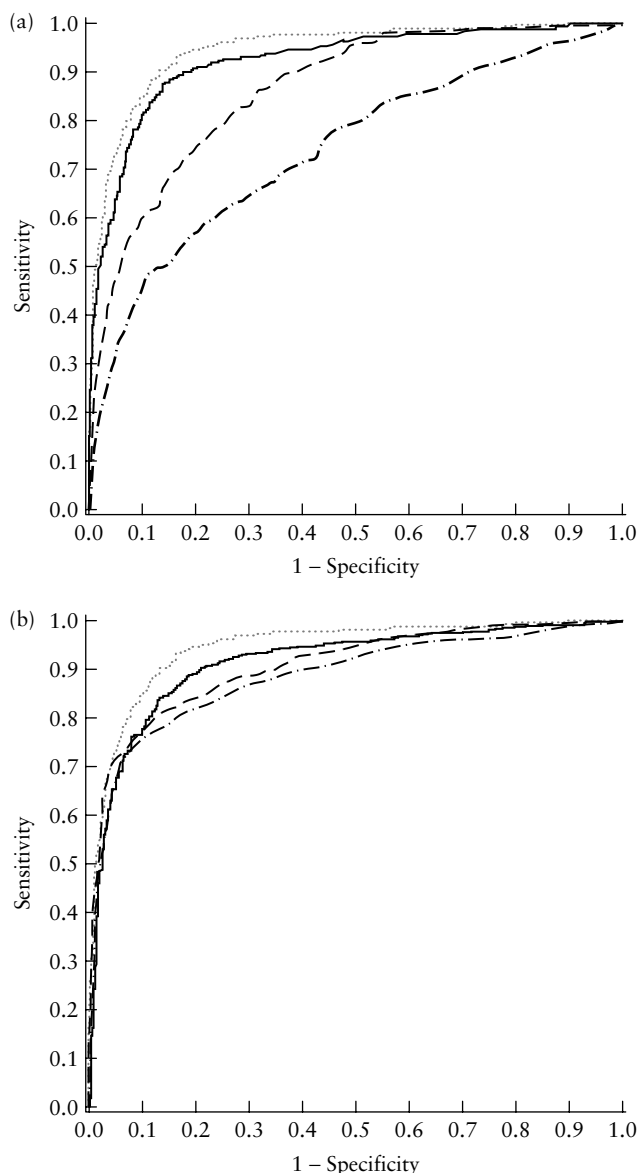


Figure 2 Receiver-operating characteristics (ROC) curves of the International Ovarian Tumor Analysis (IOTA) logistic regression model 1 (LR1;), IOTA logistic regression model 2 (LR2; —), risk of malignancy index (RMI; ---) and CA 125 (— · —) for premenopausal (a) and postmenopausal (b) patients using pooled data ($n = 2757$) from IOTA Phases 1, 1b and 2, excluding patients derived from our training set ($n = 754$). Missing values for CA 125 were handled using multiple imputation, and missing values for metastases were handled as explained in our external validation study²⁸. (a) Area under the ROC curve (AUC) of LR1, 0.945 (95% CI, 0.928–0.959); LR2, 0.922 (95% CI, 0.901–0.940); RMI, 0.865 (95% CI, 0.827–0.896); and CA 125, 0.741 (95% CI, 0.701–0.777). (b) AUC of LR1, 0.928 (95% CI, 0.910–0.942); LR2, 0.915 (95% CI, 0.895–0.931); RMI, 0.909 (95% CI, 0.889–0.926); and CA 125, 0.886 (95% CI, 0.862–0.906).

suggestive of a benign mass (B-features). These features with corresponding ultrasound images are presented in Figure 3. A mass is classified as malignant if at least one M-feature and none of the B-features are present, and vice versa³². If no B- or M-features are present, or if both B- and M-features are present, then the rules are considered inconclusive (unclassifiable mass) and a different

Table 2 False-negative test results with regard to malignancy for logistic regression model 2 (LR2), the simple rules combined with subjective assessment, risk of malignancy index (RMI) and CA 125 when applying them to International Ovarian Tumor Analysis (IOTA) Phase 1b and Phase 2 data ($n = 2445$)

Classification approach	False negative (n (%))				
	Invasive Stage I* (n = 122)	Invasive Stage II–IV† (n = 354)	Borderline (n = 131)	Metastatic (n = 78)	All malignancies (n = 685)
LR2, 10% risk threshold	9 (7.4)	13 (3.7)	33 (25.2)	4 (5.1)	59 (8.6)
Simple rules combined with subjective assessment‡	12 (9.8)	17 (4.8)	26 (19.8)	4 (5.1)	59 (8.6)
RMI, threshold 200	58 (47.5)	38 (10.7)	87 (66.4)	29 (37.2)	212 (30.9)
CA 125, thresholds 65 IU/L and 35 IU/L for pre- and postmenopausal patients	55 (45.1)	31 (8.8)	77 (58.8)	26 (33.3)	189 (27.6)

*Includes rare primary invasive Stage I tumors. †Includes rare primary invasive Stage II–IV tumors. ‡Results are shown for simple rules supplemented with subjective assessment of ultrasound findings when the rules did not apply. Missing values for CA 125 were handled using multiple imputation, and missing values for metastases were handled as explained in our external validation study²⁸.

diagnostic method should be used. The simple rules were temporally and externally validated using the IOTA Phase 2 dataset of 1938 patients³⁷. The rules could be applied to 77% of ovarian tumors. The simple rules classify tumors as benign, inconclusive or malignant. These three possible outcomes enable the construction of an ROC curve, which facilitates comparison with the prediction models LR1 and LR2. When we do this using the IOTA Phase 1b and 2 datasets (Figure 4), the performance of LR1, LR2 and simple rules are similar. We suggest subjective assessment by an experienced examiner as a second-stage test for cases where the simple rules yielded an inconclusive result. On external validation, this two-step strategy reached a sensitivity of 90% and a specificity of 93% to detect ovarian malignancy (Table 1)³⁷. Moreover, the simple rules combined with subjective assessment when the rules did not apply misclassified fewer Stage I ovarian malignancies than did RMI and measurement of serum CA 125 (Table 2). In 2011 the RCOG included the simple rules in their guideline for evaluating ovarian pathology in premenopausal women³⁸. However, in postmenopausal patients, serum CA 125 may play a role as a second-stage test, especially in centers with less-experienced ultrasound examiners. This hypothesis is currently being tested as part of the ongoing IOTA 4 project.

AN INTUITIVE APPROACH TO ULTRASOUND CHARACTERIZATION: THE USE OF ‘INSTANT’ DESCRIPTORS

An important learning point from the IOTA study is that almost half of ovarian masses have features that enable them to be characterized relatively easily (43% of the masses from Phases 1, 1b and 2). For example, ‘typical’ dermoid cysts, ‘typical’ endometriomas and late-stage ovarian cancer have very characteristic ultrasound features that should be recognized almost instantly by any ultrasound examiner. We retrospectively defined six ‘easy descriptors’³⁹ that should enable an examiner to make an ‘instant’ diagnosis of an ovarian mass without needing to use models, second-stage tests or seek a second opinion: four described features of

common benign tumors, whilst two described features of malignancies (Figure 5).

When applied retrospectively to IOTA data, each one of these six descriptors had excellent diagnostic accuracy to predict whether a mass was benign or malignant. For the masses to which the descriptors could be applied, they had a sensitivity of 98% and a specificity of 97%³⁹. If none of the six descriptors could be used, or if both a descriptor of a benign and a malignant mass were applicable, we considered the diagnosis as ‘non-instant’. In clinical practice, a second-stage test or an expert opinion is needed only for tumors where an instant diagnosis cannot be made. As a second-stage test for such masses, we retrospectively applied our previously developed simple rules with real-time subjective assessment by an expert examiner as a third step, when the simple rules could not be applied. This protocol gave a sensitivity of 92% and a specificity of 92% when retrospectively applied on all 1938 patients from IOTA Phase 2. It detected more ovarian cancers than if expert ultrasonography alone had been used in the whole study cohort, without increasing the false-positive rate (sensitivity and specificity of expert ultrasonography = 90% and 93%)³⁹. Clearly, prospective external validation is needed before we can suggest incorporation of this approach into clinical protocols.

RELEVANCE OF BIOMARKERS (CA 125 AND HUMAN EPIDIDYMIS SECRETORY PROTEIN-4) AND RISK OF OVARIAN MALIGNANCY ALGORITHM

The role of biomarkers, and in particular CA 125, in the diagnosis of ovarian cancer is controversial. Although widely used as part of the assessment of ovarian pathology, the results of the IOTA study suggest that measurements of serum CA 125 have a limited role in characterizing ovarian pathology, especially in premenopausal women⁴⁰. Incorporating serum CA 125 measurements into logistic regression models has no significant impact on performance of the model for women of any age⁴⁰. Moreover, when subjective assessment by an experienced ultrasound examiner was compared with

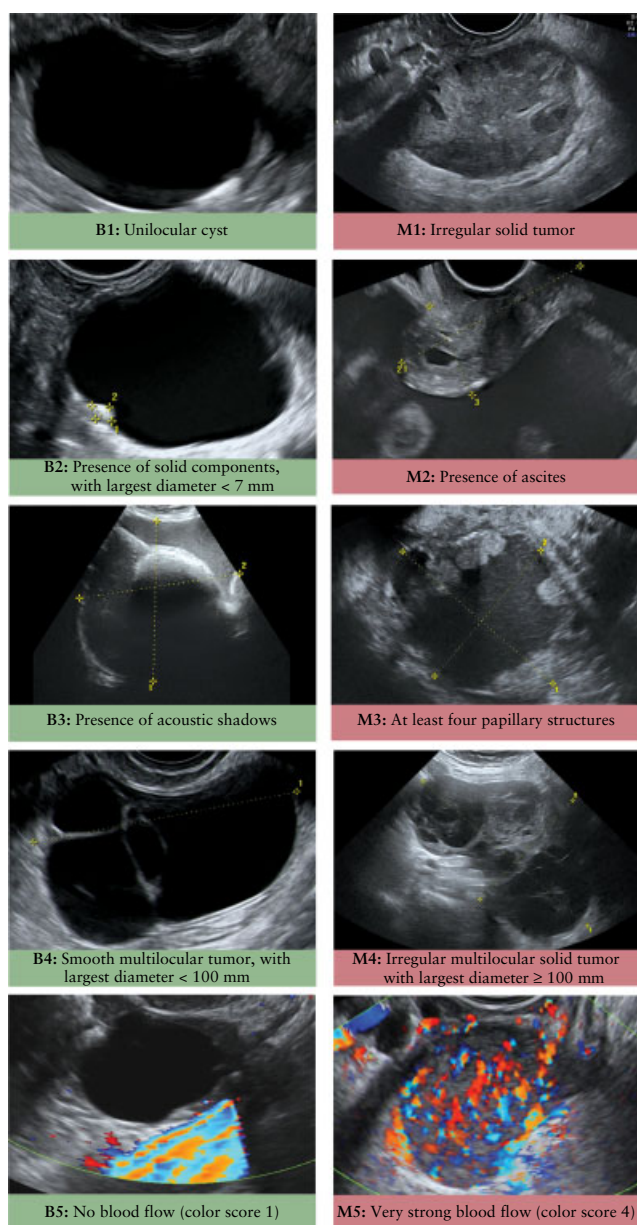


Figure 3 Ultrasound features used in the International Ovarian Tumor Analysis (IOTA) simple rules, illustrated by ultrasound images. B1–B5, benign features; M1–M5, malignant features.

serum CA 125 for discrimination between benign and malignant adnexal masses in the IOTA 1 dataset, subjective assessment performed significantly better⁴¹. This was independent of menopausal status, the specific histological diagnosis and the serum CA 125 cut-off level used. We have also shown that adding information on the serum CA 125 level to subjective assessment of ultrasound findings does not improve the diagnostic performance of experienced ultrasound examiners, irrespective of the diagnostic confidence of the examiner⁴². Upon further scrutiny, we found that single fixed cut-off values for serum CA 125 levels can reliably discriminate only between Stage II–IV invasive tumors and benign tumors that are not an abscess or endometrioma⁴³. For all other types of tumor, serum CA 125 values overlap considerably.

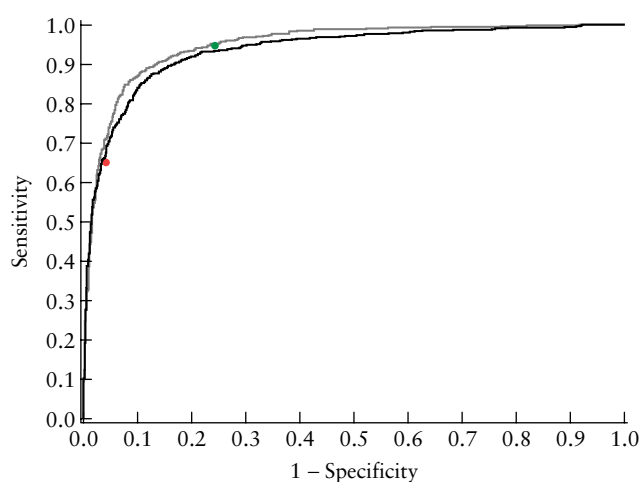


Figure 4 Receiver–operating characteristics (ROC) curves for logistic regression model 1 (LR1; —) and logistic regression model 2 (LR2; - - -), with ROC points for the simple rules superimposed. The red (simple rules: benign/inconclusive vs malignant) and green (simple rules: benign vs inconclusive/malignant) ROC points represent situations in which the ‘inconclusive tumors’ are classified as either benign or malignant, respectively. The results were obtained using pooled data ($n = 2445$) from International Ovarian Tumor Analysis (IOTA) Phases 1b and 2 ($n = 2445$).

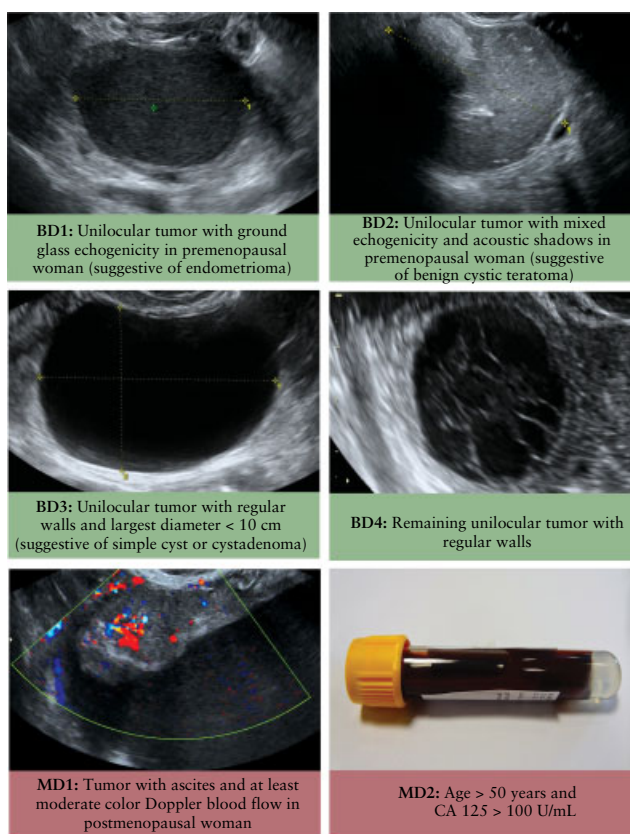


Figure 5 International Ovarian Tumor Analysis (IOTA) ‘easy descriptors’ illustrated by ultrasound images. BD1–BD4, benign descriptors; MD1–MD2, malignant descriptors.

On the other hand, if malignancy is suspected, preoperative measurement of the serum CA 125 level may be useful for postoperative follow up using serum

CA 125 as a biomarker to detect progression during chemotherapy⁴⁴.

A great deal of research has been carried out to identify new biomarkers that can be used together with, or instead of, CA 125. Human epididymis secretory protein-4 (HE4) was found to be complementary to CA 125 for the detection of malignant disease. An initial report suggested that combining these two biomarkers increased overall sensitivity and specificity compared with the use of a single biomarker⁴⁵. Based on these preliminary findings, HE4 was combined with CA 125 and menopausal status to form the Risk of Ovarian Malignancy Algorithm (ROMA)⁴⁶. This algorithm showed significantly higher sensitivity for epithelial ovarian cancer than did the original RMI at a fixed specificity level of 75%⁴⁷. Numerous studies have validated this new diagnostic method with contradicting results. Some reports have confirmed the accuracy of ROMA^{48–52}, whilst others have found that ROMA did not outperform established diagnostic tests that incorporate CA 125 or HE4 for detection of cancer in an ovarian mass^{53–55}. In a study by Van Gorp *et al.*⁵⁶, which used data from the IOTA database, it was demonstrated that expert subjective assessment of ultrasound findings was superior to ROMA for distinguishing benign from malignant adnexal masses.

FUTURE DIRECTIONS FOR THE IOTA PROJECT

Predicting subtypes of ovarian malignancy: use of polytomous models

Most ultrasound-based prediction models used in clinical practice have a dichotomous outcome (i.e. cancer or no cancer). However, different subtypes of malignancy (metastatic, primary invasive or borderline malignancy) are managed differently with implications in relation to type of surgery, length of hospitalization and financial cost. To achieve a more fine-tuned categorization, we developed polytomous (or multiclass) prediction models using logistic regression on the IOTA Phase 1 data to characterize ovarian pathology as benign, borderline malignant, primary invasive or metastatic⁵⁷. The polytomous model had a test performance (AUC=0.95) similar to that of LR1 and LR2 for discriminating between benign and malignant adnexal masses. In addition, the model was able to distinguish benign tumors from borderline, primary invasive and metastatic tumors (AUCs of 0.91, 0.95 and 0.93, respectively). These data are promising, but temporal and external validation revealed that the polytomous model could not discriminate between primary and metastatic invasive tumors⁵⁷. To address this limitation we are now focusing on developing more robust multiclass models using a larger dataset (IOTA 1, 1b and 2). We aim to also differentiate between Stage I and Stage II–IV primary invasive malignancies.

Presentation of results of risk prediction models to facilitate interpretation

Two main approaches for the prediction of malignancy have emerged from the IOTA study. The first uses mathematical models that provide risk estimates. However, it is not straightforward to understand exactly how mathematical models work to obtain risk estimates. This is particularly the case when advanced state-of-the-art algorithms (e.g. support vector machines) are used, but even in the case of a logistic regression prediction model a detailed understanding is difficult. The second approach is based on simple rules. Such rules are appealing to clinicians as their working mechanism is clearer. However, such rule-based approaches are more susceptible to oversimplification, even though validation demonstrated very good test performance for adnexal mass characterization³⁷.

Using the IOTA dataset as a case study, we combined the advantages of advanced risk modeling techniques with the interpretability and attractiveness of simple scoring systems into the novel Interval Coded Scoring (ICS) system methodology for the development of scoring systems⁵⁸. The ICS system combines elements from state-of-the-art techniques such as support vector machines, splines and L_1 regularization⁵⁹ but presents the results as color bars that are easy to interpret. The ICS can be implemented in software packages, smartphone applications or on paper, which could be useful for bedside medicines⁵⁸. The color-based representation is suitable for increasing interpretability by both the patient and the doctor, and might improve informed and shared decision making^{60,61}. Although highly promising, the ICS method should be successfully applied to several different diagnostic problems in order to demonstrate its ability to combine good performance with interpretability and user-friendliness.

Assessment of second-stage tests

In IOTA Phase 3 (2009–2012) we focus on improving diagnosis in ‘difficult adnexal tumors’^{62,63} by adding second-stage tests to conventional gray-scale and Doppler ultrasound. These include evaluation of the vascular tree of tumors using three-dimensional power Doppler⁶⁴ and novel biomarkers, such as HE-4.

Performance of IOTA rules and models in the hands of examiners with different training and levels of experience

In a validation study, Nunes *et al.*⁶⁵ demonstrated that LR2 retains its diagnostic performance (AUC=0.93) in the hands of a less-experienced operator. In IOTA Phase 4 (2012–2013), we will evaluate the performance of the IOTA model LR2, the simple rules, the easy descriptors and the RMI as first-stage tests in the hands of ultrasound examiners with less experience than those participating in the published IOTA studies and with different types of ultrasound training (sonographers and doctors).

Long-term behavior of ovarian masses managed expectantly

In IOTA Phase 5 (2012–2017), the main goal is to study the natural history of at least 3000 ovarian masses with benign ultrasound morphology managed conservatively. This should establish the risk of complications such as torsion and cyst rupture, the need for surgical intervention and give an indication of the risk of malignant transformation. These results will enable us to suggest algorithms for suitable management – expectant management or surgery – of all types of adnexal pathology.

SUMMARY

Since the start of the IOTA project in 1999 we have examined a large number of patients with ovarian masses in IOTA Phases 1, 1b and 2 using ultrasonography with a rigorously defined protocol to prospectively collect detailed information about these tumors. The study has been carried out in over 20 different centers in different countries, in both district and oncology referral hospitals. The results have been consistent and so are likely to be robust and generalizable. To date, the IOTA study is the largest study in the literature on ultrasound diagnosis of ovarian pathology.

We have found that pattern recognition of the ultrasound features of an ovarian mass by an experienced clinician is the best way of characterizing ovarian pathology. Papillary projections are characteristic of borderline tumors and Stage I primary invasive epithelial ovarian cancer. A small proportion of solid tissue at ultrasound examination makes a malignant mass more likely to be a borderline tumor or a Stage I epithelial ovarian cancer than an advanced ovarian cancer, a metastasis or a rare type of tumor⁶⁶. Our data suggest that information on serum CA 125 does not improve the diagnostic performance of subjective assessment by experienced ultrasound operators, and that it is not a necessary variable in multivariable prediction models developed to help classify ovarian masses as benign or malignant.

Two main approaches to the classification of ovarian masses have been developed using the IOTA database (Figure 6). The first uses risk-prediction models. The models LR1 and LR2, discussed above, have shown very good test performance on external validation. The second approach involves the use of either simple ultrasound-based rules or ‘easy descriptors’. These are based on ultrasound features that are virtually pathognomonic of either a benign or a malignant mass. The simple rules have been shown to apply to over 75% of masses, and have been successfully externally validated and taken up in a national protocol³⁸. The use of the easy descriptors has yet to undergo external validation. For masses to which the simple rules do not apply it seems best to refer the patient for examination by an experienced ultrasound examiner. Such an approach is also applicable to risk-prediction models, for example using the LR2-based

triage protocol described earlier³⁶. Patients classified as being at intermediate risk of malignancy (i.e. LR2 risk of a malignant tumor is between 5 and 25%)³⁶ may be referred to an experienced examiner.

We have also created multiclass models⁵⁷ that can assist clinicians to predict multiple outcomes in patients with an ovarian mass. Multiclass models can distinguish between benign, borderline, primary invasive and metastatic malignant disease and might play a central future role in the decision-making process. Their use is a logical extension of the current evolution toward an ‘individualized’ or ‘personalized’ approach to cancer treatment and healthcare in general.

RECOMMENDATIONS FOR CLINICAL PRACTICE BASED ON THE FINDINGS OF THE IOTA STUDY

These recommendations are based on validation of non-IOTA tests for classifying ovarian pathology and of tests developed from data in the IOTA study. Both non-IOTA tests and IOTA tests have undergone external validation in 12 different IOTA centers^{28,29,37}. LR2 and the simple rules have also been validated outside the frame of the IOTA study^{65,67}.

1. The IOTA simple rules can be used to classify 75% of all adnexal masses as benign or malignant. The main advantage of these rules is their ease of use. This should make it easy to implement them in clinical practice.
2. A two-step strategy with referral to a specialist in gynecological ultrasonography for subjective assessment of masses unclassifiable using the simple rules has excellent diagnostic test performance on external validation.
3. A viable alternative to the simple rules is the logistic regression model, LR2. This model provides a benefit over the simple rules in that it can be applied to an entire tumor population and produces a risk estimate for ovarian malignancy. The latter is a key element in personalized healthcare and shared decision-making. As with the simple rules, some patients can be referred to a specialist if the risk estimate is considered as intermediate or inconclusive.
4. LR2 or the simple rules should be adopted as the principal test to characterize masses as benign or malignant in premenopausal women because both perform much better than the RMI in premenopausal women.
5. Measurements of serum CA 125 are not necessary for the characterization of ovarian pathology in premenopausal women and are unlikely to improve the performance of experienced ultrasound examiners, even in the postmenopausal group. Ongoing studies are investigating its value in less-experienced hands.

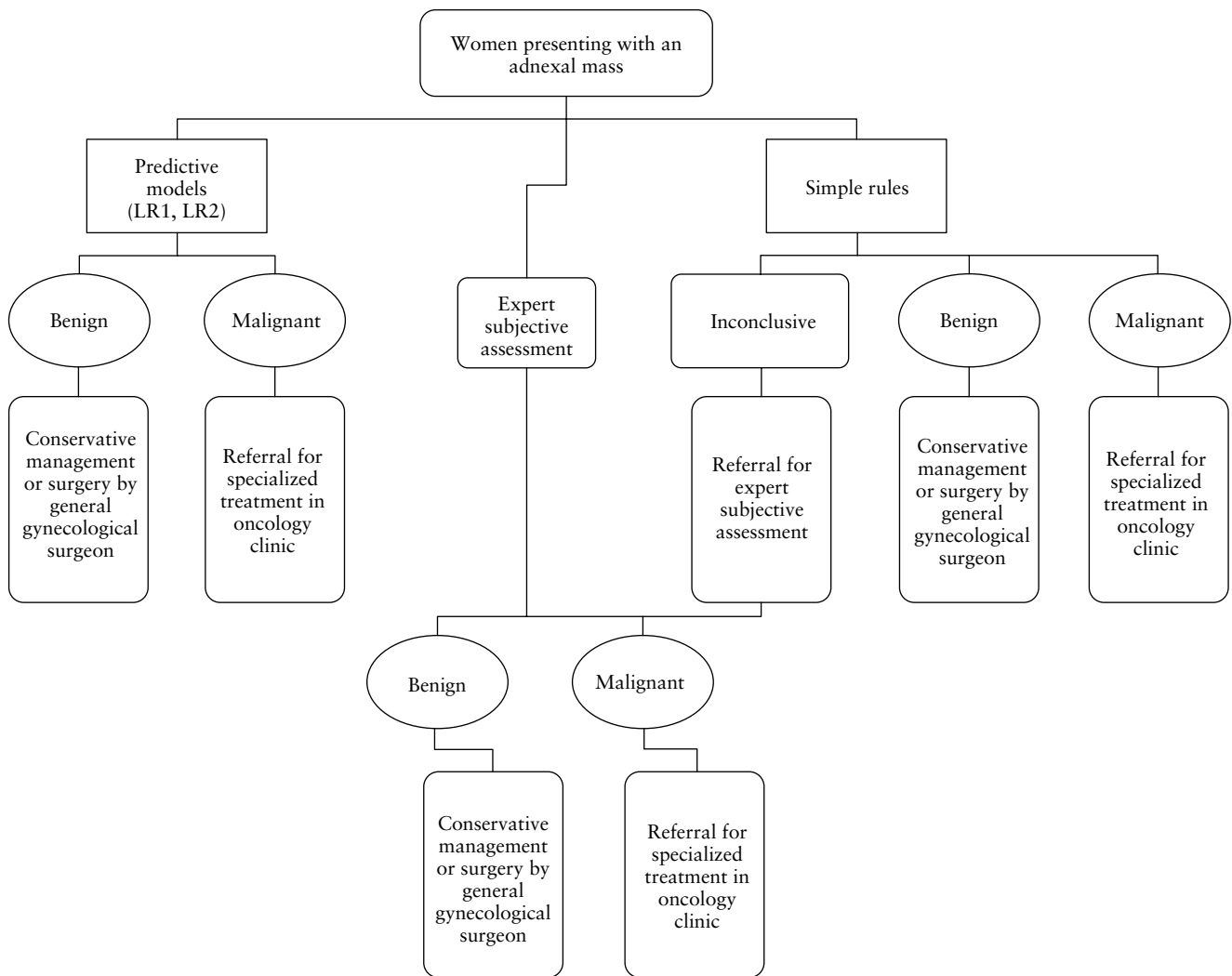


Figure 6 Flow chart showing different approaches using ultrasonography in the assessment of women with adnexal masses to estimate risk of malignancy, incorporating the evidence base of the International Ovarian Tumor Analysis (IOTA) study. LR1, IOTA logistic regression model 1; LR2, IOTA logistic regression model 2.

ACKNOWLEDGMENTS

The authors thank all participating centers, the principal investigators and the study participants for their contributions.

Recruitment Centers

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 (China); King's College Hospital London (UK); Università 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 Istituto Nazionale dei Tumori, Fondazione Pascale, Napoli (Naples B, Italy).

IOTA Steering Committee

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 (in alphabetical order)

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, 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, Rome, Italy; D. Timmerman, Leuven, Belgium; Diego Trio, Milan, Italy; L. Valentin, Malmö, Sweden; and C. Van Holsbeke, Genk, Belgium.

Details of ethics approval

The IOTA study protocol was approved by the Central Ethics Committee for Clinical Studies at the University Hospitals KU Leuven, Belgium, and by the Local Ethics Committee at each recruitment center.

Funding

The IOTA study 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: grant 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 the IOTA 5 project we received a project grant from the FWO (grant G049312N). Tom Bourne is supported by the Imperial Healthcare NHS Trust NIHR Biomedical Research Center.

REFERENCES

- Menon U, Gentry-Maharaj A, Hallett R, Ryan A, Burnell M, Sharma A, Lewis S, Davies S, Philpott S, Lopes A, Godfrey K, Oram D, Herod J, Williamson K, Seif MW, Scott I, Mould T, Woolas R, Murdoch J, Dobbs S, Amsó NN, Leeson S, Cruickshank D, McGuire A, Campbell S, Fallowfield L, Singh N, Dawney A, Skates SJ, Parmar M, Jacobs I. Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncol* 2009; **10**: 327–340.
- Carley ME, Klingee CJ, Bebhart 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.
- Vergote I, De Brabanter J, Fyles A, Bertelsen K, Einhorn N, Sevelde P, Gore ME, Kaern J, Verrelst H, Sjövall K, Timmerman D, Vandewalle J, Van Gramberen M, Tropé CG. Prognostic importance of degree of differentiation and cyst rupture in stage I invasive epithelial ovarian carcinoma. *Lancet* 2001; **357**: 176–182.
- Paulsen T, Kjaerheim K, Kaern J, Tretli S, Tropé C. Improved short-term survival for advanced ovarian, tubal, and peritoneal cancer patients operated at teaching hospitals. *Int J Gynecol Cancer* 2006; **16** (Suppl. 1): 11–17.
- Engelen MJ, van der Zee AG, de Vries EG, Willemsse PH. Debulking surgery for ovarian epithelial cancer performed by a gynaecological oncologist improved survival compared with less specialised surgeons. *Cancer Treat Rev* 2006; **32**: 320–323.
- Earle CC, Schrag D, Neville BA, Yabroff KR, Topor M, Fahey A, Trimble EL, Bodurka DC, Bristow RE, Carney M, Warren JL. Effect of surgeon specialty on process of care and outcomes for ovarian cancer patients. *J Natl Cancer Inst* 2006; **98**: 172–180.
- DePriest PD, Shenson D, Fried A, Hunter JE, Andrews SJ, Gallion HH, Pavlik EJ, Kryscio RJ, van Nagell JR Jr. A morphology index based on sonographic findings in ovarian cancer. *Gynecol Oncol* 1993; **51**: 7–11.
- 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**: 192–197.
- Finkler NJ, Benacerraf B, Lavin PT, Wojciechowski C, Knapp RC. Comparison of serum CA 125, clinical impression, and ultrasound in the preoperative evaluation of ovarian masses. *Obstet Gynecol* 1988; **72**: 659–664.
- Granberg S, Norström A, Wikland M. Tumors in the lower pelvis as imaged by vaginal sonography. *Gynecol Oncol* 1990; **37**: 224–229.
- Sassone AM, Timor-Tritsch IE, Artner A, Westhoff C, Warren WB. Transvaginal sonographic characterization of ovarian disease: evaluation of a new scoring system to predict ovarian malignancy. *Obstet Gynecol* 1991; **78**: 70–76.
- Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinski JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *Br J Obstet Gynaecol* 1990; **97**: 922–929.
- Timmerman D, Bourne TH, Taylor A, Collins WP, Verrelst H, Vandenberghe K, Vergote I. 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.
- Timmerman D, Verrelst H, Bourne TH, De Moor B, Collins WP, Vergote I, Vandewalle J. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. *Ultrasound Obstet Gynecol* 1999; **13**: 17–25.
- Lu C, Van Gestel T, Suykens JA, Van Huffel S, Vergote I, Timmerman D. Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. *Artif Intell Med* 2003; **28**: 281–306.
- Geminì P, Kruitwagen R, Bremer GL, Cnossen J, Mol BWJ. The accuracy of risk scores in predicting ovarian malignancy – a systematic review. *Obstet Gynecol* 2009; **113**: 384–394.
- Dearking AC, Aletti GD, McGree ME, Weaver AL, Sommerfield MK, Cliby WA. How relevant are ACOG and SGO guidelines for referral of adnexal mass? *Obstet Gynecol* 2007; **110**: 841–848.
- Ueland FR, Desimone CP, Seamon LG, Miller RA, Goodrich S, Podzielinski I, Sokoll L, Smith A, van Nagell JR Jr, Zhang Z. Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. *Obstet Gynecol* 2011; **117**: 1289–1297.
- Timmerman D, Van Calster B, Vergote I, Van Hoorde K, Van Gorp T, Valentin L, Bourne T. Performance of the American College of Obstetricians and Gynecologists’ ovarian tumor referral guidelines with a multivariate index assay. *Obstet Gynecol* 2011; **118**: 1179–1181.
- Mol BW, Boll D, De Kanter M, Heintz AP, Sijmons EA, Oei SG, Bal H, Brölmann HA. Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. *Gynecol Oncol* 2001; **80**: 162–167.
- 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: 192–197.
22. 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.
 23. 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; International Ovarian Tumor Analysis Group. 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: 4440–4447.
 24. 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.
 25. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group. *Ultrasound Obstet Gynecol* 2000; 16: 500–505.
 26. 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.
 27. Van Holsbeke C, Van Calster B, Testa AC, Domali E, Lu C, Van Huffel S, Valentin L, Timmerman D. Prospective internal validation of mathematical models to predict malignancy in adnexal masses: results from the international ovarian tumor analysis study. *Clin Cancer Res* 2009; 15: 684–691.
 28. Van Holsbeke C, Van Calster B, Bourne T, Ajossa S, Testa AC, Guerriero S, Fruscio R, Lissoni AA, Czekierdowski A, Savelli L, Van Huffel S, Valentin L, Timmerman D. External validation of diagnostic models to estimate the risk of malignancy in adnexal masses. *Clin Cancer Res* 2012; 18: 815–825.
 29. 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.
 30. Van Calster B, Nabney I, Timmerman D, Van Huffel S. The Bayesian approach: a natural framework for statistical modeling. *Ultrasound Obstet Gynecol* 2007; 29: 485–488.
 31. Van Calster B, Timmerman D, Lu C, Suykens JA, Valentin L, Van Holsbeke C, Amant F, Vergote I, Van Huffel S. Preoperative diagnosis of ovarian tumors using Bayesian kernel-based methods. *Ultrasound Obstet Gynecol* 2007; 29: 496–504.
 32. Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, Paladini D, Van Calster B, Vergote I, Van Huffel S, Valentin L. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol* 2008; 31: 681–690.
 33. Ameye L, Valentin L, Testa AC, Van Holsbeke C, Domali E, Van Huffel S, Vergote I, Bourne T, Timmerman D. A scoring system to differentiate malignant from benign masses in specific ultrasound-based subgroups of adnexal tumors. *Ultrasound Obstet Gynecol* 2009; 33: 92–101.
 34. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; 338: b605.
 35. König IR, Malley JD, Weimar C, Diener HC, Ziegler A; German Stroke Study Collaboration. Practical experiences on the necessity of external validation. *Stat Med* 2007; 26: 5499–5511.
 36. Van Calster B, Timmerman D, Valentin L, McIndoe A, Ghaem-Maghani S, Testa AC, Vergote I, Bourne T. 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; 119: 662–671.
 37. Timmerman D, Ameye L, Fischerova D, Epstein E, Melis GB, Guerriero S, Van Holsbeke C, Savelli L, Fruscio R, Lissoni AA, Testa AC, Veldman J, Vergote I, Van Huffel S, Bourne T, Valentin L. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. *BMJ* 2010; 341: c6839.
 38. RCOG guideline: management of suspected ovarian masses in premenopausal women. December 2011. (Green top 62). <http://www.rcog.org.uk> [Accessed 2 December 2011].
 39. Ameye L, Timmerman D, Valentin L, Paladini D, Zhang J, Van Holsbeke C, Lissoni AA, Savelli L, Veldman J, Testa AC, Amant F, Van Huffel S, Bourne T. Clinically oriented three-step strategy to the assessment of adnexal pathology. *Ultrasound Obstet Gynecol* 2012; 40: 582–591.
 40. Timmerman D, Van Calster B, Jurkovic D, Valentin L, Testa AC, Bernard JP, Van Holsbeke C, Van Huffel S, Vergote I, Bourne T. 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.
 41. Van Calster B, Timmerman D, Bourne T, Testa AC, Van Holsbeke C, Domali E, Jurkovic D, Neven P, Van Huffel S, Valentin L. Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125. *J Natl Cancer Inst* 2007; 99: 1706–1714.
 42. Valentin L, Jurkovic D, Van Calster B, Testa A, Van Holsbeke C, Bourne T, Vergote I, Van Huffel S, Timmerman D. 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: 345–354.
 43. Van Calster B, Valentin L, Zhang J, Jurkovic D, Lissoni AA, Testa AC, Czekierdowski A, Fischerová D, Domali E, Van de Putte G, Vergote I, Van Huffel S, Bourne T, Timmerman D. A novel approach to predict the likelihood of specific ovarian tumor pathology based on serum CA-125: a multicenter observational study. *Cancer Epidemiol Biomarkers Prev* 2011; 20: 2420–2428.
 44. Alexandre J, Brown C, Coeffic D, Raban N, Pfisterer J, Mäenpää J, Chalchal H, Fitzharris B, Volgger B, Vergote I, Pisano C, Ferrero A, Pujade-Lauraine E. CA-125 can be part of the tumour evaluation criteria in ovarian cancer trials: experience of the GCIG CALYPSO trial. *Br J Cancer* 2012; 106: 633–637.
 45. Moore RG, Brown AK, Miller MC, Skates S, Allard WJ, Verch T, Steinhoff M, Messerlian G, DiSilvestro P, Granai CO, Bast RC Jr. The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass. *Gynecol Oncol* 2008; 108: 402–408.
 46. Moore RG, McMeekin DS, Brown AK, DiSilvestro P, Miller MC, Allard WJ, Gajewski W, Kurman R, Bast RC Jr, Skates SJ. A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009; 112: 40–46.
 47. Moore RG, Jabre-Raughley M, Brown AK, Robison KM, Miller MC, Allard WJ, Kurman RJ, Bast RC, Skates SJ. Comparison of a novel multiple marker assay vs the Risk of Malignancy Index for the prediction of epithelial ovarian cancer in patients with a pelvic mass. *Am J Obstet Gynecol* 2010; 203: 228.e1–6.
 48. Bandiera E, Romani C, Specchia C, Zanotti L, Galli C, Ruggeri G, Tognon G, Bignotti E, Tassi RA, Odicino F, Caimi L, Sartori E, Santin AD, Pecorelli S, Ravaggi A. Serum human epididymis protein 4 and risk for ovarian malignancy algorithm as new diagnostic and prognostic tools for epithelial ovarian cancer management. *Cancer Epidemiol Biomarkers Prev* 2011; 20: 2496–2506.

49. Kim YM, Whang DH, Park J, Kim SH, Lee SW, Park HA, Ha M, Choi KH. Evaluation of the accuracy of serum human epididymis protein 4 in combination with CA125 for detecting ovarian cancer: a prospective case-control study in a Korean population. *Clin Chem Lab Med* 2011; **49**: 527–534.
50. Lenhard M, Stieber P, Hertlein L, Kirschenhofer A, Fürst S, Mayr D, Nagel D, Hofmann K, Krockner K, Burges A. The diagnostic accuracy of two human epididymis protein 4 (HE4) testing systems in combination with CA125 in the differential diagnosis of ovarian masses. *Clin Chem Lab Med* 2011; **49**: 2081–2088.
51. Molina R, Escudero JM, Augé JM, Filella X, Foj L, Torné A, Lejarcegui J, Pahisa J. HE4 a novel tumour marker for ovarian cancer: comparison with CA 125 and ROMA algorithm in patients with gynaecological diseases. *Tumor Biol* 2011; **32**: 1087–1095.
52. Ruggeri G, Bandiera E, Zanotti L, Belloli S, Ravaggi A, Romani C, Bignotti E, Tassi RA, Tognon G, Galli C, Caimi L, Pecorelli S. HE4 and epithelial ovarian cancer: comparison and clinical evaluation of two immunoassays and a combination algorithm. *Clin Chim Acta* 2011; **412**: 1447–1453.
53. Van Gorp T, Cadron I, Despierre E, Daemen A, Leunen K, Amant F, Timmerman D, De Moor B, Vergote I. HE4 and CA125 as a diagnostic test in ovarian cancer: prospective validation of the Risk of Ovarian Malignancy Algorithm. *Br J Cancer* 2011; **104**: 863–870.
54. Montagnana M, Danese E, Ruzzenente O, Bresciani V, Nuzzo T, Gelati M, Salvagno GL, Franchi M, Lippi G, Guidi GC. The ROMA (Risk of Ovarian Malignancy Algorithm) for estimating the risk of epithelial ovarian cancer in women presenting with pelvic mass: is it really useful? *Clin Chem Lab Med* 2011; **49**: 521–525.
55. Jacob F, Meier M, Caduff R, Goldstein D, Pochechueva T, Hacker N, Fink D, Heinzelmann-Schwarz V. No benefit from combining HE4 and CA125 as ovarian tumor markers in a clinical setting. *Gynecol Oncol* 2011; **121**: 487–491.
56. Van Gorp T, Veldman J, Van Calster B, Cadron I, Leunen K, Amant F, Timmerman D, Vergote I. Subjective assessment by ultrasound is superior to the risk of malignancy index (RMI) or the risk of ovarian malignancy algorithm (ROMA) in discriminating benign from malignant adnexal masses. *Eur J Cancer* 2012; **48**: 1649–1656.
57. Van Calster B, Valentin L, Van Holsbeke C, Testa AC, Bourne T, Van Huffel S, Timmerman D. Polytomous diagnosis of ovarian tumors as benign, borderline, primary invasive or metastatic: development and validation of standard and kernel-based risk prediction models. *BMC Med Res Methodol* 2010; **10**: 96.
58. Van Belle VM, Van Calster B, Timmerman D, Bourne T, Bottomley C, Valentin L, Neven P, Van Huffel S, Suykens JA, Boyd S. A mathematical model for interpretable clinical decision support with applications in gynecology. *PLoS One* 2012; **7**: e34312.
59. Hastie T, Tibshirani R. *Generalized Additive Models*. Chapman and Hall: 1990.
60. Stiggebout AM, Van der Weijden T, De Wit MP, Frosch D, Légaré F, Montori VM, Trevena L, Elwyn G. Shared decision making: really putting patients at the center of healthcare. *BMJ* 2012; **344**: e256.
61. Barry MJ, Edgman-Levitan S. Shared decision making – pinnacle of patient-centered care. *N Engl J Med* 2012; **366**: 780–781.
62. Valentin L, Ameye L, Jurkovic D, Metzger U, Lécure 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.
63. 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.
64. 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.
65. Nunes N, Yazbek J, Ambler G, Hoo W, Naftalin J, Jurkovic D. A prospective evaluation of the IOTA Logistic Regression Model (LR2) for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol* 2012; **40**: 355–359.
66. Valentin L, Ameye L, Testa A, Lécure F, Bernard JP, Paladini D, Van Huffel S, Timmerman D. Ultrasound characteristics of different types of adnexal malignancies. *Gynecol Oncol* 2006; **102**: 41–48.
67. Fathallah K, Huchon C, Bats AS, Metzger U, Lefrère-Belda MA, Bensaid C, Lécure F. External validation of simple ultrasound rules of Timmerman on 122 ovarian tumors. *Gynecol Obstet Fertil* 2011; **39**: 477–481.



This article has been selected for Journal Club.

A slide presentation, prepared by Dr Ligita Jokubkiene, one of UOG's Editors for Trainees, is available online.