Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

Clinical Cancer Research

External Validation of Diagnostic Models to Estimate the Risk of Malignancy in Adnexal Masses

Caroline Van Holsbeke, Ben Van Calster, Tom Bourne, et al.

Clin Cancer Res 2012;18:815-825. Published OnlineFirst November 23, 2011.

Updated version	Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-11-0879
Supplementary Material	Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2011/11/23/1078-0432.CCR-11-0879.DC1.html

Cited Articles	This article cites by 40 articles, 10 of which you can access for free at: http://clincancerres.aacrjournals.org/content/18/3/815.full.html#ref-list-1
Citing articles	This article has been cited by 3 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/18/3/815.full.html#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.

Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

Imaging, Diagnosis, Prognosis

Clinical Cancer Research

External Validation of Diagnostic Models to Estimate the Risk of Malignancy in Adnexal Masses

Caroline Van Holsbeke[1,4], Ben Van Calster[1,2], Tom Bourne[1,5], Silvia Ajossa[6], Antonia C. Testa[7], Stefano Guerriero[6], Robert Fruscio[8], Andrea Alberto Lissoni[8], Artur Czekierdowski[10], Luca Savelli[9], Sabine Van Huffel[2,3], Lil Valentin[11], and Dirk Timmerman[1]

Abstract

Purpose: To externally validate and compare the performance of previously published diagnostic models developed to predict malignancy in adnexal masses.

Experimental Design: We externally validated the diagnostic performance of 11 models developed by the International Ovarian Tumor Analysis (IOTA) group and 12 other (non-IOTA) models on 997 prospectively collected patients. The non-IOTA models included the original risk of malignancy index (RMI), three modified versions of the RMI, six logistic regression models, and two artificial neural networks. The ability of the models to discriminate between benign and malignant adnexal masses was expressed as the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and likelihood ratios (LR^+, LR^-).

Results: Seven hundred and forty-two (74%) benign and 255 (26%) malignant masses were included. The IOTA models did better than the non-IOTA models (AUCs between 0.941 and 0.956 vs. 0.839 and 0.928). The difference in AUC between the best IOTA and the best non-IOTA model was 0.028 [95% confidence interval (CI), 0.011–0.044]. The AUC of the RMI was 0.911 (difference with the best IOTA model, 0.044; 95% CI, 0.024–0.064). The superior performance of the IOTA models was most pronounced in premenopausal patients but was also observed in postmenopausal patients. IOTA models were better able to detect stage I ovarian cancer.

Conclusion: External validation shows that the IOTA models outperform other models, including the current reference test RMI, for discriminating between benign and malignant adnexal masses. *Clin Cancer Res; 18(3); 815–25. ©2011 AACR.*

Authors' Affiliations: [1]Department of Obstetrics and Gynaecology, University Hospitals Leuven; [2]Department of Electrical Engineering, ESAT-SCD, [3]IBBT-K.U.Leuven Future Health Department, Katholieke Universiteit Leuven, Leuven; [4]Department of Obstetrics and Gynaecology, Ziekenhuis Oost-Limburg, Genk, Belgium; [5]Department of Obstetrics and Gynaecology, Imperial College London, London, United Kingdom; [6]Department of Obstetrics and Gynecology, Ospedale San Giovanni di Dio, University of Cagliari, Cagliari; [7]Istituto di Clinica Ostetrica e Ginecologica, Università Cattolica del Sacro Cuore, Roma; [8]Clinica Ostetrica e Ginecologica, Ospedale S. Gerardo, Università di Milano Bicocca, Monza; [9]Reproductive Medicine Unit, Department of Obstetrics and Gynecology Policlinico S. Orsola-Malpighi, Bologna, Italy; [10]1st Department of Gynecological Oncology and Gynecology of the Medical University in Lublin, Lublin, Poland; and [11]Department of Obstetrics and Gynaecology, Malmö University Hospital, Lund University, Lund, Sweden

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

C. Van Holsbeke and B. Van Calster both contributed equally to this work.

Corresponding Author: Caroline Van Holsbeke, Department of Obstetrics and Gynaecology, Z.O.L. Genk, Belgium, Schiepse Bos 6, B-3600 Genk, Belgium. Phone: 32-89-327521 (office), 32-89-327524 (secretary); Fax: 32-89-327920 (office); E-mail: caroline.van.holsbeke@skynet.be

doi: 10.1158/1078-0432.CCR-11-0879

©2011 American Association for Cancer Research.

Introduction

The most important role of the sonologist examining patients with an adnexal mass is to distinguish between benign and malignant masses to optimally triage patients. Benign masses can be managed conservatively or—if technically feasible—be removed by laparoscopy. However, patients with masses that are presumed to be malignant should be referred to gynecologic oncologists for proper staging and debulking (1–3). A mistake in either direction could worsen the patient's prognosis or create unnecessary anxiety, cost, and extensive surgery.

The sonologist can use different strategies to assess the risk of malignancy but there is still no consensus as to the optimal approach.

Subjective assessment of the sonographic characteristics of the mass is an excellent method to discriminate between benign and malignant but only in the hands of experienced ultrasound examiners (4, 5).

The American College of Obstetricians and Gynecologists (ACOG) uses guidelines based on demographic,

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.

Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

Translational Relevance

A correct preoperative estimation of the benign or malignant character of an adnexal mass remains crucial because it will determine the treatment strategy and also impact significantly on the prognosis of the patient. Subjective assessment of an adnexal mass by an experienced ultrasound examiner has been shown to be the optimal approach. Unfortunately, in daily clinical practice, it is impossible for every patient to be seen by an experienced ultrasound examiner. Therefore, to help less experienced ultrasound examiners predict the risk of malignancy, several mathematical models have been developed to replicate the performance of expert operators. In several countries, the risk of malignancy index (RMI) by Jacobs and colleagues is included in clinical guidelines for the assessment of adnexal pathology. A recent review by Geomini and colleagues compared the performance of several mathematical models and concluded that the RMI had the best performance. This article externally validates the prediction models developed by the International Ovarian Tumor Analysis (IOTA) group and compares their performance with that of the best performing models published in the literature. This is the largest study conducted in this area and it shows that the performance of the IOTA models is significantly better than the performance of both the RMI and the other models tested. If adopted, these models have the potential to improve the clinical management of patients with adnexal masses.

sonographic, and biochemical parameters for the referral of pre- and postmenopausal patients with an adnexal mass to a gynecologic oncology center. These guidelines reach a high sensitivity with regard to malignancy for postmenopausal patients (93%) but a low sensitivity for premenopausal patients (79%) and poor specificity for both groups (70% and 60% for pre-and postmenopausal patients, respectively; ref. 6).

Several mathematical models or scoring systems have been developed to be used for discrimination between benign and malignant adnexal masses with encouraging results (7–14). Unfortunately, when tested prospectively, most models did worse than in the studies where they had been created, and they did not achieve the performance of experienced sonologists using subjective assessment of the ultrasound image (4, 15–18). The Royal College of Obstetricians and Gynecologists (19) and a recent review article by Geomini and colleagues (20) recommend the risk of malignancy index (RMI; ref. 7) as the method of choice for characterizing ovarian pathology as benign or malignant. The RMI is a scoring system that is derived from a logistic regression formula that combines menopausal status with serum CA125 and ultrasound variables (multilocular cysts, solid areas, metastases, ascites, bilaterality; ref. 7).

Recently, the International Ovarian Tumor Analysis (IOTA) study group developed several mathematical models to calculate the individual risk of malignancy in an adnexal mass. The details of the IOTA studies and the ultrasound protocol have been published previously (16, 21). The IOTA studies are international multicenter studies where women with adnexal masses are examined with ultrasound following a standardized research protocol with the aim of developing and validating mathematical models to estimate the risk of malignancy in adnexal masses. Patients with an adnexal mass were examined by an experienced ultrasound examiner following a standardized protocol. The ultrasound protocol required the assessment of more than 40 clinical and ultrasound variables. It was not mandatory to look for metastases, nor was it required to take a blood sample for the measurement of CA125 but this was encouraged. All patients had to undergo surgery within 4 months after the ultrasound examination to have the final histologic outcome and stage (in case of malignancy). Using this approach, the IOTA group aimed to overcome the problems of generalizability, sample size, and lack of standardized ultrasound criteria associated with previous studies. During the IOTA study phase 1, 1,066 patients were included in 9 centers from 5 countries (Belgium, Italy, Sweden, France, United Kingdom), and 11 mathematical models to estimate the risk of malignancy in adnexal masses were developed using different modeling techniques. Four different types of mathematical models were developed to evaluate which type of model did best (scoring system, logistic regression models, artificial neural networks (ANN), and vector machine models). On internal and temporal validation in the centers that previously developed the models, they had excellent diagnostic performance (16, 22–25). After temporal validation, all models had similar performance (AUCs between 0.945 and 0.950). Because logistic regression models are less complex and easier to use, the logistic regression model LR1 with 12 predictors and the simpler model LR2 with 6 predictors were from then on selected as the main IOTA models.

A crucial step before introducing a model in clinical practice involves external validation of performance in centers unrelated to where the models were developed (26, 27). Therefore, the aims of this study (IOTA phase 2) were to externally validate the ability of the IOTA models to discriminate between benign and malignant adnexal masses and to compare their performance with that of the RMI and other non-IOTA models developed to assess the risk of malignancy in adnexal masses before surgery.

Materials and Methods

Design and setting

This is a prospective cross-sectional diagnostic accuracy study of patients presenting with at least one persistent adnexal (ovarian, para-ovarian, or tubal) mass for which they were operated on. All patients underwent an ultrasound examination by one of the principal investigators who were all experienced in gynecologic ultrasound. In the

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.

Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

Table 1. Variables and cutoff values of non-IOTA and IOTA models

Reference	Type of model	Variables used	Cut-off
Non-IOTA models			
RMI Jacobs and colleagues (7)	Scoring system	(i) Menopausal status, (ii) CA125, (iii) multilocular cysts, (iv) solid areas, (v) metastases, (vi) ascites, and (vii) bilaterality	200
RMI 2 (8)	Scoring system	Same as RMI 1	200
RMI 3 (28)	Scoring system	Same as RMI 1	200
RMI 4 (29)	Scoring system	(i) Menopausal status, (ii) CA125, (iii) multilocular cysts, (iv) solid areas, (v) metastases, (vi) ascites, (vii) bilaterality, and (viii) largest diameter of lesion	450
Minaretzis and colleagues (10)	Logistic regression	(i) Tumor consistency (multilocular or solid = 1), (ii) bilaterality, (iii) maximum tumor diameter, and (iv) age	—
Tailor and colleagues (9)	Logistic regression	(i) Papillations, (ii) age, and (iii) time averaged maximum velocity in tumor vessels	50%
Timmerman and colleagues (a; ref. 11)	Logistic regression	(i) Color score, (ii) CA125, (iii) papillations, and (iv) menopausal status	25%
Timmerman and colleagues (b; ref. 12)	Logistic regression	(i) Papillations, (ii) irregular internal cyst wall, (iii) unilocular cyst, (iv) ascites, (v) bilaterality, (vi) menopausal status, and (vii) CA125	60%
Lu and colleagues (13)	Logistic regression	(i) CA125, (ii) papillations, (iii) solid tumor, (iv) color score of at least 3, (v) bilaterality, (vi) menopausal status, (vii) ascites, (viii) acoustic shadows, (ix) color score 4, and (x) irregular cyst wall	20%
Jokubkiene and colleagues (14)	Logistic regression	(i) Size of lesion (mean of 3 diameters), (ii) size of largest solid component (mean of 3 diameters), and (iii) any irregularity	12%
Timmerman and colleagues (12)	ANN (multilayer perceptron)	(i) Papillations, (ii) color score, (iii) menopausal status, and (iv) CA125	45%
Timmerman and colleagues (12)	ANN (multilayer perceptron)	(i) Papillations, (ii) smooth surface, (iii) unilocularity, (iv) ascites, (v) bilaterality, (vi) menopausal status, and (vii) CA125	60%
IOTA models			
LR1 (16)	Logistic regression	(i) Personal history of ovarian cancer, (ii) previous use of hormonal therapy, (iii) age, (iv) maximal diameter of the lesion, (v) pain, (vi) ascites, (vii) blood flow within papillary projection, (viii) solid tumor, (ix) maximal diameter of the largest solid component (bounded at 50 mm), (x) irregular internal cyst walls, (xi) acoustic shadows, and (xii) color score of intratumoral blood flow	10%
LR2 (16)	Logistic regression	(i) 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%
BMLP11-2a (24)	ANN (BMLP)	(i) Age, (ii) hormonal therapy, (iii) ascites, (iv) maximal diameter of the lesion, (v) irregular internal cyst wall, (vi) color score, (vii) blood flow within papillary projection, (viii) number of papillary projections, (ix) maximal diameter of the largest solid component, (x) multilocular solid tumor, and (xi) solid tumor	15%

(Continued on the following page)

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.

Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

Table 1. Variables and cutoff values of non-IOTA and IOTA models (Cont'd)

Reference	Type of model	Variables used	Cut-off
BMLP11-2b (24)	ANN (BMLP)	(i) Age, (ii) personal history of ovarian cancer, (iii) pelvic pain during examination, (iv) ascites, (v) maximal diameter of the lesion, (vi) irregular internal cyst wall, (vii) blood flow within papillary projection, (viii) acoustic shadowing, (ix) maximal diameter of the largest solid component, (x) unilocular tumor, and (xi) solid tumor	15%
BPER11 (24)	ANN (BPER)	(i) Age, (ii) personal history of ovarian cancer, (iii) ascites, (iv) maximal diameter of the lesion, (v) irregular internal cyst wall, (vi) color score, (vii) blood flow within papillary projection, (viii) acoustic shadowing, (ix) maximal diameter of the largest solid component, (x) unilocular tumor, and (xi) solid tumor	15%
LS-SVM lin (23)	SVM	(i) Maximal diameter of the solid component, (ii) maximal diameter of the ovary, (iii) age, (iv) color score 4, (v) multilocular solid lesion, (vi) ascites, (vii) personal history of ovarian cancer, (viii) blood flow within papillary projection, (ix) acoustic shadows, (x) previous use of hormonal therapy, (xi) irregular internal cyst walls, and (xii) tumor suspected to be of ovarian origin	15%
LS-SVM rbf (23)	SVM	Same variables as for IOTA LS-SVM lin	12%
LS-SVM add rbf (23)	SVM	Same variables as for IOTA LS-SVM lin	12%
RVM lin (23)	RVM	Same variables as for IOTA LS-SVM lin	20%
RVM rbf (23)	RVM	Same variables as for IOTA-LS SVM lin	15%
RVM add rbf (23)	RVM	Same variables as for IOTA LS-SVM lin	15%

Abbreviations: add rbf, additive radial basis function kernel; lin, linear kernel; rbf, radial basis function kernel.

case of bilateral adnexal masses, the mass with the most complex ultrasound morphology was included in our statistical analysis. If both masses had similar ultrasound morphology, the largest one or the one most easily accessible by ultrasound was included. Patients who were pregnant, refused transvaginal ultrasonography, or did not undergo surgical removal of the mass within 120 days after the ultrasound examination were excluded from analysis.

Between November 2005 and October 2007, 1,970 patients were examined according to the IOTA phase 2 protocol. Thirty-two of these were excluded because of the following reasons: 15 because surgical removal of the mass occurred more than 120 days after the ultrasound examination, 12 because they were pregnant at the time of the examination, 4 because of errors in data entry, and one because of protocol violation. Of the remaining 1,938 patients, 941 were new patients collected in 7 centers that had also participated in the IOTA phase 1 study and so were not available for external validation. Nine hundred and ninety-seven patients were collected in 12 new centers that had not participated in any IOTA study before, and these patients were used for the external validation of the models.

Seven of the new centers were oncological referral centers (465 patients, 47.6% of all patients), one was a referral center for ultrasonography (135 patients, 13.5%), and 4 were regional hospitals (387 patients, 38.8%). Ethical approval was obtained by the local committees of all centers.

The non-IOTA models

From a literature search, we identified 12 models to estimate the risk of malignancy in adnexal masses that contained variables for which information had been prospectively collected in the IOTA studies: the original RMI from the study of Jacobs and colleagues (7) and 3 variations of the RMI (8, 28, 29), 6 logistic regression models (9–14), and 2 ANNs (more specifically, multilayer perceptrons; ref. 12). The variables used in these models are listed in Table 1.

The IOTA models

During IOTA phase 1, we developed one subgroup scoring system (30), 2 logistic regression models (LR1 and LR2; ref. 16), 3 least-square support vector machines (LS-SVM; ref. 23), 3 relevance vector machines (RVM; ref. 23), 2 Bayesian multilayer perceptrons (BMLP; ref. 24), and 1 Bayesian perceptron model (BPER; ref. 24). The performance of the subgroup scoring system on external

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.

Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

validation will be analyzed and reported separately and is not included in the current study. The variables used in the IOTA models are listed in Table 1.

Subjective assessment

After completing the ultrasound examination, the ultrasound examiner classified each mass as benign or malignant according to his/her subjective assessment of the ultrasound images.

Statistical analysis

Because sensitivity and specificity are highly dependent on the cutoff point used, we regarded the area under the receiver operating characteristic (ROC) curve (AUC) as the most important measure to evaluate the discriminatory performance of the models. Confidence intervals (CI) for the AUC were computed using the logit-transform method (31, 32). CIs for the difference in AUCs were computed using the method of DeLong and colleagues (33). In addition, sensitivity and specificity as well as the positive and negative likelihood ratios (LR^+ and LR^-) were calculated for the cutoff points proposed in the original articles. These results were summarized into a single measure of test performance, the diagnostic odds ratio (DOR; refs. 34, 35). This equals LR^+ divided by LR^- and shows how much higher the odds of classifying the tumor as malignant are for patients with a malignant tumor than for patients with a benign tumor. For example, a DOR of 10 means that the odds of classifying the tumor as malignant is 10 times higher for patients with a malignant tumor than for patients with a benign tumor.

In this study, we evaluated 23 models, such that we specified *a priori* which comparisons between models were of interest to avoid excessive output. We decided to focus on comparing LR1 with LR2 and on comparing LR1, LR2, and the best IOTA model (if different from LR1 or LR2) with RMI and the best non-IOTA model (if different from the RMI). The main comparison between models involved the AUC on all data, and this was used to define the best IOTA and non-IOTA models. Secondary comparisons involved AUCs by menopausal status and the AUC for discriminating stage I ovarian cancer from benign tumors.

In the RMI, one of the variables is "presence of metastases." This was not a mandatory variable for the IOTA studies, and therefore information on metastases was sometimes missing (16%). To overcome this problem, we analyzed the differences in performance of the RMI when (i) the variable metastases were not used, (ii) the variable was imputed by 0 in case of missing information, and (iii) multiple imputation was used. The ROC curves were virtually identical for all 3 approaches. We decided to impute the variable by 0 in case of missing information because it seemed sensible to believe that in most cases where information on metastases was missing, there were no metastases.

Information on CA125 was missing in 27% of cases (272 of 997). Whenever it was missing, it was imputed using multiple imputation (36, 37) with predictive mean matching regression (38). This allowed us to use the whole database and to overcome selection bias. Missing values for CA125 were imputed 100 times, resulting in 100 complete data sets. Models using CA125 were applied on each complete data set, and results were combined using standard multiple imputation methodology to obtain final results that properly take into account the uncertainty of missing data (37). The 10 resulting ROC curves were averaged to plot a single curve. Additional details on our imputation procedure are described elsewhere (39).

Results

The data set included 742 (74%) benign and 255 (26%) malignant masses. The rate of malignancy was 40% (192 of 475) for the oncology centers, 8% (11 of 135) for the ultrasound referral centers, and 13% (52 of 387) for the regional centers. The final histologic outcome for all patients is listed in Table 2.

The performance of the models for the whole data set is presented in Table 3. LR1, the main IOTA logistic regression model using 12 variables, had the best performance with an AUC of 0.96 (95% CI, 0.94–0.97). All IOTA models had roughly similar performance with AUCs ranging between

Table 2. Demographic information and histologic diagnosis of all patients ($n = 997$)

Demographic information	
Age, median (range), y	44 (11–94)
Postmenopausal, *n* (%)	352 (35%)
Histologic diagnosis, *n* (%)	
All benign tumors	742 (74)
Endometrioma	208 (20.9%)
Dermoid/teratoma	130 (13.0%)
Serous cystadenoma	110 (11.0%)
Simple cyst/parasalpingeal cyst	75 (7.5%)
Mucinous cystadenoma	70 (7.0%)
Fibroma	37 (3.7%)
Functional cyst	50 (5.0%)
Hydrosalpinx/salpingitis	27 (2.7%)
Abscess	16 (1.6%)
Rare benign tumor[a]	7 (0.7%)
Peritoneal pseudocyst	7 (0.7%)
Uterine leiomyoma	5 (0.5%)
All malignant tumors	255 (26)
Common primary invasive	169 (16.9)
Stage I	38 (3.8)
Stage II	18 (1.8)
Stage III	97 (9.7)
Stage IV	16 (1.6)
Rare primary invasive[b]	18 (1.8)
Borderline	42 (4.2)
Metastatic	26 (2.6)

[a]Rare benign tumors, for example, struma ovarii and Brenner tumors.

[b]Rare primary invasive tumors, for example, granulosa cell tumor and Sertoli Leydig cell tumor.

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.

Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

Table 3. Performance of non-IOTA models and IOTA models for the whole tumor population ($n = 997$)

	AUC (95% CI)	
Model	**All data**	**Benign vs. primary invasive stage I[a]**
RMI variants		
RMI Jacobs	0.91 (0.88–0.93)	0.85 (0.78–0.90)
RMI2 Tingulstad 1996	0.89 (0.86–0.92)	0.85 (0.78–0.90)
RMI3 Tingulstad 1999	0.89 (0.86–0.91)	0.84 (0.77–0.89)
RMI4 Yamamoto	0.90 (0.87–0.92)	0.86 (0.80–0.91)
Non-IOTA logistic regression models		
LR Timmerman a	0.93 (0.91–0.94)	0.89 (0.83–0.93)
LR Lu	0.92 (0.89–0.94)	0.88 (0.82–0.92)
LR Jokubkiene	0.91 (0.89–0.92)	0.91 (0.88–0.94)
LR Timmerman b	0.90 (0.88–0.92)	0.85 (0.79–0.89)
LR Minaretzis	0.86 (0.83–0.88)	0.84 (0.78–0.88)
LR Tailor	0.84 (0.81–0.86)	0.86 (0.80–0.90)
Non-IOTA ANN		
ANN Timmerman 1	0.92 (0.89–0.93)	0.87 (0.81–0.91)
ANN Timmerman 2	0.87 (0.84–0.89)	0.80 (0.74–0.86)
IOTA models		
LR1	0.96 (0.94–0.97)	0.94 (0.88–0.97)
LS-SVM rbf	0.95 (0.94–0.97)	0.94 (0.88–0.97)
BPER11	0.95 (0.94–0.97)	0.92 (0.86–0.96)
BMLP 11-2b	0.95 (0.94–0.96)	0.93 (0.88–0.96)
RVM rbf	0.95 (0.93–0.96)	0.94 (0.88–0.97)
LS-SVM lin	0.95 (0.93–0.96)	0.94 (0.88–0.97)
LR2	0.95 (0.93–0.96)	0.94 (0.87–0.97)
RVM lin	0.95 (0.93–0.96)	0.94 (0.88–0.97)
BMLP 11-2a	0.94 (0.92–0.96)	0.93 (0.88–0.96)
RVM add rbf	0.94 (0.92–0.96)	0.94 (0.88–0.97)
LS-SVM add rbf	0.94 (0.92–0.96)	0.94 (0.89–0.97)

Abbreviations: add rbf, additive radial basis function; lin, linear; rbf, radial basis function.

[a]Primary invasive stage I tumors includes rare primary invasive stage I tumors.

0.94 and 0.96 (LR2, 0.95). The best non-IOTA model was a logistic regression model (11) with an AUC of 0.93 (95% CI, 0.91–0.94). The original RMI outperformed its variations with an AUC of 0.91 (95% CI, 0.88–0.93). LR1 and LR2 did significantly better than the RMI and the best non-IOTA model (Table 4). Figure 1 shows the ROC curves for LR1, LR2, the original RMI, and the best non-IOTA model.

The IOTA models were also superior to the non-IOTA models for discriminating between benign masses and stage I invasive ovarian cancers ($n = 48$), see Tables 3 and 4. The IOTA models had AUCs ranging from 0.92 to 0.94, whereas the AUCs for the non-IOTA models were at most 0.91. LR1 had an AUC of 0.94 (95% CI, 0.88–0.97), LR2 of 0.94 (95% CI, 0.87–0.97), and RMI of 0.85 (95% CI, 0.78–0.90). Table 4 shows that the advantage of LR1 and LR2 over the RMI and the best non-IOTA model is maintained when focusing on discrimination between stage I primary invasive ovarian malignancies and benign tumors.

The diagnostic performance of the models when using the cutoff values suggested in the original articles is presented in Table 5. The sensitivities and specificities of the IOTA models were all around 90%, whereas among the non-IOTA models they differed substantially. The IOTA models had DORs between 46 and 77 (DOR for LR1, 75; sensitivity, 92%; specificity, 87%), whereas the DORs of the non-IOTA models varied between 12 and 37. LR2 had a DOR of 66 (sensitivity, 92%; specificity, 86%), the RMI of 37 (sensitivity, 67%; specificity, 95%), and the best non-IOTA model of 29 (sensitivity, 92%; specificity, 86%). The best trade-off between sensitivity and specificity was achieved by the IOTA models. The detection rate of primary invasive ovarian cancer stage I ($n = 48$) of the IOTA models varied between 88% and 92% (LR1 and LR2, 92%). Detection rates of stage I cancer were lower for all non-IOTA models except for ANN 2 (94%). However, the very low specificity of ANN2 (41%) indicates that this model classified many more tumors as malignant compared with any other model. The detection rate of primary invasive ovarian cancer stage I was 51% for the RMI and 63% for the best non-IOTA model (Table 4).

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.

Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

Table 4. Comparison of the AUCs between the best IOTA model (LR1), the simpler 6-variable model (LR2), the best non-IOTA model, and the original RMI

	Difference in AUC (95% CI)			
Comparison	**All data**	**Benign vs. primary invasive stage I[a]**	**Premenopausal patients**	**Postmenopausal patients**
LR1 vs. LR2	0.006 (−0.001 to 0.014)	0.004 (−0.016 to 0.025)	0.008 (−0.007 to 0.022)	0.006 (−0.006 to 0.018)
LR1 vs. RMI Jacobs	0.045 (0.024–0.064)	0.085 (0.033–0.138)	0.084 (0.048–0.119)	0.019 (−0.015 to 0.053)
LR2 vs. RMI Jacobs	0.038 (0.018–0.058)	0.081 (0.031–0.131)	0.076 (0.043–0.109)	0.013 (−0.021 to 0.047)
LR1 vs. LR Timmerman a	0.029 (0.012–0.045)	0.048 (0.016–0.081)	0.044 (0.017–0.072)	0.014 (−0.016 to 0.045)
LR2 vs. LR Timmerman a	0.022 (0.004–0.041)	0.044 (0.003–0.085)	0.037 (0.007–0.066)	0.008 (−0.024 to 0.040)

[a]Primary invasive stage I tumors includes rare primary invasive stage I tumors.

Table 6 shows the performance of LR1, the RMI, and the best non-IOTA model in pre- and postmenopausal patients. LR1 and LR2 had better values for all performance measures in premenopausal and postmenopausal patients, although the advantage of LR1/LR2 relative to the RMI and the best non-IOTA model was larger in premenopausal patients.

Table 7 shows the number and types of malignancies that were missed by the 2 IOTA logistic regression models, the RMI of Jacobs and colleagues and subjective assessment by an experienced ultrasound examiner. The RMI missed more malignant tumors of any kind than the IOTA models and subjective assessment.

Discussion

This is the first study to externally validate and compare the diagnostic performance of IOTA models and non-IOTA models to discriminate between benign and malignant adnexal masses. We have shown that the IOTA models are robust and maintain their performance when tested in new centers in different countries with different population characteristics. Moreover, the IOTA models had highly similar performance and performed clearly better than the non-IOTA models, including the RMI. In particular, the detection rate of stage I primary invasive ovarian cancer was much higher for the IOTA models than for the very commonly used RMI. The differences in performance between the IOTA models and the non-IOTA models were smaller in postmenopausal than premenopausal patients (Table 6 and Supplementary Table). This is likely to be explained by CA125—included in the RMI and many non-IOTA models—performing better as a discriminator between benign and malignant adnexal lesions in postmenopausal women because of lower prevalence of benign lesions associated with increased CA125 levels in postmenopausal women (5, 40, 41). This is in agreement with IOTA models including CA125 not performing any better than models without CA125 (42).

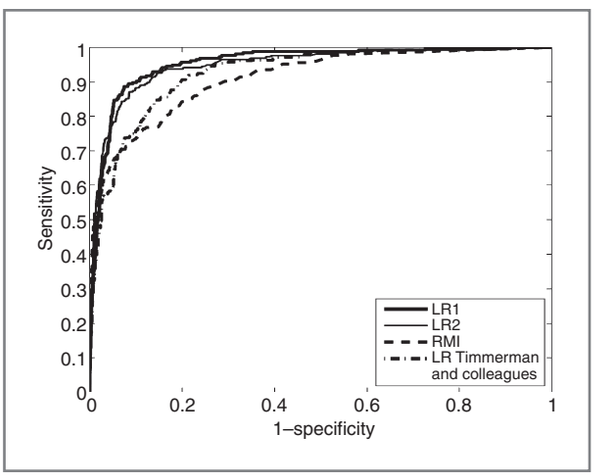


Figure 1. ROC curves of the 2 IOTA logistic regression models [LR 1 and LR 2 (16)], the RMI (7), and the best non-IOTA model [a logistic regression model by Timmerman and colleagues (11)].

A strength of our study is that it was conducted on a large number of masses. The geographic spread and the different referral profiles of the centers also suggest that the conclusions will have general applicability. Another strength is that we specifically determined the performance of the diagnostic methods to detect stage I ovarian cancer, even though there were only 48 such cancers. We think it is clinically important to achieve a high detection rate of stage I cancers, as explained below. It might be considered a weakness that in all centers, the scans were carried out by clinicians with a specific interest in gynecologic ultrasound. We need to evaluate how well the models perform in less experienced hands. However, obtaining information on the ultrasound variables required for the IOTA models should be possible for any qualified ultrasound practitioner.

Recently, Geomini and colleagues conducted a systematic review of all mathematical models that had been created for discrimination between benign and malignant adnexal masses and undergone external validation (20). They concluded that the RMI given by Jacobs and colleagues (7) is the method of choice for the preoperative assessment of an adnexal mass. However, the review of Geomini and colleagues did not include an evaluation of the IOTA models because at the time of writing that review, the results of the external validation of the IOTA models had not been

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.

Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

Table 5. Diagnostic performance of the different models when the cutoff recommended in the original article is used

Model	Cutoff	Sensitivity, %	Specificity, %	LR^+	LR^-	DOR (LR^+/LR^-; 95% CI)	Sensitivity for primary invasive stage I tumors,[a] %
RMI variants							
RMI Jacobs	200	67	95	12.7	0.34	37 (24–58)	51
RMI2 Tingulstad 1996	200	72	90	7.3	0.31	23 (16–34)	58
RMI3 Tingulstad 1999	200	67	93	9.8	0.35	28 (18–43)	51
RMI4 Yamamoto	450	67	94	11.8	0.35	34 (22–53)	52
Non-IOTA logistic regression models							
LR Timmerman a	0.25	75	91	8.2	0.28	29 (20–43)	63
LR Lu	0.20	83	85	5.4	0.20	27 (18–39)	73
LR Jokubkiene	0.12	77	88	6.6	0.26	26 (18–37)	81
LR Timmerman b	0.60	79	84	4.9	0.26	19 (13–27)	65
LR Minaretzis							
LR Tailor	0.50	20	98	9.7	0.82	12 (6.5–21)	19
Non-IOTA ANNs							
ANN Timmerman 1	0.45	75	89	6.9	0.28	25 (17–37)	65
ANN Timmerman 2	0.60	97	41	1.6	0.08	20 (10–41)	94
IOTA models							
LR1	0.10	92	87	6.8	0.09	75 (46–125)	92
LS-SVM rbf	0.12	89	90	8.8	0.12	75 (47–120)	90
BPER11	0.15	91	89	8.2	0.11	77 (48–125)	88
BMLP11-2b	0.15	91	87	6.8	0.10	65 (40–104)	90
RVM rbf	0.15	91	88	7.4	0.11	69 (43–111)	92
LS-SVM lin	0.15	87	91	9.9	0.14	70 (45–109)	85
LR2	0.10	92	86	6.4	0.10	66 (40–108)	92
RVM lin	0.20	91	89	8.0	0.11	75 (47–122)	90
BMLP11-2a	0.15	86	88	7.1	0.16	46 (30–69)	88
RVM add rbf	0.15	89	89	7.8	0.13	61 (39–96)	90
LS-SVM add rbf	0.12	87	89	7.7	0.15	53 (34–81)	92

Abbreviations: add rbf, additive radial basis function kernel; lin, linear kernel; LR^+, positive likelihood ratio; LR^-, negative likelihood ratio; rbf, radial basis function.

[a]Primary invasive stage I tumors includes rare primary invasive stage I tumors.

published. When a cutoff level of 200 for the RMI was used, the pooled estimate for sensitivity was 78% (95% CI, 71%–85%) and that of specificity was 87% (95% CI, 83%–91%). In our prospective validation, the sensitivity of RMI was lower (68%; 95% CI, 62%–73%) but the specificity was higher (95%; 95% CI, 93%–96%). Differences in tumor populations, for example, in stage distribution, might explain this discrepancy.

Table 6. Performance of the non-IOTA models and the IOTA models in pre- and postmenopausal patients

	Premenopausal patients			Postmenopausal patients		
Model	AUC	DOR	AUC stage I[a]	AUC	DOR	AUC st I[a]
RMI Jacobs	0.86	18	0.84	0.91	36	0.82
LR1	0.95	61	0.90	0.93	53	0.94
LR2	0.94	57	0.90	0.93	42	0.93
LR Timmerman a	0.90	19	0.86	0.92	23	0.87

[a]Primary invasive stage I tumors includes rare primary invasive stage I tumors.

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.

Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

Table 7. Rate of missed malignancies with regard to tumor type by 2 IOTA logistic regression models (LR1, LR2) and the RMI

	Rate of missed malignancies				
Diagnostic method	Invasive stage I[a] ($n = 48$)	Invasive stage II–IV[b] ($n = 139$)	Borderline ($n = 42$)	Metastatic ($n = 26$)	All malignancies ($n = 255$)
LR1	4 (8%)	6 (4%)	7 (17%)	3 (12%)	20 (8%)
LR2	4 (8%)	6 (4%)	9 (21%)	2 (8%)	21 (8%)
RMI Jacobs	23 (48%)	19 (14%)	31 (74%)	10 (38%)	83 (33%)
Subjective assessment	8 (17%)	13 (9%)	9 (21%)	2 (11%)	32 (13%)

Abbreviations: LR1, IOTA logistic regression model 1; LR 2, logistic regression model 2.

[a]Includes rare primary invasive stage I tumors.

[b]Includes rare primary invasive stage II–IV tumors.

A clear advantage of the IOTA models over the non-IOTA models is that they seem to have higher detection rate of stage I ovarian cancers than the other models and the RMI. Because stage I ovarian cancer is associated with a survival rate of more than 90% (43), early detection and optimal treatment might be crucial. A false-negative diagnosis of ovarian cancer will often lead to the patient undergoing laparoscopic surgery. This increases the risk of spilling cyst fluid containing malignant cells into the peritoneal cavity (44). Previous studies clearly showed that rupture of a malignant ovarian cyst worsens the prognosis for disease-free survival in stage I invasive ovarian cancers (2, 3). Spilling should also probably be avoided in borderline tumors, even though there are no prospective studies to prove this (44). Spilling and iatrogenic upstaging of the patient is not a serious problem in advanced cancer because spread throughout the peritoneal cavity is already present. Furthermore, higher stage ovarian cancers are often diagnosed clinically, making mathematical models to calculate the risk of malignancy less important. When we examined the types of malignancies that were missed, LR1 and LR2 missed the smallest number of stage I invasive cancers and borderline tumors. These models missed fewer malignancies than subjective assessment, but the latter had higher specificity (22). The RMI missed the highest number of stage I ovarian cancers and missed 73% of borderline tumors.

In conclusion, our results contradict the conclusion of Geomini and colleagues that the RMI is the method of choice to distinguish benign from malignant adnexal masses. External validation shows that the IOTA models outperform other models—including the current reference test RMI—for discriminating between benign and malignant adnexal masses.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank all participating centers, the principal investigators, and the study participants for their contribution. Data from the IOTA 2 study were presented at the 18th World Congress of Ultrasound in Obstetrics and Gynecology, organized by the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG), Chicago, 2008.

Recruitment Centers: The recruitment centers 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.

Grant Support

B. Van Calster is a postdoctoral researcher funded by the Research Foundation–Flanders (FWO), Belgium. Research supported by 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).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 18, 2011; revised October 21, 2011; accepted November 12, 2012; published OnlineFirst November 23, 2011.

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.

Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

References

1. Redman JR, Petroni GR, Saigo PE. Prognostic factors in advanced ovarian carcinoma. J Clin Oncol 1986;4:515–23.
2. 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: 176–82.
3. Bakkum-Gamez J, Richardson D, Seamon L, Aletti G, Powless C, Keeney G, et al. Influence of intraoperative capsule rupture on outcomes in stage I epithelial ovarian cancer. Obstet Gynecol 2009;113: 11–7.
4. Valentin L. Prospective cross-validation of Doppler ultrasound examination and gray-scale ultrasound imaging for discrimination of benign and malignant pelvic masses. Ultrasound Obstet Gynecol 1999;14: 273–83.
5. Van Calster B, Timmerman D, Bourne T, Testa AC, 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.
6. Dearking A, Aletti G, McGree M, Weaver A, Sommerfield M, Cliby W. How relevant are ACOG and SGO guidelines for referral of adnexal mass? Obstet Gynecol 2007;110:841–8.
7. Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal state for the accurate preoperative diagnosis of ovarian cancer. Br J Obstet Gynecol 1990;97:922–9.
8. Tingulstad S, Hagen B, Skjeldestad FE, Onsrud M, Kiserud T, Halvorsen T, et al. Evaluation of risk of malignancy index based on serum CA 125, ultrasound findings and menopausal status in the preoperative diagnosis of pelvic masses. Br J Obstet Gynecol 1996;103:826–31.
9. Tailor A, Jurkovic D, Bourne T, Collins WP, Campbell S. Sonographic prediction of malignancy in adnexal masses using multivariate logistic regression analysis. Ultrasound Obstet Gynecol 1997;10:41–47.
10. Minaretzis D, Tsionou C, Tziortziotis D, Michalas S, Aravantinos D. Ovarian tumors: prediction of the probability of malignancy by using patient's age and tumor morphologic features with a logistic model. Gynecol Obstet Invest 1994;38:140–4.
11. Timmerman D, Bourne T, Tailor A, Collins WP, Verrelst H, Vandenberghe K. A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: the development of a new logistic regression model. Am J Obstet Gynecol 1999;181: 57–65.
12. Timmerman D, Verrelst H, Bourne TH, De Moor B, Collins WP, Vergote I. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound Obstet Gynecol 1999;13:17–25.
13. Lu C, Suykens JAK, Timmerman D, Vergote I, Van Huffel S. Linear and nonlinear preoperative classification of ovarian tumors. Chapter 11. In: Ichimura T, Yoshida K, editors. Knowledge based intelligent system for health care: International series on advanced intelligence, Vol 7. Magill, Australia: Advanced Knowledge International; 2004, p. 343–82.
14. Jokubkiene L, Sladkevicius P, Valentin L. Does three-dimensional power Doppler ultrasound help in discrimination between benign and malignant ovarian masses? Ultrasound Obstet Gynecol 2007;29: 215–25.
15. 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;13:4440–7.
16. Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML. 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;23:8794–801.
17. 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.
18. Geomini P, Kruitwagen R, Bremer GL, Cnossen J, Mol BW. The accuracy of risk scores in predicting ovarian malignancy: a systematic review. Obstet Gynecol 2006;108:1167–75.
19. Ovarian cysts in postmenopausal women [monograph on the Internet]. London, UK: RCOG. RCOG guideline no. 34. [cited 2003 Oct] Available from: http://www.pelvicpain.org.uk/uploads/documents/Ovarian-Cysts2003-guidelines.pdf.
20. Geomini P, Kruitwagen R, Bremer G, Cnossen J, Mol BW. The accuracy of risk scores in predicting ovarian malignancy: a systematic review. Obstet Gynecol 2009;113:384–94.
21. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) group. Ultrasound Obstet Gynecol 2000;16:500–5.
22. 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.
23. Van Calster B, Timmerman D, Lu C, Suykens J, Valentin L, Van Holsbeke C, et al. Preoperative diagnosis of ovarian tumors using Bayesian kernel-based methods. Ultrasound Obstet Gynecol 2007;29: 496–504.
24. Van Calster B, Timmerman D, Nabney I, Valentin L, Testa AC, Van Holsbeke C, et al. Using Bayesian Neural Networks with ARD input selection to detect malignant adnexal masses prior to surgery. Neural Comput Appl 2008;17:489–500.
25. Van Holsbeke C, Van Calster B, Testa AC, Domali E, Lu C, Van Huffel S, et al. Prospective internal validation of mathematical models to predict malignancy in adnexal masses: results from the International Ovarian Tumor Analysis Study. Clin Cancer Res 2009;15:684–91.
26. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med 1999;130:515–24.
27. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338: b605.
28. Tingulstad S, Hagen B, Skjeldestad FE, Halvorsen T, Nustad K, Onsrud M. The risk-of-malignancy index to evaluate potential ovarian cancers in local hospitals. Obstet Gynecol 1999;93:448–52.
29. Yamamoto Y, Yamada R, Oguri H, Maeda N, Fukaya T. Comparison of four malignancy risk indices in the preoperative evaluation of patients with pelvic masses. Eur J Obstet Gynecol Reprod Biol 2009;144: 163–7.
30. Ameye L, Valentin L, Testa AC, Van Holsbeke C, Domali E, Van Huffel S, et al. A scoring system to differentiate malignant from benign masses in specific ultrasound-based subgroups of adnexal tumors. Ultrasound Obstet Gynecol 2009;33:92–101.
31. Pepe M. The statistical evaluation of medical tests for classification and prediction. New York: Oxford University Press; 2003.
32. Qin G, Hotilovac L. Comparison of non-parametric confidence intervals for the area under the ROC curve of a continuous-scale diagnostic test. Stat Methods Med Res 2008;17:207–21.
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves. A nonparametric approach. Biometrics 1988;44:837–45.
34. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. Br Med J 2001;323:157–62.
35. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003;56:1129–35.
36. Little JR, Rubin D. Statistical analysis with missing data. 2nd ed. New York: Wiley; 2002.
37. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. A gentle introduction to imputation of missing values. J Clin Epidemiol 2006;59: 1087–91.
38. Schenker N, Taylor JMG. Partially parametric techniques for multiple imputation. Comput Stat Data Anal 1996;22:425–46.

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.

Published OnlineFirst November 23, 2011; DOI: 10.1158/1078-0432.CCR-11-0879

39. Van Calster B, Valentin L, Van Holsbeke C, Zhang J, Jurkovic D, Lissoni AA, et al. 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;20:2420–8.
40. Bast RC, Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC. Reactivity of a monoclonal antibody with human ovarian carcinoma. J Clin Invest 1981;68:1331–7.
41. Tuxen MK. Tumor marker CA125 in ovarian cancer. J Tumor Marker Oncol 2001;16:49–68.
42. Timmerman D, Van Calster B, Jurkovic D, Valentin L, Testa AC, Bernard JP, et al. Inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and malignant adnexal tumors. J Clin Oncol 2007;20:4194–200.
43. Trimbos JB, Vergote I, Bolis G, Vermorken JB, Mangioni C, Madronal C, et al. EORTC-ACTION collaborators. European Organisation for Research and Treatment of Cancer-Adjuvant ChemoTherapy in Ovarian Neoplasm. Impact of adjuvant chemotherapy and surgical staging in early-stage ovarian carcinoma: European Organisation for Research and Treatment of Cancer-Adjuvant ChemoTherapy in Ovarian Neoplasm trial. J Natl Cancer Inst 2003;15:113–25.
44. Cadron I, Leunen K, Van Gorp T, Amant F, Neven P, Vergote I. Management of borderline ovarian neoplasms. J Clin Oncol 2007; 25:2928–37.

Downloaded from clincancerres.aacrjournals.org on February 10, 2014. © 2012 American Association for Cancer Research.