Clinical Utility of Risk Models to Refer Patients with Adnexal Masses to Specialized Oncology Care: Multicenter External Validation Using Decision Curve Analysis

Laure Wynants1, Dirk Timmerman1,2, Jan Y. Verbakel1,3, Antonia Testa4, Luca Savelli5, Daniela Fischerova6, Dorella Franchi7, Caroline Van Holsbeke8, Elisabeth Epstein9, Wouter Froyman1,2, Stefano Guerriero10, Alberto Rossi11, Robert Fruscio12, Francesco PG Leone13, Tom Bourne1,2,14, Lil Valentin15, and Ben Van Calster1

Purpose: To evaluate the utility of preoperative diagnostic models for ovarian cancer based on ultrasound and/or biomarkers for referring patients to specialized oncology care. The investigated models were RMI, ROMA, and 3 models from the International Ovarian Tumor Analysis (IOTA) group [LR2, ADNEX, and the Simple Rules risk score (SRRisk)].

Experimental Design: A secondary analysis of prospectively collected data from 2 cross-sectional cohort studies was performed to externally validate diagnostic models. A total of 2,763 patients (2,403 in dataset 1 and 360 in dataset 2) from 18 centers (11 oncology centers and 7 nononcology hospitals) in 6 countries participated. Excised tissue was histologically classified as benign or malignant. The clinical utility of the preoperative diagnostic models was assessed with net benefit (NB) at a range of risk thresholds (5%–50% risk of malignancy) to refer patients to specialized oncology care. We visualized results with decision curves and generated bootstrap confidence intervals.

Results: The prevalence of malignancy was 41% in dataset 1 and 40% in dataset 2. For thresholds up to 10% to 15%, RMI and ROMA had a lower NB than referring all patients. SRRisks and ADNEX demonstrated the highest NB. At a threshold of 20%, the NBs of ADNEX, SRRisks, and RMI were 0.348, 0.350, and 0.270, respectively. Results by menopausal status and type of center (oncology vs. nononcology) were similar.

Conclusions: All tested IOTA methods, especially ADNEX and SRRisks, are clinically more useful than RMI and ROMA to select patients with adnexal masses for specialized oncology care.

Introduction

An accurate preoperative diagnosis of an adnexal mass is pivotal to improve care, because an optimal diagnostic process improves triage and subsequent treatment decisions. In 2009, a systematic review of the ability of preoperative prediction models to correctly discriminate between benign and malignant adnexal masses recommended the use of the Risk of Malignancy Index (RMI; refs. 1, 2). However, neither the Risk of Ovarian Malignancy Algorithm (ROMA; ref. 3) nor any of the International Ovarian Tumor Analysis (IOTA) models (4–9) were included in that review. A systematic review published in 2014 updated the available evidence (10). It recommended the use of the IOTA Simple Rules, which classify masses as probably benign, probably malignant, or inconclusive (6, 7), or the IOTA logistic regression model LR2 (4, 5), because of their good discriminative ability, especially for women of reproductive age. After the publication of the 2 systematic reviews, the ADNEX risk model (Assessment of Different Neoplasias in the adnexa) was published (8). This model goes beyond the traditional distinction between benign and malignant masses by calculating the risk that an adnexal mass is benign, borderline, stage I primary cancer, stage II–IV primary cancer, or secondary metastatic cancer. At temporal and external validation, the ADNEX model showed a good discriminative ability and was well calibrated (8, 11–15). In addition, the IOTA Simple Rules have now been extended to predict the risk of malignancy.
Translational Relevance

An accurate preoperative diagnosis of an adnexal mass is important to inform decisions regarding patient triage and subsequent treatment. Prospective validation of the IOTA models (LR2, the Simple Rules, the Simple Rules Risk scoring system, and ADNEX) has shown that they discriminate well between benign and malignant masses. Direct comparisons have shown that RMI and ROMA do not discriminate as well between benign and malignant masses as the IOTA models. However, good discrimination between benign and malignant cases is not sufficient to guarantee clinical utility. ADNEX and the Simple Rules Risks have more clinical utility than RMI and ROMA as measured by the net benefit, suggesting that ADNEX and the Simple Rules Risks are the best models to decide which patients to refer to specialized oncology care. This should ultimately lead to improved patient survival, decreased morbidity, and reduced health care expenditures.

Materials and Methods
Design, setting, and patients

This is a secondary analysis of 2 cross-sectional cohort datasets containing data prospectively collected to validate models for distinguishing preoperatively between benign and malignant adnexal masses (26, 27). Dataset 1 was collected between October 2009 and May 2012 in 18 centers from 6 countries (Sweden, Belgium, Italy, Poland, Spain, and Czech Republic; ref. 27). The centers were either oncology centers (i.e., tertiary referral centers with a specific gynecology oncology unit) or general hospitals with a special interest and high level of competence in gynecologic ultrasound. Dataset 2 was collected between August 2005 and March 2009 at the University Hospitals Leuven (an oncology center in Belgium; ref. 26).

Both datasets include consecutive patients with an adnexal mass (ovarian, paraovarian, or tubal) examined with transvaginal ultrasound following a standardized research protocol by an experienced operator (principal investigator) and who subsequently underwent surgical removal of the mass. The inclusion criteria are similar to those used in the model development studies (2–4, 6, 8, 9). If multiple masses were present, the mass with the most complex ultrasound morphology was used in the statistical analysis. If masses had a similar ultrasound morphology, the largest mass or the mass most easily accessible with ultrasound was used. The excised tissues underwent histologic examination at the local center and were classified as benign or malignant. Histologic classification was done without knowledge of the ultrasound results or of the results of the diagnostic models under investigation. Details about data collection for dataset 1 and 2 are provided in the original publications (26, 27). All women gave written or oral consent as per local requirements, and data collection was approved by the Ethics Committees or Institutional Review Boards of the local centers.

Prediction models

We evaluated the clinical utility of 5 models for distinguishing preoperatively between benign and malignant adnexal masses: LR2, ADNEX, SRRisks, ROMA, and RMI. In addition, we investigated the clinical utility of the original Simple Rules (Supplementary Fig. S1). An overview of the models is presented in Table 1 (the mathematical formulae and prediction rules are presented in Supplementary Table S1). Note that all the investigated models except ROMA contain ultrasound variables. ADNEX and SRRisks include the type of center (oncology center vs. other) as a predictor to improve the calibration of risk predictions. Because RMI and the original Simple Rules do not provide risk estimates, we evaluated RMI and the original Simple Rules as dichotomous classification systems (28). For RMI, we used the cut-off value of 200 or more to identify patients at a high risk of malignancy (2). This value is often used in clinical practice, but we also investigated the utility of RMI with other cutoffs (450, 250, 100, and 25; Supplementary Fig. S1). For the original IOTA Simple Rules, masses that yielded malignant or inconclusive results were classified as malignant. For ADNEX, the total risk of malignancy is the sum of the risks for each malignant subtype, and the risk can be calculated with or without serum CA125 as a predictor (see Supplementary Fig. S1 for the results for ADNEX without CA125).
Table 1. An overview of the models and classification rules for presurgical diagnosis of adnexal tumors used in this work

<table>
<thead>
<tr>
<th>Model</th>
<th>Publication year</th>
<th>Predictors</th>
<th>Type of model</th>
</tr>
</thead>
<tbody>
<tr>
<td>RM1</td>
<td>1990</td>
<td>Menopausal status, CA125, multilocular cysts, solid areas, metastases, ascites, bilaterality</td>
<td>Numerical score of 0 or above, derived from a logistic regression model.</td>
</tr>
<tr>
<td>ROMA</td>
<td>2009</td>
<td>CA125, HE4, and menopausal status</td>
<td>Logistic regression model providing a risk of malignancy.</td>
</tr>
<tr>
<td>IOTA LR2</td>
<td>2005</td>
<td>Age, ascites, blood flow within a papillary projection, maximal diameter of the largest solid component, irregular internal cyst walls, acoustic shadows</td>
<td>Logistic regression model providing a risk of malignancy.</td>
</tr>
<tr>
<td>IOTA Simple Rules</td>
<td>2008</td>
<td>Features for malignancy: (i) Irregular solid mass; (ii) very strong intratumoral flow; (iii) irregular multilocular-solid mass with largest diameter $\geq$100 mm; (iv) ascites; and (v) $\geq$4 papillary structures. Features for a benign mass: (i) unilocular cyst; (ii) no intratumoral flow; (iii) smooth multilocular tumor with largest diameter $&lt;100$ mm; (iv) acoustic shadows; (v) solid components with largest diameter $&lt;7$ mm.</td>
<td>Classification as benign, malignant or unclassifiable.</td>
</tr>
<tr>
<td>IOTA SRRisks</td>
<td>2016</td>
<td>The 10 features used for the Simple Rules, and type of center (oncology center vs. other)</td>
<td>Logistic regression model providing a risk of malignancy.</td>
</tr>
<tr>
<td>IOTA ADNEX</td>
<td>2014</td>
<td>Age, CA125, maximal diameter of the lesion, largest diameter of the largest solid component, $&gt;10$ cyst locules, number of papillary projections, acoustic shadows, ascites, type of center.</td>
<td>Logistic regression model providing risks of 4 malignant tumor subtypes (borderline, stage I primary cancer, stage II-IV primary cancer, secondary metastatic cancer); the total predicted risk of malignancy is the sum of the risks of each malignant subtype.</td>
</tr>
</tbody>
</table>

Dataset 1 contains information that allows the clinical utility of RM1, LR2, Simple Rules, SRRisks, and ADNEX to be estimated. ROMA could not be applied in dataset 1 because information on HE4 is lacking in this dataset. Dataset 1 was originally used for temporal validation of the discriminative performance and calibration of ADNEX and SRRisks and for updating these models. Here, we use the formulae for ADNEX and SRRisks created using the development data. We do not use the updated models using both developmental and validation data (8, 9). Dataset 2 allows us to externally validate LR2, ROMA, and RM1 in terms of clinical utility, but not ADNEX and SRRisks, because a part of dataset 2 was used to develop ADNEX and SRRisks. The predictions of all models were obtained centrally by a statistician, after the data collection by clinicians was concluded.

Evaluation of clinical utility

We used NB as the key performance measure to assess the potential utility of the models for clinical decision making. NB combines the benefits of true positives and the harms of false positives on a single scale by using a weighting factor for false positives (16, 20, 21). This weighting factor corresponds to the odds of the chosen risk threshold $T$ [i.e., $T/(1-T)$] to select patients for treatment (29). In our case, treatment is equivalent to referring patients with an adnexal mass to specialized oncology care. For example, a risk threshold $T$ of 33% (odds 1:2) implies that up to 2 false positives are felt to be acceptable per true positive. In other words, if we use a risk of malignancy of 33% or higher as the threshold for referring a patient to oncology care, we consider the benefit of selecting a patient with an ovarian malignancy for specialized oncology care to be twice as large as the harm of referring one patient with a benign tumor to specialized oncology care. In this work, we consider risk thresholds between 5% and 50%. Although arbitrary, these thresholds represent clinically sensible strategies. Ideally, all patients with ovarian cancer should receive advanced care. At 5%, we would accept up to 19 false positives per true positive. This means that the benefit of selecting a patient with an adnexal malignancy for specialized oncology care is considered to be 19 times as large as the harm of referring one patient with a benign tumor for treatment to specialized oncology care. Risk thresholds close to 50% may be useful if resources are limited or waiting lists for specialized oncology care are very long. At a threshold of 50%, we would accept one false positive per true positive, that is, the benefit of selecting one patient with an ovarian malignancy for specialized oncology care is considered to be equivalent to the harm of referring one patient with a benign tumor for treatment to specialized oncology care. In clinical reality, risk thresholds of more than 50% are not sensible because this would imply that referring a patient with a benign tumor to oncology care is more harmful than not referring a patient with cancer.

Given the risk threshold $T$, the NB is calculated as follows:

$$\text{NB} = \frac{\text{Number of true positives} - \left( \frac{1}{T} \right) \times \text{number of false positives}}{\text{total sample size}}$$

The risk models that we evaluate in this work classify patients as at high risk of cancer if the predicted risk is $\geq T$; RM1 at a certain cutoff classifies patients as high risk if RM1 is at least as high as the cutoff (e.g., $\geq 200$) irrespective of $T$. When using the original Simple Rules, patients classified as having a malignant or an unclassifiable mass are considered to be high risk irrespective of $T$.

We plotted the decision curves (NB vs. $T$), for all models and for 2 default strategies: referring all patients or referring none. Referring all patients means that every patient with an adnexal mass is classified as being at high risk of ovarian cancer and is referred to specialized oncology care. Referring none means that no patient with an adnexal mass is considered to be at risk of malignancy and none are referred to specialized oncology care. If at a given risk threshold ($T$), a model has a lower NB than referring all or
referring no one, the model is considered harmful for clinical decision making because a simple default strategy yields a higher NB. We calculated the difference between the NB of each model and the NB of the default strategy with the highest NB. The maximum attainable NB equals the prevalence of the condition sought for, in this case the prevalence of malignancy (the number of positives/total sample size). We computed the difference in NB between the model with the most clinical utility (i.e., the model with a very high NB over the entire range of risk thresholds) and all other models. We generated 95% bootstrap confidence intervals (CI) for NB and the differences in NB using the percentile method with 1,000 samples.

We investigated the clinical utility of models in the following subgroups in dataset 1: premenopausal patients, postmenopausal patients, patients seen at oncology centers, and patients seen at nononcology centers. Dataset 2 was too small to allow meaningful subgroup analyses.

All analyses were performed using R version 3.3.1 (http://www.r-project.org/). NB was computed using the dca function (21).

The TRIPOD guidelines were followed for the reporting of this study (30).

Missing data for CA125

Information on serum CA125 is necessary to calculate ADNEX, RMI, and ROMA, but serum CA125 measurements were optional in the cohort study in which dataset 1 was collected. We used single imputation to deal with missing values in dataset 1. CA125 was estimated with predictive mean matching regression (31), using variables that were related to the level of CA125 or the availability of CA125 measurements. Details on the imputation procedure can be found elsewhere (8). Typically, multiple imputation is preferred over single imputation to get variance estimates that reflect uncertainty due to missingness. However, we noticed in previous studies that variance estimates were not meaningfully smaller if single imputation was used for the prediction models we assess in this study. Hence, we use single imputation in this study to reduce the computational burden.

Patient involvement

No patients or laypeople were involved in the design or conduct of this study. The main outcome measure in this work (NB) was chosen to evaluate and compare the clinical utility of models assuming various risk thresholds for referral to specialized oncology care that reflect differences in patients’ priorities and preferences.

### Results

#### Dataset 1

Data of 2,541 women with adnexal masses were available. In total, 138 women were excluded from the analyses for the following reasons: an interval of >120 days between ultrasonography and surgery; pregnancy, data errors that could not be resolved, and incomplete final histology. Of the remaining 2,403 patients used in this study, 1,423 (59%) had a benign adnexal mass and 980 (41%) had a malignant adnexal mass; 1,049 patients (44%) were postmenopausal and 1,354 (56%) were premenopausal; 1,715 (71%) were treated in oncology centers and 688 (29%) were treated in other centers. Malignancy rates are roughly comparable with those in the model development studies (2–4, 6, 8, 9). They varied by center and were generally higher in oncology centers than in other centers (Supplementary Table S2). Serum CA125 values were missing in 952 women (40%). Descriptive statistics of the data are presented in Table 2.

The NB was high for SRRisks and ADNEX and lowest for RMI (with cutoff 200) for all risk thresholds (Fig. 1; Supplementary Table S3). The NB of LR2 was intermediate. Using RMI (cutoff 200) was harmful at risk thresholds below 20%, meaning that referring all patients to specialized oncology care is clinically more useful than using RMI (cutoff 200) if one is willing to refer more than 4 benign cases per malignant case referred (Fig. 1; Supplementary Table S4). At risk threshold 20%, the NBs of ADNEX, SRRisks, LR2, and RMI (using cutoff 200) were 0.348 (95% CI, 0.328–0.369), 0.350 (95% CI, 0.329–0.372), 0.329 (0.308–0.349), and 0.270 (0.250–0.288), respectively. At this risk threshold, the NB of ADNEX was 0.078 (95% CI, 0.066–0.092) higher than the NB of RMI (using the 200 cutoff; see Supplementary

![Figure 1](https://example.com/image1.png)

Figure 1. Decision curves representing the NB of the RMI with cutoff 200 (RMI 200), the IOTA logistic regression model 2 (LR2), the IOTA ADNEX model (ADNEX), the IOTA SRRisks, referring all and referring none of the patients to specialized oncology care, for risk of malignancy thresholds between 5% and 50% (n = 2,403, 41% malignant tumors). FP, false positives; TP, true positives.

#### Table 2. Patient and tumor characteristics per dataset

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Dataset 1  (n = 2,403)</th>
<th>Dataset 2  (n = 360)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean, SD)</td>
<td>50 (16)</td>
<td>51 (16)</td>
</tr>
<tr>
<td>Postmenopausal (n %)</td>
<td>1,049 (44%)</td>
<td>187 (52%)</td>
</tr>
<tr>
<td>Ultrasound examination at oncology center (n, %)</td>
<td>1,757 (77%)</td>
<td>398 (100%)</td>
</tr>
<tr>
<td>CA125 information missing (n, %)</td>
<td>952 (40%)</td>
<td>0</td>
</tr>
<tr>
<td>Tumor histology (n, %)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benign</td>
<td>1,423 (59%)</td>
<td>216 (60%)</td>
</tr>
<tr>
<td>Malignant</td>
<td>980 (41%)</td>
<td>144 (40%)</td>
</tr>
<tr>
<td>Borderline</td>
<td>153 (6%)</td>
<td>32 (9%)</td>
</tr>
<tr>
<td>Primary invasive stage I</td>
<td>196 (8%)</td>
<td>18 (5%)</td>
</tr>
<tr>
<td>Primary invasive stage II</td>
<td>47 (2%)</td>
<td>5 (2%)</td>
</tr>
<tr>
<td>Primary invasive stage III</td>
<td>397 (17%)</td>
<td>53 (15%)</td>
</tr>
<tr>
<td>Primary invasive stage IV</td>
<td>61 (3%)</td>
<td>10 (3%)</td>
</tr>
<tr>
<td>Unknown FIGO stage</td>
<td>0 (0%)</td>
<td>2 (0.6%)</td>
</tr>
<tr>
<td>Metastatic</td>
<td>126 (5%)</td>
<td>24 (7%)</td>
</tr>
</tbody>
</table>
Table S4). This can be interpreted as follows: if one believes it is justified to refer 4 women with a benign mass to specialized oncology care per woman referred with a malignant mass (harm-to-benefit ratio 20:80 = 1:4). ADNEX is more clinically useful than RMI (cutoff 200). More specifically, when we use ADNEX, we can correctly refer a net number of 7.8 more malignant cases per 100 women than when we use RMI, for the same number of false positives (this is the number of true positives corrected for the number of false positives, using the odds of the threshold as a weighting factor for false positives). ADNEX was more clinically useful than RMI (cutoff 200) at all risk thresholds (see Supplementary Table S4). Making decisions based on ADNEX and SRRisks had similar clinical utility, except at risk thresholds close to 50% where the NB of the SRRisks showed a sudden drop (see Fig. 1 and Supplementary Table S4).

Using the RMI with a cutoff of 450 to classify patients as high risk reduced the NB compared with using a cutoff of 200. Using cutoffs lower than 200 increased NB at lower risk thresholds. For example, using the RMI with cutoff of 25 avoided harmful decision making for risk thresholds above 10%. However, whichever RMI cutoff was used, the NB for the RMI remained well below the NB of the other models (see Supplementary Fig. S1). Supplementary Figure S1 also shows that using the original IOTA Simple Rules (with inconclusive cases classified as malignant) yielded a NB similar to using SRRisks except that NB was lower for the original Simple Rules at risk thresholds above 30%. The NB of ADNEX with and without CA125 was similar, but NB was higher at risk thresholds above 30% when CA125 was used as a variable in ADNEX (Supplementary Fig. S1).

The results for pre- and postmenopausal women and for patients examined with ultrasound in oncology units and non-oncology units are shown in Figs. 2 and 3 and in Supplementary Tables S3, S5, and S6. In all subgroups, the results were similar: the RMI had the lowest NB and was harmful at low risk thresholds, RMI remained well below the NB of the other models (see Supplementary Fig. S1). Supplementary Figure S1 also shows that using the original IOTA Simple Rules (with inconclusive cases classified as malignant) yielded a NB similar to using SRRisks except that NB was lower for the original Simple Rules at risk thresholds above 30%. The NB of ADNEX with and without CA125 was similar, but NB was higher at risk thresholds above 30% when CA125 was used as a variable in ADNEX (Supplementary Fig. S1).

Discussion

This study has shown that the IOTA models ADNEX (with or without CA125) and SRRisks have clinical utility at a broad range of risk thresholds to refer patients with ovarian masses to specialized oncological care. The LR2 model also has clinical utility but less than ADNEX and SRRisks. The original Simple Rules classification model has clinical utility similar to the SRRisks. RMI has less clinical utility and is harmful at low-risk thresholds regardless of whether the commonly used cutoff of 200 or the cutoff of 25 mentioned by the RCOG was used (32). Our findings hold true for both pre- and postmenopausal patients. The clinical utility of ROMA is similar to that of the RMI for risk thresholds of 10% and higher, and lower than that of LR2.

To the best of our knowledge, this is the first study to estimate the clinical utility of models used to distinguish between benign and malignant adnexal masses before surgery. Another strength of our study is that we evaluated the clinical utility using only validation data, that is, data that were not used to develop the models. It may be regarded as a limitation of our study that serum levels for CA125, a predictor in ADNEX, RMI, and ROMA, were not available for all patients. We solved this problem by using imputation, hence avoiding bias in the results due to missing data (33). Another limitation is that we were unable to evaluate the clinical utility of the IOTA ADNEX model to distinguish between various subtypes of malignant tumors, as measures to evaluate the clinical utility with multiple outcome categories are not yet established. Some may regard it as a limitation that we did not evaluate all published models to predict malignancy in adnexal masses. We focused on the most commonly used and best performing models. OVA-1, a recently published model with a very low specificity (34–36), could not be externally validated as the algorithm is not freely available. An additional limitation of this study is that many of the sonographic measurements of predictors included in the risk models were performed by experts in ultrasound, even though the study was performed in a mix of regional centers and referral centers. Nevertheless, it is reassuring that the IOTA models have been shown to keep their excellent diagnostic performance when used by clinicians with various levels of expertise and backgrounds (11, 12, 15, 37–39).

In addition, the ADNEX model contains only ultrasound features that are relatively easy to assess and does not include any Doppler variables. Although the current study and past research (40) demonstrate the superiority of the IOTA models over ROMA in the hands of experienced investigators, it would be interesting to prospectively compare ROMA with IOTA models in the hands of less experienced sonographers in future studies, as ROMA is not based on sonographic assessment of the lesion.

Our study adds to the existing evidence that IOTA algorithms perform better than both RMI and ROMA to distinguish between benign and malignant adnexal tumors (10, 11, 40, 41). Discrimination was very good to excellent for all models (8, 9, 27, 40). In the dataset used for this study, AUCs were 0.875 for RMI, 0.918 for LR2, 0.917 for SRRisks, and 0.936 for ADNEX (8, 9, 27). Published studies have shown that LR2 substantially underestimates the risk of malignancy, whereas for ADNEX and SRRisks, only a very mild underestimation was observed (5, 8, 9, 27). In contrast to previous studies, this study goes beyond reporting statistical measures of discrimination and calibration. It incorporates the consequences of false-positive and false-negative classifications into the evaluation of models. Hence, we were able to evaluate the clinical utility of the models for deciding which patients to refer to specialized oncology care. The specification of a fixed-risk threshold to refer patients to specialized oncology care may increase the uptake of models and simplify patient management. However, decision curve analysis cannot be used to decide which threshold to choose (16). In fact,
no single threshold can be recommended, because the appropriate risk threshold depends on the clinical setting in which the model is applied. It may vary depending on the available health care resources, local referral patterns and guidelines, and the level of oncological competence in nononcology centers. Risk thresholds also depend on the decision to be made. If a model would be used to decide who needs to undergo extensive oncological surgery, the harm of a false positive would be high and the risk threshold should be set high. In addition, risk thresholds should also reflect patients’ preferences and characteristics. Different risk

![Figure 2.](image)

Decision curves representing the NB of the RMI with cutoff 200 (RMI 200), the IOTA logistic regression model 2 (LR2), the IOTA ADNEX model (ADNEX), the IOTA SRRisks, referring all and referring none of the patients to specialized oncology care, for risk thresholds between 5% and 50% in pre- and postmenopausal patients (n = 1,354, 27% malignant tumors for premenopausal patients and n = 1,049, 57% malignant tumors for postmenopausal patients). FP, false positives; TP, true positives.

![Figure 3.](image)

Decision curves representing the NB of the RMI with cutoff 200 (RMI 200), the IOTA logistic regression model 2 (LR2), the IOTA ADNEX model (ADNEX), the IOTA SRRisks, referring all and referring none of the patients to specialized oncology care, for risk thresholds between 5% and 50% in oncology centers and other centers (n = 1,715, 49% malignant tumors in oncology centers and n = 688, 18% malignant tumors in other centers). FP, false positives; TP, true positives.
require substantial ultrasound expertise.

models, ADNEX does not include Doppler variables, which are easy to assess. In contrast to the SRRisks and other IOTA population (42). This is informative for patient management by specialized oncological care, for risk thresholds between 5% and 50% (n = 360, 40% malignant tumors). FP, false positives; TP, true positives.

thresholds may be appropriate for women of reproductive age and premenopausal women. In younger women, false-negative results may have a larger impact on survival than in older women. On the other hand, younger patients more often present with borderline tumors, and if there is a reasonable level of oncological competence in nononcology centers, it may be acceptable to manage borderline tumors there. Of course, risk thresholds cannot substitute a physician’s clinical judgement; they can only be an adjunct. On the other hand, they are used. The Royal College of Obstetricians and Gynecologists recommends that postmenopausal women with an adnexal mass and an RMI score of 200 or higher be referred for assessment by an oncological multidisciplinary team (32). Our study has shown that ADNEX and SRRisks are the most promising models in terms of clinical utility when deciding who to refer for oncological care, and this is true of both pre-and postmenopausal women.

On the basis of the decision curve analysis, ADNEX and SRRisks are both very useful models to decide which patients to refer to specialized oncological care. Nevertheless, there are 2 note-worthy differences between the 2 models. First, the outcome of the ADNEX model exceeds a simple distinction between benign and malignant masses, as it offers also risk estimates for malignant subtypes (benign, borderline, stage I cancer, stage II–IV cancer, metastasis). In a first step, the ADNEX model can be used to distinguish between benign and malignant lesions. In a second step, the model reveals which malignant subtypes have an elevated risk estimate in this patient, compared with the general population (42). This is informative for patient management decisions. Second, the ultrasound variables of the ADNEX model are easy to assess. In contrast to the SRRisks and other IOTA models, ADNEX does not include Doppler variables, which require substantial ultrasound expertise.

The decision curve analysis presented in this study is a first step to assess the consequences of introducing diagnostic models into clinical practice. The clinical impact of using ADNEX or SRRisks to select women with adnexal masses for referral to oncological care could further be assessed by a formal cost-effectiveness analysis or in clinical trials, for example, in a randomized controlled trial comparing RMI or ROMA with ADNEX or SRRisks as a basis for referring women with adnexal masses to oncological care. The authors of a recently published randomized controlled trial comparing the original IOTA Simple Rules with RMI for the management of asymptomatic postmenopausal patients concluded that applying the Simple Rules lead to lower surgical intervention rates for asymptomatic women, without an increase in delayed malignant diagnoses (43).

The decision curve analysis we have presented in this study has demonstrated that IOTA models perform well, regardless of the risk threshold or menopausal status of the patients, and that IOTA ADNEX and SRRisks are the most clinically useful models available for the classification of adnexal pathology prior to surgery.

Disclosure of Potential Conflicts of Interest

T. Bourne reports receiving commercial research support from Samsung Medison and Roche Diagnostics. No potential conflicts of interest were disclosed by the other authors.

Disclaimer

The sponsors had no role in study design, in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the work for publication. The views expressed are those of the authors and not necessarily those of the NHS, NIHR, or Department of Health.

Authors’ Contributions

Conception and design: L. Wynants, D. Timmerman, A. Testa, T. Bourne, L. Valentin, B. Van Calster

Development of methodology: L. Wynants, J.Y. Verbakel, B. Van Calster

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): D. Timmerman, A. Testa, L. Savelli, D. Fischero, D. Franchi, C. Van Holbeke, E. Epstein, S. Guarnerio, A. Rossi, R. Fruscio, F.P.G. Leone, L. Valentin

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L. Wynants, J.Y. Verbakel, W. Froyman, L. Valentin, B. Van Calster

Writing, review, and/or revision of the manuscript: L. Wynants, D. Timmerman, J.Y. Verbakel, A. Testa, L. Savelli, D. Fischero, E. Epstein, W. Froyman, S. Guarnerio, T. Bourne, L. Valentin, B. Van Calster

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): D. Timmerman, C. Van Holbeke, W. Froyman, B. Van Calster

Study supervision: D. Timmerman, T. Bourne, L. Valentin, B. Van Calster

Other (guarantor): T. Bourne, D. Timmerman

Acknowledgments

We thank Andrew Vickers for his critical assessment of our work and the suggested improvements. We thank the patients who gave consent to participate in this research.

Grant Support

This study was supported by the Flemish government [Research Foundation–Flanders (FWO) projects G043212N and G084716N, Flanders’ Agency for Innovation by Science and Technology (IWT) project IWT-TBM 070706-IOTA]. Internal Funds KU Leuven for postdoctoral research mandate from Interuniversitair Fonds KU Leuven/Innereuniversitair Fonds KU Leuven. T. Bourne was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. L. Valentin was supported by the Swedish Medical Research Council (grants 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 Skåne University Hospital, Allmänna

Figure 4.

Decision curves representing the NB of the RMI with cutoff 200 (RMI 200), the IOTA logistic regression model 2 (LR2), the Risk of Ovarian Malignancy Index (ROMA), referring all and referring none of the patients to specialized oncology care, for risk thresholds between 5% and 50% (n = 360, 40% malignant tumors). Number of FP for 1 TP

Clinical Research
Clinical Utility of Diagnostic Models for Adnexal Tumors

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 December 27, 2016; revised March 1, 2017; accepted May 9, 2017; published OnlineFirst May 16, 2017.

References


Clinical Cancer Research

Clinical Utility of Risk Models to Refer Patients with Adnexal Masses to Specialized Oncology Care: Multicenter External Validation Using Decision Curve Analysis

Laure Wynants, Dirk Timmerman, Jan Y. Verbakel, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-16-3248

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2017/05/16/1078-0432.CCR-16-3248.DC1

Cited articles
This article cites 41 articles, 6 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/23/17/5082.full#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/23/17/5082.full#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, use this link
http://clincancerres.aacrjournals.org/content/23/17/5082.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.