Accepted Manuscript

Predicting the risk of malignancy in adnexal masses based on the *Simple Rules* from the International Ovarian Tumor Analysis (IOTA) group

Dirk Timmerman, MD, PhD, Ben Van Calster, MSc, PhD, Antonia Testa, MD, PhD, Luca Savelli, MD, PhD, Daniela Fischerova, MD, PhD, Wouter Froyman, MD, Laure Wynants, MSc, Caroline Van Holsbeke, MD, PhD, Elisabeth Epstein, MD, PhD, Dorella Franchi, MD, Jeroen Kaijser, MD, PhD, Artur Czekierdowksi, MD, PhD, Stefano Guerriero, MD, PhD, Robert Fruscio, MD, PhD, Francesco PG. Leone, MD, Alberto Rossi, MD, Chiara Landolfo, MD, Ignace Vergote, MD, PhD, Tom Bourne, MD, PhD, Lil Valentin, MD, PhD

PII: S0002-9378(16)00009-0

DOI: 10.1016/j.ajog.2016.01.007

Reference: YMOB 10876

To appear in: *American Journal of Obstetrics and Gynecology*

Received Date: 3 November 2015

Revised Date: 5 January 2016

Accepted Date: 5 January 2016

Please cite this article as: Timmerman D, Van Calster B, Testa A, Savelli L, Fischerova D, Froyman W, Wynants L, Van Holsbeke C, Epstein E, Franchi D, Kaijser J, Czekierdowksi A, Guerriero S, Fruscio R, Leone FP, Rossi A, Landolfo C, Vergote I, Bourne T, Valentin L, Predicting the risk of malignancy in adnexal masses based on the *Simple Rules* from the International Ovarian Tumor Analysis (IOTA) group, *American Journal of Obstetrics and Gynecology* (2016), doi: 10.1016/j.ajog.2016.01.007.

Predicting the risk of malignancy in adnexal masses based on the *Simple Rules* from the International Ovarian Tumor Analysis (IOTA) group

Dirk TIMMERMAN, MD, PhD[1,2,*], Ben VAN CALSTER, MSc, PhD[1,*], Antonia TESTA, MD, PhD[3], Luca SAVELLI, MD, PhD[4], Daniela FISCHEROVA, MD, PhD[5], Wouter FROYMAN, MD[1,2], Laure WYNANTS, MSc[6,7], Caroline VAN HOLSBEKE, MD, PhD[2,8], Elisabeth EPSTEIN, MD, PhD[9], Dorella FRANCHI, MD[10], Jeroen KAIJSER, MD, PhD[2,11], Artur CZEKIERDOWKSI, MD, PhD[12], Stefano GUERRIERO, MD, PhD[13], Robert FRUSCIO, MD, PhD[14], Francesco PG LEONE, MD[15], Alberto ROSSI, MD[16], Chiara LANDOLFO, MD[1,2], Ignace VERGOTE, MD, PhD[2,17], Tom BOURNE, MD, PhD[1,2,18], Lil VALENTIN, MD, PhD[19]

* Joint first authors

1 KU Leuven, Department of Development and Regeneration, Leuven, Belgium; 2 Department of Obstetrics and Gynecology, University Hospitals Leuven, Leuven, Belgium; 3 Department of Oncology, Catholic University of the Sacred Heart, Rome, Italy; 4 Department of Obstetrics and Gynecology, S Orsola-Malpighi Hospital, University of Bologna, Bologna, Italy; 5 Gynecological Oncology Center, Department of Obstetrics and Gynecology, Charles University, Prague, Czech Republic; 6 KU Leuven Department of Electrical Engineering-ESAT, STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Leuven, Belgium; 7 KU Leuven iMinds Medical IT Department, Leuven, Belgium; 8 Department of Obstetrics and Gynecology, Ziekenhuis Oost-Limburg, Genk, Belgium; 9 Department of Obstetrics and Gynecology, Karolinska University Hospital, Stockholm, Sweden; 10 Preventive Gynecology Unit, Division of Gynecology, European Institute of Oncology, Milan, Italy; 11 Department of Gynecology and Obstetrics, Ikazia Hospital, Rotterdam, the Netherlands; 12 1st Department of Gynecological Oncology and Gynecology, Medical University in Lublin, Lublin, Poland; 13 Department of Obstetrics and Gynecology, Azienda Ospedaliero Universitaria di Cagliari, Cagliari, Italy; 14 Clinic of Obstetrics and Gynecology, University of Milan-Bicocca, San Gerardo Hospital, Monza, Italy; 15 Department of Obstetrics and Gynecology, Clinical Sciences Institute L. Sacco, University of Milan, Italy; 16 Department of Obstetrics and Gynecology, University of Udine, Udine, Italy; 17 KU Leuven, Department of Oncology, Leuven, Belgium; 18 Queen Charlotte's and Chelsea Hospital, Imperial College, London, UK; 19 Department of Obstetrics and Gynecology, Skåne University Hospital Malmö, Lund University, Malmö, Sweden

We wish Table 10 to have published in the print issue of the Journal.

Conflicts of Interest

All authors declare: no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Source of Funding

This study was supported by the Flemish government [Research Foundation–Flanders (FWO) project G049312N, Flanders' Agency for Innovation by Science and Technology (IWT) project IWT-TBM 070706-IOTA3, and iMinds 2015] and Internal Funds KU Leuven (project C24/15/037). LW is a doctoral fellow of IWT. DT is a senior clinical investigator of FWO. TB is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. The views expressed are those of the authors and not necessarily those of the NHS, NIHR or Department of Health. LV is 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 Sjukhusets i Malmö Stiftelse för bekämpande av cancer (the Malmö General Hospital Foundation for fighting against cancer), and two Swedish governmental grants (ALF-medel and Landstingsfinansierad Regional Forskning). 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 researchers performed this work independently of the funding sources.

Corresponding author:
Dirk Timmerman, MD PhD
KU Leuven, Department of Development and Regeneration, Leuven, Belgium; Department of Obstetrics and Gynecology, University Hospitals Leuven, Leuven, Belgium
Address:
Department of Obstetrics and Gynecology,
University Hospitals Leuven
Herestraat 49
3000 Leuven
Belgium
tel. +32 16 3 44210 or +32 16 3 44216
fax +32 16 3 44205
email: dirk.timmerman@uzleuven.be

Word count abstract: 398
Word count main text: 3263

Condensation and short version of title

Condensation
Ultrasound assessment of adnexal masses using a set of *Simple Rules* allows calculation of the risk of malignancy.

Short version of title
Simple ultrasound rules to predict the risk of malignancy in adnexal masses

Abstract

Background – Accurate methods to preoperatively characterize adnexal tumors are pivotal for optimal patient management. A recent meta-analysis concluded that the International Ovarian Tumor Analysis (IOTA) algorithms such as the *Simple Rules* are the best approaches to preoperatively classify adnexal masses as benign or malignant.

Objective - To develop and validate a model to predict the risk of malignancy in adnexal masses using the ultrasound features in the *Simple Rules*.

Study Design – International cross-sectional cohort study involving 22 oncology centers, referral centers for ultrasonography, and general hospitals. We included consecutive patients with an adnexal tumor who underwent a standardized transvaginal ultrasound examination and were selected for surgery. Data on 5020 patients were recorded in three phases between 2002 and 2012. The five *Simple Rules* features indicative of a benign tumor (B-features) and the five features indicative of malignancy (M-features) are based on the presence of ascites, tumor morphology, and degree of vascularity at ultrasonography. Gold standard was the histopathologic diagnosis of the adnexal mass (pathologist blinded to ultrasound findings). Logistic regression analysis was used to estimate the risk of malignancy based on the ten ultrasound features and type of center. The diagnostic performance was evaluated by area under the receiver operating characteristic curve (AUC), sensitivity, specificity, positive and negative likelihood ratios (LR+, LR-), positive and negative predictive values (PPV, NPV) and calibration curves.

Results – Data on 4848 patients were analyzed. The malignancy rate was 43% (1402/3263) in oncology centers and 17% (263/1585) in other centers. The AUC on validation data was very similar in oncology centers (0.917, 95% CI 0.901 to 0.931) and other centers (0.916, 95% CI 0.873 to 0.945). Risk estimates showed good calibration. 23% of patients in the validation data set had a very low estimated risk (<1%), 48% had a high estimated risk (≥30%). For the 1% risk cutoff, sensitivity was 99.7%, specificity 33.7%, LR+ 1.5, LR- 0.010, PPV 44.8% and NPV 98.9%. For the 30% risk cutoff, sensitivity was 89.0%, specificity 84.7%, LR+ 5.8, LR- 0.13, PPV 75.4% and NPV 93.9%.

Conclusion – Quantification of the risk of malignancy based on the *Simple Rules* has good diagnostic performance both in oncology centers and other centers. A simple classification based on these risk estimates may form the basis of a clinical management system. Patients with a high risk may benefit from surgery by a gynecological oncologist, while patients with a lower risk may be managed locally.

Keywords: adnexa; color Doppler; diagnosis; diagnostic algorithm; IOTA; logistic regression analysis; ovarian cancer; ovarian neoplasms; preoperative evaluation; risk assessment; *Simple Rules*; ultrasonography

Introduction

Ovarian cancer is a common and lethal disease for which early detection and treatment in high volume centers and by specialized clinicians is known to improve survival.[1–4] Hence, accurate methods to preoperatively characterize the nature of an ovarian tumor are pivotal. In 2008 the International Ovarian Tumor Analysis (IOTA) group described the *Simple Rules*.[5] These are based on a set of five ultrasound features indicative of a benign tumor (B-features) and five ultrasound features indicative of a malignant tumor (M-features). When using the *Simple Rules*, tumors are classified as benign if only B-features are observed and as malignant if only M-features are observed. If no features are observed or if conflicting features are present, the *Simple Rules* cannot classify the tumor as benign or malignant (inconclusive results). Masses in which the *Simple Rules* yield an inconclusive result can be classified using subjective assessment by an experienced ultrasound operator, or given the high prevalence of malignancy in this group they can all be classified as malignant to increase the sensitivity for ovarian cancer.[6] On prospective validation both by the IOTA group (two studies including 1938 and 2403 patients, respectively)[7,8] and by other research teams (nine studies including a total of 2101 tumors)[9–17], the *Simple Rules* were applicable in 77 to 94% of tumors (range between studies). The malignancy rate ranged from 1% to 9% in cases classified as benign, from 69% to 94% in cases classified as malignant, and from 13% to 53% in inconclusive cases. In a meta-analysis comparing the ability of 19 methods to discriminate between benign and malignant adnexal masses before surgery, the *Simple Rules* had a sensitivity of 93% and a specificity of 81% when classifying inconclusive tumors as malignant.[18] In the meta-analysis the *Simple Rules* and the

IOTA logistic regression model LR2[19] were superior to all other methods. This suggests that evidence-based approaches to the preoperative characterization of adnexal masses should incorporate the use of *Simple Rules* or the LR2 model. LR2 is a mathematical risk prediction model based on age and five ultrasound variables (presence of blood flow in a papillary structure, irregular cyst walls, ascites, acoustic shadows and maximum diameter of the largest solid component).

The *Simple Rules* have been well-received by clinicians, and the Royal College of Obstetricians and Gynecologists in the United Kingdom have included the *Simple Rules* in their Green Top guideline on the assessment and management of ovarian masses in premenopausal women.[20]

Despite a combination of simplicity and excellent performance, important limitations of the *Simple Rules* are the inconclusive results in a proportion of cases and the absence of an estimated risk of malignancy. The ability to provide accurate risk estimates is highly relevant for risk stratification and individualized patient management. The objective of this study was to develop and validate a model to calculate the risk of malignancy in adnexal masses based on the ten ultrasound features in the *Simple Rules*.

Materials and Methods

Study design and setting

This international multicenter cross-sectional cohort study involves patients from 22 centers (oncology centers and other hospitals, Table 1) with at least one adnexal (ovarian, para-ovarian, or tubal) tumor selected for surgery by the managing clinician. Exclusion criteria were (1) pregnancy at the time of examination, (2) refusal of transvaginal ultrasonography, (3) declining participation, and (4) surgical intervention more than 120 days after the ultrasound examination. Data collection was carried out within the framework of the IOTA collaboration. The primary aim of the IOTA studies is to develop and validate methods for making a correct diagnosis in adnexal tumors prior to surgery. This aim is pursued by prospectively examining a large number of patients with ultrasound using a standardized examination technique and standardized terms and definitions to describe ultrasound findings.[21] Through consecutive phases, data were collected from 24 centers in ten countries. In phase 1 data were collected between 1999 and 2002, in phase 1b between 2002 and 2005, in phase 2 between 2005 and 2007, and in phase 3 between 2009 and 2012. Data from phase 1 were used to develop the *Simple Rules* and were therefore not used in the present study. The research protocols were approved by the Ethics Committees in each contributing center.

Data collection

Oral and/or written informed consent was obtained in accordance with the requirements of the local Ethics Committee. A standardized history was taken from

each patient to collect clinical information. All patients underwent a standardized transvaginal ultrasound examination by a principal investigator, who was a gynecologist or radiologist with extensive experience in gynecological ultrasound and with a special interest in adnexal masses. Transabdominal sonography was added in women with large masses that could not be visualized completely by the transvaginal approach. For women with multiple masses, the dominant mass was selected for statistical analysis.[8,19,21–24] To apply the *Simple Rules*, information on the following variables is required: the diameters of the lesion (mm), the diameters of the largest solid component (mm), type of tumor (unilocular, unilocular-solid, multilocular, multilocular-solid, solid), presence of wall irregularity, ascites, acoustic shadows, number of papillary structures, and the color score, the latter reflecting vascularization on Doppler ultrasound (1, no flow; 2, minimal flow; 3, moderate flow; 4, very strong flow). Detailed information can be found in previous reports.[8,19,21–24] The five B-features and the five M-features were not directly recorded, but were derived from the variables described above.

Reference standard

The reference standard denotes whether the tumor is benign or malignant based on the histopathologic diagnosis of the tumor following surgical removal. Surgery was performed through laparoscopy or laparotomy, as considered appropriate by the surgeon. Excised tumor tissues were histologically examined at the local center. Histological classification was performed without knowledge of the ultrasound results. Borderline tumors were classified as malignant.

Statistical analysis

Using the IOTA data from phases 1b and 2, we estimated the risk of malignancy by quantifying the predictive value of each of the ten features of the *Simple Rules* and of the type of center in which the patients underwent an ultrasound examination (oncology center vs. other hospital; the definition of oncology center being tertiary referral center with a specific gynecological oncology unit). The predictive values for malignancy were determined by the regression coefficients estimated by multivariable logistic regression. Interaction terms were not considered. The analysis included a random intercept to account for variability between centers.[25]

The risk estimates were externally validated on IOTA phase 3 data. The area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and predictive values were calculated through a meta-analysis of center-specific results[26], similar to a previous validation study using phase 3 data.[8] Positive and negative likelihood ratios were derived from these results. The risk cut-offs considered to classify a mass as malignant were 1%, 3%, 5%, 10%, 15%, 20%, 25% and 30%. Calibration plots were constructed to assess the relationship between calculated risks and observed proportions.[25,27]

After external validation, the risk calculation was updated using the same procedure but now using all available data (phases 1b, 2, and 3) to fully exploit all available information.

Results

During IOTA phases 1b, 2 and 3, data on 5020 patients were recorded at 22 centers (two centers from IOTA phase 1 did not take part in later phases). Data on 172 patients were excluded either because the patients fulfilled an exclusion criterion (n=124; 43 women were pregnant and 81 women were operated on more than 120 days after the ultrasound examination), or because of data errors or uncertain/missing final histology (n=47) or protocol violation (n=1). This leaves data on 4848 patients (Table 1, Table 2 and Table 3). The development set (phases 1b and 2) contains data on 2445 patients recruited at 11 oncology centers (n=1548) and eight other centers (n=897). The temporal validation set (phase 3) contains data on 2403 patients recruited at 11 oncology centers (n=1715) and seven other centers (n=688).

The malignancy rate was 34% (1665/4848) overall, 43% (1402/3263) in oncology centers and 17% (263/1585) in other centers. The observed malignancy rate varied between 22% and 66% at oncology centers and between 0% and 30% at other centers. The median age was 42 years (interquartile range 32 to 54) for patients with a benign tumor and 57 years (47-66) for patients with a malignant tumor. All 80 observed combinations of the ultrasound features in the *Simple Rules* are listed in Table 4. For the same combination of features, the observed malignancy rate was usually higher in oncology centers than in other centers.

Results for the development set (n=2445)

The coefficients of the regression analysis for the development data are presented in Table 5. B-features were allocated negative coefficients, and hence decrease the estimated risk of malignancy. M-features were given positive coefficients. Ultrasound examination in oncology centers was assigned a positive coefficient. The AUC of the risk estimates to predict malignancy was 0.928 (95% confidence interval (CI) 0.913 to 0.940). The AUC was similar in oncology centers (0.926, 95% CI 0.910 to 0.940) and other centers (0.937, 95% CI 0.896 to 0.963).

Results for the validation set (n=2403)

When externally validated, the AUC was 0.917 (95% CI 0.902 to 0.930) (Figure 1a). The AUC was very similar in oncology centers (0.917, 95% CI 0.901 to 0.931) and in other centers (0.916, 95% CI 0.873 to 0.945). In all but three centers, the AUC was at least 0.90. Two centers had an AUC of 0.89 and one small center had an AUC below 0.80 (Figure 2). The estimated risks were well calibrated in all validation patients (Figure 1b) and when assessed for patients from oncology centers and other hospitals separately (Figure 3).

22.8% of the patients in the validation set had a calculated risk of malignancy <1%, while 48.5% had a calculated risk ≥30%. For the 1% calculated risk cutoff, sensitivity was 99.7%, specificity 33.7%, LR+ 1.5, LR- 0.010, PPV 44.8% and NPV 98.9%. For the 30% calculated risk cutoff, sensitivity was 89.0%, specificity 84.7%, LR+ 5.8, LR- 0.13, PPV 75.4% and NPV 93.9% (Table 6). Sensitivity, specificity, PPV, NPV, LR+ and LR- for the same risk cutoff differed between oncology centers and other centers (Table 7).

Results for the total dataset

The regression coefficients for the updated analysis on all data (n=4848) are shown in Table 8. Feature B1 (unilocular cyst) was most predictive of a benign tumor (coefficient -3.4), while feature B3 (acoustic shadows) was least predictive (coefficient -1.7). Feature M2 (ascites) was most predictive of malignancy (coefficient 2.7) and feature M4 (irregular multilocular-solid tumor with largest diameter ≥ 100mm) was least predictive (coefficient 1.0). Type of center had a coefficient of 0.9.

For example, consider a patient examined at an oncology center and in whom features B3, M2, and M5 are present. This patient has a regression score of -0.97 (intercept) – 1.66 (B3) + 2.65 (M2) + 1.55 (M5) + 0.92 (oncology center) = 2.49. The estimated risk of malignancy is 92.3%. Further details on this calculation are given in Table 8.

For patients classified as benign by the original *Simple Rules* approach (i.e. only B-features present) we observed estimated risks between <0.01% and 15.2% (in oncology centers: <0.01% to 15.2%; in other hospitals: <0.01% to 6.7%), and for patients classified as malignant (only M-features present) between 50.2% and >99.9% (in oncology centers: 71.7 to >99.9%; in other hospitals: 50.2 to 99.7%). For tumors classified as inconclusive by the original *Simple Rules* approach (i.e. no features or conflicting features present), we observed estimated risks between 1.3% and 99.1% (in oncology centers: 1.3 to 99.1%; in other hospitals: 2.1 to 88.2%), demonstrating the heterogeneity of this group.

Table 9 summarizes the range of estimated risks for individual patients depending on the number of B-features and M-features present in the tumor, based on the updated analysis (n=4848). In general, the estimated risk of malignancy was at least 42.0% if more M-features than B-features were present (N=1295, 27% of all tumors) and was at most 0.29% when two or more B-features and no M-features were present (N=175, 3.6% of all tumors). The estimated risk when no feature was present was 48.7% for patients from oncology centers and 27.5% for patients from other centers (N=954, 20% of all tumors). Patients with conflicting features (one or more B-feature and one or more M-feature) were uncommon (N=161, 3.3% of all tumors). The type of feature is most important in patients with only one B-feature and no M-features: estimated risks vary between 1.2 and 15.2%. Based on these results a simple classification of adnexal masses based on the number of B-and M-features present can be used (Table 10).

Comment

Principal findings of the study

In this study we have developed a method to estimate the individual risk of malignancy in an adnexal mass using the ultrasound features in the IOTA *Simple Rules*. On prospective validation the risk estimates showed good ability to discriminate between benign and malignant tumors (AUC 0.917) and good agreement between the calculated risks of malignancy and the true prevalence of malignancy.

Implications of the work

The *Simple Rules* are intuitively attractive because of their ease of use.[9–17,20] However, when used as originally suggested they allow only a categorization of tumors into three groups: benign, malignant or inconclusive. In this study we show that the *Simple Rules* can also be used to estimate the risk of malignancy in every adnexal mass and so can be used for individualized patient management. The type of center also needed to be included in our risk estimation, because the risk of a malignant tumor is higher in oncology centers than in others. The B-feature B1 (unilocular cyst) was most predictive of a benign tumor, while the B-feature B3 (acoustic shadows) was least predictive. The M-feature M2 (ascites) was most predictive of malignancy while the M-feature M4 (irregular multilocular-solid tumor with largest diameter ≥ 100mm) was least predictive. Many clinicians would probably agree that conservative management could be an option for tumors with a very low risk of malignancy (e.g. <1%), provided that this is appropriate when taking clinical

circumstances into account. In the current study 23% of the validation patients fell into this group (16% of patients in oncology centers and 31% of patients in other centers). Some clinicians might consider conservative management also for patients with a risk of malignancy <3% (32% of the validation patients in the study), at least if the patient is asymptomatic and if she is seen in a non-oncology center. On the other hand, most clinicians would probably agree that patients with a risk of malignancy ≥30% would benefit from being referred to a gynecologic oncology center for further investigation and treatment. In the current study, 48% of the validation patients belonged to this high-risk group (61% of patients in oncology centers and 18% of patients in other centers). Patients with intermediate risks could be managed differently depending on local circumstances, e.g. depending on whether there is liberal or restricted access to ultrasound experts or gynecologic oncologic surgery. Some might want to operate on patients with intermediate risks in regional centers or refer such patients for second opinion ultrasonography by an expert.

The coefficients can be used to calculate a reliable and well-calibrated individual risk estimate. Using Table 9, this risk of malignancy can be directly read off for 97% of all patients without the need for a computer. The other 3% of patients have tumors with both M-features and B-features, for these patients the precise individual risk estimate needs to be calculated using a computer or mobile app. However, they all belong to the elevated risk and very high-risk groups. Table 10 shows an even simpler classification of patients into different risk groups. Our results may lay the basis for a clinically useful imaging and management system such as the GI-RADS (Gynecologic Imaging Reporting and Data System) system[28], as shown in Table 9 and 10. While the GI-RADS system is based on subjective assessment of ultrasound

images, this new system would be based on more objective ultrasound criteria and type of center.

The *Simple Rules* risk classification is an alternative to other algorithms such as the Risk of Malignancy Index (RMI)[29], the Risk of Ovarian Malignancy Algorithm (ROMA)[30], OVA-1[31,32] and the IOTA logistic regression models (LR1, LR2[19], ADNEX[24]). Three studies have compared the IOTA methods with RMI and ROMA on the same study population.[8,12,33,34] LR2 and the *Simple Rules* (classifying inconclusive cases as malignant) reached higher diagnostic accuracies than RMI [8,12,33] and LR2 outperformed ROMA.[34] These findings were confirmed in a systematic review and meta-analysis comparing the diagnostic performance of 19 prediction models.[18] The multivariate index assay OVA-1 has been validated by two large multicenter studies in the USA.[31,32] OVA-1 has never been compared with IOTA algorithms on the same set of patients, but it seems to have lower specificity at similar sensitivity, resulting in much higher rates of false-positives.[35,36]

When prospectively validated on IOTA phase 3 data (i.e. on the validation set in the present study), the *Simple Rules* risk estimates, LR2, and subjective assessment (using six levels of diagnostic confidence) had similar diagnostic performance in terms of discrimination between benign and malignant tumors: the AUC for LR2 was 0.918 (95% CI 0.905-0.930)[8], for subjective assessment 0.914 (95% CI 0.886-0.936)[8] and for the *Simple Rules* risk estimate 0.917 (95%CI 0.902-0.930). The discriminative ability of the ADNEX model was slightly better: AUC 0.943 (95%CI 0.934 to 0.952).[24] The ADNEX model has the advantage over the other methods of not only differentiating benign from malignant disease but also giving risk estimates for four subgroups of malignant disease (borderline tumors, stage I invasive ovarian cancer,

stage II-IV invasive ovarian cancer and metastases in the ovaries from other primary tumors).[24,37]

Because CA-125 is not used as a variable in the *Simple Rules*, it is not included in the *Simple Rules* risk classification. However, adding information on serum CA-125 levels to ultrasound information does not seem to improve mathematical models to discriminate between benign and malignant adnexal masses.[38]

Instead of using an algorithm, experienced examiners might still prefer to give an instant diagnosis using the IOTA *Easy Descriptors*. This is feasible in 42 - 46% of patients.[8,39,40] The *Easy Descriptors* apply to endometriomas, dermoid cysts, simple cysts, and obvious malignancies.[39]

In future studies, the *Simple Rules* risk estimates need to be prospectively and externally validated, and their use in a classification system for clinical management has to be investigated.

Strengths and weaknesses

The strength of this study is the use of a large multinational database in which patients were prospectively collected using well-defined terms, definitions and measurements. After development and temporal validation, the risk calculation was updated using all 4848 patients. The large sample size is likely to yield generalizable results.

The study also has limitations. First, our risk calculation model was developed and validated exclusively on patients who underwent surgery. This is because we found it necessary to use the histological diagnosis as the gold standard. Second, all ultrasound examiners in the study were experienced, and so our results may not be applicable with less experienced operators. However, published studies have shown that the *Simple Rules* retain their performance in the hands of less-experienced examiners.[10–12,14–17] This is likely to be also true of our *Simple Rules* risk calculation system, because the same ultrasound variables were used to calculate the risks.

Conclusions

We conclude that individual risk estimates can be derived from the ten ultrasound features in the *Simple Rules* with performance similar to the best previously published algorithms. A simple classification based on these risk estimates may form the basis of a clinical management system. This will hopefully facilitate choosing optimal treatment for all patients presenting with adnexal masses.

Acknowledgments

Contributors: DT and BVC conceived and designed the study, with additional support from WF, LW, TB and LV. DT, AT, LS, DFi, CVH, EE, DFr, JK, AC, SG, RF, FL, AR and LV enrolled patients and acquired data. DT, BVC and JK were involved in data cleaning. BVC analyzed the data, with support from LW. DT, BVC, AT, WF, LW, CL, TB and LV were involved in data interpretation. DT, BVC, WF, LW, CL, TB and LV wrote the first draft of the manuscript, which was then critically reviewed and revised by the other coauthors. All authors approved the final version of the manuscript for submission. DT and BVC had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Woo YL, Kyrgiou M, Bryant A, Everett T, Dickinson H. Centralisation of services for gynaecological cancers - A Cochrane systematic review. Gynecol Oncol 2012;126(2):286-290.

2. Engelen MJA, Kos HE, Willemse PHB, et al. Surgery by consultant gynecologic oncologists improves survival in patients with ovarian carcinoma. Cancer 2006;106(3):589-598.

3. Vernooij F, Heintz APM, Witteveen PO, et al. Specialized care and survival of ovarian cancer patients in The Netherlands: Nationwide cohort study. J Natl Cancer Inst 2008;100(6):399-406.

4. Earle CC, Schrag D, Neville BA, et al. Effect of surgeon specialty on processes of care and outcomes for ovarian cancer patients. J Natl Cancer Inst 2006;98(3):172-180.

5. Timmerman D, Testa AC, Bourne T, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. Ultrasound Obstet Gynecol 2008;31(6):681-690.

6. Kaijser J, Bourne T, Valentin L, et al. Improving strategies for diagnosing ovarian cancer: A summary of the International Ovarian Tumor Analysis (IOTA) studies. Ultrasound Obstet Gynecol 2013;41(1):9-20.

7. Timmerman D, Ameye L, Fischerova D, et al. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. BMJ 2010;341:c6839.

8. Testa AC, Kaijser J, Wynants L, et al. Strategies to diagnose ovarian cancer: new evidence from phase 3 of the multicentre international IOTA study. Br J Cancer 2014;111(4):680-688.

9. Fathallah K, Huchon C, Bats AS, et al. Validation externe des critères de Timmerman sur une série de 122 tumeurs ovariennes. Gynecol Obstet Fertil 2011;39(9):477-481.

10. Hartman CA, Juliato CRT, Sarian LO, et al. Ultrasound criteria and CA 125 as predictive variables of ovarian cancer in women with adnexal tumors. Ultrasound Obstet Gynecol 2012;40(3):360-366.

11. Alcázar JL, Pascual MÁ, Olartecoechea B, et al. IOTA simple rules for discriminating between benign and malignant adnexal masses: Prospective external validation. Ultrasound Obstet Gynecol 2013;42(4):467-471.

12. Sayasneh A, Wynants L, Preisler J, et al. Multicentre external validation of IOTA prediction models and RMI by operators with varied training. Br J Cancer 2013;108(12):2448-2454.

13. Tantipalakorn C, Wanapirak C, Khunamornpong S, Sukpan K, Tongsong T. IOTA Simple Rules in Differentiating between Benign and Malignant Ovarian Tumors. Asian Pacific J Cancer Prev 2014;15:5123-5126.

14. Nunes N, Ambler G, Foo X, Naftalin J, Widschwendter M, Jurkovic D. Use of IOTA simple rules for diagnosis of ovarian cancer: meta-analysis. Ultrasound Obstet Gynecol 2014;44(5):503-514.

15. Tinnangwattana D, Vichak-ururote L, Tontivuthikul P, Charoenratana C, Lerthiranwong T, Tongsong T. IOTA Simple Rules in Differentiating between Benign and Malignant Adnexal Masses by Non-expert Examiners. Asian Pacific J Cancer Prev 2015;16:3835-3838.

16. Ruiz de Gauna B, Rodriguez D, Olartecoechea B, et al. Diagnostic performance of IOTA simple rules for adnexal masses classification: a comparison between two centers with different ovarian cancer prevalence. Eur J Obstet Gynecol Reprod Biol 2015;191:10-14.

17. Knafel A, Banas T, Nocun A et al. The Prospective External Validation of International Ovarian Tumor Analysis (IOTA) Simple Rules in the Hands of Level I and II Examiners. Ultraschall Med 10.1055/s-0034-1398773

18. Kaijser J, Sayasneh A, Van hoorde K, et al. Presurgical diagnosis of adnexal tumours using mathematical models and scoring systems: A systematic review and meta-analysis. Hum Reprod Update 2014;20(3):449-462.

19. Timmerman D, Testa AC, Bourne T, et al. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: A multicenter study by the International Ovarian Tumor Analysis Group. J Clin Oncol 2005;23(34):8794-8801.

20. Royal College Of Obstetricians And Gynaecologists, British Society Of Gynaecological Endoscopy. Management of Suspected Ovarian Masses in Premenopausal Women. Green Top Guidelines 62, 2011 https://www.rcog.org.uk/en/guidelines-research-services/guidelines/gtg62/

21. Timmerman D, Valentin L, Bourne T, et al. 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(5):500-505.

22. Van Holsbeke C, Van Calster B, Testa AC, 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(2):684-691.

23. Timmerman D, Van Calster B, Testa AC, 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(2):226-234.

24. Van Calster B, Van Hoorde K, Valentin L, et al. Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study. BMJ 2014;349:g5920-g5920.

25. Bouwmeester W, Twisk JWR, Kappen TH, Van Klei WA, Moons KGM, Vergouwe Y. Prediction models for clustered data: comparison of a random intercept and standard regression model. BMC Med Res Methodol 2013;13(1):19.

26. Van Klaveren D, Steyerberg EW, Perel P, Vergouwe Y. Assessing discriminative ability of risk models in clustered data. BMC 2014;14:5(c):1-10.

27. Steyerberg EW: Clinical prediction models: a practical approach to development, validation, and updating. New York, NY, Springer, 2009

28. Amor F, Vaccaro H, Alcazar JL, Leon M, Craig JM, Martinez J. Gynecologic Imaging Reporting and Data System. J Ultrasound Med 2009;28(3):285-291.

29. Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. Br J Obs Gynaecol 1990;97:922-929.

30. Moore RG, Brown AK, Miller MC, et al. The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass. Gynecol Oncol 2008;108(2):402-408.

31. Ueland FR, Desimone CP, Seamon LG, et al. Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. Obstet Gynecol 2011;117(6):1289-1297.

32. Bristow RE, Smith A, Zhang Z, et al. Ovarian malignancy risk stratification of the adnexal mass using a multivariate index assay. Gynecol Oncol 2013;128(2):252-259.

33. Van Holsbeke C, Van Calster B, Bourne T, et al. External validation of diagnostic models to estimate the risk of malignancy in adnexal masses. Clin Cancer Res 2012;18(3):815-825.

34. Kaijser J, Van Gorp T, Van Hoorde K, et al. A comparison between an ultrasound based prediction model (LR2) and the Risk of Ovarian Malignancy

Algorithm (ROMA) to assess the risk of malignancy in women with an adnexal mass. Gynecol Oncol 2013;129(2):377-383.

35. Ware Miller R, Smith A, DeSimone CP, et al. Performance of the American College of Obstetricians and Gynecologists' ovarian tumor referral guidelines with a multivariate index assay. Obstet Gynecol 2011;117:1298–306.

36. Timmerman D, Van Calster B, Vergote I, et al. Performance of the American College of Obstetricians and Gynecologists' Ovarian Tumor Referral Guidelines With a Multivariate Index Assay. Obstet Gynecol 2011;118(5):1179-1181.

37. Van Calster B, Van Hoorde K, Froyman W, et al. Practical guidance for applying the ADNEX model from the IOTA group to discriminate between different subtypes of adnexal tumors. Facts Views Vis Obgyn 2015;7(1):32-41.

38. Timmerman D, Van Calster B, Jurkovic D, et al. Inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and malignant adnexal tumors. J Clin Oncol 2007;25:4194-4200.

39. Ameye L, Timmerman D, Valentin L, et al. Clinically oriented three-step strategy for assessment of adnexal pathology. Ultrasound Obstet Gynecol 2012;40(5):582-591.

40. Sayasneh A, Kaijser J, Preisler J, et al. A multicenter prospective external validation of the diagnostic performance of IOTA simple descriptors and rules to characterize ovarian masses. Gynecol Oncol 2013;130(1):140-146.

Table 1. Sample size, prevalence of malignancy and outcome of the *Simple Rules* in the 22 participating centers (n = 4848).

				Classification using the *Simple Rules*		
Center	**Dataset**	**Patients**	**Malignant N (%)**	**SR Benign N (%mal)**	**SR Inconclusive N (%mal)**	**SR Malignant N (%mal)**
Oncology centers		3263	1402 (43)	1436 (5)	788 (49)	1039 (90)
Leuven, Belgium	D,V	668	242 (36)	306 (4)	153 (35)	209 (85)
Rome, Italy	D,V	661	365 (55)	224 (7)	163 (59)	274 (92)
Monza, Italy	D,V	356	76 (22)	247 (4)	69 (42)	40 (95)
Prague, Czech Republic	D,V	354	234 (66)	102 (13)	109 (77)	143 (96)
Milan, Italy	D,V	312	177 (57)	112 (7)	45 (56)	155 (93)
Lublin, Poland	D,V	285	102 (36)	132 (5)	86 (45)	67 (85)
Bologna, Italy*	V	213	65 (31)	126 (3)	52 (58)	35 (89)
Stockholm, Sweden	V	120	53 (44)	38 (0)	33 (27)	49 (90)
Lund, Sweden	D,V	77	20 (26)	36 (0)	20 (10)	21 (86)
Beijing, China	D	73	16 (22)	36 (0)	20 (15)	17 (76)
London, UK	D	65	25 (38)	32 (6)	18 (50)	15 (93)
Udine, Italy	D,V	64	19 (30)	36 (3)	16 (44)	12 (92)
Naples 2, Italy	D,V	15	8 (53)	9 (22)	4 (100)	2 (100)
Other hospitals		1585	263 (17)	1021 (1)	327 (23)	237 (76)
Malmö, Sweden	D,V	462	100 (22)	205 (0)	146 (12)	111 (74)
Genk, Belgium	D,V	428	61 (14)	301 (1)	67 (21)	60 (73)
Cagliari, Italy	D,V	261	37 (14)	200 (2)	36 (33)	25 (88)
Milan 2, Italy	D,V	136	20 (15)	99 (0)	25 (40)	12 (83)
Bologna, Italy*	D	135	11 (8)	110 (0)	15 (27)	10 (70)
Naples, Italy	D,V	72	18 (25)	42 (2)	17 (35)	13 (85)
Barcelona, Spain	V	37	11 (30)	21 (10)	11 (55)	5 (60)
Milan 3, Italy	D	21	4 (19)	13 (0)	7 (43)	1 (100)
Milan 4, Italy	V	21	0 (0)	20 (0)	1 (0)	0 (-)
Hamilton, Canada	D	12	1 (8)	10 (0)	2 (50)	0 (-)

* The Bologna Center in Italy (BIT) changed from 'other hospital' to 'oncology center' during the course of the IOTA study and is therefore listed both under oncology centers and other hospitals (different patients in the two types of centers).
SR: *Simple Rules*; D: development data; V: validation data; %mal: prevalence of malignancy

Table 2. Ultrasound features of included tumors (n = 4848).

Ultrasound feature	Development (n=2445)		Validation (n=2403)	
	Benign (n=1760)	Malignant (n=685)	Benign (n=1423)	Malignant (n=980)
Maximum lesion diameter (mm)	61 (43-85)	89 (58-136)	64 (47-90)	86 (55.5-126)
Solid components				
Presence of solid components	541 (30.7%)	638 (93.1%)	474 (33.3%)	916 (93.5%)
Maximum diameter if present (mm)	25 (13-47)	54 (35-82)	28 (13-54)	59 (36.5-87)
Number of papillations				
None	1538 (87.4%)	427 (62.3%)	1243 (87.4%)	777 (79.3%)
1	137 (7.8%)	84 (12.3%)	96 (6.8%)	52 (5.3%)
2	35 (2.0%)	23 (3.4%)	31 (2.2%)	31 (3.2%)
3	22 (1.3%)	30 (4.4%)	26 (1.8%)	29 (3.0%)
More than 3	27 (1.5%)	121 (17.7%)	27 (1.9%)	91 (9.3%)
Color score				
1 (No flow)	769 (43.7%)	29 (4.2%)	574 (40.3%)	32 (3.3%)
2 (Minimal flow)	621 (35.3%)	170 (24.8%)	563 (40.0%)	199 (20.3%)
3 (Moderate flow)	331 (18.8%)	298 (43.5%)	239 (16.8%)	442 (45.1%)
4 (Very strong flow)	39 (2.2%)	188 (27.5%)	47 (3.3%)	307 (31.3%)
Type of tumor				
Unilocular	825 (47.0%)	10 (1.5%)	595 (41.8%)	5 (0.5%)
Unilocular-solid	187 (10.7%)	112(16.5%)	141 (9.9%)	117 (11.9%)
Multilocular	390 (22.2%)	37 (5.4%)	354 (24.9%)	59 (6.0%)
Multilocular-solid	196 (11.2%)	268 (39.1%)	179 (12.6%)	326 (33.3%)
Solid	158 (9.0%)	257 (37.5%)	154 (10.8%)	473 (48.3%)
Irregular cyst walls	484 (27.5%)	457 (66.7%)	385 (27.1%)	572 (58.4%)
Ultrasound features of the *Simple Rules*				
B1 (unilocular cyst)	825 (46.9%)	10 (1.5%)	595 (41.8%)	5 (0.5%)
B2 (solid components present, but < 7mm)	44 (2.5%)	5 (0.7%)	40 (2.8%)	2 (0.2%)
B3 (acoustic shadows)	307 (17.4%)	29 (4.2%)	265 (18.6%)	34 (3.5%)
B4 (smooth multilocular tumor, largest diameter < 100mm)	233 (13.2%)	3 (0.4%)	224 (15.7%)	13 (1.3%)
B5 (no blood flow; color score 1)	769 (43.7%)	29 (4.2%)	574 (40.3%)	32 (3.3%)
M1 (irregular solid tumor)	12 (0.7%)	97 (14.2%)	16 (1.1%)	189 (19.3%)
M2 (ascites)	23 (1.3%)	222 (32.4%)	18 (1.3%)	322 (32.9%)
M3 (at least 4 papillary structures)	27 (1.5%)	121 (17.7%)	27 (1.9%)	91 (9.3%)
M4 (irregular multilocular-solid tumor, largest diameter ≥ 100mm)	45 (2.6%)	144 (21.0%)	40 (2.8%)	153 (15.6%)
M5 (very strong flow; color score 4)	39 (2.2%)	188 (27.5%)	47 (3.3%)	307 (31.3%)

Results shown are median (Interquartile range) for continuous variables and N (%) for categorical variables

Table 3. Prevalence of specific pathologies in all patients (n = 4848) and separately for patients from oncology centers and other hospitals.

Tumor pathology	All patients N (%)	Patients from oncology centers N (%)	Patients from other hospitals N (%)
All benign pathologies	*3183 (65.7)*	*1861 (57.0)*	*1322 (83.4)*
Endometrioma	845 (17.4)	456 (14.0)	389 (24.5)
Benign teratoma (dermoid)	512 (10.6)	334 (10.2)	178 (11.2)
Simple/parasalpingeal cyst	285 (5.9)	147 (4.5)	138 (8.7)
Functional cyst	128 (2.6)	69 (2.1)	59 (3.7)
Hydrosalpinx	112 (2.3)	53 (1.6)	59 (3.7)
Peritoneal pseudocyst	34 (0.7)	21 (0.6)	13 (0.8)
Abscess	45 (0.9)	34 (1.0)	11 (0.7)
Fibroma	245 (5.1)	168 (5.1)	77 (4.9)
Serous cystadenoma	543 (11.2)	326 (10.0)	217 (13.7)
Mucinous cystadenoma	359 (7.4)	203 (6.2)	156 (9.8)
Rare benign pathologies	75 (1.5)	50 (1.5)	25 (1.6)
All malignant pathologies	*1665 (34.3)*	*1402 (43.0)*	*263 (16.6)*
Primary invasive stage I	222 (4.6)	184 (5.6)	38 (2.4)
Primary invasive stage II	82 (1.7)	64 (2.0)	18 (1.1)
Primary invasive stage III	658 (13.6)	579 (17.7)	79 (5.0)
Primary invasive stage IV	102 (2.1)	88 (2.7)	14 (0.9)
Rare primary invasive pathologies*	113 (2.3)	80 (2.5)	33 (2.1)
Borderline stage I	249 (5.1)	197 (6.0)	52 (3.3)
Borderline stage II	9 (0.2)	6 (0.2)	3 (0.2)
Borderline stage III	25 (0.5)	23 (0.7)	2 (0.1)
Borderline stage IV	1 (0.02)	1 (0.03)	0
Secondary metastatic cancer	204 (4.2)	180 (5.5)	24 (1.5)

*Including malignant sex cord-stromal tumors, germ cell tumors, mesenchymal tumors, lymphomas and rare malignant epithelial tumors (e.g. malignant Brenner tumor).

Table 4. All 80 observed combinations of benign and malignant ultrasound features (B-features and M-features) of the *Simple Rules* ranked by frequency (n = 4848), with their corresponding sample size and malignancy rate.

Applicable B-features (B1-B2-B3-B4-B5)	Applicable M-features (M1-M2-M3-M4-M5)	All centers N (%mal)	Oncology centers N (%mal)	Other hospitals N (%mal)
0-0-0-0-0	0-0-0-0-0	954 (42)	676 (50)	278 (22)
1-0-0-0-1	0-0-0-0-0	662 (1)	377 (1)	285 (0)
1-0-0-0-0	0-0-0-0-0	513 (2)	257 (2)	256 (1)
0-0-0-1-0	0-0-0-0-0	277 (4)	163 (6)	114 (1)
0-0-0-0-1	0-0-0-0-0	234 (12)	178 (16)	56 (0)
0-0-0-0-0	0-0-0-0-1	219 (78)	173 (83)	46 (59)
0-0-0-0-0	0-1-0-0-0	192 (95)	170 (95)	22 (95)
0-0-1-0-0	0-0-0-0-0	178 (11)	113 (14)	65 (6)
1-0-1-0-1	0-0-0-0-0	159 (1)	86 (1)	73 (0)
0-0-0-1-1	0-0-0-0-0	152 (3)	95 (3)	57 (2)
0-0-0-0-0	0-0-0-1-0	146 (74)	112 (77)	34 (65)
0-0-0-0-0	1-0-0-0-0	101 (91)	82 (90)	19 (95)
0-0-0-0-0	0-1-0-0-1	95 (100)	84 (100)	11 (100)
0-0-1-0-1	0-0-0-0-0	92 (3)	63 (5)	29 (0)
0-0-0-0-0	0-0-1-0-0	91 (80)	66 (88)	25 (60)
1-0-1-0-0	0-0-0-0-0	81 (0)	52 (0)	29 (0)
0-0-0-0-0	1-1-0-0-0	75 (96)	70 (96)	5 (100)
0-0-0-0-0	0-0-1-1-0	58 (78)	44 (86)	14 (50)
0-0-0-0-0	1-0-0-0-1	56 (95)	37 (97)	19 (89)
0-0-0-0-0	0-1-0-1-0	50 (90)	40 (90)	10 (90)
0-0-0-0-0	1-1-0-0-1	50 (100)	39 (100)	11 (100)
0-0-0-0-0	0-0-0-1-1	34 (82)	27 (85)	7 (71)
0-1-0-0-0	0-0-0-0-0	33 (3)	15 (7)	18 (0)
0-1-0-0-1	0-0-0-0-0	33 (0)	22 (0)	11 (0)
0-0-0-0-0	0-0-1-0-1	22 (86)	16 (94)	6 (67)
0-0-0-0-0	0-1-0-1-1	22 (95)	19 (95)	3 (100)
0-0-1-1-0	0-0-0-0-0	22 (0)	7 (0)	15 (0)
0-0-1-0-0	0-0-0-0-1	16 (69)	10 (70)	6 (67)
0-0-0-0-0	0-0-1-1-1	13 (100)	11 (100)	2 (100)
0-0-0-0-0	0-1-1-1-0	13 (100)	13 (100)	(0)
0-0-0-0-0	0-1-1-1-1	13 (100)	11 (100)	2 (100)
0-0-0-0-1	0-1-0-0-0	13 (77)	11 (82)	2 (50)
0-0-1-0-0	0-0-0-1-0	13 (46)	13 (46)	(0)
0-0-1-1-1	0-0-0-0-0	13 (0)	3 (0)	10 (0)
0-0-0-0-0	0-1-1-0-0	12 (100)	12 (100)	(0)
0-0-1-0-0	0-1-0-0-0	12 (50)	9 (56)	3 (33)
1-0-0-0-0	0-0-0-0-1	11 (0)	3 (0)	8 (0)
0-0-0-0-1	0-0-1-0-0	10 (30)	8 (38)	2 (0)
0-0-0-0-1	0-0-0-1-0	9 (33)	7 (29)	2 (50)
0-0-1-0-0	1-0-0-0-0	9 (22)	6 (17)	3 (33)
0-0-0-0-0	0-1-1-0-1	7 (100)	7 (100)	(0)
0-1-0-0-0	0-0-1-0-0	7 (43)	4 (50)	3 (33)
0-1-1-0-0	0-0-0-0-0	5 (0)	3 (0)	2 (0)
0-0-0-0-0	1-1-1-0-1	4 (100)	4 (100)	(0)
0-0-0-0-1	1-0-0-0-0	4 (75)	4 (75)	(0)
0-0-0-1-0	0-0-0-0-1	4 (0)	1 (0)	3 (0)
0-0-1-0-0	1-0-0-0-1	4 (100)	3 (100)	1 (100)
0-0-0-0-0	1-1-1-0-0	3 (100)	2 (100)	1 (100)
0-0-1-0-0	0-0-1-0-0	3 (33)	(0)	3 (33)
				Cont.

Table 4. Continued				
Applicable B-features (B1-B2-B3-B4-B5)	**Applicable M-features (M1-M2-M3-M4-M5)**	**All centers N (%mal)**	**Oncology centers N (%mal)**	**Other hospitals N (%mal)**
0-0-1-0-0	0-1-0-0-1	3 (100)	3 (100)	(0)
0-0-1-0-1	1-0-0-0-0	3 (0)	2 (0)	1 (0)
0-1-0-0-1	0-0-1-0-0	3 (0)	1 (0)	2 (0)
0-1-1-0-1	0-0-0-0-0	3 (0)	2 (0)	1 (0)
1-0-0-0-0	0-1-0-0-0	3 (0)	1 (0)	2 (0)
0-0-0-0-1	0-0-1-1-0	2 (0)	1 (0)	1 (0)
0-0-0-1-0	0-1-0-0-0	2 (50)	2 (50)	(0)
0-0-1-0-0	1-1-0-0-1	2 (100)	2 (100)	(0)
0-0-1-0-1	0-0-0-1-0	2 (0)	2 (0)	(0)
0-0-1-0-1	0-1-0-0-0	2 (0)	2 (0)	(0)
0-1-0-0-0	0-0-0-1-0	2 (50)	(0)	2 (50)
1-0-0-0-1	0-1-0-0-0	2 (0)	2 (0)	(0)
1-0-1-0-0	0-0-0-0-1	2 (0)	2 (0)	(0)
0-0-0-0-1	1-1-0-0-0	1 (0)	(0)	1 (0)
0-0-0-1-1	0-1-0-0-0	1 (0)	1 (0)	(0)
0-0-1-0-0	0-0-0-1-1	1 (0)	(0)	1 (0)
0-0-1-0-0	0-0-1-0-1	1 (100)	1 (100)	(0)
0-0-1-0-0	0-0-1-1-0	1 (100)	1 (100)	(0)
0-0-1-0-0	0-1-0-1-0	1 (0)	1 (0)	(0)
0-0-1-0-0	1-1-0-0-0	1 (100)	1 (100)	(0)
0-0-1-0-1	0-0-1-0-0	1 (0)	1 (0)	(0)
0-0-1-0-1	1-1-0-0-0	1 (0)	1 (0)	(0)
0-0-1-1-0	0-0-0-0-1	1 (0)	(0)	1 (0)
0-0-1-1-0	0-1-0-0-0	1 (0)	(0)	1 (0)
0-1-0-0-0	0-0-0-0-1	1 (0)	1 (0)	(0)
0-1-0-0-0	0-0-1-1-0	1 (0)	1 (0)	(0)
0-1-0-0-0	0-1-0-0-0	1 (100)	1 (100)	(0)
0-1-0-0-0	0-1-1-0-0	1 (100)	1 (100)	(0)
0-1-0-0-1	0-0-0-1-0	1 (0)	1 (0)	(0)
1-0-1-0-0	0-1-0-0-0	1 (0)	(0)	1 (0)
1-0-1-0-1	0-1-0-0-0	1 (100)	1 (100)	(0)

B-feature: benign feature; M-feature: malignant feature; %mal: prevalence of malignancy

Table 5. Model coefficients for the eleven predictors obtained on the development data (n = 2445).

Predictor	Coefficient	SE
Intercept	-1.10	0.26
B1 (unilocular cyst)	-3.10	0.34
B2 (solid components present, but < 7mm)	-1.55	0.59
B3 (acoustic shadows)	-1.58	0.27
B4 (smooth multilocular tumor with largest diameter < 100mm)	-3.59	0.60
B5 (no blood flow; color score 1)	-1.96	0.24
M1 (irregular solid tumor)	2.38	0.39
M2 (ascites)	2.87	0.29
M3 (at least 4 papillary structures)	1.72	0.28
M4 (irregular multilocular-solid tumor with largest diameter ≥ 100mm)	1.12	0.23
M5 (very strong flow; color score 4)	1.53	0.24
Oncology center	0.95	0.31

SE: standard error

Table 6. Sensitivity, specificity, likelihood ratios and predictive values for the *Simple Rules* risk estimates (different cutoffs) on the validation data (n =2403).

Cutoff for risk of malignancy	Size of high risk group n (%)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LR+	LR-
1%	1856 (77.2)	99.7 (97.8-99.9)	33.7 (25.5-43.0)	44.8 (35.4-54.7)	98.9 (97.3-99.5)	1.502	0.010
3%	1637 (68.1)	98.2 (96.9-98.9)	49.6 (41.0-58.2)	52.0 (43.6-60.2)	98.1 (96.4-99.1)	1.947	0.038
5%	1500 (62.4)	97.6 (96.0-98.6)	62.5 (52.2-71.1)	59.2 (50.9-67.1)	98.1 (96.2-99.1)	2.601	0.039
10%	1454 (60.5)	97.5 (95.8-98.5)	64.8 (53.4-74.7)	61.5 (53.9-68.6)	98.0 (96.2-99.0)	2.771	0.039
15%	1376 (57.3)	95.7 (93.2-97.3)	70.9 (61.7-78.6)	64.7 (56.0-72.5)	97.3 (94.8-98.7)	3.289	0.061
20%	1299 (54.1)	94.9 (92.2-96.7)	75.8 (69.0-81.5)	68.8 (59.4-76.8)	97.0 (94.0-98.5)	3.924	0.068
25%	1294 (53.8)	94.8 (92.3-96.5)	75.8 (69.1-81.5)	68.6 (59.2-76.8)	96.8 (93.9-98.3)	3.919	0.069
30%	1165 (48.5)	89.0 (78.2-94.8)	84.7 (75.2-91.0)	75.4 (68.3-81.3)	93.9 (90.8-96.0)	5.811	0.130

LR+: positive likelihood ratio; LR-: negative likelihood ratio; PPV: positive predictive value; NPV: negative predictive value; CI: Confidence Interval

Sensitivities, specificities, positive and negative predictive values have been computed using meta-analysis of center-specific results.

Table 7. Sensitivity, specificity, likelihood ratios and predictive values for the *Simple Rules* risk estimates (different cutoffs) on the validation data in oncology centers (n =1715) and other centers (n=688).

Cutoff for risk of malignancy	Center type	Size of high risk group n (%)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LR+	LR-
1%	Oncology	1439 (83.9)	99.7 (99.0-99.9)	27.3 (20.3-35.5)	51.5 (41.0-61.8)	98.9 (96.5-99.7)	1.370	0.012
	Other	417 (60.6)	98.3 (84.5-99.8)	48.0 (37.4-58.8)	29.7 (25.4-34.4)	99.3 (91.4-100.0)	1.890	0.035
3%	Oncology	1312 (76.5)	98.4 (97.3-99.1)	41.3 (34.8-48.1)	56.3 (46.0-66.1)	97.1 (94.1-98.6)	1.678	0.038
	Other	325 (47.2)	98.5 (85.0-99.9)	66.4 (52.6-77.9)	38.4 (33.0-44.2)	99.5 (93.6-100.0)	2.934	0.023
5%	Oncology	1201 (70.0)	97.8 (96.3-98.7)	57.0 (46.9-66.5)	64.7 (57.0-71.7)	97.0 (94.6-98.4)	2.272	0.039
	Other	299 (43.5)	98.4 (84.9-99.9)	72.5 (57.5-83.7)	44.2 (34.6-54.1)	99.5 (94.0-100.0)	3.583	0.022
10%	Oncology	1199 (69.9)	97.8 (96.4-98.7)	57.2 (47.3-66.4)	64.8 (57.0-71.8)	97.0 (94.6-98.4)	2.283	0.038
	Other	255 (37.1)	96.7 (90.1-98.9)	80.1 (67.7-88.6)	51.4 (42.0-60.8)	99.2 (96.0-99.8)	4.868	0.041
15%	Oncology	1121 (65.4)	96.1 (93.3-97.7)	65.6 (56.6-73.7)	69.1 (60.7-76.7)	95.7 (92.3-97.6)	2.796	0.060
	Other	255 (37.1)	96.7 (90.1-98.9)	80.1 (67.7-88.6)	51.4 (42.0-60.8)	99.2 (96.0-99.8)	4.868	0.041
20%	Oncology	1045 (60.9)	94.9 (92.0-96.8)	73.4 (66.9-79.1)	74.2 (65.8-81.1)	95.0 (91.4-97.2)	3.573	0.069
	Other	254 (36.9)	96.7 (90.1-98.9)	80.2 (67.9-88.6)	51.6 (42.2-60.9)	99.2 (96.0-99.8)	4.895	0.041
25%	Oncology	1045 (60.9)	94.9 (92.0-96.8)	73.4 (66.9-79.1)	74.2 (65.8-81.1)	94.0 (91.4-97.2)	3.573	0.069
	Other	249 (36.2)	95.8 (90.1-98.3)	80.2 (67.9-88.6)	51.3 (41.3-61.2)	98.8 (96.9-99.5)	4.845	0.053
30%	Oncology	1042 (60.8)	94.9 (91.8-96.9)	73.7 (67.2-79.3)	74.4 (65.7-81.5)	95.0 (91.2-97.2)	3.607	0.069
	Other	123 (17.9)	63.3 (44.5-78.8)	94.8 (91.0-97.1)	71.4 (62.7-78.8)	92.1 (87.1-95.3)	12.280	0.387

LR+: positive likelihood ratio; LR-: negative likelihood ratio; PPV: positive predictive value; NPV: negative predictive value; CI: Confidence Interval

Sensitivities, specificities, positive and negative predictive values have been computed using meta-analysis of center-specific results.

Table 8. Model coefficients for the eleven predictors updated using all data (n=4848).

Predictor	Coefficient	SE
Intercept	-0.97	0.24
B1 (unilocular cyst)	-3.41	0.27
B2 (solid components present, but < 7mm)	-2.25	0.46
B3 (acoustic shadows)	-1.66	0.18
B4 (smooth multilocular tumor with largest diameter < 100mm)	-2.75	0.27
B5 (no blood flow; color score 1)	-1.86	0.17
M1 (irregular solid tumor)	2.19	0.24
M2 (ascites)	2.65	0.21
M3 (at least 4 papillary structures)	1.53	0.20
M4 (irregular multilocular-solid tumor with largest diameter ≥ 100mm)	0.98	0.16
M5 (very strong flow; color score 4)	1.55	0.16
Ultrasound examination at oncology center	0.92	0.27

SE: standard error

To use this model to estimate the risk of malignancy, add -0.97 (intercept) to the applicable coefficients to obtain the regression score (RS). Conversion of RS into a risk estimate is done using the formula (exp(RS)/(1+exp(RS)).

Table 9. Summary figure of *Simple Rules* features combinations and the associated risk of malignancy (in %) when updated using all data (n=4848).

Oncology centers	**Number of M features**							
Number of B features	**0**	**1 (M4)**	**1 (M3)**	**1 (M5)**	**1 (M1)**	**1 (M2)**	**2**	**>2**
0	48.7	71.7	81.4	81.7	89.5	93.1	92.1-99.2	98.2- ≥99.9
1 (B3)	15.2	Specific combinations are rare, consider suspicious (risks estimated to be between 12.9 and 71.9% depending on which B- and M-feature)					Rare finding, consider highly suspicious	
1 (B5)	12.8							
1 (B2)	9.1							
1 (B4)	5.7							
1 (B1)	3.1							
2	0.49-2.7	Rare finding, consider suspicious						
>2	0.09-0.29							

Other centers	**Number of M features**							
Number of B features	**0**	**1 (M4)**	**1 (M3)**	**1 (M5)**	**1 (M1)**	**1 (M2)**	**2**	**>2**
0	27.5	50.2	63.6	64	77.2	84.3	82.3-98.0	95.6-99.7
1 (B3)	6.7	Specific combinations are rare, consider suspicious (risks estimated to be between 5.6 and 50.5% depending on which B- and M-feature)					Rare finding, consider highly suspicious	
1 (B5)	5.6							
1 (B2)	3.8							
1 (B4)	2.4							
1 (B1)	1.2							
2	0.19-1.1	Rare finding, consider suspicious						
>2	≤0.01-0.12							

B-feature: benign feature; M-feature: malignant feature

This table shows the risk of malignancy (in %) for the number of B and M features present. If only one feature applies, the risk for the specific B or M feature is depicted.

The upper table applies to oncology centers, the lower table to other centers. Dark green color indicates very low risk of malignancy, green color low risk, yellow color moderate risk, orange color elevated risk, and red color very high risk. These Tables are intended to be used together with the original *Simple Rules* form.[5]

Table 10. Summary classification of Simple Rules risk calculation based on all data (n=4848).

Features	Observed malignancy rate	Estimated individual risk of malignancy	Classification
No M-features AND >2 B-features	1/175 (0.6%)	<0.01-0.29%	Very low risk
- No M-features AND two B-features - No M-features AND feature B1 present	20/1560 (1.3%)	0.19-2.7% 1.2-3.1%	Low risk
No M-features AND one B-feature present (except B1)	60/722 (8.3%)	2.4 -15.2%	Intermediate risk
- No features - Equal number of M and B-features - >0 M-features, but more B than M-features	451/1096 (41.1%)	27.5-48.7% 5.6-78.1% 1.3-28.4%	Elevated risk
More M than B-features present	1133/1295 (87.5%)	42.0- >99.9%	Very high risk

B-feature: benign feature; M-feature: malignant feature

This simplified system only provides risk ranges for the number of B-and M-features present, but facilitates clinical triaging in the absence of electronic devices. Personalized risk estimates can be obtained in a second step.

Figure legend

Figure 1. Validation ROC (Receiver Operating Characteristic) curve (Panel A) and calibration curve (Panel B) for the calculated risk of malignancy (n = 2403). In the ROC curve the results for cut-offs 20% and 25% nearly coincide. Gray line: ideal calibration, black line: calibration curve, gray area: 95% confidence band. In the calibration plot, the distribution of estimated risks of malignancy is depicted in a histogram at the bottom, the positive bins showing the number of patients with malignant tumors, and the negative bins showing the number of patients with benign tumors.

Figure 2. Forest plot with center-specific validation areas under the receiver operating characteristic curve (AUCs) (total n = 2403).

Figure 3. Validation calibration curves by type of center (total n = 2403). Gray line: ideal calibration, black line: calibration curve, gray area: 95% confidence band. In the calibration plots, the distribution of estimated risks of malignancy is depicted in a histogram at the bottom, the positive bins showing the number of patients with malignant tumors, and the negative bins showing the number of patients with benign tumors.