# BMJ

# Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group

Dirk Timmerman, professor in obstetrics and gynaecology ,[1] Lieveke Ameye, postdoctoral researcher in biostatistics,[2] Daniela Fischerova, consultant gynaecologist,[3] Elisabeth Epstein, associate professor in obstetrics and gynaecology,[4] Gian Benedetto Melis, professor in obstetrics and gynaecology,[5] Stefano Guerriero, associate professor in obstetrics and gynaecology,[5] Caroline Van Holsbeke, consultant gynaecologist,[6] Luca Savelli, consultant gynaecologist,[7] Robert Fruscio, consultant gynaecologist,[8] Andrea Alberto Lissoni, consultant gynaecologist,[8] Antonia Carla Testa, assistant professor in gynaecology,[9] Joan Veldman, research fellow in gynaecology,[1] Ignace Vergote, professor in obstetrics and gynaecology,[1] Sabine Van Huffel, professor in biomedical data processing,[2] Tom Bourne, consultant gynaecologist,[10] visiting professor in obstetrics and gynaecology,[1] Lil Valentin, professor in obstetrics and gynaecology[11]

[1]Department of Obstetrics and Gynaecology, University Hospitals KU Leuven, 3000 Leuven, Belgium

[2]Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium

[3]Oncogynaecological Center, First Faculty of Medicine and General University Hospital of Charles University, Prague, Czech Republic

[4]Department of Obstetrics and Gynaecology, Lund University, Lund, Sweden

[5]Department of Obstetrics and Gynaecology, University of Cagliari, Ospedale San Giovanni di Dio, Cagliari, Sardinia, Italy

[6]Department of Obstetrics and Gynaecology, Ziekenhuis Oost-Limburg, Genk (ZOL) 3600 Genk, Belgium

[7]Reproductive Medicine Unit, Department of Obstetrics and Gynaecology, University of Bologna, Italy

[8]Clinica Ostetrica e Ginecologica, Ospedale S Gerardo, Università di Milano Bicocca, Monza, Italy

[9]Istituto di Clinica Ostetrica e Ginecologica, Università Cattolica di Sacro Cuore, Roma, Italy

[10]Department of Obstetrics and Gynaecology, Imperial College London, Queen Charlotte's and Chelsea Hospital, London, UK

[11]Department of Obstetrics and Gynaecology, Malmö University Hospital, Lund University, Malmö, Sweden

Correspondence to: D Timmerman
dirk.timmerman@uzleuven.be

## ABSTRACT

**Objectives** To prospectively assess the diagnostic performance of simple ultrasound rules to predict benignity/malignancy in an adnexal mass and to test the performance of the risk of malignancy index, two logistic regression models, and subjective assessment of ultrasonic findings by an experienced ultrasound examiner in adnexal masses for which the simple rules yield an inconclusive result.

**Design** Prospective temporal and external validation of simple ultrasound rules to distinguish benign from malignant adnexal masses. The rules comprised five ultrasonic features (including shape, size, solidity, and results of colour Doppler examination) to predict a malignant tumour (M features) and five to predict a benign tumour (B features). If one or more M features were present in the absence of a B feature, the mass was classified as malignant. If one or more B features were present in the absence of an M feature, it was classified as benign. If both M features and B features were present, or if none of the features was present, the simple rules were inconclusive.

**Setting** 19 ultrasound centres in eight countries.

**Participants** 1938 women with an adnexal mass examined with ultrasound by the principal investigator at each centre with a standardised research protocol.

**Reference standard** Histological classification of the excised adnexal mass as benign or malignant.

**Main outcome measures** Diagnostic sensitivity and specificity.

**Results** Of the 1938 patients with an adnexal mass, 1396 (72%) had benign tumours, 373 (19.2%) had primary invasive tumours, 111 (5.7%) had borderline malignant tumours, and 58 (3%) had metastatic tumours in the ovary. The simple rules yielded a conclusive result in 1501 (77%) masses, for which they resulted in a sensitivity of 92% (95% confidence interval 89% to 94%) and a specificity of 96% (94% to 97%). The corresponding sensitivity and specificity of subjective assessment were 91% (88% to 94%) and 96% (94% to 97%). In the 357 masses for which the simple rules yielded an inconclusive result and with available results of CA-125 measurements, the sensitivities were 89% (83% to 93%) for subjective assessment, 50% (42% to 58%) for the risk of malignancy index, 89% (83% to 93%) for logistic regression model 1, and 82% (75% to 87%) for logistic regression model 2; the corresponding specificities were 78% (72% to 83%), 84% (78% to 88%), 44% (38% to 51%), and 48% (42% to 55%). Use of the simple rules as a triage test and subjective assessment for those masses for which the simple rules yielded an inconclusive result gave a sensitivity of 91% (88% to 93%) and a specificity of 93% (91% to 94%), compared with a sensitivity of 90% (88% to 93%) and a specificity of 93% (91% to 94%) when subjective assessment was used in all masses.

**Conclusions** The use of the simple rules has the potential to improve the management of women with adnexal masses. In adnexal masses for which the rules yielded an inconclusive result, subjective assessment of ultrasonic findings by an experienced ultrasound examiner was the most accurate diagnostic test; the risk of malignancy index and the two regression models were not useful.

## INTRODUCTION

When deciding on the type of surgery for a patient with an adnexal mass, estimating the risk of malignancy is essential. Benign masses can be managed conservatively or with laparoscopy, avoiding unnecessary costs and morbidity. On the other hand, peri-operative rupture of a stage I ovarian cancer may worsen the prognosis.[1] When malignancy is suspected, referral to a gynaecological oncologist is needed for proper staging and debulking surgery.

Transvaginal ultrasonography is an excellent tool for discriminating between benign and malignant adnexal masses. Several studies have shown that the risk of malignancy is very low in unilocular ovarian cysts.[2-5] The presence of morphological features other than a unilocular cyst, such as papillary structures and solid areas, as well as increased vascularity as determined by Doppler ultrasound, is associated with a variably increased risk of malignancy.[2 5] Optimisation of the diagnostic performance of transvaginal sonography by creating predictive models with the use of scoring systems, logistic regression analysis, neural networks, and support vector machines has been attempted. However, when these models were tested prospectively, they performed less well than was originally reported.[6-11] The risk of malignancy index is the test recommended by the Royal College of Obstetricians and Gynaecologists and in a recent review by Geomini et al.[11 12] On the other hand, two logistic regression models (logistic regression model 1 and logistic regression model 2) developed in the International Ovarian Tumour Analysis (IOTA) study performed as well in new centres as in the units where the models were first developed,[13] but no model or biochemical marker of ovarian malignancy has been shown to be superior to subjective assessment of grey scale and colour Doppler ultrasonic findings by an experienced ultrasound examiner.[8 14 15] Unfortunately, the expertise of experienced ultrasound examiners is not easily transferred to less experienced examiners. Less experienced examiners might be helped by scoring systems and risk calculation models, but a criticism has been that the ultrasonic information required in some ultrasound based risk calculation models is too difficult to obtain outside specialist centres.[16]

In a previous report, we used data collected in the first phase of the IOTA study to develop simple and clinically useful ultrasound based rules for discriminating between benign and malignant adnexal masses.[5] We developed simple rules and then temporally validated them prospectively in a small group of patients (n=507). On temporal validation, the simple rules yielded a conclusive result in 76% of all tumours, for which they resulted in a sensitivity with regard to malignancy of 95%, a specificity of 91%, a positive likelihood ratio of 10.5 and negative likelihood ratio of 0.06.[5] We concluded that "most adnexal tumours in an ordinary tumour population can be correctly classified as benign or malignant using simple ultrasound-based rules. For tumours that cannot be classified using simple rules, ultrasound examination by an expert examiner might be useful."[5]

The aim of the study reported here was to do a prospective temporal and external validation in a large study population to assess the ability of the previously published simple ultrasound based rules to distinguish between benign and malignant adnexal masses before surgery. A secondary aim was to determine the diagnostic performance of subjective assessment of ultrasonic findings by an experienced ultrasound examiner, the risk of malignancy index,[12] and the logistic regression models 1 and 2 when used in tumours for which the simple rules yield an inconclusive result.

## METHODS

In this prospective study, the IOTA phase 2 study, we examined the performance of the simple rules in a population of women who had surgery for an adnexal mass. Local clinicians made the decision to operate on the basis of local rules and clinical judgment. We followed the guidelines of the standards for the reporting of diagnostic accuracy studies initiative.[17]

### Recruitment centres

We tested the rules both in the same seven centres where they had been developed (old centres[5 18]) and in a further 12 centres that had not participated in any IOTA study before (new centres). In total, 19 centres from eight countries participated.

### Patients

We included patients who presented with at least one adnexal mass and who had an ultrasound examination by a principal investigator at one of the participating centres. In the case of bilateral adnexal masses, we included the mass with the most complex ultrasonic morphology in our statistical analysis. If both masses had similar ultrasonic morphology, we included the largest one or the one most easily accessible by transvaginal ultrasound. We excluded patients who were pregnant or refused transvaginal ultrasonography and those who did not have surgical removal of the mass within 120 days after the ultrasound examination.

### Data collection

A dedicated, secure data collection system was developed for the study (IOTA 2 study screen, astraia,

**Table 1** | Predictive value of each ultrasonic feature used in simple rules among masses for which simple rules yielded conclusive result (benign or malignant). Values are percentages (95% confidence interval); numbers

| Ultrasonic features | Predictive value |
|---|---|
| **For predicting a malignant tumour (M features)** | |
| M1—Irregular solid tumour | 96 (88 to 98); 64/67 |
| M2—Presence of ascites | 97 (93 to 99); 157/162 |
| M3—At least four papillary structures | 88 (80 to 93); 75/85 |
| M4—Irregular multilocular solid tumour with largest diameter ≥100 mm | 84 (77 to 90); 103/122 |
| M5—Very strong blood flow (colour score 4) | 88 (82 to 92); 131/149 |
| At least one M feature | 87 (84 to 90); 340/389 |
| **For predicting a benign tumour (B features)** | |
| B1—Unilocular | 99 (98 to 99); 673/681 |
| B2—Presence of solid components, of which largest solid component has largest diameter <7 mm | 100 (90 to 100); 33/33 |
| B3—Presence of acoustic shadows | 95 (92 to 97); 223/234 |
| B4—Smooth multilocular tumour with largest diameter <100 mm | 99 (97 to 100); 190/191 |
| B5—No blood flow (colour score 1) | 98 (96 to 99); 615/629 |
| At least one B feature | 97 (96 to 98); 1083/1112 |

Rule 1: If one or more M features are present in absence of B feature, mass is classified as malignant.
Rule 2: If one or more B features are present in absence of M feature, mass is classified as benign.
Rule 3: If both M features and B features are present, or if no B or M features are present, result is inconclusive and second stage test is recommended.

Munich, Germany). A unique identifier was generated automatically for each patient's record. Clinicians at each centre could view or update only patients' records from their own centre. We ensured data security by not transferring the patient's name and by encrypting all data communication. Data integrity and completeness were ensured by client side checks in the astraia system and manual checks by one biostatistician and two expert ultrasound examiners.

### Clinical variables

A standardised history was taken in the same manner as in the IOTA phase 1 study.[18] It included information on personal history of ovarian cancer and breast cancer; the number of first degree relatives with ovarian cancer or breast cancer; and the patient's age, menopausal status, and current hormone treatment. Women aged 50 years or more who had undergone hysterectomy before menopause were defined as postmenopausal.

### Ultrasound examination

In all cases, a principal investigator at the participating centres did a transvaginal ultrasound scan in the same standardised manner as in the IOTA phase 1 study.[18] The principal investigators were fully trained gynaecologists or radiologists with a special interest in gynaecological ultrasound and more than five years' experience in this field. They used a variety of ultrasound machines with transvaginal probe frequencies ranging between 5 and 12 MHz. The investigators also used transabdominal ultrasonography to examine large masses that could not be seen in their entirety by using a transvaginal probe. They used grey scale and colour Doppler ultrasound images to obtain

morphological and blood flow variables to characterise each adnexal mass. Details of the ultrasound examination technique and the ultrasound terms and definitions used have been described elsewhere.[18 19] Finally, the investigator stated whether the mass was likely to be malignant or benign on the basis of subjective evaluation of ultrasonic findings ("subjective assessment"). The ultrasonic information was recorded prospectively and locked at the time of the examination and so could not be changed after surgery. We calculated the risk of malignancy by using the IOTA logistic regression models 1 and 2 centrally after the conclusion of the study, ensuring that these logistic regression models had no role in the decision making process. The same was true of the simple rules.

### Reference standard

The reference standard was the histological diagnosis and, in case of malignancy, the surgical stage. Surgery was carried out by laparoscopy or laparotomy, according to the surgeon's judgment. The excised tissues were examined histologically at the local centre. The pathologists had no knowledge of the ultrasound results. We classified tumours according to the criteria recommended by the International Federation of Gynaecology and Obstetrics.[20]

### Simple rules

We applied to the tumours the simple ultrasound based rules that have been described in detail in a previous report.[5] Briefly, we used five ultrasonic features to predict a malignant tumour (M features): irregular solid tumour (M1), ascites (M2), at least four papillary structures (M3), irregular multilocular solid tumour with a
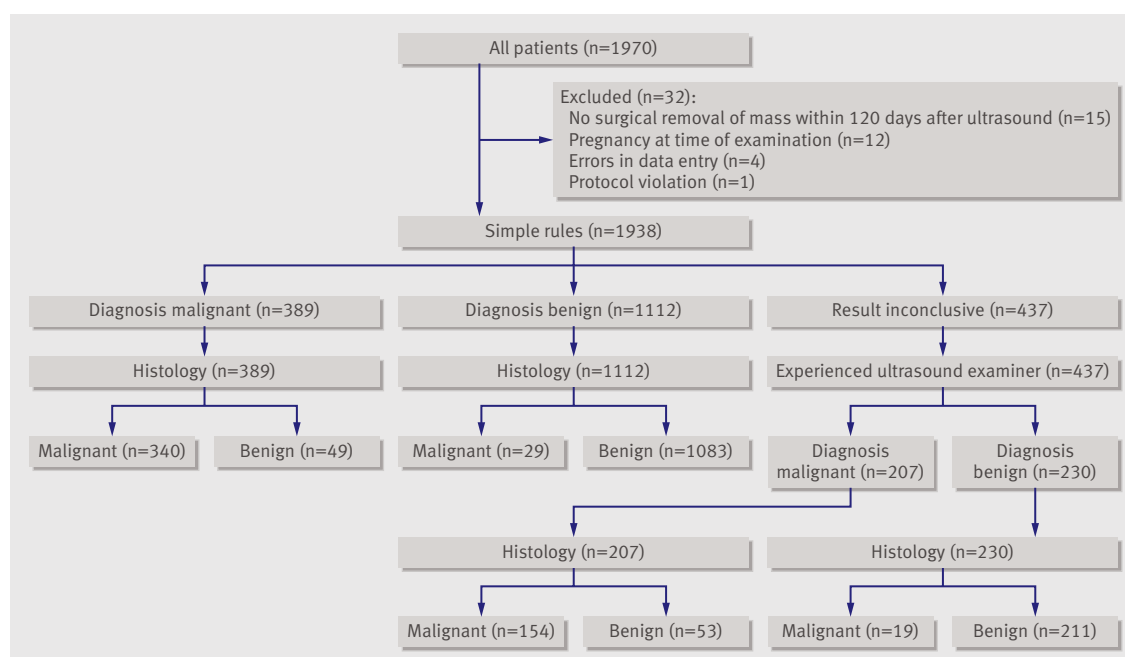


**Flow diagram of simple rules as triage test and subjective assessment by experienced ultrasound examiner as second stage test**

Table 2 | Sensitivity and specificity of simple rules, subjective assessment, logistic regression models 1 and 2, and risk of malignancy index in cases in which simple rules yielded conclusive result (benign or malignant), stratified by menopausal status

| | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|
| **All cases** | | |
| Total (n=1501): | | |
| Simple rules | 92 (89 to 94) | 96 (94 to 97) |
| Subjective assessment | 91 (88 to 94) | 96 (94 to 97) |
| Logistic regression model 1 | 94 (91 to 96) | 92 (91 to 94) |
| Logistic regression model 2 | 95 (92 to 97) | 91 (89 to 92) |
| Premenopausal (n=969): | | |
| Simple rules | 91 (84 to 95) | 97 (95 to 98) |
| Subjective assessment | 90 (83 to 94) | 97 (96 to 98) |
| Logistic regression model 1 | 90 (83 to 94) | 96 (94 to 97) |
| Logistic regression model 2 | 92 (85 to 96) | 95 (93 to 96) |
| Postmenopausal (n=532): | | |
| Simple rules | 93 (89 to 95) | 92 (89 to 95) |
| Subjective assessment | 91 (87 to 94) | 90 (86 to 93) |
| Logistic regression model 1 | 96 (93 to 98) | 83 (78 to 87) |
| Logistic regression model 2 | 97 (93 to 98) | 77 (72 to 82) |
| **Cases with CA 125 available** | | |
| Total (n=1147): | | |
| Simple rules | 92 (89 to 95) | 95 (93 to 96) |
| Subjective assessment | 91 (87 to 93) | 95 (93 to 96) |
| Logistic regression model 1 | 94 (91 to 96) | 91 (89 to 93) |
| Logistic regression model 2 | 95 (93 to 97) | 89 (87 to 91) |
| Risk of malignancy index | 75 (71 to 80) | 95 (93 to 96) |
| Premenopausal (n=688): | | |
| Simple rules | 91 (84 to 95) | 96 (94 to 98) |
| Subjective assessment | 90 (83 to 95) | 97 (95 to 98) |
| Logistic regression model 1 | 90 (83 to 95) | 94 (92 to 96) |
| Logistic regression model 2 | 92 (85 to 96) | 94 (92 to 96) |
| Risk of malignancy index | 55 (45 to 64) | 96 (94 to 97) |
| Postmenopausal (n=459): | | |
| Simple rules | 93 (88 to 95) | 92 (87 to 95) |
| Subjective assessment | 91 (87 to 94) | 90 (86 to 94) |
| Logistic regression model 1 | 96 (93 to 98) | 81 (76 to 86) |
| Logistic regression model 2 | 97 (94 to 98) | 78 (72 to 83) |
| Risk of malignancy index | 84 (79 to 88) | 91 (87 to 94) |

largest diameter of at least 100 mm (M4), and very high colour content on colour Doppler examination (M5). We used five ultrasonic features to predict a benign tumour (B features): unilocular cyst (B1), presence of solid components for which the largest solid component is <7 mm in largest diameter (B2), acoustic shadows (B3), smooth multilocular tumour (B4), and no detectable blood flow on Doppler examination (B5). If one or more M features were present in the absence of a B feature, we classified the mass as malignant (rule 1). If one or more B features were present in the absence of an M feature, we classified the mass as benign (rule 2). If both M features and B features were present, or if none of the features was present, the simple rules were inconclusive (rule 3).

## Logistic regression models 1 and 2

The logistic regression model 1 was based on the age of the patient (in years), the presence of ascites, the presence of blood flow within a papillary projection, the largest diameter of the solid component (in millimetres, but with no increase >50 mm), the presence of irregular internal cyst walls, the presence of acoustic shadows, personal history of ovarian cancer, current hormonal treatment, the largest diameter of the lesion (mm), tenderness of the lesion during the examination, the presence of a purely solid tumour, and the colour score (1, 2, 3, or 4). The simpler logistic regression model (model 2) used only the first six variables. As suggested in the original publication, an estimated probability of malignancy above 0.10 by logistic regression model 1 or 2 classified the mass as malignant.[18]

## Risk of malignancy index

We determined the risk of malignancy index by using the ultrasonic findings, the menopausal status, and the serum CA 125 concentration.[12] We assessed five ultrasonic features suggestive of cancer in an ultrasound score (U): multilocularity, solid areas, bilateral masses, ascites, and evidence of metastases. U was 0 when none of these features was present, 1 if one feature was present, and 3 if two or more features were present. We assigned a score (M) of 1 to premenopausal women and a score of 3 to postmenopausal women. We defined the risk of malignancy index as U×M×serum CA 125 concentration (U/mL). As suggested in the original publication, a risk of malignancy index of more than 200 classified the mass as malignant.[12]

## Statistical analysis

We compared the simple rules with subjective assessment by an experienced ultrasound examiner, the risk of malignancy index, and the logistic regression models 1 and 2 in cases in which the simple rules yielded a conclusive result. We also assessed the performance of a strategy in which the simple rules were used as a triage test,[21] with a second stage test (subjective assessment, risk of malignancy index, or logistic regression model 1 or 2) being used for masses for which the simple rules yielded an inconclusive result.

We expressed diagnostic performance in terms of sensitivity and specificity. We used Wilson's method to calculate the 95% confidence limits of binomial proportions. We used McNemar's test to determine the statistical significance of differences in paired binomial proportions: sensitivity and specificity. We determined the statistical significance of differences in categorical data for unpaired comparisons by using the $\chi^2$ test. We used SAS system release 9.2 for statistical analyses.

## RESULTS

We enrolled 1970 patients between November 2005 and October 2007. Of these, we excluded 32 (1.6%) for the following reasons: no surgical removal of the mass within 120 days after the ultrasound examination (n=15), pregnant at the time of the examination (n=12),

**Table 3 | Sensitivity and specificity of subjective assessment, logistic regression models 1 and 2, and risk of malignancy index in cases in which simple rules yielded inconclusive result, stratified by menopausal status**

| | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|
| **All cases** | | |
| Total (n=437): | | |
|   Subjective assessment | 89 (83 to 93) | 80 (75 to 84) |
|   Logistic regression model 1 | 88 (83 to 92) | 47 (41 to 53) |
|   Logistic regression model 2 | 80 (74 to 86) | 50 (44 to 56) |
| Premenopausal (n=227): | | |
|   Subjective assessment | 84 (74 to 91) | 85 (78 to 90) |
|   Logistic regression model 1 | 81 (71 to 89) | 64 (56 to 71) |
|   Logistic regression model 2 | 70 (58 to 79) | 63 (55 to 70) |
| Postmenopausal (n=210): | | |
|   Subjective assessment | 92 (85 to 96) | 73 (64 to 80) |
|   Logistic regression model 1 | 93 (87 to 97) | 21 (15 to 30) |
|   Logistic regression model 2 | 87 (80 to 92) | 30 (22 to 39) |
| **Cases with CA 125 available** | | |
| Total (n=357): | | |
|   Subjective assessment | 89 (83 to 93) | 78 (72 to 83) |
|   Logistic regression model 1 | 89 (83 to 93) | 44 (38 to 51) |
|   Logistic regression model 2 | 82 (75 to 87) | 48 (42 to 55) |
|   Risk of malignancy index | 50 (42 to 58) | 84 (78 to 88) |
| Premenopausal (n=182): | | |
| Subjective assessment | 84 (73 to 91) | 83 (75 to 88) |
|   Logistic regression model 1 | 84 (73 to 91) | 63 (54 to 71) |
|   Logistic regression model 2 | 74 (62 to 83) | 63 (54 to 71) |
|   Risk of malignancy index | 32 (22 to 45) | 87 (79 to 92) |
| Postmenopausal (n=175): | | |
|   Subjective assessment | 92 (84 to 96) | 71 (61 to 80) |
|   Logistic regression model 1 | 93 (86 to 97) | 20 (13 to 29) |
|   Logistic regression model 2 | 88 (79 to 93) | 28 (19 to 38) |
|   Risk of malignancy index | 63 (52 to 72) | 80 (71 to 87) |

errors in data entry (n=4), and protocol violation (n=1) (figure). We thus included data from 1938 patients. The mean age was 46 (range 11-94) years, 38% (742) of the patients were postmenopausal, 41% (793) were nulliparous, and 11% (214) were receiving hormonal treatment. Of the tumours, 542 (28%) were malignant, including 111 (20%) borderline masses, 373 (69%) primary invasive masses, and 58 (11%) metastatic masses.

Table 1 shows the predictive value of each ultrasonic feature used in the simple rules. In total, the simple rules yielded a conclusive result (rule 1=malignant, rule 2=benign) for 1501 of the tumours. This corresponds to 77% (1501) of all masses in the dataset. The malignancy rate was 25% (369/1501) in masses for which the simple rules yielded a conclusive result, compared with 40% (173/437) in the remainder (P<0.001). In 456 cases at least one feature for a malignant tumour (M feature) was present, and in 389 (85%) of these no B feature was present. Of the 389 masses predicted to be malignant by the simple rules, 87% (340) were malignant according to histology. In 1179 cases at least one B feature was present, and in 1112 (94%) of these no M features were present. Of the

1112 masses predicted to be benign by the simple rules, 97% (1083) were benign according to histology.

Among the tumours for which the simple rules yielded a conclusive result, they had a sensitivity of 92% (340/369) and a specificity of 96% (1083/1132) (table 2). Among these tumours, the sensitivity and specificity of subjective assessment were similar to those of the simple rules: 91% (336/369) (P=0.35) and 96% (1083/1132) (P=1.0). Subjective assessment missed 33 cancers (false negative) and gave 49 false positive diagnoses. The simple rules missed 29 cancers and gave 49 false positive diagnoses. The simple rules performed similarly in "old" and "new" centres: the sensitivity was 93% (179/192) in the old centres and 91% (161/177) in the new centres (P=0.42), and the specificity was 95% (487/513) in the old centres and 96% (596/619) in the new centres (P=0.27). The sensitivity of the simple rules was similar in premenopausal and postmenopausal patients (91% (102/112) v 93% (238/257); P=0.62) but the specificity was higher in the premenopausal patients (97% (829/857) v 92% (254/275); P=0.004). The simple rules yielded a conclusive result more often in premenopausal patients than in postmenopausal patients (81% (969/1196) v 72% (532/742); P<0.001).

Table 3 shows the diagnostic performance of subjective assessment, logistic regression models 1 and 2, and the risk of malignancy index among the tumours for which the simple rules yielded an inconclusive result (rule 3). Among these tumours, the diagnostic performance of subjective assessment by the ultrasound examiner was superior to that of logistic regression model 1, logistic regression model 2, and the risk of malignancy index in both premenopausal and postmenopausal patients. Logistic regression model 1 and logistic regression model 2 had low specificity, whereas the risk of malignancy index had low sensitivity. The specificity of logistic regression models 1 and 2 was significantly lower than that of subjective assessment (47% and 50% v 80%, P<0.001 for both comparisons among all patients; 64% and 63% v 85%, P<0.001 for both comparisons among premenopausal patients; 21% and 30% v 73%, P<0.001 for both comparisons among postmenopausal patients). The sensitivity of the risk of malignancy index was significantly lower than that of subjective assessment (50% v 89% among all patients, P<0.001; 32% v 84%, P<0.001 among premenopausal patients; 63% v 92%, P<0.001 among postmenopausal patients).

If the simple rules were used in all tumours, the sensitivity was 63% (340/542) and the specificity was 78% (1083/1396). If the simple rules were used as a triage test, and subjective assessment of ultrasound findings was used for those masses for which the simple rules yielded an inconclusive result (figure), the test performance was as follows: sensitivity 91% (494/542) (63% (340/542) by simple rules and 28% (154/542) by subjective assessment), and specificity 93% (1294/1396) (78% (1083/1396) by simple rules and 15% (211/1396) by subjective assessment). This performance was similar to that of using subjective assessment in

Table 4 | Sensitivity and specificity of strategy in which simple rules are used as triage test and subjective assessment, logistic regression model 1 or 2, or risk of malignancy index is used in tumours for which simple rules yield inconclusive result, and sensitivity and specificity of using subjective assessment, logistic regression model 1 or 2, or risk of malignancy index in all tumours

| | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|
| **Total (n=1938)** | | |
| Simple rules + subjective assessment | 91 (88 to 93) | 93 (91 to 94) |
| Simple rules + logistic regression model 1 | 91 (88 to 93) | 86 (84 to 88) |
| Simple rules + logistic regression model 2 | 88 (85 to 91) | 87 (85 to 89) |
| Simple rules + inconclusive="malignant"* | 95 (92 to 96) | 78 (75 to 80) |
| Subjective assessment | 90 (88 to 93) | 93 (91 to 94) |
| Logistic regression model 1 | 92 (90 to 94) | 84 (82 to 86) |
| Logistic regression model 2 | 90 (88 to 93) | 83 (81 to 85) |
| **Cases with CA 125 available (n=1504)** | | |
| Simple rules + subjective assessment | 91 (88 to 93) | 92 (90 to 93) |
| Simple rules + logistic regression model 1 | 91 (88 to 93) | 85 (82 to 87) |
| Simple rules + logistic regression model 2 | 89 (86 to 91) | 85 (83 to 88) |
| Simple rules + risk of malignancy index | 79 (75 to 83) | 93 (91 to 94) |
| Simple rules + inconclusive="malignant"* | 95 (92 to 96) | 76 (73 to 78) |
| Subjective assessment | 90 (87 to 92) | 92 (90 to 93) |
| Logistic regression model 1 | 93 (90 to 95) | 81 (79 to 84) |
| Logistic regression model 2 | 91 (88 to 93) | 81 (79 to 83) |
| Risk of malignancy index | 68 (63 to 72) | 93 (91 to 94) |

*Cases in which simple rules yielded inconclusive result are classified as malignant.

all tumours, which had a sensitivity of 90% (490/542) (P=0.35) and a specificity of 93% (1294/1396) (P=1.0) (table 4).

The simple rules yielded a conclusive result in most benign tumours (81%, 1132/1396) and in most primary invasive tumours (74%, 275/373) but in only half of the borderline tumours (50%, 56/111) (table 5). The performance of the simple rules was poor for abscesses, fibromas, and serous borderline stage I tumours.

## DISCUSSION

In this study, we have prospectively validated the ability of the IOTA simple ultrasound rules to discriminate between benign and malignant adnexal masses. The results of this study confirmed that when the rules yielded a conclusive result, they reliably discriminated between benign and malignant adnexal masses. They did so just as well as did subjective assessment by an experienced ultrasound examiner. The rules worked well both on temporal validation in the centres where they had been developed and on external validation in the new centres. This confirms that the rules are generalisable. The test performance of a strategy in which the simple rules were used as a triage test and subjective assessment of ultrasonic findings was used as a second stage test in those masses for which the rules yielded an inconclusive result (sensitivity 91% and specificity 93%) was similar to that of using subjective assessment by an experienced examiner in all tumours (sensitivity 90% and specificity 93%). Because few clinicians have special skills in the ultrasound examination of ovarian pathology, a reliable test that can be used effectively by all ultrasound examiners is needed. The simple rules have the potential to become that test.

### Strengths and weaknesses

A strength of this study was its prospective and multicentre design. As the data were collected in different countries with patients with different characteristics, the simple rules are likely to prove applicable and to perform well in other populations. Another strength is the large number of patients studied with a detailed predefined protocol with agreed terms, measurement technique, and definitions. We did both a temporal and an external validation. Because the results were virtually identical in the old and new centres, we can justify reporting the results for the old and new centres together to take advantage of a larger study population and be able to estimate the measures of performance, such as sensitivity and specificity, with greater precision.

A limitation of the study is that all the examinations were done by experienced ultrasound examiners. Validation of the simple rules by less experienced examiners is needed. The purpose is to use the rules for triaging patients for referral to an examiner specialised in gynaecological ultrasound.

### Comparison with other studies

Previous studies on pre-operative characterisation of adnexal masses as benign or malignant were mostly small and single centre.[7-9 12] We have previously developed and tested the simple rules in a multicentre study.[5] However, Altman and colleagues wrote that neither internal nor temporal validation examines the generalisability of a model, for which using new data collected from an appropriate patient population in a different centre is necessary.[22] This study contains the first prospective temporal and external validation of the simple rules, and the results of our external validation confirm the generalisability of these rules.

The main advantage of the simple rules is their simplicity. The ultrasonic variables are straightforward to obtain, and the rules are easier to use in clinical practice than are many mathematical models. By using a simple tick box system, a result can be produced rapidly without the need for computer software. Moreover, contrary to when the risk of malignancy index is used, no blood sample for a serum CA 125 is needed. The simple rules are therefore likely to be an ideal tool to help less experienced ultrasound examiners to differentiate between benign and malignant tumours. The disadvantage of the simple rules is that they yield an inconclusive result in about 25% of all tumours, whereas mathematical models yield a useful result in all masses.[18] When using the simple rules, therefore, having established appropriate referral patterns to a specialist in gynaecological ultrasound for cases in which the simple rules yield an inconclusive result is important. The risk of malignancy index or the two IOTA logistic regression models 1 and 2 are not good enough for discrimination between benign and malignant tumours when the simple rules yield an inconclusive result. The rules seem to work less well for abscesses,

**Table 5 | Ability of simple rules to classify tumours with different histology as benign or malignant. Values are numbers (percentages)**

| Pathology | Total No | Conclusive diagnosis using simple rules | Correctly classified if simple rules yielded conclusive result |
|---|---|---|---|
| Benign: | 1396 | 1132/1396 (81) | 1083/1132 (96) |
| Endometrioma | 400 | 347/400 (87) | 339/347 (98) |
| Teratoma | 226 | 211/226 (93) | 211/211 (100) |
| Simple cyst + parasalpingeal cyst | 131 | 122/131 (93) | 122/122 (100) |
| Functional cyst | 77 | 62/77 (81) | 62/62 (100) |
| Hydrosalpinx + salpingitis | 49 | 37/49 (76) | 34/37 (92) |
| Peritoneal pseudocyst | 11 | 11/11 (100) | 11/11 (100) |
| Abscess | 24 | 13/24 (54) | 10/13 (77) |
| Fibroma | 81 | 44/81 (54) | 34/44 (77) |
| Serous cystadenoma | 236 | 181/236 (77) | 170/181 (94) |
| Mucinous cystadenoma | 138 | 91/138 (66) | 81/91 (89) |
| Rare benign | 18 | 10/18 (56) | 8/10 (80) |
| Uterine fibroid | 5 | 3/5 | 1/3 |
| Borderline: | 111 | 56/111 (50) | 45/56 (80) |
| Borderline serous stage I | 55 | 33/55 (60) | 25/33 (76) |
| Borderline serous stage II | 3 | 0/3 | – |
| Borderline serous stage III | 8 | 4/8 | 3/4 |
| Borderline mucinous stage I | 41 | 16/41 (39) | 14/16 (88) |
| Borderline mucinous stage IV | 1 | 1/1 | 1/1 |
| Borderline endometroid stage I | 1 | 0/1 | – |
| Rare borderline | 2 | 2/2 | 2/2 |
| Primary invasive: | 373 | 275/373 (74) | 259/275 (94) |
| Primary invasive stage I | 70 | 46/70 (66) | 45/46 (98) |
| Primary invasive stage II | 30 | 20/30 (67) | 19/20 (95) |
| Primary invasive stage III | 202 | 157/202 (78) | 149/157 (95) |
| Primary invasive stage IV | 30 | 24/30 (80) | 23/24 (96) |
| Rare primary invasive | 41 | 28/41 (68) | 23/28 (82) |
| Metastatic | 58 | 38/58 (66 ) | 36/38 (95) |

fibromas, and stage I serous borderline tumours. These conditions are also difficult to classify with subjective assessment of ultrasonic findings.[23] Future research needs to determine the performance of the simple rules when used by less experienced ultrasound examiners and whether the use of the simple rules will improve care of patients and reduce costs.

## WHAT IS ALREADY KNOWN ON THIS TOPIC

Preoperative characterisation of adnexal tumours determines the management of the patient, and appropriate management determines the prognosis

Subjective assessment of ultrasound examination is the most reliable method to distinguish between benign and malignant adnexal masses before surgery, but it requires expertise

Simple rules have been proposed to discriminate between benign and malignant masses, but they require external validation

## WHAT THIS STUDY ADDS

The simple rules were conclusive in about 75% of adnexal masses

When conclusive, they performed as well as subjective assessment by an experienced examiner for discrimination between benign and malignant masses

Their use may change clinical practice by providing an accurate instant classification of most adnexal masses while reducing the number of patients that need to be referred for expert scanning

## Conclusion

Because the simple rules offer a straightforward approach to correctly characterise about 75% of adnexal masses, their use should enable all sonographers and general gynaecologists to reliably distinguish between benign and malignant adnexal masses in most cases. Where the rules yield an inconclusive result, we propose referring the patient for subjective assessment of ultrasonic findings by an experienced ultrasound examiner, because this provides the most accurate diagnosis. If we use the simple rules as a triage test and subjective assessment by an experienced ultrasound examiner as a second stage test in those masses for which the simple rules yield an inconclusive result, we obtain the same diagnostic performance as when subjective assessment is used in all masses. In this way, the use of the simple rules has the potential to improve the management of women with adnexal masses.

1  Vergote I, De Brabanter J, Fyles A, Bertelsen K, Einhorn N, Sevelda P, et al. Prognostic importance of degree of differentiation and cyst rupture in stage I invasive epithelial ovarian carcinoma. *Lancet* 2001;357:176-82.

2 Granberg S, Wikland M, Jansson I. Macroscopic characterization of ovarian tumors and the relation to the histological diagnosis: criteria to be used for ultrasound evaluation. *Gynecol Oncol* 1989;35:139-44.

3 Roman LD. Small cystic pelvic masses in older women: is surgical removal necessary? *Gynecol Oncol* 1998;69:1-2.

4 Ekerhovd E, Wienerroith H, Staudach A, Granberg S. Preoperative assessment of unilocular adnexal cysts by transvaginal ultrasonography: a comparison between ultrasonographic morphologic imaging and histopathologic diagnosis. *Am J Obstet Gynecol* 2001;184:48-54.

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

6 Ferrazzi E, Zanetta G, Dordoni D, Berlanda N, Mezzopane R, Lissoni AA. Transvaginal ultrasonographic characterization of ovarian masses: comparison of five scoring systems in a multicenter study. *Ultrasound Obstet Gynecol* 1997;10:192-7.

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

8 Valentin L, Hagen B, Tingulstad S, Eik-Nes S. Comparison of "pattern recognition" and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross-validation. *Ultrasound Obstet Gynecol* 2001;18:357-65.

9 Mol BW, Boll D, De Kanter M, Heintz AP, Sijmons EA, Oei SG, et al. Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. *Gynecol Oncol* 2001;80:162-7.

10 Van Holsbeke C, Van Calster B, Valentin L, Testa AC, Ferrazzi E, Dimou I, et al. External validation of mathematical models to distinguish between benign and malignant adnexal tumors: a multicenter study by the International Ovarian Tumor Analysis Group. *Clin Cancer Res* 2007;13:4440-7.

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

12 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 Obstet Gynaecol* 1990;97:922-9.

13 Timmerman D, Van Calster B, Testa AC, Guerriero S, Fischerova D, Lissoni AA, et al. Ovarian cancer prediction in adnexal masses using ultrasound based logistic regression models: a temporal and external validation study by the IOTA group. *Ultrasound Obstet Gynecol* 2010;36:226-34.

14 Timmerman D. The use of mathematical models to evaluate pelvic masses: can they beat an expert operator? *Best Pract Res Clin Obstet Gynaecol* 2004;18:91-104.

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

16 Moore RG, Bast RC Jr. How do you distinguish a malignant pelvic mass from a benign pelvic mass? Imaging, biomarkers or none of the above? *J Clin Oncol* 2007;25:4159-61.

17 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41-4.

18 Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, et al. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group. *J Clin Oncol* 2005;23:8794-801.

19 Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I. Terms, definitions and measurements to describe the ultrasonographic features of adnexal tumors: a consensus opinion from the international ovarian tumor analysis (IOTA) group. *Ultrasound Obstet Gynecol* 2000;16:500-5.

20 Heintz AP, Odicino F, Maisonneuve P, Quinn MA, Benedet JL, Creasman WT, et al. Carcinoma of the ovary: FIGO 6th annual report on the results of treatment in gynecological cancer. *Int J Gynaecol Obstet* 2006;95:S161-92.

21 Hayen A, Macaskill P, Irwig L, Bossuyt P. Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. *J Clin Epidemiol* 2010;63:883-91.

22 Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.

23 Valentin L, Ameye L, Jurkovic D, Metzger U, Lécuru F, Van Huffel S, et al. Which extrauterine pelvic masses are difficult to correctly classify as benign or malignant on the basis of ultrasound findings and is there a way of making a correct diagnosis? *Ultrasound Obstet Gynecol* 2006;27:438-44.