European Journal of Cancer (2012) **48**, 1649–1656

Available at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.ejconline.com

Subjective assessment by ultrasound is superior to the risk of malignancy index (RMI) or the risk of ovarian malignancy algorithm (ROMA) in discriminating benign from malignant adnexal masses

Toon Van Gorp [a,b,*], Joan Veldman [a,c], Ben Van Calster [a], Isabelle Cadron [a,d], Karin Leunen [a], Frederic Amant [a], Dirk Timmerman [a], Ignace Vergote [a]

[a] *Department of Obstetrics and Gynaecology, Leuven Cancer Institute, Universitaire Ziekenhuizen Leuven, Katholieke Universiteit Leuven, Herestraat 49, 3000 Leuven, Belgium*
[b] *Department of Obstetrics and Gynaecology, MUMC+, GROW – School for Oncology and Developmental Biology, PO Box 5800, 6202AZ Maastricht, The Netherlands*
[c] *Department of Obstetrics and Gynaecology, Virga Jesse Ziekenhuis, Stadsomvaart 11, 3500 Hasselt, Belgium*
[d] *Department of Obstetrics and Gynaecology, AZ Turnhout, Steenweg op Merksplas 44, 2300 Turnhout, Belgium*

Available online 5 January 2012

KEYWORDS
HE4
CA125
Ultrasound
Ovarian neoplasm
Risk of malignancy index
Subjective assessment
Risk of ovarian malignancy algorithm
Sensitivity and specificity

Abstract *Purpose:* The combination of two tumour markers, CA125 and HE4, in the risk of ovarian malignancy assay (ROMA) has been shown to be successful in classifying patients into those who have a high or low risk of epithelial ovarian cancer. In the present study, the diagnostic accuracy of ROMA was assessed and compared to the diagnostic accuracy of the two most widely used ultrasound methods, namely the risk of malignancy index (RMI) and subjective assessment by ultrasound.
Methods: From August, 2005 to March, 2009, 432 women with a pelvic mass who were scheduled to have surgery were enrolled in a single-centre prospective cohort study. A preoperative ultrasound was performed and preoperative CA125 and HE4 serum levels were measured. Once the final surgical pathology reports were obtained, the diagnostic accuracy and performance indices of ROMA, RMI and subjective assessment were calculated.
Results: Of the 432 eligible patients, 374 could be analysed. Subjective assessment had the highest area under the receiver operator characteristic curve (AUC) (0.968, 95% CI:0.945–0.984),

* *Corresponding author at:* Division of Gynaecological Oncology, Department of Obstetrics and Gynaecology, Maastricht University Medical Centre – MUMC+, GROW – School for Oncology and Developmental Biology, PO Box 5800, 6202AZ Maastricht, The Netherlands. Tel.: +31 43 3874767; fax: +31 43 3874765.
E-mail address: toon.van.gorp@mumc.nl (T. Van Gorp).

doi:10.1016/j.ejca.2011.12.003

followed by the RMI (0.931, 95% CI:0.901–0.955). The subjective assessment and RMI both had significantly higher AUCs than the ROMA (0.893, 95% CI:0.857–0.922; $P < 0.0001$ and $P = 0.0030$, respectively). The pre- and postmenopausal populations generated similar results. *Conclusion:* Although new tumour markers models are promising, they do not contribute significantly to the diagnosis of ovarian cancer. Ultrasound, especially subjective assessment by ultrasound, remains superior in discriminating malignant from benign ovarian masses.

1. Introduction

Ovarian cancer is the sixth most common cause of cancer-related death among women in Europe.[1] Differentiating between benign and malignant pelvic masses is difficult due to the anatomical localisation of the ovaries. Generally, women are evaluated on the basis of their personal history, a clinical examination, ultrasound and tumour marker levels. CA125 is the most widely used tumour marker in ovarian cancer.[2] A significant problem associated with CA125 is that it can be expressed in numerous benign and malignant conditions, which leads to false positive results; moreover, it is only expressed by about 50% of early stage ovarian cancers, which leads to false negative results.[3] Another tumour marker which gained attention is the human epididymis secretory protein 4 (HE4). HE4 is overexpressed by ovarian and endometrial cancer.[4–6] Moore et al. developed an algorithm, the risk of ovarian malignancy algorithm (ROMA), which is based on both CA125 and HE4. The ROMA was suggested to be superior to CA125 alone.[4] Some validation studies confirmed the superiority of the ROMA to CA125 alone,[6] while others did not.[7,8]

Sonography by greyscale and colour Doppler imaging is also used widely to classify ovarian masses. While many scoring systems and models have been described, the risk of malignancy index (RMI) is probably the most widely used model at present.[9,10] The RMI is calculated by an algorithm based on several ultrasound variables, the menopausal status and the CA125 level. Its relative simplicity makes it easy to use. Another ultrasound method used to evaluate ovarian masses is the subjective impression of a sonographer, the so-called subjective assessment or pattern recognition. Subjective assessment is a highly accurate method for discriminating benign from malignant ovarian masses.[11–15]

Previous studies have compared the ROMA to RMI,[8,16] but comparison between subjective assessment and the ROMA has never been made. In 2005, we initiated a prospective cohort study to validate newly discovered biomarkers such as HE4. In a previous study, this cohort was used to compare the accuracy of ROMA to that of CA125.[7] The aim of this study was to determine whether ultrasound models are similar or superior to the ROMA.

2. Materials and methods

2.1. Patients

From August, 2005 until March, 2009, 432 consecutive women were found to be eligible to participate in a prospective single-centre cohort study conducted at the University Hospitals Leuven. Patients were considered to be eligible if they were diagnosed with a pelvic mass that was suspected to be of ovarian origin and they were to undergo surgery. Prior to surgery, imaging by pelvic US was performed and a serum sample was taken for tumour marker analysis. Patients with a prior bilateral oophorectomy were not eligible. Patients who were diagnosed with ovarian cancer were completely surgically staged. Prior to enrolment in the study, all patients were required to give fully informed consent. The protocol was approved by the Local Ethics Committee (reference: OG032/ML3132). Patient participation in the study was concluded once the final surgical pathology reports were obtained.

2.2. Ultrasound

All ultrasound examinations were performed in the same department by a standardised examination technique that employed standardised terms and definitions and high-quality ultrasound equipment.[17] The examiner was an experienced sonographer or a trainee supervised by an experienced sonographer. Transvaginal sonography was performed in all cases. Transabdominal sonography was added to examine large masses that could not be seen in their entirety by using a transvaginal probe.

2.2.1. RMI

Defined as $U \times M \times$ CA125, where $U =$ the ultrasound score, $M =$ menopausal status, and CA125 = the level of this marker.[9] U was calculated as follows: multilocularity, solid areas, bilaterality, ascites and intra-abdominal metastases each scored one point and total scores of 0, 1 and $\geqslant 2$ points yielded U values of 0, 1 and 3, respectively. Postmenopausal status was associated with an M score of 3 and was defined as more than 1 year of amenorrhoea, or an age of 50 years or older if the woman had had a prior hysterectomy.

A premenopausal status yielded an *M* score of 1. Since the CA125 level (U/mL) was not disclosed to the sonographer at the time of the ultrasound examination, this was entered into the equation after the ultrasound report was finished. A cut-off of 200 was used to differentiate between benign and malignant, as suggested in the literature.[9]

2.2.2. Subjective assessment

On the basis of greyscale and colour Doppler findings, the ultrasound examiner was obliged to give his/her subjective impression in two ways: (a) classification of each mass as benign or malignant, and (b) expressing his/her level of confidence as follows: benign, probably benign, uncertain, probably malignant, or malignant. The category 'uncertain' was split into two subcategories: uncertain but initially classified as benign, and uncertain but initially classified as malignant.

2.3. Serum samples and marker assays

Immediately before surgery, blood samples were obtained in 10 ml clotting tubes (BD Vacutainer® Serum Tube, ref. 369033). Serum tubes were centrifuged at 800*g* for 10 min. The serum was collected, dispensed into cryotubes and frozen at –80 °C. CA125 and HE4 concentrations were measured by using the CanAg CA125 EIA assay and HE4 EIA assay (Fujirebio Diagnostics, Göteborg, Sweden), according to the manufacturer's instructions. Each enzyme-linked immunosorbent assay (ELISA) was performed manually in duplicate. CA125 and HE4 were combined in the ROMA, as described previously.[7] For premenopausal and postmenopausal patients, cut-offs of 12.5% and 14.4%, respectively, were used.[18]

2.4. Histology

The histology of the tumours was classified according the World Health Organisation classification of tumours.[19] Borderline tumours were not excluded from the present analysis and were classified as malignant tumours.

2.5. Statistical analysis

Statistical analysis was performed with MedCalc v11.5.1.0 (MedCalc Software, Mariakerke, Belgium). Mean patient ages were compared by using the independent Student's *t*-test (Welch-test) and menopausal status was compared by using the Chi-square test. Receiver operator characteristic (ROC) curves were constructed and the areas under the curve (AUC) with binomial exact 95% confidence intervals (95% CI) were calculated.[20] Using the six levels of diagnostic confidence as different cut-offs, an ROC curve could be constructed for the subjective assessment as well. The method described by DeLong et al.[21] was used to calculate the difference between two AUCs.

The diagnostic performance of the models was also expressed as sensitivity, specificity, positive and negative predictive values and positive and negative likelihood ratios when using the recommended cut-off values for the ROMA and the RMI. Since the sonographer had to distinguish between benign and malignant, this was also used as a cut-off. However, the sensitivity and specificity of a model depend on the chosen cut-off, whereas the AUC reflects overall test performance. Therefore, the AUC was considered to be the most important measure of diagnostic performance.

Since both the RMI and the ROMA include the menopausal status in their algorithm, the statistical analysis was performed on the whole population and after stratification for menopausal status. Other exploratory subset analyses are provided in a Supplementary file.

For all statistical comparisons, a level of $P < 0.05$ was accepted as being statistically significant.

3. Results

3.1. Patient and tumour characteristics

Of the 432 eligible patients, 374 could be analysed (Fig. 1). The reasons for patient exclusion are detailed in Table 1. Of the 374 analysed patients, 224 (59.9%) and 150 (40.1%) patients had benign and malignant disease, respectively. Patients with benign disease were younger (mean age = 46.2 [95% CI:44.1–48.3] versus 57.7 [95% CI:55.7–59.8] years; $P < 0.0001$). Of the patients with benign disease, 37.9% (95% CI:31.6–44.3) were postmenopausal, while 74.0% (95% CI:67.0–81.0) of the patients with malignant disease were postmenopausal ($P < 0.0001$).

The most common benign ovarian tumours were endometriomas, cystadenomas, mature teratomas, cystadenofibromas, fibromas/thecomas and functional cysts (Fig. 1). Mixed tumours ($n = 14$) contain two or more different histological subtypes, making it impossible to categorise these tumours into a specific subtype. The cystadenomas and cystadenofibromas included 46 serous, 25 mucinous, and five other histological types or mixed cystadenomas/cystadenofibromas. The majority of the malignant tumours were epithelial ovarian cancers (Table 2). Most of the epithelial ovarian cancers were of high grade and diagnosed at an advanced stage (Table 2 and Fig. 1). The non-epithelial primary ovarian tumours were sex cord stromal tumours ($n = 2$) and sarcomas ($n = 2$). Along with the primary ovarian tumours, 25 extra-ovarian primary tumours with metastases to the ovary were diagnosed. The metastatic tumours were mainly of an endometrial or gastrointestinal origin.

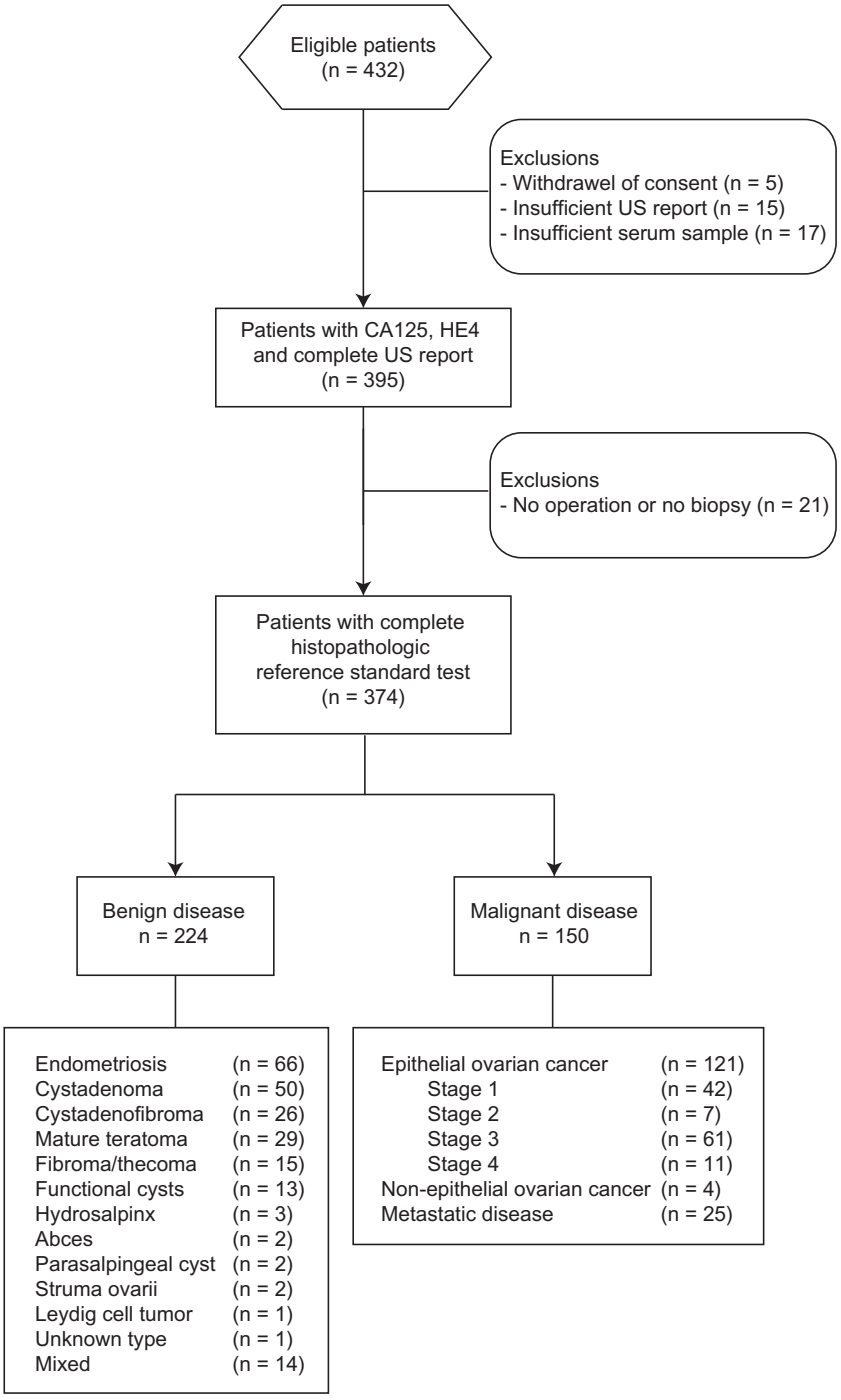

Fig. 1. Flow diagram indicating the inclusion and exclusion of eligible patients.

3.2. ROC curves

For the whole study population, subjective assessment was associated with the highest AUC, followed by the RMI, and the ROMA (Fig. 2). Pairwise comparison of ROC curves (Table 3) indicated that the AUCs of both ultrasound models (subjective assessment and RMI) were significantly larger than the AUC of the ROMA. Similar results were found after stratification according to menopausal status. When subjective assessment and the RMI were compared to each other, subjective assessment performed significantly better, even after stratification according to menopausal status (Table 3). Exploratory subset analyses (excluding borderline tumours, metastatic tumours, non-epithelial ovarian tumours and/or advanced stage disease are provided in a Supplementary file. Subjective assessment was consistently associated with the highest AUC.

Table 1
Reasons for patient exclusion or non-eligibility and their final histological diagnosis.

	n	Final diagnosis		
		Benign	Malignant	Unknown
Withdrawal/refusal of consent	5	–	–	5
Insufficient ultrasound report				
No ultrasound performed	13	3	10	0
Data missing	2	1	1	0
Insufficient serum sample				
No sample taken	5	2	3	0
Insufficient volume of sample	11	3	8	0
Problem with processing of sample in lab	1	1	0	0
No operation or no biopsy				
Conservative management due to poor prognosis	4	0	4	0
Conservative management for a presumed benign cyst	6	6	0	0
Operated in other hospital: no pathology report obtained	5	0	0	5
No cyst at the time of operation	6	6	0	0
Total	58	22	26	10

For patients without a proven histological diagnosis (no operation or biopsy), the presumed diagnosis was based on the patient's clinical course.

Table 2
Histological types and subtypes and the differentiation grade of malignant disease.

	n	%
Histological type		
Epithelial	121	80.7
Serous	76	50.7
Mucinous	21	14.0
Endometrioid	6	4.0
Clear cell	6	4.0
Mixed	5	3.3
Carcinosarcoma	4	2.7
Undifferentiated	3	2.0
Granulosa cell	2	1.3
Sarcoma	2	1.3
Metastatic	25	16.7
Endometrium	11	7.3
Colon	5	3.3
Appendix	3	2.0
Mesothelioma	1	0.7
Breast	1	0.7
Lung	1	0.7
Lymphoma	1	0.7
Pancreas	1	0.7
Stomach	1	0.7
Total	150	100.0
Differentiation grade[a]		
Borderline	31	25.6
1 – Well differentiated	12	9.9
2 – Moderately differentiated	14	11.6
3 – Poorly differentiated	64	52.9
Total	121	100.0

[a] For epithelial ovarian cancer only.

3.3. Performance indices

The calculated sensitivities and specificities at the recommended cut-off values are shown in Table 4. Subjective assessment scored the highest overall in terms of sensitivity for the whole study population as well as the postmenopausal and premenopausal populations. The RMI had the highest specificity for the whole study population and the postmenopausal population. For the premenopausal population, all diagnostic tests had a high specificity but this was accompanied by a sensitivity below 70% for the RMI, and ROMA.

4. Discussion

Over the past few years, the performance of HE4 and the ROMA to classify ovarian masses has been studied many times. We showed previously that CA125, HE4 and ROMA perform equally well.[7] In the present study, the ability of the ROMA to diagnose ovarian cancer was compared to that of greyscale and colour Doppler ultrasound. The present data suggest that ultrasound methods are superior to ROMA to classify ovarian masses. Moreover, subjective assessment was superior to the RMI.

The RMI is very popular because of its simplicity: little experience is needed to detect the different ultrasound features that have to be scored (multilocularity, solid areas, bilaterality, ascites and intra-abdominal metastases) and the algorithm can be memorised readily. It also enables general gynaecologists to refer patients on an objective basis to gynaecological oncologists.[22] With an AUC of 0.931, it seems that the overall performance of the RMI is good. However, due to a high false negative rate at the suggested cut-off of 200, there was a low sensitivity of only 76.0%. This is in accordance with a recent review that calculated a pooled estimate of sensitivity of 78%.[10] This means that in one of four cases, the tumour will be wrongly diagnosed as benign. In the worst case scenario, these patients will not be referred to a gynaecological oncologist and will be operated on by laparoscopy, which is associated with an increased

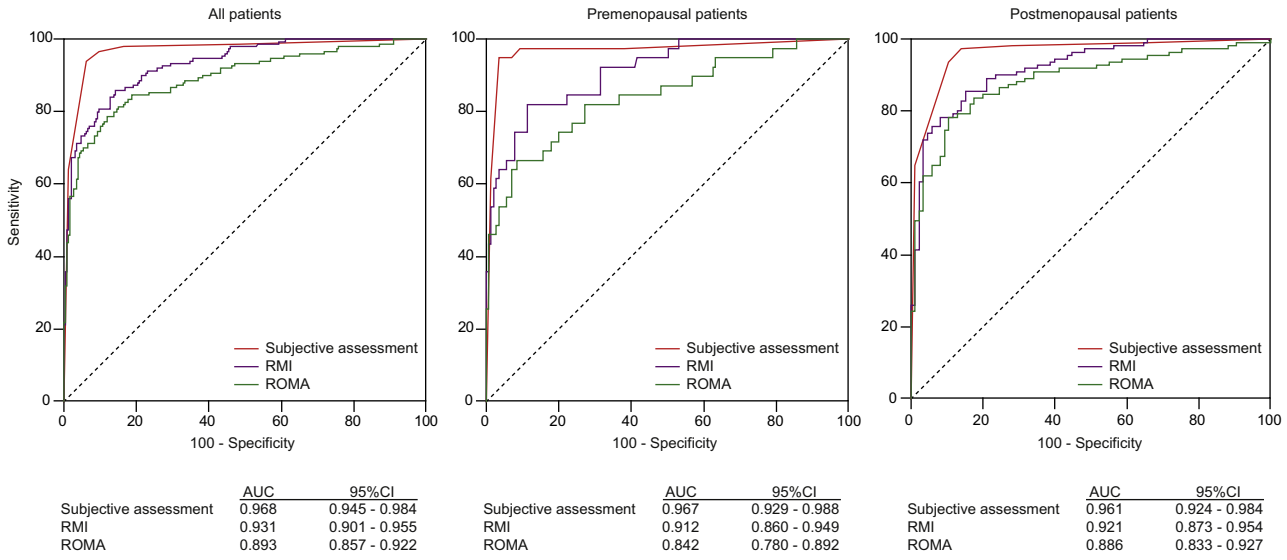

Fig. 2. ROC curves for the detection of malignant disease (including borderline ovarian tumours) for subjective assessment with sonography, risk of malignancy index (RMI) and risk of ovarian malignancy algorithm (ROMA) in the whole population, the pre-menopausal population and the post-menopausal population. Total area under the curve (AUC) values with corresponding 95% confidence intervals are listed below the curves.

Table 3
Differences in the area under the curve (AUC) of the receiver operating characteristic (ROC) curves for the diagnosis of malignant disease (including borderline ovarian tumours) with the corresponding 95% confidence intervals (95% CI) and *P*-values. Pairwise ROC curve comparisons were calculated for the whole study population, for the postmenopausal population and for the premenopausal population. The method described by DeLong et al. was used to calculate the difference between two AUCs.[19]

All patients			Premenopausal		Postmenopausal	
Difference	95% Confidence interval	*P*-value	Difference	95% CI	Difference	95% CI
Subjective assessment versus ROMA						
0.076	0.043–0.109	<0.0001	0.125	0.054–0.197	0.075	0.028–0.121
RMI versus ROMA						
0.039	0.013–0.064	0.0030	0.070	0.003–0.136	0.034	0.007–0.061
Subjective assessment versus RMI						
0.037	0.011–0.063	0.0058	0.056	–0.004–0.115	0.041	0.003–0.078

Abbreviations: RMI = risk of malignancy index; ROMA = risk of ovarian malignancy algorithm.

risk of spillage of cyst fluid.[23] Both the non referral and cyst fluid spillage might decrease the overall survival.[24,25]

The RMI and ROMA have been compared previously in two studies.[8,16] One was a multicentre study by Moore et al. that determined the RMI and ROMA values for 457 patients.[16] At a fixed specificity of 75%, the ROMA had a sensitivity of 94.3% while the RMI had a sensitivity of 84.6% ($P = 0.0029$). However, despite the fact that there was an important quality control for the ROMA (central lab measurements and quality assurance), there was no central radiology review or standardisation of the imaging reports. Moreover, the RMI was calculated by using information from a variety of imaging techniques ultrasound, computer tomography (CT) scan and magnetic resonance imaging (MRI), with only 85.3% of patients having received a standard ultrasound scan. The RMI was developed to be used with ultrasound and was not validated for other imaging techniques. CT scans are usually not indicated in the evaluation of adnexal masses because of poor soft tissue discrimination.[26] In a second study that compared the ROMA to the RMI, the results of both algorithms in 127 patients with benign or malignant ovarian disease were compared.[8] Although different kinds of subgroups were analysed in the article, the authors did not describe the overall performance of ROMA and RMI in the comparison of benign versus malignant disease, even though this is the most important comparison to be made in a diagnostic setting.

With an AUC of 0.968, subjective assessment appears to be an excellent method to discriminate between benign and malignant adnexal masses. Moreover, its sensitivity is 96.7%, which means that only one of 30 cases will be misdiagnosed as a benign mass. Subjective assessment has been validated by numerous studies.[11–14]

Table 4
Sensitivity, specificity, positive likelihood ratio (+LR), negative likelihood ratio (–LR), positive predictive value (PPV) and negative predictive value (NPV) of subjective assessment, the risk of malignancy index (RMI) and the risk of ovarian malignancy algorithm (ROMA) for malignant disease (including borderline ovarian tumours). The diagnostic performance indices are calculated for the whole study population, the postmenopausal population and the premenopausal population. The 95% confidence intervals (95% CI) are indicated between brackets.

	Subjective assessment		RMI		ROMA	
All patients						
Sensitivity	96.7%	(92.4–98.9%)	76.0%	(68.4–82.6%)	84.7%	(77.9–90.0%)
Specificity	90.2%	(85.5–93.7%)	92.4%	(88.1–95.5%)	76.8%	(70.7–82.2%)
+LR	9.84	(6.61–14.65)	10.01	(6.29–15.95)	3.65	(2.85–4.67)
–LR	0.04	(0.02–0.09)	0.26	(0.19–0.35)	0.20	(0.14–0.29)
PPV	86.8%	(80.7–91.6%)	87.0%	(80.0–92.3%)	71.0%	(63.7–77.5%)
NPV	97.6%	(94.5–99.2%)	85.2%	(80.1–89.4%)	88.2%	(82.8–92.4%)
Postmenopausal						
Sensitivity	97.3%	(92.3–99.4%)	80.2%	(71.5–87.1%)	91.0%	(84.1–95.6%)
Specificity	85.9%	(76.6–92.5%)	87.1%	(78.0–93.4%)	58.8%	(47.6–69.4%)
+LR	6.89	(4.08–11.65)	6.20	(3.54–10.84)	2.21	(1.70–2.87)
–LR	0.03	(0.01–0.10)	0.23	(0.16–0.33)	0.15	(0.08–0.28)
PPV	90.0%	(83.2–94.7%)	89.0%	(81.2–94.4%)	74.3%	(66.1–81.4%)
NPV	96.1%	(88.9–99.2%)	77.1%	(67.4–85.1%)	83.3%	(71.5–91.7%)
Premenopausal						
Sensitivity	94.9%	(82.7–99.4%)	64.1%	(47.2–78.8%)	66.7%	(49.8–80.9%)
Specificity	92.8%	(87.2–96.5%)	95.7%	(90.8–98.4%)	87.8%	(81.1–92.7%)
+LR	13.19	(7.23–24.07)	14.85	(6.56–33.62)	5.45	(3.31–8.97)
–LR	0.06	(0.01–0.21)	0.38	(0.25–0.57)	0.38	(0.24–0.59)
PPV	78.7%	(64.3–89.3%)	80.7%	(62.5–92.6%)	60.5%	(44.4–75.2%)
NPV	98.5%	(94.6–99.8%)	90.5%	(84.5–94.7%)	90.4%	(84.1–94.8%)

The cut-off values used were 200 for the RMI, and 12.5% and 14.4% for the premenopausal and postmenopausal patients with the ROMA, respectively.
The prevalence of malignancy: all patients 40.1% (95% CI: 35.1–45.3%), postmenopausal patients 56.6% (95% CI: 49.4–63.7%) and premenopausal patients 21.9% (95% CI: 16.1–28.7%).

Ultrasound examiners take demographic, clinical and ultrasound information into account when they evaluate an adnexal mass and they subconsciously apply their experience from previous examinations during subsequent evaluations of the adnexal masses. The level of ultrasound expertise therefore has a marked influence on the quality of an ultrasound examination.[27] The International Ovarian Tumour Analysis (IOTA) study group has sought to tackle this problem by developing a two-tiered diagnostic setup. Ten simple ultrasound rules were developed for discriminating between benign and malignant adnexal masses.[28] These rules were applicable in three-quarters of patients with an ovarian mass with a sensitivity of 92% and a specificity of 96%.[29] When these simple rules do not apply the patient should be referred to an expert ultrasound centre. This two-tiered diagnostic setup yielded a sensitivity of 91% and a specificity of 93%.

The present study had certain limitations. One was that 13.4% of all eligible patients were excluded; the majority had malignant adnexal masses. We presume that in these cases, ultrasound was not performed due to time constraints. The serum volume was also more frequently insufficient due to a larger number of blood tubes taken for other analyses. Second, the data were obtained in a tertiary referral centre with a specialised gynaecological ultrasound unit. The high prevalence of malignant disease in this centre will have influenced the predictive values in this study. A lower prevalence of ovarian cancer would have decreased the positive predictive value and increased the negative predictive value. However, this would also have been true for all of the diagnostic tests examined in this study. Moreover, in smaller hospitals with a lower prevalence of ovarian cancer, a test with the same sensitivity and specificity would yield an even higher negative predictive value. Third, the fact that these results were obtained within a tertiary referral centre will also have increased the experience of the sonographers and thereby improved the performance of the ultrasound methods, in particular the subjective assessment. Nevertheless, we hope that the present study will encourage sonographers to increase their knowledge and experience by training in gynaecological sonography.

In summary, although much energy has been put into the discovery and validation of new tumour marker algorithms, such as the ROMA, the present study suggests that the diagnostic value of these algorithms is limited compared to sonography. In particular, subjective assessment seems to be highly accurate. The fact that subjective assessment is influenced by experience should not discourage sonographers in non-expert centres to increase their knowledge and expertise.

Role of the funding source

The supporting organisations had no role in the study design or in the collection, analysis and interpretation of data.

Financial disclosures

There are no financial disclosures.

Conflict of interest statement

None declared.

Acknowledgements

This work was supported, in part, the Belgian Federation against Cancer, a non-profit organisation (SCIE2004-42), the Research Foundation – Flanders (FWO) (G.0457.05 and 1.2.516.09N) and IWT-TBM 070706 (IOTA); B.V.C. is a postdoctoral fellow of the Research Foundation – Flanders (FWO).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejca.2011.12.003.

References

1. Ferlay J, Parkin DM, Steliarova-Foucher E. Estimates of cancer incidence and mortality in Europe in 2008. *Eur J Cancer* 2010;**46**(4):765–81.
2. Bast Jr RC, Klug TL, St John E, et al. A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *N Engl J Med* 1983;**309**(15):883–7.
3. Jacobs I, Bast Jr RC. The CA 125 tumour-associated antigen: a review of the literature. *Human Reprod* 1989;**4**(1):1–12.
4. 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):402–8.
5. Hellstrom I, Raycraft J, Hayden-Ledbetter M, et al. The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. *Cancer Res* 2003;**63**(13):3695–700.
6. Moore RG, McMeekin DS, Brown AK, et al. A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol* 2009;**112**(1):40–6.
7. Van Gorp T, Cadron I, Despierre E, et al. HE4 and CA125 as a diagnostic test in ovarian cancer: prospective validation of the Risk of Ovarian Malignancy Algorithm. *Br J Cancer* 2011;**104**(5):863–70.
8. Jacob F, Meier M, Caduff R, et al. No benefit from combining HE4 and CA125 as ovarian tumor markers in a clinical setting. *Gynecol Oncol* 2011;**121**(3):487–91.
9. Jacobs I, Oram D, Fairbanks J, et al. 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**(10):922–9.
10. Geomini P, Kruitwagen R, Bremer GL, Cnossen J, Mol BW. The accuracy of risk scores in predicting ovarian malignancy: a systematic review. *Obstet Gynecol* 2009;**113**(2 Pt 1):384–94.
11. Timmerman D, Schwarzler P, Collins WP, et al. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. *Ultrasound Obstet Gynecol* 1999;**13**(1):11–6.
12. Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. *Ultrasound Obstet Gynecol* 1999;**14**(5):338–47.
13. 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**(4):357–65.
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**(1):91–104.
15. Van Calster B, Timmerman D, Bourne T, et al. Discrimination between benign and malignant adnexal masses by specialist ultrasound examination versus serum CA-125. *J Natl Cancer Inst* 2007;**99**(22):1706–14.
16. Moore RG, Jabre-Raughley M, Brown AK, et al. Comparison of a novel multiple marker assay vs the Risk of Malignancy Index for the prediction of epithelial ovarian cancer in patients with a pelvic mass. *Am J Obstet Gynecol* 2010;**203**(3):228.e1–6.
17. Timmerman D, Valentin L, Bourne TH, et al. 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.
18. HE4 EIA product insert, version 2008-09. Available from: http://www.fdi.com/documents/products/inserts/eia/HE4%20EIA%20404-10,%202008-09,%20r1.pdf [accessed April 1, 2010].
19. World Health Organization classification of tumours. Pathology and genetics of the breast and female genital organs. Lyon, IARC Press, 2003.
20. Hilgers RA. Distribution-free confidence bounds for ROC curves. *Methods Inf Med* 1991;**30**(2):96–101.
21. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**(3):837–45.
22. van den Akker PA, Aalders AL, Snijders MP, et al. Evaluation of the Risk of Malignancy Index in daily clinical management of adnexal masses. *Gynecol Oncol* 2010;**116**(3):384–8.
23. Panici PB, Palaia I, Bellati F, et al. Laparoscopy compared with laparoscopically guided minilaparotomy for large adnexal masses: a randomized controlled trial. *Obstet Gynecol* 2007;**110**(2 Pt 1):241–8.
24. du Bois A, Rochon J, Pfisterer J, Hoskins WJ. Variations in institutional infrastructure, physician specialization and experience, and outcome in ovarian cancer: a systematic review. *Gynecol Oncol* 2009;**112**(2):422–36.
25. Vergote I, De Brabanter J, Fyles A, et al. Prognostic importance of degree of differentiation and cyst rupture in stage I invasive epithelial ovarian carcinoma. *Lancet* 2001;**357**(9251):176–82.
26. Togashi K. Ovarian cancer: the clinical role of US, CT, and MRI. *Eur Radiol* 2003;**13**(Suppl 4):L87–104.
27. Yazbek J, Raju SK, Ben-Nagi J, et al. Effect of quality of gynaecological ultrasonography on management of patients with suspected ovarian cancer: a randomised controlled trial. *Lancet Oncol* 2008;**9**(2):124–31.
28. Timmerman D, Testa AC, Bourne T, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol* 2008;**31**(6):681–90.
29. 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.