Discrimination Between Benign and Malignant Adnexal Masses by Specialist Ultrasound Examination Versus Serum CA-125

Ben Van Calster, Dirk Timmerman, Tom Bourne, Antonia Carla Testa, Caroline Van Holsbeke, Ekaterini Domali, Davor Jurkovic, Patrick Neven, Sabine Van Huffel, Lil Valentin

Background Subjective evaluation of gray-scale and Doppler ultrasound findings (i.e., pattern recognition) by an experienced examiner and preoperative serum levels of CA-125 can both discriminate benign from malignant adnexal (i.e., ovarian, paraovarian, or tubal) masses. We compared the diagnostic performance of these methods in a large multicenter study.

Methods In a prospective multicenter study—the International Ovarian Tumor Analysis—1066 women with a persistent adnexal mass underwent transvaginal gray-scale and color Doppler ultrasound examinations by an experienced examiner within 120 days of surgery. Pattern recognition was used to classify a mass as benign or malignant. Of these women, 809 also had blood collected preoperatively for measurement of serum CA-125. Various levels of CA-125 were used as cutoffs to classify masses. Results from both assays were then compared with histologic findings after surgery.

Results Pattern recognition correctly classified 93% (95% confidence interval [CI] = 90.9% to 94.6%) of the tumors as benign or malignant. Serum CA-125 correctly classified at best 83% (95% CI = 80.3% to 85.6%) of the masses. Histologic diagnoses that were most often misclassified by CA-125 were fibroma, endometrioma, and abscess (false-positive results) and borderline tumor (false-negative results). Pattern recognition correctly classified 86% (95% CI = 81.1% to 90.4%) of masses of these four histologic types as being benign or malignant, whereas a serum CA-125 at a cutoff of 30 U/mL correctly classified 41% (95% CI = 34.4% to 47.5%) of them. Pattern recognition assigned a correct specific histologic diagnosis to 333 (59%, 95% CI = 54.5% to 62.8%) of the 567 benign lesions.

Conclusion Pattern recognition was superior to serum CA-125 for discrimination between benign and malignant adnexal masses.

J Natl Cancer Inst 2007;99:1706–14

Preoperative disease classification for patients with ovarian masses, in particular discrimination between benign and malignant ovarian tumors, is important for optimal patient management. At present it is not clear whether subjective evaluation of an ultrasound image (i.e., pattern recognition) of an adnexal (i.e., ovarian, paraovarian, or tubal) mass or determination of serum CA-125 levels is the best method to distinguish benign from malignant tumors.

CA-125 is a glycoprotein that is assessed by the monoclonal antibody OC 125. An elevated CA-125 serum level of at least 30 U/mL is often considered to indicate the presence of malignancy (1–3), although other CA-125 cutoff values have also been used (e.g., ≥35 U/mL in postmenopausal women and ≥65 U/mL in premenopausal women) (4–8). Serum CA-125 levels can also be misleading—false-positive results can be obtained in women who have conditions that affect the peritoneal surface, such as endometriosis (9), and false-negative results can be obtained in women who have early-stage invasive disease and borderline ovarian tumors (9–12). Notwithstanding these limitations, measurement of CA-125 serum levels has become standard practice for the preoperative evaluation of ovarian masses. Although CA-125 serum levels can discriminate between benign and malignant adnexal masses,

Affiliations of authors: Department of Electrical Engineering, ESAT-SISTA, Katholieke Universiteit Leuven, Belgium (BVC, SVH); Department of Obstetrics and Gynecology, University Hospitals, KU Leuven, Leuven, Belgium (DT, CVH, ED, PN); Department of Obstetrics and Gynaecology, St. George's Hospital Medical School, University of London, London, U.K. (TB); Istituto di Clinica Ostetrica e Ginecologica, Università Cattolica del Sacro Cuore, Rome, Italy (ACT); Department of Obstetrics and Gynaecology, King's College Hospital, London, U.K. (DJ); Department of Obstetrics and Gynaecology, Malmö University Hospital, Lund University, Malmö, Sweden (LV).

Correspondence to: Dirk Timmerman, MD, PhD, Department of Obstetrics and Gynecology, University Hospitals, Katholieke Universiteit Leuven, Herestraat 49, B-3000 Leuven, Belgium (e-mail: dirk.timmerman@uz.kuleuven.ac.be).

See "Funding" and "Notes" following "References."

DOI: 10.1093/jnci/djm199

little is known about the ability of CA-125 serum levels to discriminate among histologic subgroups of adnexal lesions (9,13–15). Moreover, most published studies on histologic subgroups are based on CA-125 data from small numbers of patients that were analyzed as categorical data by use of a defined cutoff value, rather than as a continuous data.

Gray-scale and Doppler ultrasound examination of adnexal masses can also be used to discriminate between benign and malignant tumors (4,16–21). An experienced examiner's subjective evaluation of gray-scale ultrasound morphology and color Doppler findings of an ovarian mass (i.e., pattern recognition) has been shown to be highly accurate and to be better than mathematical models at predicting whether adnexal tumors are malignant or benign (20,22). An experienced ultrasound examiner can also frequently identify the specific histologic type of an adnexal mass (21,23–26).

To our knowledge, there are no published reports comparing the diagnostic performance of CA-125 with that of ultrasound pattern recognition. The aim of this study was to compare the diagnostic performance of pattern recognition by an experienced examiner with that of preoperative levels of serum CA-125 with regard to discrimination between benign and malignant adnexal masses and also to discriminate among subgroups of adnexal masses.

Patients and Methods

Study Participants

We used information from the International Ovarian Tumor Analysis (IOTA) database. The IOTA study was a multicenter study that included nine European ultrasound centers. It was approved by the local ethics committees and has been described in detail previously (27). Oral informed consent was obtained. Women were recruited to the study between June 1, 1999, and June 30, 2002. Women with at least one persistent adnexal mass underwent gray-scale and color Doppler ultrasonography that was performed by experienced ultrasound examiners who used a standardized examination technique and standardized terms and definitions (28). All women received a transvaginal scan; transabdominal sonography was also performed when large masses could not be fully visualized by use of a transvaginal probe. Whenever possible, ultrasound examiners provided a specific histologic diagnosis (e.g., endometrioma, dermoid cyst, or hydrosalpinx [i.e., fluid collection in the Fallopian tube]) by use of pattern recognition data (24). The ultrasound examiner had no knowledge of the CA-125 serum levels when suggesting a diagnosis.

In this analysis, a woman was considered to be postmenopausal if she reported no menstruation for at least 1 year after the age of 40 years, provided that the amenorrhea was not explained by medication or disease. Women who were aged 50 years or older and had undergone a hysterectomy, so that the time of menopause could not be determined, were also defined as postmenopausal. Women with menstrual periods during the year before the examination and women who were younger than 50 years and had undergone a hysterectomy without bilateral oophorectomy were defined as premenopausal.

CONTEXT AND CAVEATS

Prior knowledge

Subjective evaluation of gray-scale and Doppler ultrasound findings (i.e., pattern recognition) by an experienced examiner and preoperative serum levels of CA-125 can both discriminate benign from malignant adnexal (i.e., ovarian, paraovarian, or tubal) masses.

Study design

Multicenter prospective study of 1066 women with a persistent adnexal mass that compared the abilities of pattern recognition and of CA-125 serum levels to classify masses as benign or malignant and to classify them further by histologic subtype. After its resection, each mass was examined histologically.

Contribution

Pattern recognition by an experienced examiner was superior to serum CA-125 for discrimination between benign and malignant adnexal masses and correctly classified more than half of the masses by subtype.

Implications

Measurement of CA-125 is not as helpful as pattern recognition by an experienced examiner for the correct classification of adnexal masses. Because the accuracy of pattern recognition depends on the level of experience of the examiner, more effort should be expended to educate and train ultrasound examiners.

Limitations

CA-125 serum levels were not available for 24% of patients, most of whom had benign masses by pattern recognition with a high degree of confidence. These results reveal a bias whereby serum CA-125 was more likely to be measured in women with masses that were suspected of being malignant.

CA-125 Assay

Centers were encouraged to measure the level of serum CA-125 in peripheral blood from all patients, but the availability of this measurement was not an essential requirement for participation in the study. Second-generation immunoradiometric assay kits for CA-125 (i.e., CA-125 II) (29) were obtained from five companies (Centocor, Malvern, PA; Cis-Bio, Gif-sur-Yvette, France; Abbott Axsym system, REF 3B41-22, Abbott Laboratories Diagnostic Division, Abbott Park, IL; Immuno-l-analyser, Bayer Diagnostics, Tarrytown, NY; or Vidas, bioMérieux, Marcy l'Etoile, France). All kits used the OC 125 antibody.

Histology

The histology of surgically removed ovarian tissues was the final endpoint of the study. Tumors were classified according to the criteria recommended by the International Federation of Gynecology and Obstetrics (30). When the tumor was malignant, the surgical stage was recorded. Histologic diagnoses were grouped into 15 categories—11 of which were benign and four of which were malignant. For subjective assessment of the mass on the basis of ultrasound findings, in addition to these 15 categories, a 16th category was added—i.e., "don't know." The four specific diagnoses of malignancy were primary invasive ovarian cancer, metastasis, borderline tumor, and rare type of malignancy (e.g., granulosa

cell tumor or dysgerminoma). The 11 benign diagnoses were endometrioma, teratoma or dermoid cyst, simple cyst, functional cyst, hydrosalpinx, peritoneal pseudocyst, abscess, fibroma, serous cystadenoma, mucinous cystadenoma, and rare type of benign tumor (e.g., Brenner tumor).

Statistical Analysis

Because CA-125 is a continuous variable and pattern recognition is binary, we did not perform a power calculation. Instead, we aimed for a large database to produce reliable estimates of statistical quantities and their confidence intervals (CIs). A large database also allowed us to investigate CA-125 levels for all 15 diagnostic subgroups.

All statistical analyses were performed with SAS version 9.1 (SAS Institute, Cary, NC). The 95% confidence interval for a single proportion was computed by use of Wilson's score confidence interval method with continuity correction (31). For a difference between two proportions, the 95% confidence interval was computed with method 10 as described by Newcombe for paired (32) and unpaired (33) proportions, respectively. The methods used for differences between proportions are based on Wilson's score interval. In addition, we report paired *P* values from McNemar's test and unpaired *P* values from Fisher's exact test, where appropriate.

The percentages of correctly classified benign and malignant tumors represent the specificity and the sensitivity, respectively. The percentage of correctly classified tumors in general represents the accuracy, which is influenced by the prevalence of malignant tumors. The positive likelihood ratio (LR+) is computed as the sensitivity divided by 100 minus the specificity. The negative likelihood ratio (LR–) is computed as 100 minus the sensitivity divided by the specificity.

One test (e.g., pattern recognition or serum CA-125) was considered to be superior to another test if it was associated with a higher sensitivity at the same or similar specificity or with a higher specificity at the same or similar sensitivity. If a test had a sensitivity for malignancy of less than 50%, that test was considered to be clinically useless.

To determine the difference in CA-125 serum levels between any two diagnostic groups, the θ index for effect size (with its 95%

Table 1. Demographic background data, histologic diagnoses, estimation of risk of malignancy by experienced ultrasound examiners who used pattern recognition, and rate of correct diagnoses with regard to malignancy when using pattern recognition in women included and excluded in this analysis*

Characteristic	Excluded women (n = 257)	Included women		
		All (n = 809)	Premenopausal (n = 445)	Postmenopausal (n = 364)
Mean age, y (SD)	41.9 (14.5)	48.8 (15.6)	37.4 (9.1)	62.8 (9.4)
Postmenopausal, No. (%)	68 (26.5)	364 (45.0)		
Histologic diagnosis, No. (%)				
All benign tumors	233 (90.7)	567 (70.1)	359 (80.7)	208 (57.1)
Endometrioma	84 (32.7)	128 (15.8)	123 (27.6)	5 (1.4)
Dermoid/teratoma	44 (17.1)	83 (10.3)	68 (15.3)	15 (4.1)
Simple cyst	15 (5.8)	84 (10.4)	39 (8.8)	45 (12.4)
Functional cyst	13 (5.1)	15 (1.9)	12 (2.7)	3 (0.8)
Hydrosalpinx	9 (3.5)	15 (1.9)	12 (2.7)	3 (0.8)
Peritoneal pseudocyst	4 (1.6)	4 (0.5)	3 (0.7)	1 (0.3)
Abscess	6 (2.3)	19 (2.3)	13 (2.9)	6 (1.6)
Fibroma	8 (3.1)	29 (3.6)	9 (2.0)	20 (5.5)
Cystadenoma	34 (13.2)	102 (12.6)	38 (8.5)	64 (17.6)
Mucinous cystadenoma	14 (5.4)	80 (9.9)	40 (9.0)	40 (11.0)
Rare benign tumor	2 (0.8)	8 (1.0)	2 (0.4)	6 (1.6)
All malignant tumors	24 (9.3)	242 (29.9)	86 (19.3)	156 (42.9)
Primary invasive	17 (6.6)	127 (15.7)	32 (7.2)*	95 (26.1)*
Stage I	9 (3.5)	33 (4.1)	10 (2.2)	23 (6.3)
Stage II	2 (0.8)	10 (1.2)	2 (0.4)	8 (2.2)
Stage III	4 (1.6)	69 (8.5)	17 (3.8)	52 (14.3)
Stage IV	2 (0.8)	15 (1.9)	3 (0.7)	12 (3.3)
Borderline	3 (1.2)	52 (6.4)	27 (6.1)	25 (6.9)
Metastatic	4 (1.6)	38 (4.7)	13 (2.9)	25 (6.9)
Rare primary invasive	0 (0.0)	25 (3.1)	14 (3.1)	11 (3.0)
Risk estimation by US examiner, No. (%)				
Certainly benign	168 (65.4)	384 (47.5)	276 (62.0)	108 (29.7)
Probably benign	41 (16.0)	140 (17.3)	75 (16.9)	65 (17.9)
Unclassifiable	24 (9.3)	66 (8.2)	26 (5.8)	40 (11.0)
Probably malignant	14 (5.4)	69 (8.5)	26 (5.8)	43 (11.8)
Certainly malignant	10 (3.9)	150 (18.5)	42 (9.4)	108 (29.7)
Correctly classified with regard to malignancy by the US examiner, No. (%)	244 (94.9)	752 (93.0)	421 (94.6)	331 (90.9)
Benign masses, No. (% all benign)	224 (96.1)	538 (94.9)	351 (97.8)	187 (89.9)
Malignant masses, No. (% all malignant)	20 (83.3)	214 (88.4)	70 (81.4)	144 (92.3)

* Women included in this study had CA-125 serum levels available. SD = standard deviation; US = ultrasound.

confidence interval) was used as the main indicator (34). The θ index can take on any value between 0.5 and 1 and can be interpreted as the degree of overlap between two CA-125 distributions. A θ value of 1 means no overlap; i.e., patients from one group had higher CA-125 levels than any patient from the other group. A θ value of 0.5 means maximal overlap; i.e., CA-125 levels from patients in one group were not generally higher or lower than those in the other group. The θ index is mathematically identical to the area under the receiver operating characteristic curve, which may help nonstatisticians with the interpretation of θ, although a receiver operating characteristic curve analysis has different objectives (34). We also report *P* values from the Mann–Whitney test. All statistical tests were two-sided.

Results

Patient Demographic and Tumor Characteristics

Of the 1066 patients in the IOTA database, 809 had CA-125 serum levels available and were included in this analysis. Table 1 gives an overview of the patients who were included and excluded from this study. Some centers did not measure serum CA-125 if the tumor was judged to be clearly benign on the basis of pattern recognition. Therefore, the group of excluded patients contained more young women, more premenopausal women, and more women with benign tumors than the group included in this study. Pattern recognition correctly classified 95% of the benign tumors in the patients included and in 96% of those excluded (difference = −1%, 95% CI = −4.1% to 2.4%, *P* = .583) and 88% versus 83% of the malignant tumors (difference = 5%; 95% CI = −5.9% to 24.6%, *P* = .507).

CA-125 Levels and Histologic Diagnosis

Preoperative CA-125 serum levels differed depending on histology (Fig. 1). CA-125 serum levels were 30 U/mL or higher in 183 of the 242 women with malignancies (76%, 95% CI = 69.6% to 80.8%), 81 of the 128 with endometriomas (63%, 95% CI = 54.3% to 71.5%), 11 of the 19 with abscesses (58%, 95% CI = 34.0% to 78.9%), and 13 of the 29 with fibromas (45%, 95% CI = 27.0% to 64.0%). In contrast, CA-125 serum levels were 30 U/mL or higher in only nine of the 83 women with dermoid cysts (11%, 95% CI = 5.4% to 20.1%), 11 of the 84 with simple cysts (13%, 95% CI = 7.0% to 22.6%), and one of the 15 with hydrosalpinges (7%, 95% CI = 0.3% to 34.0%). The CA-125 serum levels in women with endometriomas, fibromas, or abscesses were similar to those in women with borderline ovarian malignancies. CA-125 serum levels were higher in women with stage I primary invasive ovarian cancers, rare malignancies, or metastases than in women with benign or borderline tumors. CA-125 levels were highest in women with stage II–IV primary invasive ovarian cancers. CA-125 serum levels were lower in women with stage I primary invasive ovarian cancers than in those with stage II (θ = 0.75, 95% CI = 0.55 to 0.88, *P* = .017), stage III (θ = 0.80, 95% CI = 0.69 to 0.87, *P*<.001), or stage IV (θ = 0.87, 95% CI = 0.71 to 0.95, *P*<.001) cancers.

Pattern Recognition and Histologic Diagnosis

By use of pattern recognition, ultrasound examiners assigned a correct specific diagnosis to at least 80% of endometriomas (80%, 95% CI = 72.3% to 86.7%), dermoid cysts (84%, 95% CI = 74.3% to 91.1%), hydrosalpinges (93%, 95% CI = 66.0% to 99.7%), and primary invasive ovarian cancers (80%, 95% CI = 71.3% to 86.0%) (Table 2). The examiners assigned a correct specific diagnosis to 333 (59%, 95% CI = 54.5% to 62.8%) of the 567 benign tumors. Among premenopausal women, 67% (95% CI = 62.3% to 72.2%) of the benign lesions and 48% (95% CI = 36.9% to 58.7%) of the malignant lesions were assigned a correct specific diagnosis. Among postmenopausal women, 44% (95% CI = 37.0% to 50.8%) of the benign lesions and 58% (95% CI = 49.5% to 65.5%) of the malignant lesions were assigned a correct specific diagnosis.

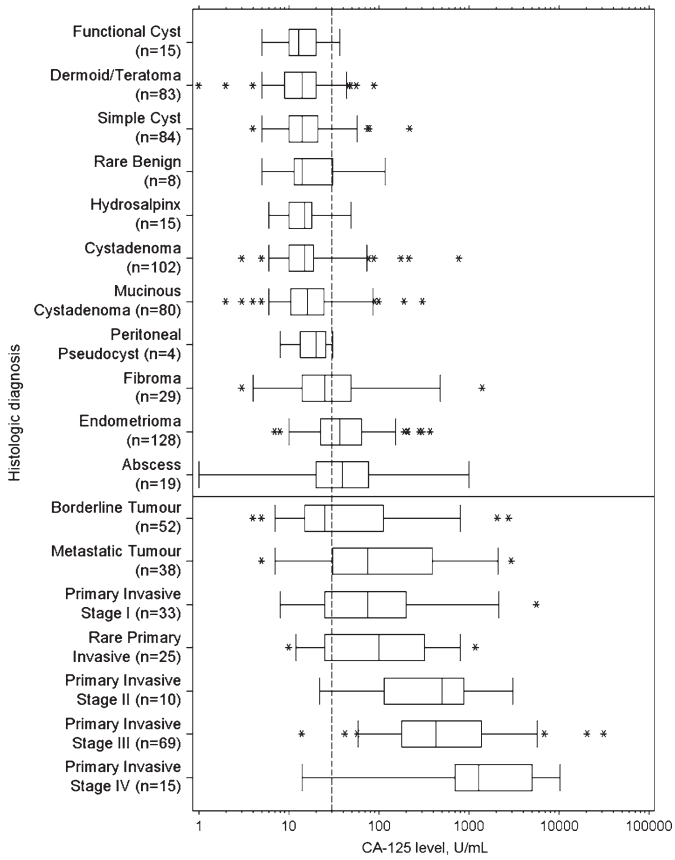

Fig. 1. Box plots showing CA-125 serum levels by histologic diagnosis of adnexal masses. **Box** = range of the middle 50% of the CA-125 levels; **line inside the box** = median; **whiskers** = the 5th and 95th percentile; **asterisks** = data points that lie outside the whiskers.

Comparison of Diagnostic Abilities Between Pattern Recognition and CA-125 Serum Levels

Pattern recognition correctly classified 93% (95% CI = 90.9% to 94.6%) of the tumors as benign or malignant, whereas a CA-125 serum cutoff value of 30 U/mL classified 72% (95% CI = 69.2% to 75.5%) of the tumors correctly (Table 2). By the use of CA-125 serum levels, at best 83% (95% CI = 80.3% to 85.6%) of the tumors were correctly classified with a cutoff value of 100 U/mL. However, when a CA-125 serum cutoff of 100 U/mL, instead of 30 U/mL, was used, CA-125 correctly classified as many as 94% (95% CI = 91.8% to 95.9%) of the benign tumors as benign but only 57% (95% CI = 50.5% to 63.3%) of the malignant tumors as malignant.

Table 2. Rate of correct diagnoses in histologic subgroups, when using ultrasound findings (pattern recognition) versus when using CA-125, and rate of correct specific ultrasound diagnoses, when using pattern recognition

Histologic diagnosis	No.	Median CA-125, U/mL	Correct specific US diagnosis*, No. (%)	Correctly classified diagnoses as benign or malignant, No. (%)						
				Pattern recognition	CA-125 cutoff†					
					30 U/mL	65 U/mL	100 U/mL	200 U/mL	400 U/mL	1000 U/mL
Endometrioma	128	37	103 (80)	125 (98)	47 (37)	96 (75)	112 (88)	123 (96)	128 (100)	128 (100)
Dermoid/teratoma	83	14	70 (84)	82 (99)	74 (89)	82 (99)	83 (100)	83 (100)	83 (100)	83 (100)
Simple cyst	84	14	28 (33)	83 (99)	73 (87)	80 (95)	83 (99)	83 (99)	84 (100)	84 (100)
Functional cyst	15	13	2 (13)	15 (100)	12 (80)	15 (100)	15 (100)	15 (100)	15 (100)	15 (100)
Hydrosalpinx	15	15	14 (93)	14 (93)	14 (93)	15 (100)	15 (100)	15 (100)	15 (100)	15 (100)
Peritoneal pseudocyst	4	20	2 (50)	4 (100)	3 (75)	4 (100)	4 (100)	4 (100)	4 (100)	4 (100)
Abscess	19	39	9 (47)	15 (79)	8 (42)	14 (74)	15 (79)	17 (89)	18 (95)	18 (95)
Fibroma	29	25	15 (52)	25 (86)	16 (55)	23 (79)	24 (83)	26 (90)	27 (93)	28 (97)
Cystadenoma	102	15	59 (58)	92 (90)	85 (83)	95 (93)	99 (97)	100 (98)	101 (99)	102 (100)
Mucinous cystadenoma	80	16	30 (38)	75 (94)	65 (81)	75 (94)	77 (96)	79 (99)	80 (100)	80 (100)
Rare benign tumor	8	14	1 (13)	8 (100)	6 (75)	7 (88)	7 (88)	8 (100)	8 (100)	8 (100)
All benign tumors	567	17	333 (59)	538 (95)	403 (71)	506 (89)	534 (94)	553 (98)	563 (99)	565 (99.6)
Premenopausal	359	19	242 (67)	351 (98)	239 (67)	314 (87)	337 (94)	351 (98)	359 (100)	359 (100)
Postmenopausal	208	15	91 (44)	187 (90)	164 (79)	192 (92)	197 (95)	202 (97)	204 (98)	206 (99)
Primary invasive	127	366	101 (80)	125 (98)	113 (89)	104 (82)	94 (74)	79 (62)	60 (47)	36 (28)
Stage I	33	76	23 (70)	32 (97)	22 (67)	17 (52)	13 (39)	9 (27)	6 (18)	3 (9)
Stage II	10	508.5	10 (100)	10 (100)	9 (90)	9 (90)	9 (90)	6 (60)	5 (50)	2 (20)
Stage III	69	431	56 (81)	68 (99)	68 (99)	64 (93)	58 (84)	50 (72)	36 (52)	22 (32)
Stage IV	15	1292	12 (80)	15 (100)	14 (93)	14 (93)	14 (93)	14 (93)	13 (87)	9 (60)
Borderline	52	25	14 (27)	32 (62)	22 (42)	17 (33)	13 (25)	7 (13)	5 (10)	2 (4)
Metastatic	38	76	10 (26)	35 (92)	30 (79)	20 (53)	18 (47)	14 (37)	9 (24)	3 (8)
Rare primary invasive	25	100	6 (24)	22 (88)	18 (72)	14 (56)	13 (52)	10 (25)	5 (20)	1 (4)
All malignant tumors	242	167	131 (54)	214 (88)	183 (76)	155 (64)	138 (57)	110 (45)	79 (33)	42 (17)
Premenopausal	86	62.5	41 (48)	70 (81)	57 (66)	43 (50)	36 (42)	25 (29)	15 (17)	9 (10)
Postmenopausal	156	244.5	90 (58)	144 (92)	126 (81)	112 (72)	102 (65)	85 (54)	64 (41)	33 (21)
All masses	809	23	464 (57)	752 (93)	586 (72)	661 (82)	672 (83)	663 (82)	642 (79)	607 (75)
Premenopausal	445	22	283 (64)	421 (95)	296 (67)	357 (80)	373 (84)	376 (84)	374 (84)	368 (83)
Postmenopausal	364	25	181 (50)	331 (91)	290 (80)	304 (84)	299 (82)	287 (79)	268 (74)	239 (66)

* US = ultrasound. The 95% confidence intervals are not shown in this table for readability; relevant 95% confidence intervals, however, are given in the text.

† Values equal to or above the cutoff were assumed to indicate malignancy.

Pattern recognition was superior to serum CA-125 (with a cutoff value of 30 U/mL) for correctly classifying both benign and malignant tumors (95% of benign tumors correctly classified versus 71%: difference = 24%, 95% CI = 19.7% to 27.9%, $P<.001$; 88% of malignant tumors correctly classified versus 76%: difference = 12%, 95% CI = 7.2% to 18.5%, $P<.001$). Pattern recognition remained superior after stratifying for menopausal status—98% of benign tumors in premenopausal women being correctly classified by pattern recognition versus 67% by serum CA-125 (difference = 31%, 95% CI = 26.1% to 36.4%, $P<.001$), 81% of malignant tumors in premenopausal women being correctly classified by pattern recognition versus 66% by CA-125 (difference = 15%, 95% CI = 3.3% to 26.5%, $P = .012$), 90% of benign tumors in postmenopausal women being correctly classified by pattern recognition versus 79% by CA-125 (difference = 11%, 95% CI = 4.5% to 17.6%, $P = .001$), and 92% of malignant tumors in postmenopausal women being correctly classified by pattern recognition versus 81% by CA-125 (difference = 11%, 95% CI = 5.6% to 18.0%, $P<.001$).

For each specific histologic diagnosis, pattern recognition was superior to serum CA-125. Pattern recognition correctly classified most benign tumors associated with high levels of serum CA-125 as benign (i.e., endometriomas, fibromas, and abscesses), and it correctly classified more borderline tumors as malignant than did serum CA-125 at any cutoff value (Table 2). For example, pattern recognition correctly classified 86% (95% CI = 81.1% to 90.4%) of these four histologies, but a CA-125 cutoff of 30 U/mL correctly classified only 41% (95% CI = 34.4% to 47.5%). Moreover, pattern recognition was superior to serum CA-125 for discriminating between benign tumors and borderline tumors (Table 3) and between benign tumors and stage I primary invasive malignancies (Table 4) in both pre- and postmenopausal women, irrespective of the CA-125 cutoff chosen to indicate malignancy.

CA-125 Levels in Premenopausal and Postmenopausal Women

To investigate the effect of menopause, we determined CA-125 serum levels in pre- and postmenopausal patients with regard to histology of adnexal masses. Among patients with benign masses, CA-125 serum levels were lower in postmenopausal women than in premenopausal women (Fig. 2; median = 15 U/mL and 19 U/mL, respectively; $\theta = 0.61$, 95% CI for $\theta = 0.56$ to 0.66, $P<.001$). Among patients with malignant masses, CA-125 serum levels were higher for postmenopausal women than for premenopausal women (244.5 U/mL and 62.5, respectively; $\theta = 0.66$, 95% CI for $\theta = 0.58$ to 0.72,

Table 3. Diagnostic performance of subjective evaluation of the ultrasound image by the ultrasound examiner (pattern recognition) and of serum CA-125 level in a population of women with benign tumors or borderline tumors*

Population	Diagnostic method	% Acc (95% CI)	% Sens (95% CI)	% Spec (95% CI)	LR+ (95% CI)	LR– (95% CI)
All	Pattern recognition	92.1 (89.6 to 94.0)	61.5 (47.0 to 74.4)	94.9 (92.6 to 96.5)	12.0 (7.95 to 18.2)	0.41 (0.29 to 0.57)
	CA-125 (30)†	68.7 (64.8 to 72.3)	42.3 (29.0 to 56.7)	71.1 (67.1 to 74.7)	1.46 (1.04 to 2.06)	0.81 (0.64 to 1.03)
	CA-125 (65)	84.5 (81.3 to 87.2)	32.7 (20.7 to 47.2)	89.2 (86.3 to 91.6)	3.04 (1.93 to 4.80)	0.75 (0.62 to 0.91)
	CA-125 (100)	88.4 (85.5 to 90.7)	25.0 (14.5 to 39.2)	94.2 (91.8 to 95.9)	4.30 (2.42 to 7.64)	0.80 (0.68 to 0.93)
	CA-125 (200)	90.5 (87.8 to 92.6)	13.5 (6.0 to 26.4)	97.5 (95.8 to 98.6)	5.45 (2.30 to 12.9)	0.89 (0.80 to 0.99)
	CA-125 (400)	91.8 (89.2 to 93.7)	9.6 (3.6 to 21.8)	99.3 (98.1 to 99.8)	13.6 (3.78 to 49.2)	0.91 (0.83 to 0.99)
	CA-125 (1000)	91.6 (89.1 to 93.6)	3.8 (0.7 to 14.3)	99.7 (98.6 to 99.9)	10.9 (1.57 to 75.8)	0.96 (0.91 to 1.02)
Premenopausal	Pattern recognition	94.8 (92.0 to 96.7)	55.6 (35.6 to 74.0)	97.8 (95.5 to 99.0)	24.9 (11.6 to 53.5)	0.45 (0.30 to 0.69)
	CA-125 (30)	64.5 (59.5 to 69.2)	37.0 (20.1 to 57.5)	66.6 (61.4 to 71.4)	1.11 (0.66 to 1.85)	0.95 (0.71 to 1.27)
	CA-125 (65)	82.9 (78.7 to 86.4)	22.2 (9.4 to 42.7)	87.5 (83.5 to 90.6)	1.77 (0.83 to 3.78)	0.89 (0.73 to 1.09)
	CA-125 (100)	88.3 (84.6 to 91.3)	14.8 (4.9 to 34.6)	93.9 (90.7 to 96.0)	2.42 (0.90 to 6.51)	0.91 (0.78 to 1.06)
	CA-125 (200)	91.5 (88.1 to 94.0)	7.4 (1.3 to 25.8)	97.8 (95.5 to 99.0)	3.32 (0.74 to 14.9)	0.95 (0.85 to 1.05)
	CA-125 (400)	93.5 (90.5 to 95.7)	7.4 (1.3 to 25.8)	100.0 (98.7 to 100)	–	0.93 (0.83 to 1.03)
	CA-125 (1000)	93.3 (90.2 to 95.5)	3.7 (0.2 to 20.9)	100.0 (98.7 to 100)	–	0.96 (0.89 to 1.04)
Postmenopausal	Pattern recognition	87.6 (82.5 to 91.4)	68.0 (46.4 to 84.3)	89.9 (84.8 to 93.5)	6.74 (4.14 to 11.0)	0.36 (0.20 to 0.63)
	CA-125 (30)	75.5 (69.4 to 80.8)	48.0 (28.3 to 68.2)	78.8 (72.5 to 84.1)	2.27 (1.40 to 3.69)	0.66 (0.45 to 0.96)
	CA-125 (65)	87.1 (82.0 to 91.0)	44.0 (25.0 to 64.7)	92.3 (87.6 to 95.4)	5.72 (3.00 to 10.9)	0.61 (0.43 to 0.86)
	CA-125 (100)	88.4 (83.4 to 92.1)	36.0 (18.7 to 57.4)	94.7 (90.5 to 97.2)	6.81 (3.13 to 14.8)	0.68 (0.50 to 0.91)
	CA-125 (200)	88.8 (83.9 to 92.4)	20.0 (7.6 to 41.3)	97.1 (93.5 to 98.8)	6.93 (2.28 to 21.1)	0.82 (0.68 to 1.00)
	CA-125 (400)	88.8 (83.9 to 92.4)	12.0 (3.2 to 32.3)	98.1 (94.8 to 99.4)	6.24 (1.48 to 26.3)	0.90 (0.78 to 1.04)
	CA-125 (1000)	88.8 (83.9 to 92.4)	4.0 (0.2 to 22.3)	99.0 (96.2 to 99.8)	4.16 (0.39 to 44.2)	0.97 (0.89 to 1.05)

* The numbers of benign and borderline tumors in this table come from Table 1. A total of 619 women were in all populations (386 of whom were premenopausal and 233 of whom were postmenopausal) Acc = accuracy; Sens = sensitivity for detection of malignant tumor; Spec = specificity; LR+ = positive likelihood ratio; LR– = negative likelihood ratio; CI = confidence interval; – = infinity (division by zero).

† Values equal to or higher than the cutoff value in parentheses (expressed as units per milliliter) were assumed to indicate malignancy.

$P<.001$). Accordingly, the θ value for serum CA-125 when comparing benign and malignant masses was 0.73 (95% CI for θ = 0.67 to 0.78) in premenopausal patients and 0.89 (95% CI for θ = 0.85 to 0.92) in postmenopausal patients. Values overlapped substantially between pre- and postmenopausal women with the same histopathologic diagnosis, but for many benign diagnoses, the values tended to be lower in postmenopausal women than in premenopausal women, whereas they tended to be higher for abscesses and malignant diagnoses in postmenopausal women than in premenopausal women (Fig. 3). We observed that, among premenopausal patients, 52% (95% CI = 31.1% to 72.6%) of borderline tumors were mucinous and, among postmenopausal patients, 40% (95% CI = 21.8% to 61.1%) of borderline tumors were mucinous. Consequently, differences in CA-125 between pre- and postmenopausal patients may be explained by differences in the mixture of tumor types and possibly by differences in expression of CA-125 between these groups of patients (Fig. 3).

Discussion

We found that pattern recognition by an experienced examiner was superior to assessing the level of serum CA-125 for discrimination between benign and malignant adnexal masses. We also elucidated the distribution of CA-125 serum levels among patients with various types of adnexal masses and investigated the potential utility of serum CA-125 in the preoperative characterization of these tumors. Other publications have described increased levels of serum CA-125 in patients with malignant tumors (9,13,35,36) and described an association between CA-125 serum levels and the stage and histology of tumors. We have also confirmed that women with ovarian carcinoma, primarily late stage disease, have increased levels of serum CA-125 (9,13,36,37) and that women with borderline and stage I tumors, which are more difficult to detect clinically and to diagnose correctly as malignant than advanced stage cancers, have a relatively low level of serum CA-125 (38–40). Patients with mucinous cancers have elevated CA-125 serum levels less often than patients with serous and undifferentiated epithelial cancers (37). Patients with other primary cancers, such as those of the endometrium, pancreas, liver, or lung, may also have elevated levels of serum CA-125 (9,13).

Few published studies have focused on CA-125 serum levels in women with benign adnexal tumors. However, it has been reported that the level of serum CA-125 is often elevated in women with endometriosis or endometriomas (6,9,13,38,41–43). For example, Jacobs and Bast (13) found that CA-125 serum levels exceeded 35 U/mL in approximately 10% of women with benign tumors and in a higher percentage of women with serous benign tumors than among those with cystic teratomas. Our results agree with those of others in that CA-125 serum levels were elevated more often among premenopausal women with benign tumors than among postmenopausal women with benign tumors (9,41) and thus were more useful in distinguishing between benign and malignant tumors in postmenopausal patients (36,44,45). These different results between pre- and postmenopausal patients could be explained by the different mixtures of tumor types in pre- and postmenopausal patients—34% of all benign masses among premenopausal women were endometriomas (which are associated with increased levels of serum CA-125), compared with only 2% among postmenopausal women, and 50% of all malignant masses among postmenopausal women were primary invasive cancers of stage II–IV (which are associated with very high levels of serum

Table 4. Diagnostic performance of subjective evaluation of the ultrasound image by the ultrasound examiner (pattern recognition) and of serum CA-125 in a population of women with benign or stage I primary invasive tumors*

Population	Diagnostic method	% Acc (95% CI)	% Sens (95% CI)	% Spec (95% CI)	LR+ (95% CI)	LR– (95% CI)
All	Pattern recognition	95.0 (92.9 to 96.5)	97.0 (82.5 to 99.8)	94.9 (92.6 to 96.5)	19.0 (13.2 to 27.2)	0.03 (0.005 to 0.22)
	CA-125 (30)†	70.8 (67.0 to 74.4)	66.7 (48.1 to 81.4)	71.1 (67.1 to 74.7)	2.30 (1.75 to 3.03)	0.47 (0.29 to 0.76)
	CA-125 (65)	87.2 (84.2 to 89.7)	51.5 (33.9 to 68.8)	89.2 (86.3 to 91.6)	4.79 (3.19 to 7.19)	0.54 (0.38 to 0.77)
	CA-125 (100)	91.2 (88.5 to 93.3)	39.4 (23.4 to 57.8)	94.2 (91.8 to 95.9)	6.77 (3.95 to 11.6)	0.64 (0.49 to 0.85)
	CA-125 (200)	93.7 (91.3 to 95.4)	27.3 (13.9 to 45.8)	97.5 (95.8 to 98.6)	11.0 (5.16 to 23.6)	0.75 (0.61 to 0.92)
	CA-125 (400)	94.8 (92.7 to 96.4)	18.2 (7.6 to 36.1)	99.3 (98.1 to 99.8)	25.8 (7.64 to 86.9)	0.82 (0.70 to 0.97)
	CA-125 (1000)	94.7 (92.5 to 96.3)	9.1 (2.4 to 25.5)	99.7 (98.6 to 99.9)	25.8 (4.46 to 149)	0.91 (0.82 to 1.02)
Premenopausal	Pattern recognition	97.8 (95.6 to 99.0)	100.0 (65.5 to 100)	97.8 (95.5 to 99.0)	44.9 (22.6 to 89.0)	0.00 (0.00 to –)
	CA-125 (30)	66.1 (61.0 to 70.9)	50.0 (20.1 to 79.9)	66.6 (61.4 to 71.4)	1.50 (0.79 to 2.83)	0.75 (0.40 to 1.40)
	CA-125 (65)	86.2 (82.1 to 89.4)	40.0 (13.7 to 72.6)	87.5 (83.5 to 90.6)	3.19 (1.42 to 7.15)	0.69 (0.41 to 1.14)
	CA-125 (100)	92.4 (89.1 to 94.8)	40.0 (13.7 to 72.6)	93.9 (90.7 to 96.0)	6.53 (2.76 to 15.4)	0.64 (0.39 to 1.06)
	CA-125 (200)	95.4 (92.6 to 97.2)	10.0 (0.5 to 45.9)	97.8 (95.5 to 99.0)	4.49 (0.62 to 32.6)	0.92 (0.75 to 1.13)
	CA-125 (400)	97.3 (94.9 to 98.6)	0.0 (0.0 to 34.5)	100.0 (98.7 to 100)	–	1.00 (1.00 to 1.00)
	CA-125 (1000)	97.3 (94.9 to 98.6)	0.0 (0.0 to 34.5)	100.0 (98.7 to 100)	–	1.00 (1.00 to 1.00)
Postmenopausal	Pattern recognition	90.5 (85.8 to 93.8)	95.7 (76.0 to 99.8)	89.9 (84.8 to 93.5)	9.47 (6.26 to 14.3)	0.05 (0.007 to 0.33)
	CA-125 (30)	78.4 (72.4 to 83.4)	73.9 (51.3 to 88.9)	78.8 (72.5 to 84.1)	3.49 (2.44 to 5.00)	0.33 (0.17 to 0.66)
	CA-125 (65)	88.7 (83.8 to 92.4)	56.5 (34.9 to 76.1)	92.3 (87.6 to 95.4)	7.35 (4.07 to 13.3)	0.47 (0.30 to 0.75)
	CA-125 (100)	89.2 (84.3 to 92.7)	39.1 (20.5 to 61.2)	94.7 (90.5 to 97.2)	7.40 (3.43 to 16.0)	0.64 (0.46 to 0.89)
	CA-125 (200)	90.9 (86.3 to 94.2)	34.8 (17.2 to 57.2)	97.1 (93.5 to 98.8)	12.1 (4.59 to 31.7)	0.67 (0.50 to 0.91)
	CA-125 (400)	90.9 (86.3 to 94.2)	26.1 (11.1 to 48.7)	98.1 (94.8 to 99.4)	13.6 (4.13 to 44.6)	0.75 (0.59 to 0.96)
	CA-125 (1000)	90.5 (85.8 to 93.8)	13.0 (3.4 to 34.7)	99.0 (96.2 to 99.8)	13.6 (2.39 to 77.0)	0.88 (0.75 to 1.03)

* The numbers of benign and stage I primary invasive tumors in this table come from Table 1. A total of 600 women were in all populations (369 of whom were premenopausal and 231 of whom were postmenopausal). Acc = accuracy; Sens = sensitivity for detection of malignant tumor; Spec = specificity; LR+ = positive likelihood ratio; LR– = negative likelihood ratio; CI = confidence interval; – = infinity (division by zero).

† Values equal to or higher than the cutoff (the cutoff is given in brackets and is expressed in units per milliliter) were assumed to indicate malignancy.

CA-125), compared with only 31% of malignant masses among premenopausal women. Surprisingly, however, there were also some differences in CA-125 serum levels between pre- and postmenopausal women with the same histologic diagnosis. Compared with premenopausal women with similar tumor types, postmenopausal women with various benign tumor types tended to have lower CA-125 serum levels, whereas postmenopausal women with malignant tumors had higher CA-125 serum levels, irrespective of tumor type and stage. These observations might be explained by differential tumor behaviors in pre- and postmenopausal women with the same tumor types and differences in type of borderline tumors in these groups of women (i.e., borderline tumors tended to be mucinous more often in premenopausal women than in postmenopausal women), in tumor size, or in the spread of malignant tumors of the same stage.

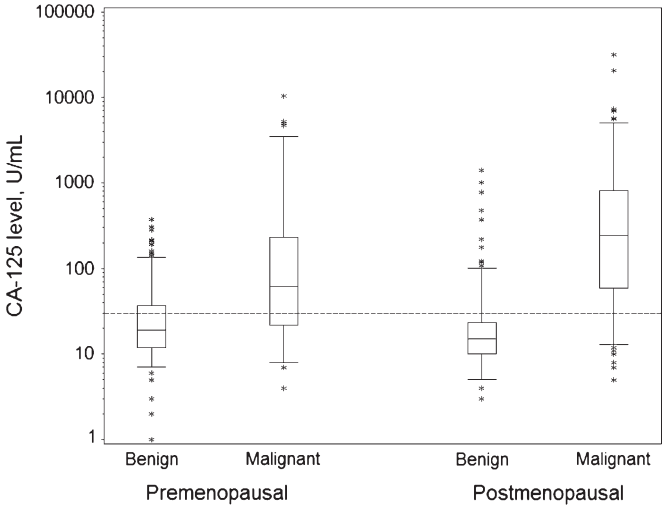

Fig. 2. Box plots showing CA-125 serum levels by menopausal status and tumor outcome. **Box** = the range of the middle 50% of the CA-125 levels; **line inside the box** = median; **whiskers** = the 5th and 95th percentile; **asterisks** = data points that lie outside the whiskers; dotted horizontal line = a CA-125 level of 30 U/mL.

We have shown that 752 (93%) of the 809 tumors were correctly classified as malignant or benign by pattern recognition, and in 464 (57%) of these tumors, a correct specific histologic diagnosis was proposed. At any serum CA-125 cutoff level used, serum CA-125 resulted in more misclassifications with regard to malignancy than pattern recognition. Our findings are in line with those in our previous report from the IOTA study (46), in which serum CA-125 was not retained as a variable in a logistic regression model that was designed to discriminate between benign and malignant adnexal masses. Other studies (23–26) have also shown that experienced ultrasound examiners can generally make a correct diagnosis with regard to the nature of a mass by the subjective evaluation of gray-scale and Doppler ultrasound findings, i.e., pattern recognition. These findings are clinically relevant for patients with an ovarian mass because clinicians often use the CA-125 serum level in such patients to estimate the likelihood that the mass is malignant. As a result of this analysis, we suggest that this approach may not be as helpful as pattern recognition.

This study has several limitations. One is that CA-125 serum levels were not available for 24% of case patients in the original IOTA study population. An analysis of the masses without information on serum CA-125 (Table 1) showed that most masses were benign and were classified correctly as benign with a high level of

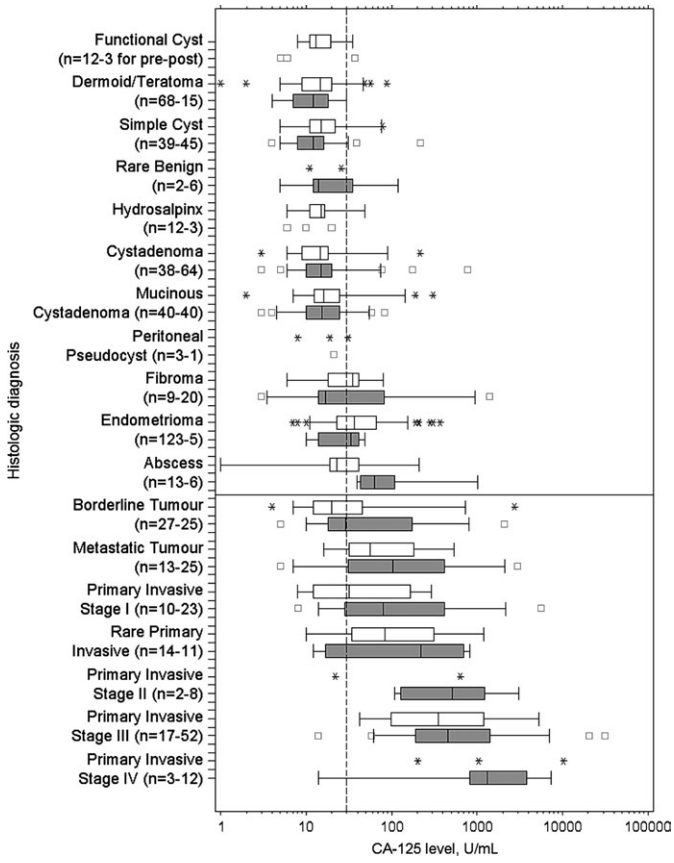

Fig. 3. Box plots showing CA-125 serum levels by histologic diagnosis and menopausal status of the patients with adnexal tumors. **Box** = the range of the middle 50% of the CA-125 levels; **line inside the box** = median; **whiskers** = the 5th and 95th percentile; **asterisks** or **squares** = data points that lie outside the whiskers for premenopausal women or postmenopausal women, respectively; **open bars** = premenopausal women; **solid bars** = postmenopausal women. The values for n are the number of premenopausal patients (first value) and postmenopausal patients (second value) with that tumor. All data points are presented individually if there are fewer than four individuals in a group.

confidence by use of pattern recognition. This finding reveals a bias to the study whereby serum CA-125 is more likely to have been measured in women with lesions that were suspected of being malignant by the ultrasound examination than in women with lesions that were suspected of being benign. We do not believe that this bias invalidates our conclusions, because in all likelihood, serum CA-125 would have performed more poorly in the patients excluded in this study than in the included patients, given the large proportion of endometriomas and the small proportion of stage III and IV cancers in the excluded group. Another limitation is that five CA-125 kits were used to assess the level of serum CA-125. The use of multiple kits may have introduced more variation in the CA-125 results than if only a single kit had been used. However, this procedure reflects clinical reality, and previous studies have found that variation in the results from the different kits is not large (47,48).

In light of our results, it is difficult to understand why measurement of CA-125 serum levels is such an entrenched part of the preoperative evaluation of adnexal masses. It probably reflects a lack of confidence in the use of ultrasonography despite numerous reports in the literature that transvaginal ultrasound can be used to classify most ovarian masses with a high degree of accuracy (23–25). The accuracy, however, depends on the level of experience of the examiner; little variation in accuracy has been found between experienced examiners (23). Our results were obtained by specialized gynecologic ultrasound units. The quality of ultrasound outside such centers is almost certainly more variable. We hope our results will encourage responsible parties to expend more effort and more resources to educate and train those who perform gynecologic ultrasound examinations, so that the potential of the ultrasound technique can be maximized.

References

(1) Jacobs IJ, Skates S, Davies AP, Woolas RP, Jeyerajah A, Weidemann P, et al. Risk of diagnosis of ovarian cancer after raised serum CA 125 concentration: a prospective cohort study. BMJ 1996;313:1355–8.

(2) Royal College of Obstetricians and Gynaecologists. Guideline no. 34. Ovarian cysts in postmenopausal women. London: RCOG press; 2003.

(3) Paramasivam S, Tripcony L, Crandon A, Quinn M, Hammond I, Marsden D, et al. Prognostic importance of preoperative CA 125 in International Federation of Gynecology and Obstetrics stage I epithelial ovarian cancer: an Australian multicenter study. J Clin Oncol 2005;23:5938–42.

(4) Bast RC Jr, Klug TL, St John E, Jenison E, Niloff JM, Lazarus H, et al. A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. N Engl J Med 1983;309:883–7.

(5) Bon GG, Kenemans P, Verstraeten R, van Kamp GJ, Hilgers J. Serum tumor marker immunoassays in gynecologic oncology: establishment of reference values. Am J Obstet Gynecol 1996;174:107–14.

(6) Gadducci A, Baicchi U, Marai R, Ferdeghini M, Bianchi R, Facchini V. Preoperative evaluation of D-dimer and CA 125 levels in differentiating benign from malignant ovarian masses. Gynecol Oncol 1996;60: 197–202.

(7) Predanic M, Vlahos N, Pennisi JA, Moukhtar M, Aleem FA. Color and pulsed Doppler sonography, gray-scale imaging, and serum CA 125 in the assessment of adnexal disease. Obstet Gynecol 1996;88:283–8.

(8) Eltabbakh GH, Belinson JL, Kennedy AW, Gupta M, Webster K, Blumenson LE. Serum CA-125 measurements > 65 U/ml. clinical value. J Reprod Med 1998;43:635–6.

(9) Tuxen MK. Tumor marker CA125 in ovarian cancer. J Tumor Marker Oncol 2001;16:49–68.

(10) Einhorn N, Bast RC Jr, Knapp RC, Tjernberg B, Zurawski VR Jr. Preoperative evaluation of serum CA125 levels in patients with primary epithelial ovarian cancer. Obstet Gynecol 1986;67:414–6.

(11) Tamakoshi K, Kikkawa F, Shibata K, Tomoda K, Obata NH, Wakahara F. Clinical value of CA125, CA19-9, CEA, CA72-4, and TPA in borderline ovarian tumor. Gynecol Oncol 1996;62:67–72.

(12) Engelen MJ, de Bruijn HW, Hollema H, ten Hoor KA, Willemse PH, Aalders JG, et al. Serum CA 125, carcinoembryonic antigen, and CA 19-9 as tumor markers in borderline ovarian tumors. Gynecol Oncol 2000;78:16–20.

(13) Jacobs I, Bast RC Jr. The CA 125 tumour-associated antigen: a review of the literature. Hum Reprod 1989;4:1–12.

(14) Vasilev SA, Schlaerth JB, Campeau J, Morrow CP. Serum CA 125 levels in preoperative evaluation of pelvic masses. Obstet Gynecol 1988;71:751–6.

(15) Buamah PK, Skillen AW. Serum CA 125 concentrations in patients with benign ovarian tumours. J Surg Oncol 1994;56:71–4.

(16) 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.

(17) Bourne T, Campbell S, Steer C, Whitehead MI, Collins WP. Transvaginal colour flow imaging: a possible new screening technique for ovarian cancer. Br Med J 1989;299:1367–70.

(18) 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.

(19) Lerner JP, Timor-Tritsch IE, Federman A, Abramovich G. Transvaginal ultrasonographic characterization of ovarian masses with an improved, weighted scoring system. Am J Obstet Gynecol 1994;170:81–5.

(20) 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.
(21) Valentin L. Use of morphology to characterize and manage common adnexal masses. Best Pract Res Clin Obstet Gynaecol 2004;18: 105–23.
(22) de Kroon CD, van der Sandt HA, van Houwelingen JC, Jansen FW. Sonographic assessment of non-malignant ovarian cysts: does sonohistology exist? Hum Reprod 2004;19:2138–43.
(23) Timmerman D, Schwärzler P, Collins WP, Claerhout F, Coenen M, Amant F. Subjective assessment of adnexal masses using ultrasonography: an analysis of interobserver variability and experience. Ultrasound Obstet Gynecol 1999;13:11–6.
(24) Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. Ultrasound Obstet Gynecol 1999;14:338–47.
(25) Jermy K, Luise C, Bourne T. The characterization of common ovarian cysts in premenopausal women. Ultrasound Obstet Gynecol 2001;17: 140–4.
(26) 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.
(27) 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.
(28) 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;6:500–5.
(29) Kenemans P, van Kamp GJ, Oehr P, Verstraeten RA. Heterologous double-determinant immunoradiometric assay CA 125 II: reliable second-generation immunoassay for determining CA 125 in serum. Clin Chem 1993;39:2509–13.
(30) Heintz APM, Odicino F, Maisonneuve P, Beller U, Benedet JL, Creasman WT, et al. Carcinoma of the ovary, in 25th Annual Report on the Results of Treatment in Gynecological Cancer. Int J Gynaecol Obstet 2003;83: S135–7.
(31) Newcombe RG. Two-sided confidence intervals for the single proportion comparison of seven methods. Stat Med 1998;17:857–72.
(32) Newcombe RG. Improved confidence intervals for the difference between binomial proportions based on paired data. Stat Med 1998;17:2635–50.
(33) Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med 1998;17: 873–90.
(34) Newcombe RG. Confidence intervals for an effect size measure based on the Mann-Whitney statistic: Part 2: asymptotic methods and evaluation. Stat Med 2006;25:559–73.
(35) Duffy MJ, Bonfrer JM, Kulpa J, Rustin GJ, Soletormos G, Torre GC, et al. CA125 in ovarian cancer: European Group on tumor markers guidelines for clinical use. Int J Gynecol Cancer 2005;15:679–91.
(36) Tuxen MK, Soletormos G, Dombernowsky P. Tumor markers in the management of patients with ovarian cancer. Cancer Treat Rev 1995;21: 215–45.
(37) Cooper BC, Sood AK, Davis CS, Ritchie JM, Sorosky JI, Anderson B, et al. Preoperative CA125 levels: an independent prognostic factor for epithelial ovarian cancer. Obstet Gynecol 2002;100:59–64.
(38) Gotlieb WH, Soriano D, Achiron R, Zalel Y, Davidson B, Kopolovic J, et al. CA 125 measurement and ultrasonography in borderline tumors of the ovary. Am J Obstet Gynecol 2000;83:541–6.
(39) Makar AP, Kaern J, Kristensen GB, Vergote I, Bormer OP, Trope CG. Evaluation of serum CA 125 level as a tumor marker in borderline tumors of the ovary. Int J Gynecol Cancer 1993;3:299–303.
(40) Rice LW, Lage JM, Berkowitz RS, Goodman A, Muto MG, Knapp RC, et al. Preoperative serum CA-125 levels in borderline tumors of the ovary. Gynecol Oncol 1992;46:226–9.
(41) Somigliana E, Vigano P, Tirelli AS, Felicetta I, Torresani E, Vignali M, et al. Use of the concomitant serum dosage of CA 125, CA 19-9 and interleukin-6 to detect the presence of endometriosis. Results from a series of reproductive age women undergoing laparoscopic surgery for benign gynaecological conditions. Hum Reprod 2004;19:1871–6.
(42) Kitawaki J, Ishihara H, Koshiba H, Kiyomizu M, Teramoto M, Kitaoka Y, et al. Usefulness and limits of CA-125 in diagnosis of endometriosis without associated ovarian endometriomas. Hum Reprod 2005;20: 1999–2003.
(43) Phupong V, Chen O, Ultchaswadi P. High level of CA 125 due to large endometrioma. J Med Assoc Thai 2004;87:1108–11.
(44) Cane PA, Colleen-Lopez E, Platt LD, Karlan BY. Tumor marker trends in asymptomatic women at risk for ovarian cancer: relevance for ovarian cancer screening. Gynecol Oncol 1995;57:240–5.
(45) Malkasian GD Jr, Knapp RC, Lavin PT, Zurawski VR Jr, Podratz KC, Stanhope CR, et al. Preoperative evaluation of serum CA 125 levels in premenopausal and postmenopausal patients with pelvic masses: discrimination of benign from malignant disease. Am J Obstet Gynecol 1988; 159:341–6.
(46) Timmerman D, Van Calster B, Jurkovic D, Valentin L, Testa AC, Bernard JP, et al. The inclusion of CA-125 does not improve mathematical models developed to distinguish between benign and malignant adnexal tumors. J Clin Oncol 2007;25:4194–200.
(47) Bonfrer J, Baan A, Jansen E, Lentfer D, Kenemans P. Technical evaluation of three second generation CA 125 assays. Eur J Clin Chem Clin Biochem 1994;32:201–7.
(48) Davelaar E, van Kamp G, Verstraeten R, Kenemans P. Comparison of seven immunoassays for the quantification of CA 125 antigen in serum. Clin Chem 1198;44:1417–22.

Funding

The research council of the Katholieke Universiteit Leuven, Belgium (GOA-AMBioRICS, CoE EF/05/006 Optimization in Engineering OPTEC); the Belgian Federal Science Policy Office (IUAP P6/04 (DYSCO); the EU: BIOPATTERN (FP6-2002-IST 508803); ETUMOUR (FP6-2002-LIFESCIHEALTH 503094); Healthagents (IST–2004–27214); the Swedish Medical Research Council (K2001-72X-11605-06A, K2002-72X-11605-07B, K2004-73X-11605-09A, K2006-73X-11605-11-3); funds administered by Malmö University Hospital; Allmänna Sjukhusets i Malmö Stiftelse för bekämpande av cancer (the Malmö General Hospital Foundation for fighting against cancer); ALF-medel (a Swedish governmental grant from the region of Scania).

Notes

The funding agencies had no role in the design of the study, the collection of the data, the analysis and interpretation of the data, the writing of the manuscript or the decision to submit the manuscript for publication. The authors had full responsibility for all these activities.

Manuscript received May 22, 2007; revised September 4, 2007; accepted September 20, 2007.