EJGO European Journal of Gynaecological Oncology

Comparison of nine morphological scoring systems to detect ovarian malignancy

Erhan Aktürk, Murat Dede, Müfit C. Yenen, Y. Kemal Koçyiğit, Ali Ergün

Department of Obstetrics and Gynecology, Gülhane Military Medical Faculty, Ankara (Turkey)

Summary

Purpose of investigation: The aim of this study was to prospectively compare the diagnostic performances of nine gray-scale sonographic prediction models to detect ovarian malignancy. *Materials and Methods:* Clinical data of 322 women presenting with an adnexal mass were obtained and used in nine scoring systems. For each model a ROC curve demonstrating the capacity of the model to diagnose malignancy was constructed for all cases and for the subgroups of premenopause and postmenopause. The performance of each model was expressed as area under the ROC curve, sensitivity, and specificity. *Results:* The area under the ROC curve, sensitivity, and specificity of these models in the present study varied between 0.737 and 0.929, 70.7% and 87.9%, 60.2% and 80.3%, respectively. *Conclusions:* This study has revealed the usefulness of morphological scoring systems to correctly discriminate between benign and malignant pelvic masses.

Key words: Ovarian; Ultrasonography; Mass; Model.

Introduction

Adequate preoperative assessment of an adnexal mass is needed since the accurate diagnosis directs management. The preoperative definitive diagnosis of a malignant mass cannot always be made with current diagnostic modalities. However prior to surgery, it is important to differentiate between malignant and benign pathology of the adnexal mass in order not to face a malignancy unexpectedly and to avoid an inappropriate surgery.

Today, several parameters are available to distinguish between benign and malignant masses. For this purpose, gray-scale sonographic parameters are frequently used to evaluate the risk of malignancy. Other parameters that are used in discrimination between benign and malignant masses are Doppler ultrasonography, patient characteristics (menopausal state or age), and biochemical markers (CA 125, human epididymis protein 4 and prealbumin). Some authors [1-9] used only gray-scale sonographic markers as a preoperative diagnostic tool to assess the risk of malignancy while the others [10-12] combined these parameters for the same purpose. In most cases subjective assessment with ultrasonographic views by experienced sonologists is a main method for detecting the malignant mass [13]. However less experienced sonologists should still use morphological scoring systems that can be helpful to classify adnexal masses as benign or malignant. The present authors hence decided to reveal which gray-scale scoring model or models should be used by less experienced sonographers.

Overall, the aim of this study was to prospectively compare the diagnostic performances of nine gray-scale sonographic prediction models to detect ovarian malignancy.

Materials and Methods

A Medline search was performed to detect diagnostic models that are principally based on sonographic characteristics and are used to distinguish benign masses from malignant ones. “Ovarian”, “ultrasonography”, “mass”, and “model” were the key words. The authors also checked the references of the detected articles. Overall, they detected nine scoring systems six of which were morphological indices [3-5, 7-9] and three of which were morphological classifications [1, 2, 6]. All parameters used in these models were only gray-scale ultrasonographic signs.

The clinical data of 322 women with pelvic masses appointed for laparotomy or laparoscopy between October 1, 2008, and October 7, 2012, to the present hospital were obtained prospectively and used in nine scoring systems. Preoperative examination included vaginal examination and transvaginal sonography. The ultrasound was performed transvaginally by a 7.5-MHz transducer. A transabdominal repeat examination with full bladder was obtained if a mass was found to be too large to be observed completely transvaginally. The histopathological diagnosis was considered the gold standard for definitive outcome. Malignancy was defined according to criteria of histologic typing of ovarian tumors, according to WHO criteria [14], and staged according to criteria recommended by the International Federation of Gynecology and Obstetrics (FIGO). The authors included low malignant potential tumors as malignant. Cut-off levels we performed were 9, 4, 9, 5, 7, 3, 2, 8, and 5 for the scoring systems of Sassone *et al.*, Lerner *et al.*, Ferrazzi *et al.*, DePriest *et al.*, Finkler *et al.*, Maggino *et al.*, Granberg *et al.*, Szpurek *et al.*, and Ueland *et al.*,

Revised manuscript accepted for publication December 30. 2013

Eur. J. Gynaecol. Oncol. - ISSN: 0392-2936
XXXVI, n. 3, 2015
doi: 10.12892/ejgo2579.2015

7847050 Canada Inc.
www.irog.net

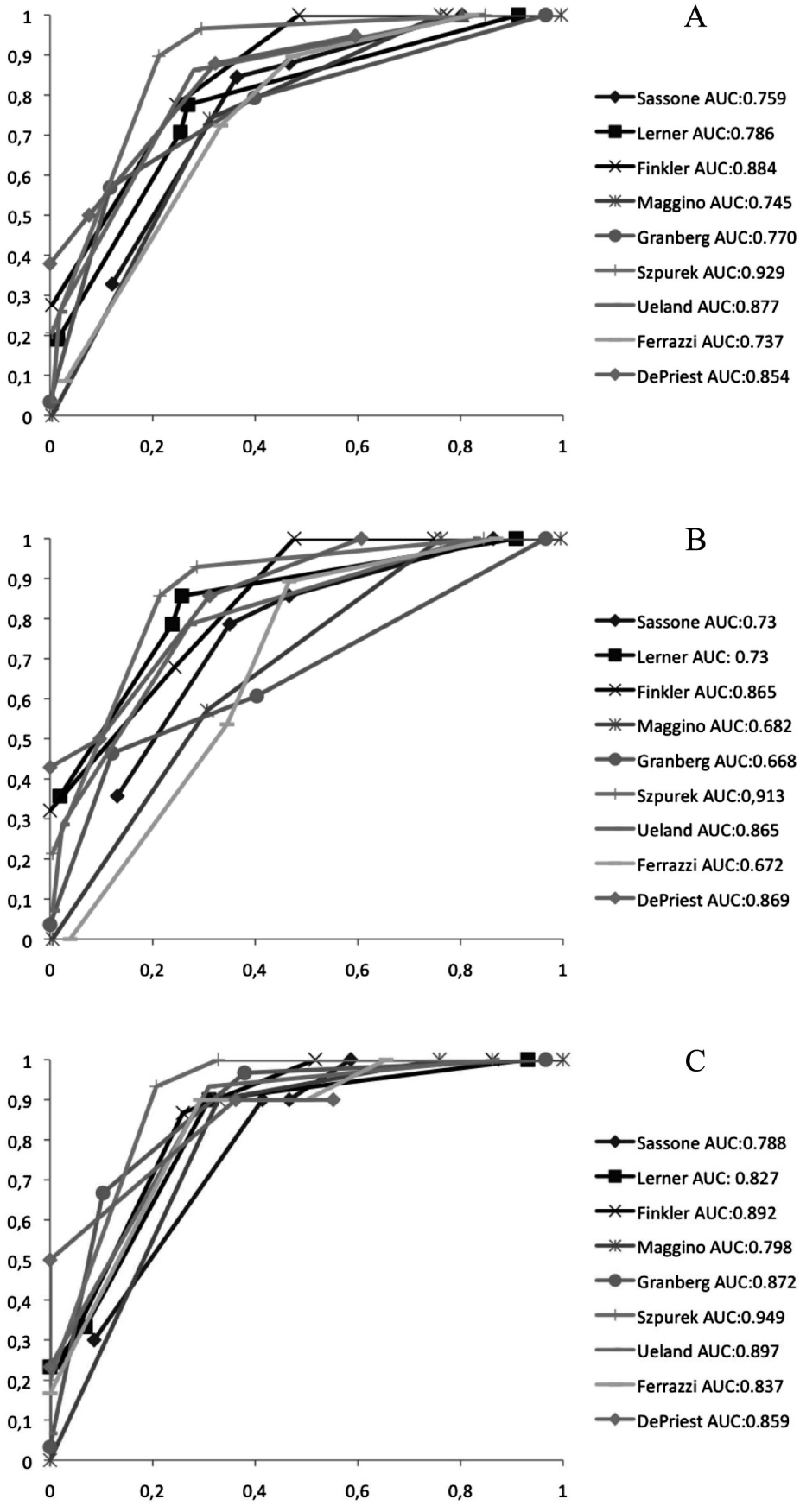

Figure 1. — A: Comparison of ROC curves for all patients B: Comparison of ROC curves for premenopausal subgroup. C: Comparison of ROC curves for postmenopausal subgroup.

Table 1. — *Accuracies reported in the original and present studies.*

Author	Original studies			Present study		
	Sensitivity (%)	Specificity (%)	Area ROC curve (%)	Sensitivity (%)	Specificity (%)	Area ROC curve (%)
Sassone [3]	100	83	-	84.5	63.6	75.9
Lerner [5]	97	77	-	70.7	74.6	78.6
Ferrazzi [7]	87	67	84	72.4	66.7	73.7
DePriest [4]	89	70	-	87.9	67.8	85.4
Finkler [1]	62	95	-	75.9	75.4	88.4
Maggino [6]	85	78	-	75.9	68.9	74.5
Granberg [2]	82	92	-	77.6	60	77
Szpurek [9]	86.7	77	89	86.2	80.3	92.9
Ueland [8]	98.1	80.7	-	84.5	72	87.7

Table 2. — *The sensitivity, specificity, and positive (PPV) and negative (NPV) predictive values, and diagnostic accuracy (DA) of nine scoring systems in all cases.*

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	DA (%)
Sassone [3]	84.5	63.6	33.8	94.9	67
Lerner [5]	70.7	74.6	38	92.1	74
Ferrazzi [7]	72.4	66.7	32.3	91.7	67.7
DePriest [4]	87.9	67.8	37.5	96.2	71
Finkler [1]	75.9	75.4	40.4	93.4	75
Maggino [6]	75.9	68.9	34.9	92.9	70
Granberg [2]	77.6	60.2	30	92.4	63
Szpurek [9]	86.2	80.3	49	96.4	81
Ueland [8]	84.5	72	39.8	95.5	74

respectively [1-9]. All statistical analyses were performed using the Statistical Package for Social Sciences (SPSS), Version 15.0. For each model a ROC curve demonstrating the capacity of the model to diagnose malignancy was constructed for all cases and for the subgroups of premenopause and postmenopause. The performance of each model was expressed as an area under the ROC curve, sensitivity, and specificity. The McNemar's test was used for testing differences in performances between nine scoring systems. When each score was compared with the other eight in the series, the authors obtained thirty-six comparisons of paired samples. [Score 1 was compared with scores 2 ,3, 4, 5, 6, 7, 8, 9 (eight comparisons); score 2 was compared with scores 3, 4, 5, 6, 7, 8, 9 (seven comparisons); score 3 was compared with scores 4, 5, 6, 7,8 , 9 (six comparisons); score 4 was compared with scores 5, 6, 7, 8, 9 (five comparisons); score 5 was compared with scores 6, 7, 8, 9 (four comparisons); score 6 was compared with scores 7, 8, 9 (three comparisons); score 7 was compared with scores 8, 9 (two comparisons), and score 8 was compared with scores 9 (one comparison)]. The sensitivity was defined as the percentage of patients with malignant disease having a positive test result. The specificity was defined as the percentage with benign disease having a negative test result. The positive predictive value was defined as the percentage of patients with a positive test result having malignant disease and the negative predictive value was defined as the percentage of patients with a negative test result having benign disease. Finally, the diagnostic accuracy was expressed as a proportion of correctly classified subjects (true positive + true negative) among all subjects. In the present study the authors assumed a concluding error risk of 5% and (correlated with the risk) significance level of $p < 0.05$ indicating statistically significant differences.

Results

Between 2008 and 2012, 322 patients were operated for an adnexal mass. The median age of these patients was 41.64 years, with a range from 16 to 79 years (SD: 13.262). Eighty-eight (27.3%) of the 322 patients were postmenopausal. Overall, 58 patients (18%) turned out to have a malignancy. Among the malignant diagnosis, there were 26 serous carcinomas, 16 mucinous carcinomas, one ovarian lymphoma, three Krukenberg tumors, one dysgerminoma, one granulosa cell tumor, three borderline serous tumors, and seven borderline mucinous tumors. Thirty (51.7%) of the patients had FIGO Stage I disease, ten (17.2%) had FIGO Stage II disease, 11 (19%) had FIGO Stage III disease, and seven (12.1%) had FIGO Stage IV disease; 264 cases were found to be benign. Among the benign diagnosis there were 18 simple cysts, 91 endometriosis, 51 dermoid cysts, 28 serous cystadenomas, 16 mucinous cystadenomas, nine fibromas, five thecomas, 12 corpus luteum cysts, 17 paratubal cysts, seven leiomyomas, four struma ovarii tumors, and six tuboovarian abscesses. The area under the ROC curves of these models in the present study varied between 0.737 and 0.929 (Figure 1A). When the sensitivity was fixed at 0.90, specificity varied between 0.42 and 0.79. The highest area under ROC for all cases was 0.929 with Szpurek index. For premenopausal subgroup, the highest area under ROC was 0.913 for Szpurek index (Figure 1B). For postmenopausal group the highest area under ROC were 0.949 for Szpurek index (Figure 1C). Corresponding ROC curves and other values of area under ROC curves are highlighted in Figure 1A, B, and C. Accuracies reported in the present and the original studies are shown in Table 1.

In all patients, the highest sensitivity was 87.9% for DePriest index with a relatively low specificity value of 67.8% and positive and negative predictive values of respectively 37.5% and 96.2%. Diagnostic accuracy of DePriest index was 71%. Sensitivities, specificities, positive and negative predictive values, and diagnostic accuracies of all other scoring systems in all cases are presented in Table 2.

In the premenopausal subgroup, sensitivity, specificity, positive and negative predictive values, and diagnostic accuracy for Szpurek index were 82.1%, 80.6%, 36.5%, 97.1%, and 80%, respectively. Sensitivities, specificities, positive and negative predictive values, and diagnostic accuracies of all other scoring systems in the premenopausal subgroup are presented in Table 3.

In the postmenopausal subgroup, sensitivity, specificity, positive and negative predictive values, and diagnostic ac-

Table 3. — *The sensitivity, specificity, and positive (PPV) and negative (NPV) predictive values, and diagnostic accuracy (DA) of nine scoring systems in the premenopausal subgroup.*

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	DA (%)
Sassone [3]	78.6	65	23.4	95.7	66
Lerner [5]	50	76.2	22.2	91.8	73
Ferrazzi [7]	53.6	65.5	17.4	91.2	64
DePriest [4]	85.7	68.9	27.3	97.3	71
Finkler [1]	64.3	75.7	26.5	94	74
Maggino [6]	57.1	68.9	20	92.2	67
Granberg [2]	57.1	59.7	16.2	91.1	59
Szpurek [9]	82.1	80.6	36.5	97.1	80
Ueland [8]	75	72.3	26.9	95.5	72

Table 4. — *The sensitivity, specificity, and positive (PPV) and negative (NPV) predictive values, and diagnostic accuracy (DA) of nine scoring systems in the postmenopausal subgroup.*

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	DA (%)
Sassone [3]	90	58.6	52.9	91.9	69.3
Lerner [5]	90	69	60	93	76.1
Ferrazzi [7]	90	70.7	61.4	93.2	77.2
DePriest [4]	90	63.8	56.3	92.5	72.7
Finkler [1]	86.7	74.1	63.4	91.5	78.4
Maggino [6]	93.3	69	60.9	95.2	77.2
Granberg [2]	96.7	62.1	56.9	97.3	73.8
Szpurek [9]	90	79.3	69.2	93.9	82.9
Ueland [8]	93.3	70.7	62.2	95.3	78.4

curacy for Szpurek index were 90%, 79.3%, 69.2%, 93.9%, and 73%, respectively. Sensitivities, specificities, positive and negative predictive values, and diagnostic accuracies of all other scoring systems in the postmenopausal subgroup are presented in Table 4.

According to McNemar's test, there were no statistically significant differences for paired samples of Sassone and Lerner, Sassone and Granberg, Lerner and Finkler, Lerner and Szpurek, Lerner and Ueland, Ferrazzi and DePriest, DePriest and Maggino, DePriest and Granberg, DePriest and Ueland, Finkler and Szpurek, Finkler and Ueland and Maggino and Ueland [1-9]. All thirty-six comparisons of paired samples with *p*-values are shown in Table 5.

Discussion

Assessing an ovarian mass with an ultrasonography is generally subjective and depends on the performer's experience and skill. However in the preoperative evaluation of an ovarian mass, the correct way that takes us to the diagnosis of malignancy is still made by clinical impression and ultrasound examination. For this reason explicit criteria should be established and defined rigidly for adnexal mass morphology and anyone should be able to apply the same scoring system.

Interpretation of sonographic findings is a diagnostic challenge. Transvaginal ultrasound examination in comparison to the transabdominal route has improved the results [2, 3]. All scoring systems that have been described are based on semiquantitative parameters of the appearance of the tumor. The present authors studied three morphological classification systems and six morphologic indices here.

The diagnostic performances of the models were poorer in this study than in those in which the models were created except for DePriest and Szpurek indices results of which were similar to the present [4, 9].

Subtle differences of examination technique and definitions may partly illuminate the lower performances in the present study than those in the original ones. The most important contributing factor to the differences in results is likely to be true differences in the tumor populations studied. In the present study, certain tumor types tended to be over-represented among the true negative, and false positive diagnoses. Simple benign cysts and endometriomas were usually truly classified as benign and were over-represented among the true-negative diagnoses. In the false positive group, cystadenomas, tuboovarian abscesses, leiomyomas, and fibromas were often misclassified as malignant.

Although it is believed that tumor volume estimation lowers index specificity [3], in the present study, indices using tumor volume as a parameter [4, 8, 9], had higher performances due to large sample of benign cysts. The present authors obtained low positive predictive values due to the high number of false-positive cases and they met high negative predictive values because of the fact that the number of the false-negative cases was low. The reason of this fact is that, in the tumor populations studied, low malignant potential tumors were not much and mostly classified as malignant because of their large volumes.

This study has revealed the usefulness of scoring systems to correctly discriminate between benign and malignant pelvic masses. Morphological indices or classification systems studied here can be used for the discrimination between benign and malignant pelvic masses and for selection of cases for optimal therapy. Although some degree of simple mathematical calculations are mandatory for gray-scale morphological indices, they appear to be very accurate, are useful in clinical practice, are easy to perform, and should therefore be the test of choice in the preoperative evaluation of the adnexal mass. Higher diagnostic performance of Szpurek Index is making it attractive to be used for detecting malignancy.

References

[1] Finkler N.J., Benacerraf B., Lavin P.T., Wojciechowski C., Knapp R.C.: "Comparison of serum CA125, clinical impression, and ultrasound in the preoperative evaluation of ovarian masses". *Obstet. Gynecol.*, 1988, *72*, 659.

[2] Granberg S, Norstro¨m A, Wikland M: Tumors in the lower pelvis as imaged by vaginal sonography. Gynecol Oncol 37:224 –229, 1990.

[3] Sassone A.M., Timor-Tritsch I.E., Artner A., Westhoff C., Warren W.B.: "Transvaginal sonographic characterization of ovarian disease: evaluation of a new scoring system to predict ovarian malignancy". *Obstet. Gynecol.*, 1991, *78*, 70.

[4] DePriest P.D., Shenson D., Fried A., Hunter J.E., Andrews S.J., Gallion H.H., et al.: "A morphology index based on sonographic findings in ovarian cancer". *Gynecol. Oncol.*, 1993, *51*, 7.

[5] Lerner J.P., Timor-Tritsch I.E., Federman A., Abramovich G.: "Transvaginal ultrasonographic characterization of ovarian masses with an improved, weighted scoring system". *Am. J. Obstet. Gynecol.*, 1994, *170*, 81.

[6] Maggino T., Gadducci A., D'Addario V., Pecorelli S., Lissoni A., Stella M., *et al.*: "Prospective multicenter study on CA 125 in postmenopausal pelvic masses". *Gynecol. Oncol.*, 1994, *54*, 117.

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

[8] Ueland F.R., DePriest P.D., Pavlik E.J., Kryscio R.J., van Nagell J.R. Jr.: "Preoperative differentiation of malignant from benign ovarian tumors: the efficacy of morphology indexing and Doppler flow sonography". *Gynecol. Oncol.*, 2003, *91*, 46.

[9] Szpurek D., Moszynski R., Zietkowiak W., Spaczynski M., Sajdak S.: "An ultrasonographic morphological index for prediction of ovarian tumor malignancy". *Eur. J. Gynaecol. Oncol.*, 2005, *26*, 51.

[10] Jacobs I., Oram D., Fairbanks J., Turner J., Frost C., Grudzinskas J.: "A risk of malignancy index incorporating CA125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer". *BJOG*, 1990, *97*, 922.

[11] Ronco D.A., Manahan K.J., Geisler J.P.: "Ovarian cancer risk assessment: a tool for preoperative assessment". *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 2011, *158*, 325. doi: 10.1016/j.ejogrb.2011.05.018. Epub 2011 Jul 28.

[12] Moore R.G., McMeekin D.S., Brown A.K., DiSilvestro P., Miller M.C., Allard W.J., *et al.*: "A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer inpatients with a pelvic mass". *Gynecol. Oncol.*, 2009, *112*, 40.

[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*, 357.

[14] Serov S.F., Scully R.E., Sobin L.H.: "International histological classification of tumors, no. 9. Histological typing of ovarian tumors". Geneva: World Health Organization, 1973.

Address reprint requests to:
E. AKTÜRK, M.D.
Department of Obstetrics and Gynecology
Gülhane Military Medical Faculty
Etlik, 06010, Ankara (Turkey)
e-mail: erhnakturk@gmail.com