Original Article

Access this article online

Quick Response Code:

Website: www.pogsjournal.org

DOI: 10.4103/pjog.pjog_37_22

Assessing the diagnostic performance of four ovarian malignancy prediction risk models in differentiating benign and malignant ovarian masses in a tertiary hospital

Mea Janelle F. Sarmiento-Babiera[1], Erlidia F. Llamas-Clark[1]

Abstract:

INTRODUCTION: Ovarian cancer is considered the most lethal gynecologic malignancy because it is difficult to diagnose in its early stages. Ovarian malignancy prediction models may be useful in discriminating between benign and malignant masses, allowing for accurate and timely referral as well as proper therapeutic care.

OBJECTIVE: To evaluate the diagnostic performance of the four ovarian prediction models: Risk of Malignancy Index-4 (RMI-4), Risk of Ovarian Malignancy Algorithm (ROMA), Copenhagen Index (CPH-I), and International Ovarian Tumor Analysis (IOTA)-Assessment of Different NEoplasias in the AdneXa (ADNEX) in identifying malignant and benign ovarian masses.

MATERIALS AND METHODS: This was a retrospective, cross-sectional, analytical diagnostic study in a tertiary hospital between January 2017 and December 2020. Receiver operating characteristic (ROC) curves, area under the curves (AUCs), sensitivities, specificities, positive and negative predictive values, and positive and negative likelihood ratios were used to assess the diagnostic performance of the prediction models.

RESULTS: We analyzed a total of 248 patients. One hundred and sixty-one (65%) had benign tumors, 28 (11%) had borderline, and 59 (24%) had malignant tumors. The AUCs of all models were all above 90%, but when compared to the other models, CPH-I had the best estimate. RMI-4 had the highest sensitivity (98.3%) in diagnosing malignancy. For appropriately diagnosing benign disease, the IOTA-ADNEX model exhibited the highest specificity (92.1%). Overall, RMI-4 had the lowest diagnostic accuracy (74.6%), whereas IOTA-ADNEX had the greatest (93.2%).

CONCLUSION: The four malignancy prediction models in this study were all useful tools in discriminating between benign and malignant ovarian tumors. IOTA-ADNEX, CPH-I, and ROMA all demonstrated overlapping diagnostic performances indicating that they are equal in that regard. In terms of sensitivity in predicting malignancy, RMI-4 was the most sensitive. CPH-I is the predictor with the best overall estimate. Lastly, IOTA-ADNEX was the most specific, and displayed highest diagnostic accuracy among the four.

Keywords:

Copenhagen Index, HE4, International Ovarian Tumor Analysis-Assessment of Different NEoplasias in the AdneXa, ovarian cancer, Risk of Malignancy Index, ROMA

[1]*Department of Obstetrics and Gynecology, Manila Doctors Hospital, Manila, Philippines*

Address for correspondence:
Dr. Mea Janelle F. Sarmiento-Babiera, Department of Obstetrics and Gynecology, Manila Doctors Hospital, 667 United Nations Ave, Ermita, Manila, 1000 Metro Manila, Philippines.
E-mail: mjfsarmiento@gmail.com

Submitted: 12-Sep-2022
Revised: 12-Sep-2022
Accepted: 12-Sep-2022
Published: 06-Dec-2022

*First Place, 2022 PHILIPPINE OBSTETRICAL AND GYNECOLOGICAL SOCIETY (Foundation), INC., Midyear Research Paper Contest, July 17, 2022, Online Platform: ZOOM Webinar

How to cite this article: Sarmiento-Babiera MJ, Llamas-Clark EF. Assessing the diagnostic performance of four ovarian malignancy prediction risk models in differentiating benign and malignant ovarian masses in a tertiary hospital. Philipp J Obstet Gynecol 2022;46:193-201.

Introduction

More women die from ovarian cancer than from any other malignancy of the female reproductive system.[1] In the Philippines, the latest data show that ovarian cancer ranks as 12th overall, and the 5th most common cancer among Filipino women.[2] However, in the latest GLOBOCAN 2018, ovarian cancer ranks 8th worldwide as the leading cause of death of women due to cancer.[3] The 5-year survival rate among these patients is only 35% to 57%.[4]

Ovarian cancer may be considered the most lethal gynecologic malignancy as it is often diagnosed at an advanced stage.[5] As a result, it is critical to develop tests that can predict and correctly identify benign from malignant masses. According to the American College of Obstetrics and Gynecology and the Society of Gynecologic Oncology, women should be referred to a gynecologic oncologist if they have an elevated score on a formal risk assessment test, such as the Risk of Ovarian Malignancy Algorithm (ROMA), multivariate index assay, Risk of Malignancy Index (RMI), or one of the ultrasound-based scoring systems.[6]

For that reason, this study was made to compare the predictive performance of four malignancy ovarian prediction models, namely, RMI-4, Risk of Malignancy Algorithm (ROMA), Copenhagen Index (CPH-I), and International Ovarian Tumor Analysis-Assessment of Different NEoplasias in the AdneXa (IOTA-ADNEX) in differentiating benign and malignant ovarian masses. The comparison of the ovarian prediction models done in the local setting will help in determining the best and most cost-effective tool to be utilized and possibly adopted in the tertiary hospital. While histopathology is still the only definitive diagnosis, these models may help and guide referral decisions for women suspected of having ovarian cancer.

Review of Related Literature

Risk of malignancy index

The very first RMI was proposed in 1990 by Jacobs *et al*., using Cancer antigen 125 (CA 125), ultrasound findings, and menopausal status.[7] Formula is computed as follows: RMI = U × M × CA125, where:

- U: ultrasound score –1 point scored for each of the following features: multilocular cysts, solid areas, metastases, ascites, and bilateral lesions. *U* =1 (1 point) or *U* = 3 (2–5 points)
- M: menopause score, *M* = 1 (premenopausal) or *M* = 3 (postmenopausal)
- CA125: serum CA125 measured in international units/ml.

Over the years, however, there have been amendments to the RMI scoring. The RMI-4, described by Yamamoto *et al.* now included another variable, S, which represented size of mass in centimeters.[8] Ultrasound and menopausal status were also scored differently than the original RMI. Updated formula shows: RMI-4 = U × M × S × CA125, where:

- *M* = 1 (premenopausal), or *M* = 4 (postmenopausal)
- *U* = 1 for ultrasound score of 0 or 1, *U* = 4 for ultrasound score of >1
- *S* = 1 for single greatest diameter of tumor size <7 cm, *S* = 2 for single greatest diameter of tumor size ≥7 cm.

To establish an association with malignancy under RMI-4, score should be above 450.[8] According to a study comparing RMI-4, RMI-4 was noted to be more sensitive, specific, and accurate in identifying malignancy than the other three (McNemar test, *P* = 0.063).[9]

Risk of ovarian malignancy algorithm

ROMA was proposed by Moore *et al*. where they associated HE4 and CA125 levels according to the menopausal status.[10] A predictive index (PI) value is calculated, and formula depends on whether the woman is pre- or postmenopausal to determine the score.[10]

- Premenopausal: PI = −12.0 + 2.38 × LN (HE4) +0.0626 × LN (CA125)
- Postmenopausal: PI = −8.09 + 1.04 × LN (HE4) +0.732 × LN (CA125)
- LN = Natural (logarithm)
- ROMA score (%) = exp (PI)/(1 + exp [PI]) ×100%.

Patients were classified as high or low risk for epithelial ovarian cancer based on the following criteria: [11]

- Premenopausal ROMA score:
  - ≥11.4% = high risk for epithelial ovarian cancer
  - <11.4% = low risk for epithelial ovarian cancer.
- Postmenopausal ROMA score:
  - ≥29.9% = high risk for epithelial ovarian cancer
  - <29.9% = low risk for epithelial ovarian cancer.

For Chen *et al*., ROMA was more sensitive than HE4 (96.7% vs. 73.3%) but with less specificity (80% vs. 98.6%).[12] In that same study, AUC for ROMA and HE4 were not significantly different (0.97 and 0.96, respectively).

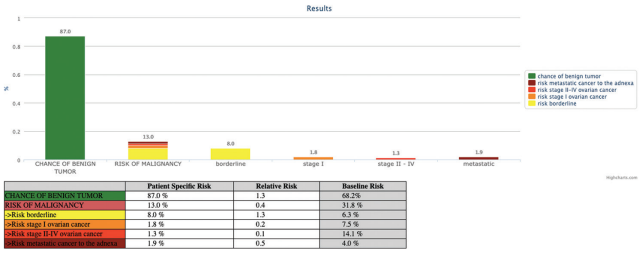
Copenhagen index

The CPH-I was reported by Karlsen *et al.* as a novel diagnostic score index in ovarian tumors.[13] It uses the same mathematical method of ROMA using predicted probability (PP). CPH-I was calculated using HE4, CA125, and age rather than menopausal status. The formula for CPH-I = −14.0647 + 1.0649 × log2 (HE4) +0.6050 × log2 (CA125) +0.2672 × age/10 with

PP = e (CPH-/(1 + e[CPH-I]). An optimal cutoff of ≥0.070 was established in their Danish development sample of 2665 patients.[13] A patient's age is objective and easily obtained, in contrast to menopausal status (used in ROMA and RMI) and ultrasound information (used in RMI). Thus, CPH-I may have a role in improving triage of women with suspected ovarian cancer.

International ovarian tumor analysis-assessment of different neoplasias in the adnexa

In 2014, Van Calster *et al*. proposed a model with a better diagnostic performance which was later adapted by the IOTA group.[14] The IOTA-ADNEX is the first model to distinguish if a mass has the risk of being borderline, Stage I, Stage II-IV, or is a secondary metastasis to the adnexa.[14] The model is composed of three clinical predictors and six ultrasonographic predictors. The IOTA group have produced Apple, Android, and web applications for calculating the risk score[15] and a sample of the result is depicted below.



	Patient Specific Risk	Relative Risk	Baseline Risk
CHANCE OF BENIGN TUMOR	87.0 %	1.3	68.2%
RISK OF MALIGNANCY	13.0 %	0.4	31.8 %
->Risk borderline	8.0 %	1.3	6.3 %
->Risk stage I ovarian cancer	1.8 %	0.2	7.5 %
->Risk stage II-IV ovarian cancer	1.3 %	0.1	14.1 %
->Risk metastatic cancer to the adnexa	1.9 %	0.5	4.0 %

The European Society of Gynecological Oncology, International Society of Ultrasound in Obstetrics and Gynecology, IOTA group, and the European Society for Gynecological Endoscopy jointly developed clinically relevant and evidence-based statements on the preoperative diagnosis of ovarian tumors wherein they have recommended the use of the IOTA-ADNEX model and IOTA simple rules as they outperform existing scoring systems, including the RMI.[16] This has also been reinforced by the Philippine Obstetrical and Gynecological Society in the management of adnexal masses, wherein the IOTA-ADNEX model was recommended to be used for discriminating benign and malignant masses as well as preoperative staging of these masses.[17]

Research Objectives

General

To evaluate the diagnostic performance of the four ovarian prediction models: RMI-4, ROMA, CPH-I, and IOTA-ADNEX model in identifying benign and malignant ovarian masses.

Specific

1. To determine the prevalence of malignant and benign ovarian masses among gynecologic patients in a tertiary hospital
2. To determine the sensitivity, specificity, likelihood ratios (LRs), positive predictive value (PPV) and negative predictive value (NPV), and accuracy of RMI-4 in detecting benign versus malignant ovarian mass comparing against histopathology as the gold standard
3. To determine the sensitivity, specificity, LRs, PPV and NPV, and accuracy of ROMA in detecting benign versus malignant ovarian mass comparing against histopathology as the gold standard
4. To determine the sensitivity, specificity, LRs, PPV and NPV, and accuracy of CPH-I in detecting benign versus malignant ovarian mass comparing against histopathology as the gold standard
5. To determine the sensitivity, specificity, LRs, PPV and NPV, and accuracy of IOTA-ADNEX in detecting benign versus malignant ovarian mass comparing against histopathology as the gold standard
6. To determine the proportion of benign and borderline ovarian masses confirmed by histopathological biopsy among those tumors considered to be benign or borderline by IOTA-ADNEX model
7. To analyze the four models in accurately identifying malignant or benign ovarian masses by comparing their sensitivities, specificities, receiver operating characteristic (ROC), and area under the curve (AUC) using paired tests.
8. To discuss the comparison of the four models in terms of cost-effectiveness, prediction completeness, ease of use, and results with other studies.

Materials and Methods

Research design and setting

This was a retrospective single-center, analytical, observational, and cross-sectional diagnostic study utilizing review of data obtained from medical records of patients admitted at a tertiary hospital from January 2017 to December 2020.

Research population

This study included patients who presented with ovarian masses in a tertiary hospital.

Inclusion criteria

- Patients admitted with adnexal or ovarian mass who underwent surgery in the tertiary hospital
- Ultrasound confirmation with IOTA-ADNEX scoring of the ovarian masses done by Obstetrics and Gynecology sonologists who are fellows of the Philippine Society of Ultrasound in Obstetrics and Gynecology.
- Availability of CA125, HE4, and histopathology results.

Exclusion criteria

- Incomplete patient data and profile
- No ultrasound confirmation of ovarian masses and no surgery done
- Previous bilateral oophorectomy or known ovarian cancer
- Incomplete tumor markers: CA125 and HE4
- No histopathological report.

Sample size computation

A minimum of 246 patients with ovarian masses were required for this study based on a level of significance of 5%, an area under the curve of 0.882, and a width of the confidence interval (CI) of 0.10.

Legend:

n = minimum sample size AUC = 0.882

L = width of the CI (precision) ±0.05 = 0.10 Zα/2 = 1.96

Sample size formula:[18]

$$n = 2\left(\frac{Z^2_{a/2}\ V(\text{AUC})}{d^2}\right)$$

$$V(\text{AUC}) = (0.0099 \times e^{-a^2/2}) \times (6 \times a^2 + 16)$$

$$a = \varphi^{-1}\ (\text{AUC}) \times 1.414$$

$$a = 1.676$$

$$V(\text{AUC}) = (0.0099 \times e^{-1.676^2/2}) \times (6 \times 1.676^2 + 16)$$

$$V(\text{AUC}) = 0.080$$

$$n = 2\left(\frac{(1.96)^2\ (0.080)}{0.05^2}\right)$$

$$n \geq 246$$

Data collection procedure

The list of patients who fulfilled the inclusion criteria generated from the database of all patients who had presented with ovarian mass from January 2017 to December 2020 was retrieved. The researcher sought ethical clearance from the Institutional Review Board (IRB). Permission to retrieve medical charts was obtained from the Data Privacy Officer of the institution, as well as the Chair of the IRB. Data were collected from medical charts using a data collection form where the demographic profile of each patient, as well as the diagnostic and histopathological descriptions, was transcribed. All patients were given a specific number code, and the data obtained from the medical charts was encoded under each patient number code. A summarized index and reference table was utilized in determining the malignancy score for each patient. No patient identifier including names and patient ID was collected. The overall procedure is shown in the diagrammatic workflow.

Data processing and analysis

For all objectives

Descriptive statistics was used to summarize the general and clinical characteristics of the participants. Frequency and proportion were used for categorical variables. Shapiro–Wilk test was used to determine the normality distribution of continuous variables. Continuous quantitative data that met the normality assumption were summarized using mean and standard deviation (SD), while those that do not were described using median and range.

For specific objectives 1, 2, 3, 4, 5, and 6

Sensitivity, specificity, PPV, NPV, positive LR, negative LR, and diagnostic accuracy were used to assess the diagnostic performance of the four ovarian prediction models (RMI-4, ROMA, CPH-I, and IOTA-ADNEX) in predicting benign or malignant ovarian masses, with histopathology results as the gold standard.

For specific objective 6

Cohen's kappa was used to determine the agreement of diagnosis between IOTA-ADNEX and histopathological results in identifying benign, borderline, and malignant ovarian masses.

For specific objective 7

ROC curve was constructed to determine the optimal cutoff value of the four ovarian prediction models (RMI-4, ROMA, CPH-I, and IOTA-ADNEX) to predict benign or malignant ovarian masses, with histopathology results as the gold standard. Youden's J index was defined for all points along the ROC curve, and the maximum value of the index was used as a criterion for selecting the best cut point. DeLong test was used to compare the AUCs.

Agreement

All valid data were included in the analysis. Missing variables were neither replaced nor estimated. Null hypothesis was rejected at 0.05α-level of significance. STATA 15.0 (StataCorp SE, College Station, TX, USA) was used for data analysis.

Results

We analyzed 248 women with ovarian masses using the four malignancy risk models. The median age at the time of presentation (± SD) was 44 ± 15 years old. One hundred and fifty-three (61.7%) of the 248 patients were premenopausal, while 95 (38.3%) were postmenopausal. The majority were multigravid (50%). Palpable

Table 1: Demographic and clinical profile of women (*n*=248)

	Mean±SD; median (range); frequency (%)
Age (years)	44.39±15.01
Obstetric	
Gravidity	
G0	84 (33.87)
G1	40 (16.13)
≥G2	124 (50)
Parity	
P0	90 (36.29)
P1	44 (17.74)
≥P2	114 (45.97)
Term births (*n*=158)	2 (0-6)
Premature births (*n*=158)	0 (0-3)
Abortions	0 (0-3)
Menopausal stage	
Premenopause	153 (61.69)
Postmenopause	95 (38.31)
Chief complaint	
Palpable abdominal mass	84 (33.87)
Increased abdominal girth	80 (32.26)
Incidental finding of ovarian cyst	47 (18.95)
Hypogastric pain	20 (8.06)
Progressive dysmenorrhea	11 (4.44)
Postmenopausal bleeding	3 (1.21)
Dyspnea	1 (0.4)
Heavy and prolonged menses	1 (0.4)
Pelvic heaviness	1 (0.4)
Comorbidities (*n*=108)	
Hypertension	59 (54.63)
Diabetes mellitus	24 (22.22)
Bronchial asthma	13 (12.04)
Breast cancer	3 (2.78)
Colon cancer	3 (2.78)
Hyperthyroidism	3 (2.78)
Hypothyroidism	3 (2.78)
PCOS	3 (2.78)
Pneumonia	3 (2.78)
Abnormal uterine bleeding	2 (1.85)
CKD	2 (1.85)
Pregnant	2 (1.85)
Previous admission	19 (7.66)
CA125 level, (U/mL)	67.00 (4.80-1364.00)
HE4 level (pmol/L)	36.80 (10.15-1216.13)

CKD: Chronic kidney disease, PCOS: Polycystic ovary syndrome, SD: Standard deviation

Table 2: Ultrasound features of women (*n*=248)

	Median (range) frequency (%)
Maximum lesion diameter (mm)	140 (58-380)
Proportion of solid tissue (%)	0 (0-100)
Number of papillary projections	
0	239 (96.37)
1	3 (1.21)
2	3 (1.21)
3	1 (0.40)
>3	2 (0.81)
Presence of more than 10 locules	38 (15.32)
Acoustic shadows	31 (12.50)
Ascites	134 (54.03)
Bilateral lesions	38 (15.32)
Metastasis	25 (10.08)

Table 3: Histologic types and cancer staging (*n*=248)

	Frequency (%)
Benign	161 (64.92)
Mucinous cystadenoma	42 (26.09)
Mature cystic teratoma	37 (22.98)
Serous cystadenoma	24 (14.91)
Endometriotic cyst	23 (14.29)
Fibroma	7 (4.35)
Serous cystadenofibroma	7 (4.35)
Seromucinous cystadenoma	4 (2.48)
Struma ovarii	4 (2.48)
Fibrothecoma	2 (1.24)
Tubo-ovarian abscess	2 (1.24)
Borderline	28 (11.29)
Borderline mucinous tumor	26 (92.86)
Borderline serous tumor	2 (7.14)
Malignant	59 (23.79)
Serous adenocarcinoma	18 (30.51)
Endometrioid adenocarcinoma	14 (23.73)
Mucinous adenocarcinoma	8 (13.56)
Clear cell adenocarcinoma	4 (6.78)
Seromucinous carcinoma	2 (3.39)
Yolk sac tumor	2 (3.39)
Granulosa cell tumor	2 (3.39)
Immature teratoma	2 (3.39)
Mixed Mullerian tumor	2 (3.39)
Staging of malignancy	
I	18 (31.58)
II	8 (14.04)
III	23 (40.35)
IV	8 (14.04)

abdominal mass (34%) and increased abdominal girth (32%) were the most common chief complaints while an incidental finding of an ovarian cyst prompted further workup in 19% of cases [Table 1].

The data in Table 2 summarize the ultrasound findings and characteristics of the ovarian tumors. The median maximum diameter of the lesion was 140 mm (range: 58–380). Only 9 (4%) were detected with papillary projection, while more than 10 cyst locules were noted in 15% of patients. More than half of the patients had ascites (54%). Fifteen percent had bilateral ovarian lesions, and 10% had metastatic cancer.

The histopathological findings are shown in Table 3. A final diagnosis of benign pathology was made for 65% (95% CI: 58.63%–70.85%), malignant for 24% (95% CI: 18.63%–29.59%), and borderline for 11% (95% CI: 7.63%–15.90%).

No secondary cancer of the ovary was determined. Among benign lesions, mucinous cystadenoma (26%), mature cystic teratoma (23%), serous cystadenoma (15%), and endometriotic cyst (14%) made up the majority of findings. Borderline tumors were mostly mucinous (93%). The predominant malignant histopathological diagnoses were serous adenocarcinoma (31%), endometrioid adenocarcinoma (24%), and mucinous adenocarcinoma (14%). Most cancers were either Stage III (40%) or Stage I (32%).

Based on this receiver operating characteristic (ROC) curve, all had their area under the curve (AUC) above 90%, which indicates high discriminative power, meaning all models were very good in differentiating between benign versus malignant tumors. CPH-I had the best estimate of AUC at 0.978 while ROMA had the lowest. [Table 4]. CPH-I had the best estimate of AUC at 0.978, while ROMA had the lowest at 0.946. The AUCs of RMI-4 and ADNEX were very similar at 0.952 and 0.955, respectively [Figure 1].

Table 5 summarizes the diagnostic performances of the ovarian models. RMI-4 displayed the highest sensitivity (98.3% [95% CI: 90.9%–100%]) among the four malignancy risk indices. ROMA and CPH-I had slightly lower and equal sensitivities (94.9% [95% CI: 85.9%–98.9%]), while IOTA-ADNEX had the lowest sensitivity (86.4% [95% CI: 75.0%–94.0%]) in detecting ovarian malignancy.

In terms of correctly labeling benign pathology, RMI-4 had the lowest specificity (67.2% [95% CI: 60.0%–73.8%]), ROMA and CPH-I had higher and equal specificities (89.4% [95% CI: 84.1%–93.4%]), while IOTA-ADNEX displayed the highest specificity at 92.1% (95% CI: 87.2%–95.5%). For the other diagnostic parameters, IOTA-ADNEX demonstrated the highest PPV (85.0%) and positive LR (18.20), while RMI-4 had the highest NPV (99.2%) and lowest negative LR (0.025). The overall accuracy was lowest for RMI-4 (74.6% [95% CI: 68.7%–79.9%]) and highest for ADNEX at 93.2% (95% CI: 89.3%–96.0%).

IOTA-ADNEX, CPH-I, and ROMA all had overlapping diagnostic performances (see 95% CIs), so they are equivalent in that regard. RMI-4 had lower values when compared to the other three.

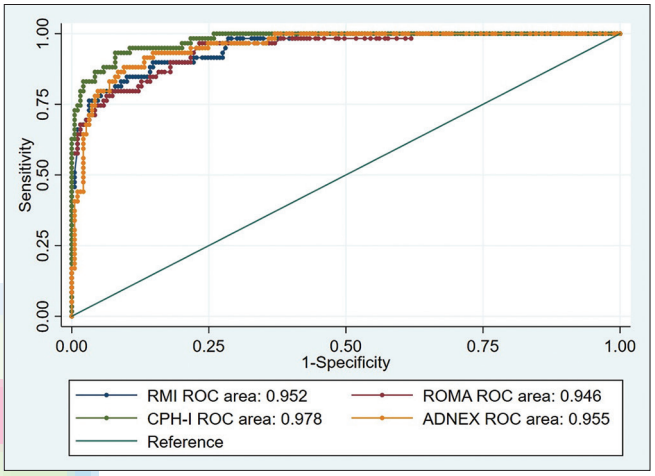


Figure 1: Receiver operating characteristic curves of Risk of Malignancy Index-4, ROMA, Copenhagen Index-I, and International Ovarian Tumor Analysis-Assessment of Different NEoplasias in the AdneXa for predicting ovarian malignancy

Table 4: Areas under the receiver operating characteristic curves of the four models to predict malignant ovarian mass

RMI-4	ROMA	CPH-I	IOTA-ADNEX	*P*
Area under the ROC curve (95% CI)				
0.952 (0.926-0.979)	0.946 (0.915-0.977)	0.978 (0.962-0.994)	0.955 (0.929-0.980)	<0.001

Statistical test used: DeLong's test. CI: Confidence interval, IOTA: International Ovarian Tumor Analysis, ADNEX: Assessment of Different NEoplasias in the AdneXa, RMI: Risk of Malignancy Index, ROMA: Risk of Ovarian Malignancy Algorithm, CPH-I: Copenhagen Index, ROC: Receiver-operating characteristic

Table 5: Diagnostic performance of ovarian prediction models (*n*=248)

	RMI-4	ROMA	CPH-I	IOTA-ADNEX
True positives (*n*)	58	56	56	51
True negatives (*n*)	127	169	169	180
False positives (*n*)	62	20	20	9
False negatives (*n*)	1	3	3	8
Sensitivity (%)	98.3 (90.9-100)	94.9 (85.9-98.9)	94.9 (85.9-98.9)	86.4 (75.0-94.0)
Specificity (%)	67.2 (60.0-73.8)	89.4 (84.1-93.4)	89.4 (84.1-93.4)	92.1 (91.2-97.8)
PPV (%)	48.3 (39.1-57.6)	73.7 (62.3-83.1)	73.7 (62.3-83.1)	85.0 (73.4-92.9)
NPV (%)	99.2 (95.7-100)	98.3 (95.0-99.6)	98.3 (95.0-99.6)	95.7 (91.8-98.1)
Positive LR	3.00 (2.44-3.69)	8.97 (5.90-13.60)	8.97 (5.90-13.60)	18.20 (9.52-34.6)
Negative LR	0.025 (0.004-0.177)	0.057 (0.019-0.17)	0.057 (0.019-0.17)	0.142 (0.075-0.271)
Accuracy (%)	74.6 (68.7-79.9)	90.7 (86.4-94.0)	90.7 (86.4-94.0)	93.2 (89.3-96.0)

LR: Likelihood ratio, IOTA: International Ovarian Tumor Analysis, ADNEX: Assessment of Different NEoplasias in the AdneXa, RMI: Risk of Malignancy Index, ROMA: Risk of Ovarian Malignancy Algorithm, CPH-I: Copenhagen Index, PPV: Positive predictive value, NPV: Negative predictive value

Table 6: International Ovarian Tumor Analysis-Assessment of Different NEoplasias in the AdneXa and histopathology results

IOTA-ADNEX	Histopathology results, frequency (%)			Agreement (%)	Kappa	*P*
	Benign (*n*=161)	Borderline (*n*=28)	Malignant (*n*=59)			
Benign	154 (95.65)	26 (92.86)	8 (13.56)	82.66	0.615	<0.001
Borderline	0	0	0			
Malignant	7 (4.35)	2 (7.14)	51 (86.44)			

IOTA: International Ovarian Tumor Analysis, ADNEX: Assessment of Different NEoplasias in the AdneXa

This table showed that the IOTA-ADNEX categories significantly agreed with histopath classification (kappa=0.615) and the agreement was unlikely to be due to chance alone ($P < 0.001$) [Table 6].

Discussion

Assessment of the diagnostic performance of risk of malignancy index-4

In this study, RMI-4 had the highest sensitivity in predicting ovarian malignancy and the lowest specificity in correctly labeling benign pathology. As a result, it had the lowest overall accuracy (74.6%) out of the four prediction models. This might be because of the role of CA125 in ovarian masses. Because CA125 is a nonspecific tumor marker, it may also be increased in some benign tumors as well.

There had been various comparisons among the different RMI scores. The study's original cutoff was 200, and with a sample size of 143, the sensitivity and specificity were 85.4 and 96.9, respectively.[7] RMI-4, created by Yamamoto *et al*., had sensitivity and specificity of 86.8 and 91, respectively, in their retrospective study of 253 cases.[8] The AUCs of RMI-4 and IOTA-ADNEX in this study were very similar at 0.952 and 0.955. Because of that, it still proves to be a good tool in distinguishing benign and malignant ovarian masses.

Assessment of the diagnostic performance of ROMA and CPH-I

This study showed that both ROMA and CPH-I had similarities in both their sensitivities in detecting malignancy and specificities in identifying benign disease. Moreover, ROMA and CPH-I have almost similar diagnostic performances as well. The results might be because of the almost similar algorithm of ROMA and CPH-I, their difference being the use of menopausal status versus age in their respective formulas. In addition to that, it should be noted that CPH-I, being the novel model among the four, had the highest AUC estimate of 0.978 (95% CI: 0.962–0.994). The diagnostic performance of CPH-I was also validated in another study wherein CPH-I performed as well as ROMA and RMI in differentiating benign and malignant ovarian masses.[19]

Assessment of the diagnostic performance of international ovarian tumor analysis-assessment of different neoplasias in the adnexa

In this investigation, IOTA-ADNEX exhibited the lowest sensitivity in diagnosing ovarian cancer (86.4%). However, in terms of specificity, IOTA-ADNEX displayed the highest specificity at 92.1%. The strength of the original study of the IOTA-ADNEX showed an AUC of 0.94 (0.93–0.95), leading Van Calster *et al.* in acknowledging its potential in discriminating benign and malignant ovarian tumors.[14]

It should be noted that in this study, IOTA-ADNEX exhibited the highest PPV (85.0%) and positive LR (18.20), resulting in the highest overall accuracy of 93.2%. The investigation also showed that the IOTA-ADNEX categories substantially agreed with the classification confirmed by histopathological biopsy among those tumors considered to be benign or borderline and the agreement was unlikely to be due to chance alone.

Comparison of the risk of malignancy index, roma, copenhagen index, and international ovarian tumor analysis-assessment of different neoplasias in the adnexa

To provide a comprehensive assessment, we attempted to compare the four prediction models in the following categories:

Cost-effectiveness

In comparing the four, expenses were broken down to determine which is most cost-effective. Ultrasound is used in both IOTA-ADNEX and RMI-4. In our setting, ultrasound costs about P2900-P4800, and can be susceptible to additional charges. Both also use CA-125 which costs roughly P1,470.

For IOTA-ADNEX and RMI-4, a price range of P4-6 thousand may be charged. On the other hand, ROMA, and CPH-I, will cost about P4,095 for both CA-125 and HE4, saving the patient roughly P200-P2,200. Therefore, it's reasonable to conclude that ROMA and CPH-I are more cost-effective in assessing malignancy risk. Both are good additional test options if IOTA-ADNEX or trained sonographers are not available in a facility (particularly in the rural context).

Prediction completeness

The IOTA ADNEX appears to be a better option in terms of risk prediction completeness. Whereas RMI-4, ROMA, and CPH-I only determine whether a tumor is benign or malignant, IOTA-ADNEX also determines whether the tumor is borderline, has a risk of stage I, stage II-IV, or metastasis.

Ease of use

The inventors and third-party developers of ROMA, RMI-4, and IOTA-ADNEX have addressed ease of use as a factor by creating online and mobile applications. CPH-I, on the other hand, has yet to create an app or an online tool for doing the calculations. To obtain results from this tool, users will have to solve logarithmic equations.

Comparison with other studies

A study was done comparing IOTA-ADNEX and ROMA as a tool to predict malignancy showing that IOTA-ADNEX was superior to ROMA in terms of sensitivity (94% vs. 84%) and specificity (82% and 80%, respectively).[20] Another study was performed by Auekitrungrueng (2019) comparing the accuracy of IOTA versus RMI-1 and RMI-2 in diagnosing ovarian tumors, and it was found out the IOTA-ADNEX had a significantly higher sensitivity and specificity (83.8% and 92.0%, respectively) compared to the sensitivity and specificity of RMI-1 (77.2% and 86.8%, respectively) and RMI-2 (82.1% and 82.6%, respectively).[21] Another study in Sweden showed that ROMA and RMI were equally effective in identifying between malignant and benign tumors, with an equal specificity of 75%.[22] In another study, it was found out that CPH-I and ROMA were almost very equal in terms of sensitivity and specificity, both approaching 89% and 85%, respectively.[23]

In comparison with the other studies previously stated, according to this study, all four ovarian prediction models were useful in distinguishing between benign and malignant ovarian masses, as evidenced by their AUCs. Their AUCs were all above 90%, which indicates high discriminative power. CPH-I, on the other hand, had a highest AUC estimate of 0.978, while ROMA had the lowest at 0946. In terms of the diagnostic performance, IOTA-ADNEX, CPH-I, and ROMA all demonstrated overlapping diagnostic performances, indicating that they are equal in that regard. When compared to the other three, RMI-4 exhibited lower values.

Conclusion

The four ovarian malignancy prediction models, namely, RMI-4, ROMA, CPH-I, and IOTA-ADNEX, were all useful tools in discriminating between benign and malignant ovarian tumors. Based on this study, all the ovarian models demonstrated overlapping diagnostic performances indicating that they are equal in that regard. In terms of sensitivity in predicting malignancy, RMI-4 was the most sensitive. CPH-I is the predictor with the best overall estimate. Lastly, IOTA-ADNEX was the most specific, and displayed highest diagnostic accuracy among the four.

There is no one-size-fits-all approach to using a specific model in clinical practice as there is with any prediction tool. The goal of risk models is only to provide accurate risk estimates for individual patients. Because the four indices in this study were found to be useful, it may be up to the physician's discretion to decide which is best to use in their own clinical setting.

Limitations

This study was limited to a retrospective single-center study and data gathered were limited to review of patient charts from January 2017 to December 2020. Since this is a retrospective study, there may be some selection bias. All ultrasound data were obtained by various sonographers who did not work for a single center but had substantial gynecologic ultrasound experience.

Recommendations

Therefore, it is recommended that all the predictive models may be used. However, the best option is the IOTA-ADNEX because it has the highest accuracy. However, if IOTA-ADNEX may not be done in a facility or there's no trained sonologist, we may use ROMA or CPH-I. Aside from that, since this is a restrospective study, more prospective studies may be conducted to further determine the diagnostic performance of each model.

The modest number of malignant ovarian masses hampered the subanalysis of the accuracy of the four ovarian malignancy prediction models, even though the study was able to attain its goal sample size. As a result, it is suggested that the sample size be increased even more to better test the accuracy of distinguishing between the various malignancy subgroups.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. United States Cancer Statistics. Available from: https://www.cdc.gov/cancer/ovarian/statistics/. [Last accessed on 2021 Oct 22].
2. Ovarian Cancer. Available from: https://doh.gov.ph/node/361. [Last accessed on 2021 Oct 22].
3. Cancer Today. Available from: https://gco.iarc.fr/today/

online-analysis-multi-bars?v=2020 and mode=cancer and mode_population=countries and population=900 and populations=900 and key=total and sex=2 and cancer=39 and type=0 and statistic=5 and prevalence=0 and population_group=0 and ages_group%5B%5D=0 and ages_group%5B%5D=17 and nb_items=10 and group_cancer=1 and include_nmsc=1 and include_nmsc_other=1 and type_multiple=%257B%2522inc%2522%253Atrue%252C%2522mort%2522%253Atrue%252C%2522prev%2522%253Afalse%257D and orientation=horizontal and type_sort=0 and type_nb_items=%257B%2522top%2522%253Atrue%252C%2522bottom%2522%253Afalse%257D and population_group_globocan_id. [Last accessed on 2021 Oct 22].

4. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, *et al*. Ovarian cancer statistics, 2018. CA Cancer J Clin 2018;68:284-96.
5. Doubeni CA, Doubeni AR, Myers AE. Diagnosis and management of ovarian cancer. Am Fam Phys 2016;93:937-94.
6. American College of Obstetricians and Gynecologists. Evaluation and management of adnexal masses. Pract Bull 2016;174:1-17.
7. 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.
8. Yamamoto Y, Yamada R, Oguri H, Maeda N, Fukaya T. Comparison of four malignancy risk indices in the preoperative evaluation of patients with pelvic masses. Eur J Obstet Gynecol Reprod Biol 2009;144:163-7.
9. Hada A, Han LP, Chen Y, Hu QH, Yuan Y, Liu L. Comparison of the predictive performance of risk of malignancy indexes 1-4, HE4 and risk of malignancy algorithm in the triage of adnexal masses. J Ovarian Res 2020;13:46.
10. Moore RG, McMeekin DS, Brown AK, DiSilvestro P, Miller MC, Allard WJ, *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;112:40-6.
11. Moore RG, Miller MC, Disilvestro P, Landrum LM, Gajewski W, Ball JJ, *et al*. Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass. Obstet Gynecol 2011;118:280-8.
12. Chen X, Zhou H, Chen R, He J, Wang Y, Huang L, *et al*. Development of a multimarker assay for differential diagnosis of benign and malignant pelvic masses. Clin Chim Acta 2015;440:57-63.
13. Karlsen MA, Høgdall EV, Christensen IJ, Borgfeldt C, Kalapotharakos G, Zdrazilova-Dubska L, *et al*. A novel diagnostic index combining HE4, CA125 and age may improve triage of women with suspected ovarian cancer – An international multicenter study in women with an ovarian mass. Gynecol Oncol 2015;138:640-6.
14. Van Calster B, Van Hoorde K, Valentin L, Testa AC, Fischerova D, Van Holsbeke C, *et al*. Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: Prospective multicentre diagnostic study. BMJ 2014;349:g5920.
15. ADNEX Risk Model. Available from: https://www.iotagroup.org/iota-models-software/adnex-risk-model. [Last accessed on 2021 Oct 22].
16. Timmerman D, Planchamp F, Bourne T, Landolfo C, du Bois A, Chiva L, *et al*. ESGO/ISUOG/IOTA/ESGE Consensus Statement on pre-operative diagnosis of ovarian tumors. Int J Gynecol Cancer 2021;31:961-82.
17. Philippine Obstetrical and Gynecological Society (POGS). Clinical Practice Guidelines on Adnexal Masses. 1st ed. Manila, Philippines: Philippine Obstetrical and Gynecological Society (POGS); 2018. p. 18.
18. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform 2014;48:193-204.
19. Wang Z, Tao X, Ying C. CPH-I and HE4 Are More Favorable Than CA125 in differentiating borderline ovarian tumors from epithelial ovarian cancer at early stages. Dis Markers 2019;2019:6241743.
20. Kaijser J, Van Gorp T, Van Hoorde K, Van Holsbeke C, Sayasneh A, Vergote I, *et al*. A comparison between an ultrasound based prediction model (LR2) and the risk of ovarian malignancy algorithm (ROMA) to assess the risk of malignancy in women with an adnexal mass. Gynecol Oncol 2013;129:377-83.
21. Auekitrungrueng R, Tinnangwattana D, Tantipalakorn C, Charoenratana C, Lerthiranwong T, Wanapirak C, *et al*. Comparison of the diagnostic accuracy of International Ovarian Tumor Analysis simple rules and the risk of malignancy index to discriminate between benign and malignant adnexal masses. Int J Gynaecol Obstet 2019;146:364-9.
22. Liest AL, Omran AS, Mikiver R, Rosenberg P, Uppugunduri S. RMI and ROMA are equally effective in discriminating between benign and malignant gynecological tumors: A prospective population-based study. Acta Obstet Gynecol Scand 2019;98:24-33.
23. Yoshida A, Derchain SF, Pitta DR, Andrade LA, Sarian LO. Comparing the Copenhagen Index (CPH-I) and Risk of Ovarian Malignancy Algorithm (ROMA): Two equivalent ways to differentiate malignant from benign ovarian tumors before surgery? Gynecol Oncol 2016;140:481-5.