Original Article

Diagnostic Performance of International Ovarian Tumor Analysis Logistic Regression Model LR2 for Adnexal Masses Classification at a Tertiary Gynecology Center in Singapore

Pamela Partanaa, b, Sook Ling Leea, Wei Ching Tan,a

Abstract

Background: The International Ovarian Tumor Analysis (IOTA) LR2 model has been shown to provide a reasonably accurate preoperative classification of ovarian tumors. The purpose of this study was to evaluate the diagnostic performance of the IOTA LR2 model in distinguishing benign and malignant adnexal masses in the Singapore population.

Methods: This was a retrospective study in a tertiary referral center. Women who attended the Gynecology Unit at Singapore General Hospital with evidence of adnexal tumor on ultrasound examination were evaluated using the IOTA LR2 protocol. The LR2 model was then used to calculate the probability of malignancy. Likelihood ratio of malignancy greater than 10% classifies the mass as malignant. The preoperative diagnosis of women who underwent surgery within 120 days of ultrasound examination was correlated with the final histopathological result.

Results: Of the 353 women included in the final study population, 223 had benign disease, 29 had borderline malignant, and 101 had invasive cancer. The IOTA LR2 model had a sensitivity of 79.2% (95% confidence interval (CI), 71.2-85.8%) and a specificity of 79.4% (95% CI, 73.5-84.5%). The area under the receiver-operating characteristics curve was 0.84 (95% CI, 0.80 - 0.89).

Conclusions: The IOTA LR2 model maintained its overall diagnostic accuracy when used in our local population. Although it is useful as a first-step test for triaging women with ovarian masses for surgery, a second-stage test would be required to minimize the number of women with benign disease being offered surgery for suspected ovarian malignancy.

Keywords: Adnexal mass; IOTA; Logistic regression model; Ovarian cancer; Ultrasonography

Introduction

Accurate preoperative characterization of adnexal masses is crucial and often determines where patients will be best managed, which in turn has significant impact on their disease outcome. Studies have shown that patients with suspected ovarian malignancy who are operated on by a sub-specialist gynecological oncologist at a designated cancer center with multidisciplinary team support have a better prognosis and improved survival compared to that managed in a non-cancer center [1, 2]. On the other hand, adopting a conservative approach to benign ovarian masses helps to avoid morbidity associated with major abdominal surgery. Laparoscopic surgery for benign ovarian masses is associated with less pain, a more rapid recovery, and shorter hospital stay [3, 4].

Differentiation between benign and malignant ovarian tumors preoperatively presents a clinical challenge. At present, subjective assessment of ultrasound findings by examiners who are experts in imaging of adnexal pathology or “pattern recognition” provides the most accurate distinction [5-7]. However, operators performing ultrasound examination have undergone different training programs and are at different levels of experience.

Since 2004, the International Ovarian Tumor Analysis (IOTA) Group has developed several ultrasound-based prediction models to help standardize the way adnexal masses are classified, with the aim of achieving the standard performance of experienced examiners [8]. The logistic regression model LR2 incorporating one clinical variable and five sonographic features showed excellent performance in various prospective and validation studies taking place across centers in Europe, Canada, and China [9]. Its use in this part of the world, however, is limited. The risk of malignancy index (RMI) recommended by the Royal College of Obstetricians and Gynaecologists (RCOG) for triaging women with ovarian mass is a more commonly used protocol amongst gynecologists in Singapore [10]. Recent studies however suggest a significantly better di-
agnostic performance for the IOTA model LR2 over RMI, with fewer borderline tumor and stage 1 invasive ovarian cancer being missed when applied during preoperative ultrasound assess-
ment [11].

To date, our hospital is the only institution in the coun-
try that has adopted the ultrasound-based protocol IOTA lo-
gistic regression model LR2 for preoperative characterization of ovarian masses. The purpose of this study was to evaluate
the performance of this protocol in discriminating benign and
malignant ovarian tumors in Singapore.

Materials and Methods

This was a retrospective single-center study conducted over a
4-year period from December 2014 to December 2018. All the
patients included in the study attended the Gynecology Unit
at Singapore General Hospital, a tertiary gynecological oncol-
ogy referral center. Women with an adnexal mass detected on
ultrasound examination performed in our unit during the study
period were included. Women with an adnexal mass detected
during pregnancy and those who did not undergo surgical re-
moval of the mass within 120 days of the ultrasound exami-
ation were excluded from this study.

Transvaginal ultrasonography was utilized in most cases.
Patients who are virgo intacta underwent transabdominal or
transrectal ultrasonography. In a proportion of patients with
a large mass that cannot be seen in its entirety using a trans-
vaginal approach, transabdominal ultrasonography was used
as an adjunct. In women with more than one adnexal mass, the
lesion with higher risk of malignancy according to the calcu-
lated LR2 score was included in the analysis. All the women
included in this study were examined by sonographers who are
certified members of IOTA and trained in the use of the LR2
protocol. Ethical approval for this study was obtained from the
Singhealth Centralized Institutional Review Board (CIRB Ref.
No. 2020/2819). This study was conducted in compliance with
the ethical standards of the responsible institution on human
subjects as well as with the Helsinki Declaration.

Sonographic evaluation of an adnexal mass was done
in accordance with the IOTA logistic regression model LR2
protocol, as described by Timmerman et al [8]. Women with
ultrasound evidence of adnexal masses underwent systematic
examination of the tumor characteristics using the six varia-
tives described in this protocol: (1) age of the patient, in years;
(2) presence of ascites (yes = 1, no = 0); (3) presence of blood
flow within a papillary projection (yes = 1, no = 0); (4) largest
diameter of the solid component (measured in mm, but with
no increase after > 50 mm); (5) irregular internal cyst walls
(yes = 1, no = 0); (6) presence of acoustic shadows (yes = 1,
no = 0). The logistic regression model LR2 calculates the
likelihood of an adnexal mass being malignant as $1/(1 + e^{-z})$,
where $z = -5.3718 + 0.0354 \times (1) + 1.6159 \times (2) + 1.1768 \times
(3) + 0.0697 \times (4) + 0.9586 \times (5) - 2.9486 \times (6)$. A risk cut-off
of 0.10 (10%) was used to predict masses that are malignant,
consistent with that suggested for clinical use in the original
IOTA study [8].

Histopathology was the primary reference standard used.

Specimens removed during the surgery underwent histo-
logical examination at the Singapore General Hospital Department
of Pathology. Tumors were classified according to the criteria
recommended by the World Health Organization guidelines
and malignancies were staged according to the International
Federation of Gynecology and Obstetrics criteria [12, 13]. Pri-
mary invasive cancer, borderline tumor, and metastatic tumor
were grouped together as malignant disease for the purpose
of data analysis. Approaches to surgery, laparoscopy or lapa-
rotomy were based on surgeon’s judgement.

Statistical analysis

Patients were cross-tabulated based on their histological out-
come (benign or malignant) and their pre-operative malign-
nancy risk (< 10% being benign, ≥ 10% classifies the mass
as malignant). Diagnostic performance measures calculated
in this study were sensitivity, specificity, positive and negative
likelihood ratio. Sensitivity and specificity of the model were
presented with 95% confidence intervals (CIs). Discriminatory
performance was evaluated and reported using receiver-oper-
ating characteristic (ROC) curve and the area under the curve
(AUC). Statistical analysis was performed using the software
R 3.6.2 (https://www.r-project.org).

Results

Between December 2014 and December 2018, 1003 patients
were diagnosed with at least one adnexal mass on ultrasound
examination in our center. Of these, 407 (40.6%) underwent
surgical treatment. Fifty-four patients were excluded, 53 did
not have surgery within 120 days of their ultrasound assess-
ment and one was pregnant (Fig. 1). The median age of the
patients in our population was 54 (mean age 52, range 18 -84)
years. In our final study population of 353 women, 223
(63.2%) had benign tumors while 130 (36.8%) were confirmed
to have malignant tumors. Of the malignant lesions, 29 were
borderline tumors, 87 were primary invasive ovarian carcino-
ma, 11 were metastatic tumors of the ovary, and three were of
non-ovarian malignancy. Details on histological outcomes are
presented in Table 1.

The diagnostic performance of IOTA logistic regression
model LR2 in our center is summarized in Figure 2. The IOTA
logistic regression model LR2 had a sensitivity of 79.2% (95%
CI, 71.2-85.8%) and specificity of 79.4% (95% CI, 73.5-
84.5%) in our study. The positive predictive value was 69.1%
(95% CI, 61.0-76.4%) and the negative predictive value was
86.8% (95% CI, 81.3-91.1%). The AUC of LR2 in our center
was 0.84 (95% CI, 0.80 - 0.89). Using a probability cut-off of
10.0% to predict malignant adnexal masses, LR2 missed 24
cancers (12 borderline tumors, two immature teratomas, and
10 primary invasive ovarian carcinomas) and yielded 60 false
positive diagnoses (Table 2), resulting in positive likelihood
ratio (LR+) of 3.84 (95% CI, 2.93 - 5.04) and negative likely-
hood ratio (LR-) of 0.26 (95% CI, 0.19 - 0.37). All cases of
primary invasive ovarian malignancies that were missed were
of stage I disease, except for one case of stage 3C high-grade serous carcinoma originating from the Fallopian tube.

Specificity in our study was better than in the original study, but it was similar to the specificity in the temporal validation study. Sensitivities in the original IOTA report and prospective validation studies were higher than in our study. The AUC of the current study was lower than that reported in the original IOTA study, the difference of which was statistically significant ($P = 0.0034$) (Table 3).

Discussion

To our knowledge, this is the first reported study evaluating the diagnostic accuracy of the IOTA Group logistic regression model LR2 in Singapore. Our study shows that the IOTA logistic regression model LR2 has excellent diagnostic performance in distinguishing benign and malignant adnexal masses in our center, albeit with a lower AUC value than that reported in the original IOTA study [8, 9]. The specificity in our study is similar to that reported in the original IOTA study. The sensitivity in the original and subsequent validation studies, however, was significantly higher than in our study.

The prevalence of adnexal malignancies in this study population is 36.8% (130 malignancies vs. 223 benign tumors). This is higher than that reported in the original, temporal, and external validation IOTA studies (24%, 30%, and 26%, respectively) [8, 9] and might be explained by the fact that the study was conducted in a single tertiary referral center for gynecological oncology.

All the false-negative cases in our study (24/130 (18.5%)) were contributed by borderline (n = 12) and stage I tumors (n = 11), except for one case of stage IIIC cancer originating from the fallopian tube. No other higher stage disease or metastatic tumor of the ovary was missed. It is well established that borderline and stage I tumors are more difficult to classify, which may explain the lower sensitivity reported in our study compared to that reported in the original study. The other two cancers that were missed were teratomas with predominantly mature solid cystic component but containing focal areas of...
immature teratoma.

The lower sensitivity reported in this study could partly be explained by the level of familiarity our operators with using the IOTA logistic regression model LR2. A prospective study evaluating the performance of the IOTA logistic regression model LR2 in the hands of a non-expert ultrasound operator showed that the experience of the level II examiner increases with time and the performance of the model was better in the second half of the study [14]. This current study reports the performance of the model during its first few years of implementation in our center during which the operators may be at the beginning of the learning curve. We believe that with increased familiarity, continual standardized training, and regular audit, the use of accurate definitions, terminology, and measurements will be improved resulting in better pre-operative differentiation between benign and malignant adnexal masses.

Table 1. Histological Diagnosis of Ovarian Masses Included in the Study (n = 353)

<table>
<thead>
<tr>
<th>Histological diagnosis</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign</td>
<td>223 (63.2)</td>
</tr>
<tr>
<td>Mature cystic teratoma</td>
<td>56 (15.9)</td>
</tr>
<tr>
<td>Fibroma</td>
<td>6 (1.7)</td>
</tr>
<tr>
<td>Endometrioma</td>
<td>30 (8.5)</td>
</tr>
<tr>
<td>Cystadenoma (serous, mucinous, seromucinous)</td>
<td>74 (21.0)</td>
</tr>
<tr>
<td>Cystadenofibroma (serous, mucinous, seromucinous)</td>
<td>15 (4.2)</td>
</tr>
<tr>
<td>Hemorrhagic cyst</td>
<td>5 (1.4)</td>
</tr>
<tr>
<td>Simple ovarian cyst</td>
<td>2 (0.6)</td>
</tr>
<tr>
<td>Hydrosalpinx</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>Tubo-ovarian abscess</td>
<td>5 (1.4)</td>
</tr>
<tr>
<td>Paraovarian/paratubal cyst</td>
<td>6 (1.7)</td>
</tr>
<tr>
<td>Functional cyst</td>
<td>8 (2.3)</td>
</tr>
<tr>
<td>Fibrothecoma</td>
<td>3 (0.8)</td>
</tr>
<tr>
<td>Rare benign tumors&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5 (1.4)</td>
</tr>
<tr>
<td>Peritoneal inclusion cyst</td>
<td>3 (0.8)</td>
</tr>
<tr>
<td>Borderline</td>
<td>29 (8.2)</td>
</tr>
<tr>
<td>Mucinous</td>
<td>19 (5.4)</td>
</tr>
<tr>
<td>Serous</td>
<td>6 (1.7)</td>
</tr>
<tr>
<td>Seromucinous</td>
<td>3 (0.8)</td>
</tr>
<tr>
<td>Endometrioid</td>
<td>1 (0.2)</td>
</tr>
<tr>
<td>Primary invasive ovarian carcinoma</td>
<td>87 (24.6)</td>
</tr>
<tr>
<td>Epithelial carcinoma</td>
<td></td>
</tr>
<tr>
<td>High-grade serous carcinoma</td>
<td>19 (5.4)</td>
</tr>
<tr>
<td>Endometrioid carcinoma</td>
<td>26 (7.4)</td>
</tr>
<tr>
<td>Clear cell carcinoma</td>
<td>18 (5.1)</td>
</tr>
<tr>
<td>Mucinous carcinoma</td>
<td>14 (4.0)</td>
</tr>
<tr>
<td>Low-grade serous carcinoma</td>
<td>3 (0.8)</td>
</tr>
<tr>
<td>Germ cell tumor</td>
<td>4 (1.1)</td>
</tr>
<tr>
<td>Immature teratoma</td>
<td></td>
</tr>
<tr>
<td>Sex cord stromal tumor</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>Adult granulosa cell tumor</td>
<td></td>
</tr>
<tr>
<td>Carcinosarcoma</td>
<td>2 (0.6)</td>
</tr>
<tr>
<td>Primary uterine carcinoma&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3 (0.8)</td>
</tr>
<tr>
<td>Metastatic</td>
<td>11 (3.1)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Two cases of mesothelial inclusion cyst, one Brenner tumor, one right ovarian leiomyoma, and one adenomatoid tumor of the ovary. <sup>b</sup>Carcinosarcoma (n = 1), leiomyosarcoma (n = 1), Mullerian adenosarcoma (n = 1).
One-third of the false positive cases in our study (60/223 (26.9%)) were contributed by benign cystic teratomas (n = 16) and stromal tumors (fibroma, fibrothecoma (n = 4). This might be due to misinterpretation of the solid component within these benign entities leading to inaccuracy in entering the ultrasound variable to this prediction model. The positive predictive value was 69.1%. This means that a third of benign lesions were incorrectly classified as malignant and if decision to intervene is determined by LR2 score alone, a significant proportion of asymptomatic women with benign tumor will undergo unnecessary surgery for suspected cancer. The relatively low specificity when using this model has also been reported in other studies. Nunes et al reported LR2 specificity of 76.7% (95% CI, 71.9-81.0%) for LR2 in diagnosing malignancy in their study [15]. They concluded that IOTA models are useful as a first-stage screening test to diagnose ovarian cancer, but an additional second-stage diagnostic tool is needed to decrease the numbers of false positive.

In conclusion, our study suggests that IOTA logistic regression model LR2 maintained its excellent discriminatory performance in classifying benign and malignant ovarian tumors in our population. It is an excellent first-stage test that can be reliably used by non-expert examiners of average ability and experience in gynecological ultrasound, but a secondary test would be required in those suspected to have ovarian cancer based on this model to minimize the number of false positive findings. Dynamic contrast-enhanced, diffusion-weighted pelvic magnetic resonance imaging (MRI) shows promising potential as a second line tool after ultrasoundography and we look forward to the result of the ongoing IOTA-MRI study (Clinical-Trials.gov NCT02836275) [16]. For the time being, pattern recognition by an expert (level-III) ultrasound examiner would likely be most helpful as a secondary test in women with suspected malignancy on the LR2 model.

**Table 2.** Histological Outcomes in Women With a False-Positive Finding Using IOTA Model LR2 (n = 60)

<table>
<thead>
<tr>
<th>Histological diagnosis</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cystadenofibroma</td>
<td>4 (6.7)</td>
</tr>
<tr>
<td>Cystadenoma</td>
<td>16 (26.7)</td>
</tr>
<tr>
<td>Endometrioma</td>
<td>7 (11.7)</td>
</tr>
<tr>
<td>Fibroma</td>
<td>3 (5.0)</td>
</tr>
<tr>
<td>Fibrothecoma</td>
<td>1 (1.7)</td>
</tr>
<tr>
<td>Functional cyst</td>
<td>5 (8.3)</td>
</tr>
<tr>
<td>Hemorrhagic cyst</td>
<td>3 (5.0)</td>
</tr>
<tr>
<td>Broad ligament leiomyoma</td>
<td>2 (3.3)</td>
</tr>
<tr>
<td>Peritoneal inclusion cyst</td>
<td>1 (1.7)</td>
</tr>
<tr>
<td>Rare benign</td>
<td>1 (1.7)</td>
</tr>
<tr>
<td>Teratoma</td>
<td>16 (26.7)</td>
</tr>
<tr>
<td>Tubo-ovarian abscess</td>
<td>1 (1.7)</td>
</tr>
</tbody>
</table>

IOTA: International Ovarian Tumor Analysis.

**Acknowledgments**

We would like to thank our statistician Dr. Li HuiHua of Singhealth Health Services Research Unit for reviewing the statistics of the study.

**Financial Disclosure**

None to declare.

**Conflict of Interest**

We have no conflict of interest to disclose.

**Informed Consent**

Waiver of informed consent was obtained from Singhealth Centralized Institutional Review Board.

**Author Contributions**

TWC contributed to the conception and design of the study. PP was involved in data acquisition, data analysis, data interpretation, and drafting the manuscript. LSL and TWC had full access to all the data in the study, provided critical revision to the manuscript, and gave the final approval of the version of the manuscript to be published.
IOTA LR2 for Adnexal Mass Classification

Data Availability

Any inquiries regarding supporting data availability of this study should be directed to the corresponding author.

References


Table 3. Comparison of the Results of Diagnostic Performance of IOTA Logistic Regression Model LR2 in Our Study With the Original and Validation IOTA Studies (Temporal and External Validation Studies) [8, 9]

<table>
<thead>
<tr>
<th>Study</th>
<th>AUC (95% CI)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>LR+</th>
<th>LR-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current study (n = 353)</td>
<td>0.84 (0.80 - 0.89)</td>
<td>79.2</td>
<td>79.4</td>
<td>3.84</td>
<td>0.26</td>
</tr>
<tr>
<td>Original IOTA study (n = 312) [8]</td>
<td>0.92</td>
<td>89.0</td>
<td>73.0</td>
<td>3.3</td>
<td>0.15</td>
</tr>
<tr>
<td>Temporal validation study (n = 941) [9]</td>
<td>0.92 (0.90 - 0.94)</td>
<td>89.2</td>
<td>79.8</td>
<td>4.4</td>
<td>0.14</td>
</tr>
<tr>
<td>External validation study (n = 997) [9]</td>
<td>0.95 (0.93 - 0.96)</td>
<td>91.8</td>
<td>85.6</td>
<td>6.36</td>
<td>0.10</td>
</tr>
</tbody>
</table>

IOTA: International Ovarian Tumor Analysis; AUC: area under receiver-operating characteristics; CI: confidence interval; LR-: negative likelihood ratio; LR+: positive likelihood ratio.