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

Rapeepat Auekitrungrueng | Dangcheewan Tinnangwattana | Charuwan Tantipalakorn | Cholaros Charoenratana | Thitikarn Lerthiranwong | Chanane Wanapirak | Theera Tongsong*

Department of Obstetrics and Gynecology, Chiang Mai University, Chiang Mai, Thailand

*Correspondence
Theera Tongsong, Department of Obstetrics and Gynecology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand.
Email: theera.t@cmu.ac.th

Funding Information
Chiang Mai University Research Fund

Abstract

Objective: To compare the diagnostic accuracy of International Ovarian Tumor Analysis (IOTA) simple rules and risk of malignancy index (RMI 1/RMI 2) scoring to discriminate between benign and malignant adnexal masses.

Methods: Secondary analysis of a cohort of patients scheduled for surgery for adnexal masses in a tertiary center between April 2010 and March 2018. Ultrasound examinations were performed by general gynecologists within 24 hours prior to surgery to evaluate sonographic features. Demographic data and preoperative CA 125 levels were recorded. IOTA rules and RMI scoring were applied to predict malignancy and prospectively recorded. Final diagnosis was based on pathological or intraoperative diagnosis.

Results: A total of 479 masses met the inclusion criteria and were retrieved from the database: 334 (69.7%) benign and 145 (30.3%) malignant. IOTA rules could be applied to 392 (81.8%) masses and were inconclusive in 87 (18.2%). Sensitivity and specificity of IOTA rules (83.8% and 92.0%, respectively) were significantly higher than RMI 1 (77.2% and 86.8%, respectively) and RMI 2 (82.1% and 82.6%, respectively).

Conclusion: IOTA simple rules had higher diagnostic accuracy compared with RMI to discriminate between benign and malignant adnexal masses; however, nearly 20% of IOTA results were inconclusive and needed expert consultation.

KEYWORDS
Benign adnexal mass; International Ovarian Tumor Analysis (IOTA) simple rules; Malignant adnexal mass; Risk of malignancy index (RMI); Ultrasound

1 | INTRODUCTION

Preoperative differentiation between malignant and benign adnexal masses is essential for proper management and counseling of patients as approaches to treat the two conditions are often different. Functional ovarian cysts are usually treated with expectant management and benign ovarian tumor-like dermoid cysts or serous cysts may need simple cystectomy via either laparoscopic surgery or laparotomy that can be performed by general surgeons; in contrast, ovarian cancers require extensive surgery or complete surgical staging, which typically requires expert consultation or referral to a tertiary center employing gynecologic oncologists.

Several systems have been proposed and developed to differentiate between malignant and benign adnexal masses, including clinical...
parameters, tumor markers, and pelvic ultrasound. One of the most common methods used in low-income countries is the risk of malignancy index (RMI), which has been in use for many years.

RMI is a scoring system for the combination of various clinical characteristics. It was developed to improve diagnostic accuracy in predicting ovarian malignancy. Originally, the RMI system was developed by Jacob et al., based on the combination of sonographic findings, menopausal status, and serum levels of CA 125. RMI has a sensitivity and specificity of 85.4% and 96.9%, respectively. Subsequently, Tinguulstad et al. developed RMI 2 to improve diagnostic accuracy. They showed that RMI 2 was superior to RMI 1 in predicting malignancy, with sensitivity of 80% and specificity of 92% compared with RMI 1 that had sensitivity of 71% and specificity of 96%.

In 2008, Timmerman et al. proposed the International Ovarian Tumor Analysis (IOTA) Simple Rules to differentiate benign from malignant ovarian masses. The IOTA rules were created for clinicians to assist in the identification of ultrasound features suggesting malignancy (M-features) or benignity (B-features). We have found that, in our population, the IOTA rules can be used effectively by nonspecialist sonographers. Recently, the IOTA rules have increased in popularity. Our prospective database of preoperative ultrasound adnexal mass images has been ongoing for more than eight years. However, performance of the IOTA simple rules has rarely been compared with other systems, such as RMI, in the same study group.

The aim of the present study was to compare the diagnostic accuracy of IOTA simple rules and RMI scoring to discriminate between benign and malignant adnexal masses on ultrasound performed by general gynecologists.

2 | MATERIALS AND METHODS

The present study was a secondary analysis of our prospective database project ongoing at Maharaj Nakorn Chiang Mai Hospital, Chiang Mai University, with ethical approval from the institutional review board of Chiang Mai University (OBG-2561-05364). The database includes the records of patients admitted for elective pelvic surgery for adnexal masses between April 2010 and March 2018. All patients were counseled and provided written informed consent to join the project.

Recruited patients met the following inclusion criteria: (1) diagnosis of an adnexal mass either by previous pelvic ultrasound examination or by vaginal examination; and (2) no known diagnosis of the ovarian mass prior to surgery, either by prior diagnostic laparoscopy or previous history of pelvic surgery. Participants were not recruited consecutively, but instead based on the availability of the research team.

All recruited participants underwent either transabdominal or transvaginal ultrasound examination, or both, within 24 hours prior to surgery, using a real-time 5–7.5 MHz transvaginal or 3.5–5 MHz transabdominal curvilinear transducer connected to an Aloka model SSD alpha-10 (Aloka, Tokyo, Japan), a Voluson E8 (GE Medical Systems, Zipf, Austria), or a real-time ProSound F37 (Hitachi-Aloka Medical Ltd, Inc., Tokyo, Japan). Ultrasound examinations were performed by the authors (certified gynecologists) and second- and third-year residents who had undertaken a two-week training course in ultrasound examination of adnexal masses, trained by an experienced sonographer (TT). Those performing the imaging were blinded to the patients’ clinical information.

On ultrasound examination, sonographic morphology of the adnexal masses was characterized by 2D real-time and color Doppler ultrasound. Demographic data of the patients such as age, tumor marker levels, if available, and sonographic features of the adnexal masses used in the IOTA simple rules and RMI scoring were prospectively recorded in the research forms and stored in the computerized database. The IOTA simple rules to characterize whether the features were benign (B) or malignant (M) were based on the descriptions proposed by Timmerman et al., as presented in Table 1 and Figure 1. If one or more M-features applied in the absence of a B-feature, the mass was categorized as malignant. If one or more B-features applied in the absence of an M-feature, the mass was categorized as benign. If both M-rules and B-features applied or no rule applied, the mass was categorized as inconclusive.

For RMI scoring, ultrasound features of the mass, menopausal status, and serum CA 125 levels were incorporated to score the mass. The ultrasound findings of solid area, multilocular cyst, bilateral masses, ascites, and intra-abdominal metastases were scored 1 point for each. A total ultrasound score (U) was calculated for each patient. Postmenopausal status (M) was defined as one year or more of missed menstruation; if this did not apply, the alternative status was classified as premenopausal. RMI 1 and RMI 2 were calculated for all patients, using a cut-off level of 200 for predicting malignancy.

RMI 1 scoring, described by Jacobs et al., was calculated as follows: scores = U x M x serum CA 125, where a total ultrasound score of 0 gave U=0, a score of 1 gave U=1, and a score of

\[ \text{Table 1} \] International Ovarian Tumor Analysis (IOTA) simple rules for identifying benign or malignant tumors.

<table>
<thead>
<tr>
<th>Features typical for benign and malignant tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Malignant tumor (M-features)</strong></td>
</tr>
<tr>
<td>M1</td>
</tr>
<tr>
<td>M2</td>
</tr>
<tr>
<td>M3</td>
</tr>
<tr>
<td>M4</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>M5</td>
</tr>
<tr>
<td><strong>Benign tumor (B-features)</strong></td>
</tr>
<tr>
<td>B1</td>
</tr>
<tr>
<td>B2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>B3</td>
</tr>
<tr>
<td>B4</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>B5</td>
</tr>
</tbody>
</table>

\[ \text{a} \text{Source: Timmerman et al.}^{3} \]
FIGURE 1  International Ovarian Tumor Analysis (IOTA) simple rules for classifying benign tumors (B-features: B1–B5) and malignant tumors (M-features: M1–M5) on ultrasound (see Table 1 for descriptions of each feature).
greater than or equal to 2 gave $U=3$; premenopausal status gave $M=1$, postmenopausal status gave $M=3$; and serum CA 125 level was multiplied directly into the formula. RMI 2 scoring, described by Tingulstad et al.,\textsuperscript{2} was calculated as follows: scores = $U \times M \times \text{serum CA 125}$, where a total ultrasound score of 0 gave $U=0$, a score of 1 gave $U=1$, and a score of greater than or equal to 2 gave $U=4$; premenopausal status gave $M=1$, postmenopausal status gave $M=4$; and serum CA 125 level was the absolute measured CA 125 level (U/mL) of each patient.

The definite diagnoses of the adnexal masses, used as a gold standard, were based on pathological reports or intraoperative diagnosis made by the surgeons for cases of some benign masses without pathological specimens. All masses were classified into two groups: benign or malignant. Masses with pathological diagnosis of borderline tumors were classified as malignant. Patients who underwent surgery beyond 24 hours after ultrasound examination and patients with no result of CA 125 levels prior to surgery were excluded.

The diagnostic accuracy of the IOTA simple rules, RMI 1, and RMI 2 was calculated for sensitivity, specificity, and positive/negative predictive values. A comparison of the accuracy to predict benign or malignant masses between the IOTA and RMI methods was done using the McNemar test. Statistical analyses were performed using SPSS version 21.0 (IBM, Armonk, NY, USA). $P<0.05$ was considered statistically significant.

3 | RESULTS

During the eight-year study period, information on 721 masses was collected and 242 masses were excluded (Fig. 2). A total of 479 adnexal masses from 442 women met the inclusion criteria and were included in the analysis. The mean age of the patients was $42.1 \pm 12.5$ years (range, 12–80 years). Most of the masses were from women of premenopausal age ($n=364, 76.0\%$), whereas $24\%$ ($n=115$) were from postmenopausal women. Almost half of the women ($n=212, 48.0\%$) were nulliparous.

Of 479 adnexal masses, 334 (69.7\%) were pathologically benign and 145 (30.3\%) were malignant, including 130 (27.1\%) cases of cancer and 15 (3.1\%) masses of low malignant potential. Endometrioma was the most common adnexal mass, accounting for 22.1\% ($n=106$) of all masses. The pathological diagnoses of all ovarian tumors are presented in Table 2 under subgroups. IOTA simple rules could be applied to 392 (81.8\%) masses and were inconclusive for 87 (18.2\%), whereas RMI 1 and RMI 2 could be applied in all cases (100\%). Of the inconclusive results by IOTA rules, 28 (19.3\% of malignant masses) and 59 (17.7\% of benign masses) masses were finally diagnosed as malignant and benign, respectively.

The sensitivity and specificity of IOTA rules (83.8\% and 92.0\%, respectively) were significantly higher than RMI 1 (77.2\% and 86.8\%; $P=0.013$ and $P=0.006$, respectively) and RMI 2 (82.1\% and 82.6\%; $P=0.013$ and $P=0.006$, respectively).
P=0.065 and P=0.011, respectively) (Table 3). The accuracy of RMI 1 and RMI 2 was not significantly different (P=0.189).

### 4 | DISCUSSION

The main purpose of our study was to identify an accurate, simple, inexpensive, and practical method for low-income countries where resources are limited. Our findings showed that, in the hands of non-specialists, the IOTA simple rules were more effective than RMI 1 and RMI 2 scoring at discriminating between benign and malignant adnexal masses. This finding was consistent with previous studies, although it may not be as effective as has been reported by the pioneer investigators. Because of its high diagnostic performance, simplicity, and ease to learn with a short training course or practice under supervision, our study suggests that the IOTA rules should be used instead of RMI, which is traditionally used in Thailand. However, the IOTA rules had a relatively high rate of inconclusive results, which must be taken into consideration for its wider use.

To date, many studies of the IOTA simple rules have been published. However, most studies are confined to a few groups of investigators, and the method is rarely tested by other groups for external validation, especially in other regions of the world. Our team has been studying the IOTA rules for more than eight years and has found the method to be reproducible in our population of Thai women, even among non-specialist examiners, although performance may not be as high as reported by the pioneers.

Few studies have compared the diagnostic performance of the IOTA simple rules with other conventional techniques, particularly RMI or the risk of ovarian malignancy algorithm (ROMA). In addition, most ultrasound examiners in those studies were experienced sonographers, which would not represent examiners in practice in low-income countries where resources are limited. Accordingly, studies comparing diagnostic performance between the new method (IOTA) and RMI or ROMA still need to be conducted to accumulate enough data to evaluate their suitability before proper implementation. Our results were consistent with Sayasneh et al., who showed that the IOTA rules gave a better result than RMI. Likewise, Kaijser et al. showed that an ultrasound-based prediction model (developed by IOTA) had better prediction than ROMA for the characterization of a pelvic mass in both pre- and postmenopausal women.

Although the IOTA simple rules yielded high diagnostic accuracy, approximately 20% of examinations were inconclusive, whereas there were no inconclusive results when evaluated by RMI scoring; this finding was consistent with other studies. Inconclusive results are the main disadvantage of the IOTA simple rules. This indicates that specialist sonographers would need to be consulted in a significant number of cases. This disadvantage may raise concern for its wider use.

However, a disadvantage of RMI 1/RMI 2 is that this scoring needs to measure tumor marker levels. This must be considered for clinical use. We suggest that, for low-income countries with limited resources, the IOTA simple rules should be the first-line method adopted by general gynecologists to differentiate between benign and malignant adnexal masses since it requires only ultrasound examination and has high accuracy. However, in cases of inconclusive results, consultation from a specialist sonographer should be sought; if the mass is considered to be malignant, consultation with an oncologist or referral to a specialist center is recommended.

The strengths of the study included that comparisons of all techniques were performed on the same patients and the same ultrasound scans, which allowed ideal comparison in terms of patient characteristics, examiners, ultrasound scanners, and the same settings for ultrasound examinations. In addition, the sample size was large enough to gain power to show a small but significant difference in performance of the three methods. Furthermore, the ultrasound examiners identifying sonographic features using IOTA simple rules and RMI could represent clinicians in real practice because they were on the residency training program and not specialists in gynecologic ultrasound. This suggests that the high effectiveness of the IOTA rules, as well as RMI, would likely be reproducible when widely used by non-specialist clinicians.

The weaknesses of this study included that ultrasound examinations were performed by general gynecologists including second- and third-year residents, who were at different levels in their skills development; as such, they might not accurately represent clinicians in real practice.

### TABLE 3 Diagnostic indices of subjective assessment in predicting malignant ovarian masses.

<table>
<thead>
<tr>
<th>IOTA rules</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>PPV, % (95% CI)</th>
<th>NPV, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malignant</td>
<td>83.8 (77.1–90.4)</td>
<td>2.0 (88.8–95.2)</td>
<td>81.7 (74.7–88.6)</td>
<td>93.0 (88.5–97.6)</td>
</tr>
<tr>
<td>Benign</td>
<td>19</td>
<td>253</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMI 1</td>
<td>77.2 (70.4–84.1)</td>
<td>86.8 (83.2–90.5)</td>
<td>71.8 (64.7–78.9)</td>
<td>89.8 (85.0–94.5)</td>
</tr>
<tr>
<td>Malignant</td>
<td>82.1 (75.8–88.3)</td>
<td>82.6 (78.6–86.7)</td>
<td>67.2 (60.3–74.1)</td>
<td>91.4 (87.3–95.5)</td>
</tr>
<tr>
<td>Benign</td>
<td>26</td>
<td>276</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: NPV, negative predictive value; PPV, positive predictive value; RMI, risk of malignancy index.
In conclusion, the IOTA simple rules had higher diagnostic accuracy compared with RMI to discriminate between benign and malignant adnexal masses, although approximately 20% of IOTA results were inconclusive and needed expert consultation, referral to a center where oncologists were available, or RMI assessment. Owing to its high effectiveness, simplicity, and lack of need for specialist sonographers or measurement of tumor markers, the IOTA rules should be applied in daily practice by general gynecologists. That said, when the results are inconclusive, expert consultation is needed.

AUTHOR CONTRIBUTIONS

RA was responsible for conceptualization, design of the research, data validation, drafting, and revising the work. DT and CW were responsible for conceptualization, data curation, and final approval of the manuscript. CT contributed to the conceptualization, manuscript revision, and final approval. CC and TL were responsible for data curation, manuscript revision, and final approval. TT contributed to conceptualization, design of the research, analysis, and manuscript revision.

ACKNOWLEDGMENTS

We wish to thank the Chiang Mai University Research Fund (CMU-2562), for financial support.

CONFLICTS OF INTEREST

The authors have no conflicts of interest.

REFERENCES