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

Pamela Partana[a, b], Sook Ling Lee[a], Wei Ching Tan[a]

Abstract

Background: The International Ovarian Tumor Analysis (IOTA) LR2 model has been shown to provide a reasonably accurate pre-operative 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.

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 assessment [11].

To date, our hospital is the only institution in the country that has adopted the ultrasound-based protocol IOTA logistic 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 oncology 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 removal of the mass within 120 days of the ultrasound examination 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 transvaginal 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 calculated 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 variables 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 histological 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]. Primary invasive cancer, borderline tumor, and metastatic tumor were grouped together as malignant disease for the purpose of data analysis. Approaches to surgery, laparoscopy or laparotomy were based on surgeon's judgement.

Statistical analysis

Patients were cross-tabulated based on their histological outcome (benign or malignant) and their pre-operative malignancy 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-operating 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 assessment 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 carcinoma, 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 likelihood ratio (LR-) of 0.26 (95% CI, 0.19 - 0.37). All cases of primary invasive ovarian malignancies that were missed were
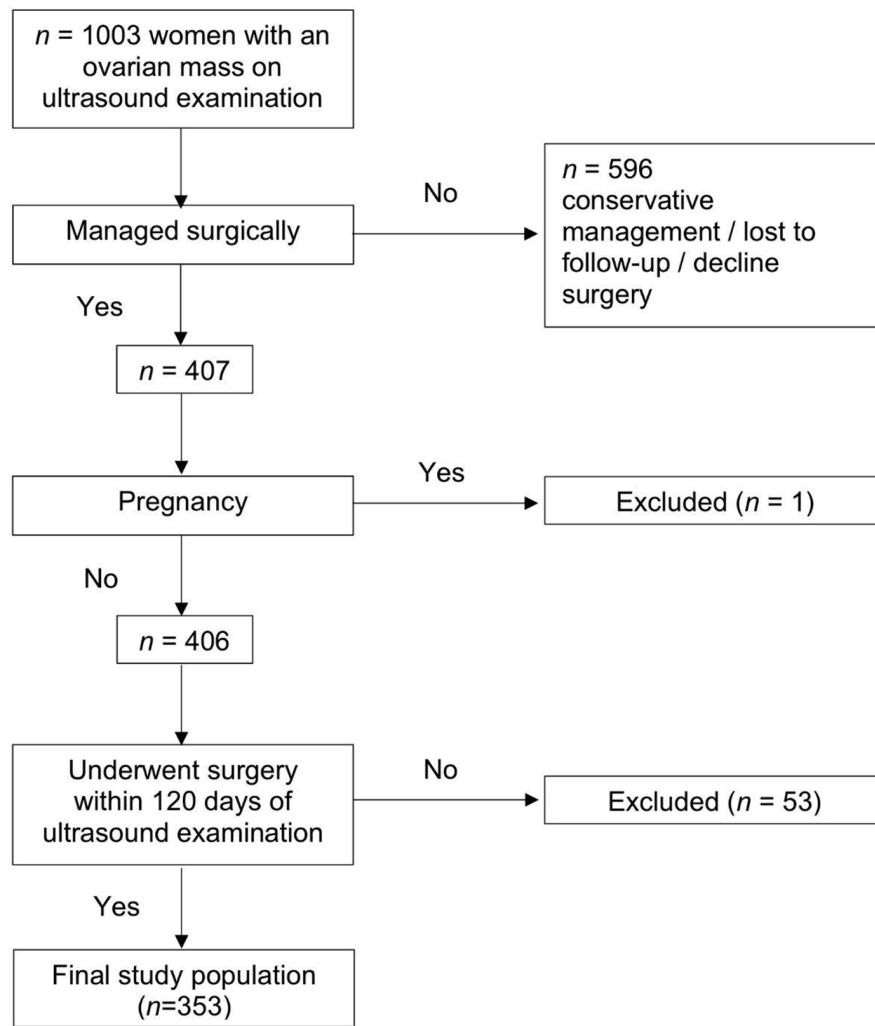
Figure 1. Flowchart of patient numbers.

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

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

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

[a]Two cases of mesothelial inclusion cyst, one Brenner tumor, one right ovarian leiomyoma, and one adenomatoid tumor of the ovary. [b]Carcinosarcoma (n = 1), leiomyosarcoma (n = 1), Mullerian adenosarcoma (n = 1).

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.
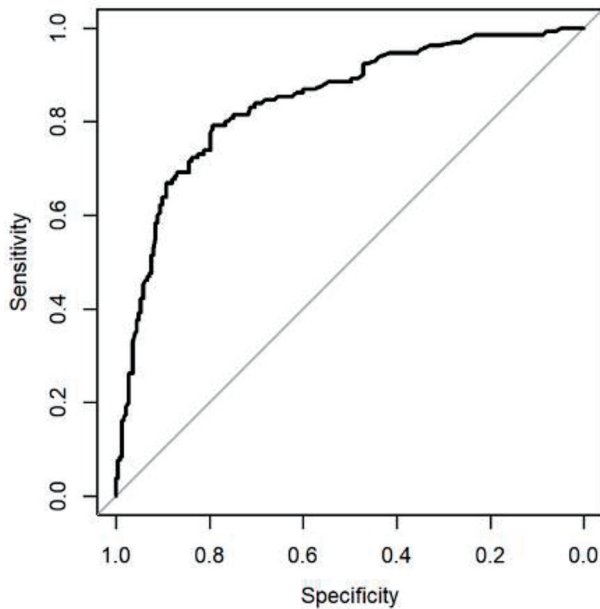
Figure 2. Receiver-operating characteristics (ROC) curve, showing the ability of the International Ovarian Tumor Analysis (IOTA) logistic regression model LR2 in predicting the presence of ovarian malignancy in 353 women with an adnexal mass. Area under the curve was 0.84.

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

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

IOTA: International Ovarian Tumor Analysis.

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 ultrasonography 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.

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.

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]

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

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

Data Availability

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

References

1. Earle CC, Schrag D, Neville BA, Yabroff KR, Topor M, Fahey A, Trimble EL, et al. Effect of surgeon specialty on processes of care and outcomes for ovarian cancer patients. J Natl Cancer Inst. 2006;98(3):172-180.
2. Engelen MJ, van der Zee AG, de Vries EG, Willemse PH. Debulking surgery for ovarian epithelial cancer performed by a gynaecological oncologist improved survival compared with less specialised surgeons. Cancer Treat Rev. 2006;32(4):320-323.
3. Medeiros LR, Stein AT, Fachel J, Garry R, Furness S. Laparoscopy versus laparotomy for benign ovarian tumor: a systematic review and meta-analysis. Int J Gynecol Cancer. 2008;18(3):387-399.
4. Carley ME, Klingele CJ, Gebhart JB, Webb MJ, Wilson TO. Laparoscopy versus laparotomy in the management of benign unilateral adnexal masses. J Am Assoc Gynecol Laparosc. 2002;9(3):321-326.
5. Timmerman D, Schwarzler P, Collins WP, Claerhout F, Coenen M, Amant F, Vergote I, et al. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. Ultrasound Obstet Gynecol. 1999;13(1):11-16.
6. Timmerman D. The use of mathematical models to evaluate pelvic masses; can they beat an expert operator? Best Pract Res Clin Obstet Gynaecol. 2004;18(1):91-104.
7. Valentin L, Jurkovic D, Van Calster B, Testa A, Van Holsbeke C, Bourne T, Vergote I, et al. Adding a single CA 125 measurement to ultrasound imaging performed by an experienced examiner does not improve preoperative discrimination between benign and malignant adnexal masses. Ultrasound Obstet Gynecol. 2009;34(3):345-354.
8. Timmerman D, Testa AC, Bourne T, Ferrazzi E, Ameye L, Konstantinovic ML, Van Calster B, et al. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group. J Clin Oncol. 2005;23(34):8794-8801.
9. Timmerman D, Van Calster B, Testa AC, Guerriero S, Fischerova D, Lissoni AA, Van Holsbeke C, et al. Ovarian cancer prediction in adnexal masses using ultrasound-based logistic regression models: a temporal and external validation study by the IOTA group. Ultrasound Obstet Gynecol. 2010;36(2):226-234.
10. Royal College of Obstetricians and Gynaecologists. The management of ovarian cysts in postmenopausal women. RCOG Green-top Guideline No. 34. London: RCOG. 2016.
11. Van Holsbeke C, Van Calster B, Bourne T, Ajossa S, Testa AC, Guerriero S, Fruscio R, et al. External validation of diagnostic models to estimate the risk of malignancy in adnexal masses. Clin Cancer Res. 2012;18(3):815-825.
12. Serov SF, Scully RE, Sobin LH, World Health Organization. Histological typing of ovarian tumours (WHO International Histological Classification of Tumours No. 9). World Health Organization: Geneva, Switzerland. 1973.
13. Prat J, Oncology FCoG. Staging classification for cancer of the ovary, fallopian tube, and peritoneum. Int J Gynaecol Obstet. 2014;124(1):1-5.
14. Nunes N, Yazbek J, Ambler G, Hoo W, Naftalin J, Jurkovic D. Prospective evaluation of the IOTA logistic regression model LR2 for the diagnosis of ovarian cancer. Ultrasound Obstet Gynecol. 2012;40(3):355-359.
15. Nunes N, Ambler G, Foo X, Widschwendter M, Jurkovic D. Prospective evaluation of IOTA logistic regression models LR1 and LR2 in comparison with subjective pattern recognition for diagnosis of ovarian cancer in an outpatient setting. Ultrasound Obstet Gynecol. 2018;51(6):829-835.
16. Timmerman D, Planchamp F, Bourne T, Landolfo C, du Bois A, Chiva L, Cibula D, et al. ESGO/ISUOG/IOTA/ESGE Consensus Statement on preoperative diagnosis of ovarian tumors. Ultrasound Obstet Gynecol. 2021;58(1):148-168.