GI-RADS CLASSIFICATION AND IOTA CLASSIFICATION: COMPLEMENTARY TO ADNEXAL MASS DIAGNOSES?

Katiuska Ríos Calderón*

Ultrasound is, by excellence, the method of choice to evaluate the benign or malignant nature of the adnexal masses. This assessment is interpreted by the operator to generate a report that suggests a probable diagnosis that may be useful to the treating physician to make decisions. Accordingly, different working groups have attempted to establish standardized terminologies, in order to create reproducible ultrasound nomenclature in specific therapeutic behaviors. They have been raised based on the subjective impression of the examiner scoring systems, logistic regression models and scales.

In 1999 Timmerman and his team at the University Hospital of Leuven in Belgium created the International Ovarian Tumor Analysis (IOTA), with the aim of developing a tool to be used by non-expert sonographers for the presumptive diagnosis of ovarian cancer. In 2000 the IOTA Group published a consensus statement on definitions to describe the ultrasound characteristics of adnexal tumors, developing later mathematical models of prediction. IOTA developed the so-called Simple Rules and mathematical models based on logistic regression to estimate the eventual risk of malignancy. These models were validated prospectively and externally, demonstrating a good correlation very close to the subjective evaluation performed by expert sonographers. In 2014, the IOTA group published the ADNEX model, a multiclass predictive model, capable of differentiating between four subgroups of malignant tumors.

The consensus proposed by the IOTA group has provided very useful information and tools in the morphological description of tumoral adnexal masses. However, false positive rates of up to 24% have
been reported, which could be due to operator experience or difficulty to effectively transmit information from the findings to the clinician\textsuperscript{7}.

Looking for a more accurate diagnostic approach, the initiative of Amor F and collaborators in 2007 published a document with the proposal Gynecology Imaging Reporting and data System (GI-RADS) \textsuperscript{8}, which suggests a systematization of ultrasound language consisting of a nomenclature with a system of classification according to the images, based on the same systematization of the information proposed by the American College of Radiologists for the interpretation of images in mammography research under the BI-RADS system. This is how GI-RADS arises, which includes for its morphological evaluation the description provided by IOTA and the analysis of recognition patterns\textsuperscript{3,8}.

The suggested scale for GIRADS was:

- **GI-RADS 1**: Normal ovaries identified, no adnexal masses were observed. Diagnosis: Definitely benign. Estimated Probability of Malignancy (EPM): 0%.
- **GI-RADS 2**: Adjuvant lesions of probable functional origin: follicles, corpus luteum, hemorrhagic cyst. Diagnosis: Most likely benign. EPM: < 1%
- **GI-RADS 3**: Neoplastic adnexal lesions that are considered benign, such as endometrioma, teratoma, simple cyst, hydrosalpinx, paraovarian cyst, peritoneal pseudocyst, pedicled myoma, or findings suggestive of pelvic inflammatory disease. Diagnosis: probably benign. EPM 1-5%.
- **GI-RADS 4**: Any adnexal lesion not included in GI-RADS 1 to 3 and with one or two findings suggestive of malignancy such as thick papillary projections, thick septa, solid areas and / or ascites, vascularization within solid areas, defined according to IOTA criteria. Diagnosis: Suspected malignancy. EPM 5-20%
- **GI-RADS 5**: Adnexal mass with 3 or more features suggestive of malignancy. Diagnosis: High probability of malignancy. EPM > 20%

The system described above, is based on the subjective characterization of the adnexal images, by a trained operator, a method currently validated by several publications\textsuperscript{3,4,9} and that surpasses various scoring systems.

After presenting its document, this team publishes in 2009 the first report of this system with a group of 171 patients\textsuperscript{3}. Two physicians with more than 20 years of experience performed the morphological assessment of the adnexal masses of these patients and subsequently performed an ultrasound with
Doppler effect in search of signs of vascularization within the tumor. Finally, the lesion was explored with pulsed Doppler ultrasound and resistance indices (RI).

Patients classified as GI-RADS 1 and 2 were managed with expectant supervision, patients with reports classified as GI-RADS 3, 4 and 5, were surgically resolved.

The results showed an inter-observer agreement index, evaluated by the kappa index of 0.846, which was considered highly satisfactory. The sensitivity of the study was 92%, the specificity was 95%. Likewise, the negative predictive value (NPV) of 85% and the positive predictive value (PPV) of 99%.

The authors conclude that the reports issued with the GI-RADS classification were useful in decision-making for physicians, further suggesting that a subclassification in category 4 could be discussed as used in the BI-RADS system. Similarly, consider the proposal for this new system to report the sonographic findings of adnexal masses, would use the same terminology for reports that will be read by doctors, the therapeutic behavior can be set to follow.

Undoubtedly, in order to establish the true value of this scale, it should be tested prospectively and through multicenter studies. In support of this external validation, in 2011 the Amor and Vaccaro group published a prospective, multicenter study with 372 patients to evaluate the clinical utility of GI-RADS as a reporting system for the ultrasound of adnexal masses.

Additionally, to determine the usefulness of ultrasound findings and the transmission of information to the physician in decision-making, a survey was applied with a single question:

How useful is the GI-RADS reporting system to understand the results of ultrasound and rely on it for clinical decision-making regarding your patient? Answers: a) totally useful, b) quite useful, c) no useful nor useless, d) useless and e) completely useless. The doctors who referred their patients considered it useful. The results showed a sensitivity of 99%, specificity of 85.9%, PPV: 71.1% and NPV: 99.6%, interobserver reproducibility resulted in a kappa index of 0.846.

Patients with GI-RADS 1 were not included in the study. Patients with GI-RADS 2, expectant management, GI-RADS 3 laparoscopy or laparotomy and patients GI-RADS 4 and 5 were referred to gynecologists oncologists for clinical evaluation, extension studies with magnetic resonance and computed tomography as for surgical management. Examining physicians were not part of the decision-making process, limited only to the report of the findings.

According to Timmerman, the examiner's subjective impression is the best approach, no other method has proven to be superior in prospective testing, nor can any of the current models...
outperform an expert sonographer.

Other authors have tested GI-RADS with similar results, considering this scale as a useful classification for ultrasound examination of malignant adnexal masses $^{9,11}$.

In summary, prospective and retrospective studies, analyzing the information provided by expert ultrasonographers, have demonstrated that the GI-RADS classification functions efficiently as a reporting system in adnexal masses and seems to be very useful for decision making. Currently, a Venezuelan team is about to publish its results, together with the University of Navarra, regarding the external validation of this scale, with similar results in the preliminary reports.

The insistence on the resolution of malignant suspected adnexal tumors in the hands of oncologists is a guideline that will make a difference in the improvement of survival and prognosis of patients.

In a recent letter to the editor of AJGO magazine, Goldstein$^{12}$ commented on IOTA's "Simple Rules", noting that, in addition to the importance of a proper diagnosis to differentiate between benign and malignant adnexal masses, the need to refine those diagnoses in order to avoid bringing all the suspicious patients into the operating room due to their involvement in the increase in the costs of care and the possible sequelae for the patient. It is also necessary to improve accuracy and precision in the diagnosis in order to reliably indicate a surgical exploration.

It is interesting to evaluate the results published by IOTA in 2016$^{12}$, where their findings on sensitivity and negative predictive values were similar to those published for GI-RADS. This finding was not reproduced with specificity and positive predictive value, ranging from 33% to 75% and between 44.8% and 75.4%, respectively. On this fact is that is pronounced Goldstein and in response to this call, Amor proposes the methodology GI-RADS $^{13}$.

This classification, which is presented as an option for decision making, does not exclude other tools described as IOTA, but incorporates them. However, external validations are still lacking and the pros and cons that may justify their use. The discussion about knowledge, scientific evidence and validation of experience will increasingly bring us closer to the diagnostic accuracy in making highly reliable decisions with these tools in favor of patients.

*Gynecology and Human Reproduction. Endocrine gynecology. Bogotá - Colombia*
References: