European Journal of Obstetrics & Gynecology and Reproductive Biology 171 (2013) 157–160

Contents lists available at ScienceDirect

European Journal of Obstetrics & Gynecology and Reproductive Biology

journal homepage: www.elsevier.com/locate/ejogrb

Assessing the reproducibility of the IOTA simple ultrasound rules for classifying adnexal masses as benign or malignant using stored 3D volumes

Stefano Guerriero [a], Luca Saba [b], Silvia Ajossa [a], Cristina Peddes [a], Federica Sedda [a], Alessandra Piras [a], Begoña Olartecoechea [c], María Aubá [c], Juan Luis Alcázar [c,*]

[a] Department of Obstetrics and Gynecology, Ospedale San Gionvanni di Dio, University of Cagliari, Cagliari, Italy
[b] Department of Radiology, Azienda Ospedaliero Universitaria di Cagliari – Polo di Monserrato, Cagliari, Italy
[c] Department of Obstetrics and Gynecology, Clínica Universidad de Navarra, University of Navarra, Spain

ARTICLE INFO

Article history:
Received 28 February 2013
Received in revised form 17 June 2013
Accepted 14 August 2013

Keywords:
Adnexal mass
Diagnosis
Three-dimensional ultrasound
Reproducibility

ABSTRACT

Objective: To analyze the reproducibility of the IOTA simple ultrasound rules for classifying adnexal masses as benign or malignant among examiners with different level of expertise using stored 3D volumes of adnexal masses.
Study design: Five examiners, with different levels of experience and blinded to each other, evaluated 100 stored 3D volumes from adnexal masses and looked for the presence or absence of malignant or benign features according to the IOTA definitions. Multiplanar view and virtual navigation were used. All examiners had to assess the 3D volume of each adnexal mass and classify it as benign or malignant. To analyze intra-observer agreement each examiner performed the assessment twice with a two-week interval between the first and second assessments. To analyze the inter-observer agreement, the second assessment from each examiner was used. Reproducibility was assessed calculating the weighted Kappa index.
Results: Intra-observer reproducibility was moderate or good for all observers (Kappa index ranging from 0.59 to 0.74). Inter-observer reproducibility was moderate to good (Kappa index range: 0.46–0.67).
Conclusions: The simple rules are reasonably reproducible among observers with different level of expertise when assessed in stored 3D volumes.

1. Introduction

Traditionally, ultrasound diagnosis of adnexal masses is based on the examiner's subjective impression [1], but examiner's experience significantly affects diagnostic performance [2–4]. In 2008, the international ovarian tumor analysis group (IOTA) proposed the use of ultrasound simple rules for the diagnosis of ovarian malignancy [5]. These rules are based on the simple identification of some findings on ultrasound examination: some of them are characteristic of malignancy (malignant or M features) and others are characteristic of benign lesions (benign or B features). To date, however, no study has assessed how reproducible these simple rules are when used by examiners with different levels of experience.

Three-dimensional (3D) ultrasound has been introduced into clinical practice in the last 15 years. This technique allows the acquisition and storage of 3D volumes for subsequent analysis. Some studies show that stored 3D volumes may be reliably used for assessing anatomical structures in fetal ultrasound [6–8]. A few studies have demonstrated that the assessment of 3D volumes from adnexal masses may render similar results in terms of diagnostic performance to real-time ultrasound when evaluated by expert examiners [9] and that the assessment of features in adnexal masses using 3D volumes may be reproducible among expert examiners [10,11].

We hypothesized that 3D volumes from adnexal masses could be used for describing the ultrasound features of adnexal masses using the IOTA simple rules. The aim of this study was to analyze the reproducibility of the simple rules among examiners with different level of expertise using stored 3D volumes. It was not our intention to assess diagnostic performance of these simple rules.

* Corresponding author at: Department of Obstetrics and Gynecology, Clinica Universidad de Navarra, Avenida Pio XII, 36, 31008 Pamplona, Spain. Tel.: +34 948296234; fax: +34 948296500.
E-mail address: jlalcazar@unav.es (J.L. Alcázar).

0301-2115/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ejogrb.2013.08.031

2. Materials and methods

Five examiners with different levels of experience in gynecological ultrasound were provided with a set of 100 stored 3D volumes of adnexal masses. All volumes had been acquired using a Voluson 730 Expert or Voluson E8 (GE Healthcare, Milwaukee, Il, USA) equipped with a 5–9 MHz endovaginal probe, after a complete 2D transvaginal ultrasound assessment of the mass. These volumes were selected from one of the authors (JLA) center database which contains more than 900 cases. All selected 3D volumes had been acquired by the author, who performed the selection. Selection criteria were based on the fact that the mass was entirely or almost entirely included in the 3D volume and that the volume contained gray-scale and power Doppler information (Figs. 1 and 2). The benign/malignant ratio of the masses was 3:2. These masses had been diagnosed in women evaluated and treated at the first author's center and all of them had been surgically removed. Definitive histological diagnosis was available for all these tumors.

The five examiners who assessed the 3D volumes were blinded to histological results and patient data. Two examiners, A (SG) and B (SA), were considered expert, being gynecologists specially devoted to gynecological ultrasound with over 15 years of experience. Examiner C (AP) was considered a moderate expert, as a gynecologist with no special dedication to gynecological ultrasound. Two examiners, D (FS) and E (CP), were trainees in obstetrics and gynecology, with a very low level of experience in gynecological ultrasound.

Each examiner had to assess the 3D volume using the dedicated software (4D View, GE Healthcare, Zipf, Austria) and look for the presence or absence of malignant or benign features according to the IOTA definitions [5]. Malignant (M) features were as follows. M1: Irregular solid tumor, M2: Presence of ascites, M3: At least 4 papillary projections, M4: Irregular multilocular solid tumor with largest diameter >100 mm, M5: Very strong blood flow (IOTA color score 4). Benign (B) features were as follows. B1: Unilocular tumor, B2: Presence of solid components where solid component largest diameter is <7 mm, B3: Presence of acoustic shadows, B4: Smooth multilocular tumor with largest diameter <100 mm, B5: No blood flow (IOTA color score 1).

Assessment was performed using the multiplanar view and virtual navigation through the mass and mass's size as well as solid component measurement in the three orthogonal planes. Each examiner had to determine the presence or absence of each malignant (M1–M5) and benign (B1–B5) feature and to establish a diagnosis of inconclusive, benign or malignant lesion. A mass was classified as malignant if at least one M feature was present in the absence of any B feature. A mass was classified as benign if at least one B feature was present in absence of any M feature. When both M and B features were present or neither M nor B features were detected, the mass was classified as inconclusive.

To analyze intra-observer agreement each examiner performed the assessment twice with a two-week interval between the first and second assessments. To analyze the inter-observer agreement, the second assessment from each examiner was used.

Reproducibility was assessed calculating the weighted Kappa index [11].

3. Results

Among the 100 stored 3D volumes, 40 masses were malignant (5 borderline tumors, 32 primary invasive cancers, 3 metastatic tumors) and 60 were benign (15 endometriomas, 12 serous cystadenomas, 9 hemorrhagic cysts, 8 dermoid cysts, 5 mucinous cysts, 4 ovarian fibromas, 3 pedunculated leiomyomas, 2 hysdrosalpinges, 2 peritoneal cysts).

The rate of inconclusive masses according to the first reading for different observers ranged from 7% to 19% (Table 1).

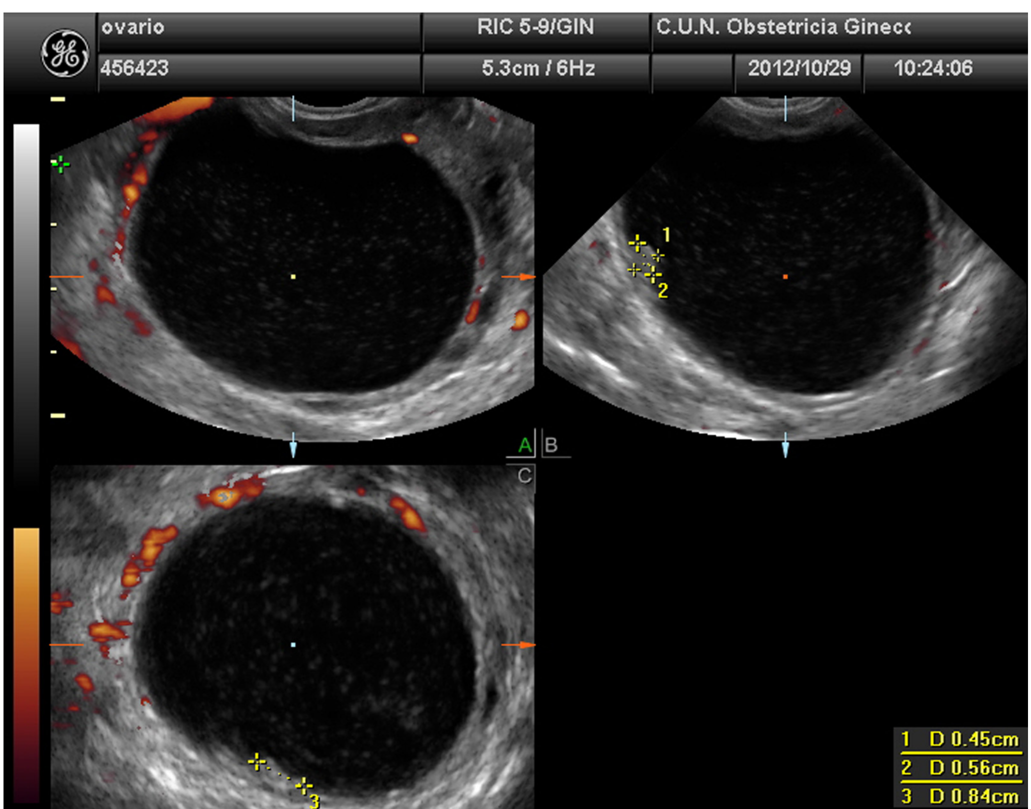

Fig. 1. Three-dimensional images as seen on multiplanar display from a unilocular cyst with a papillary projection.

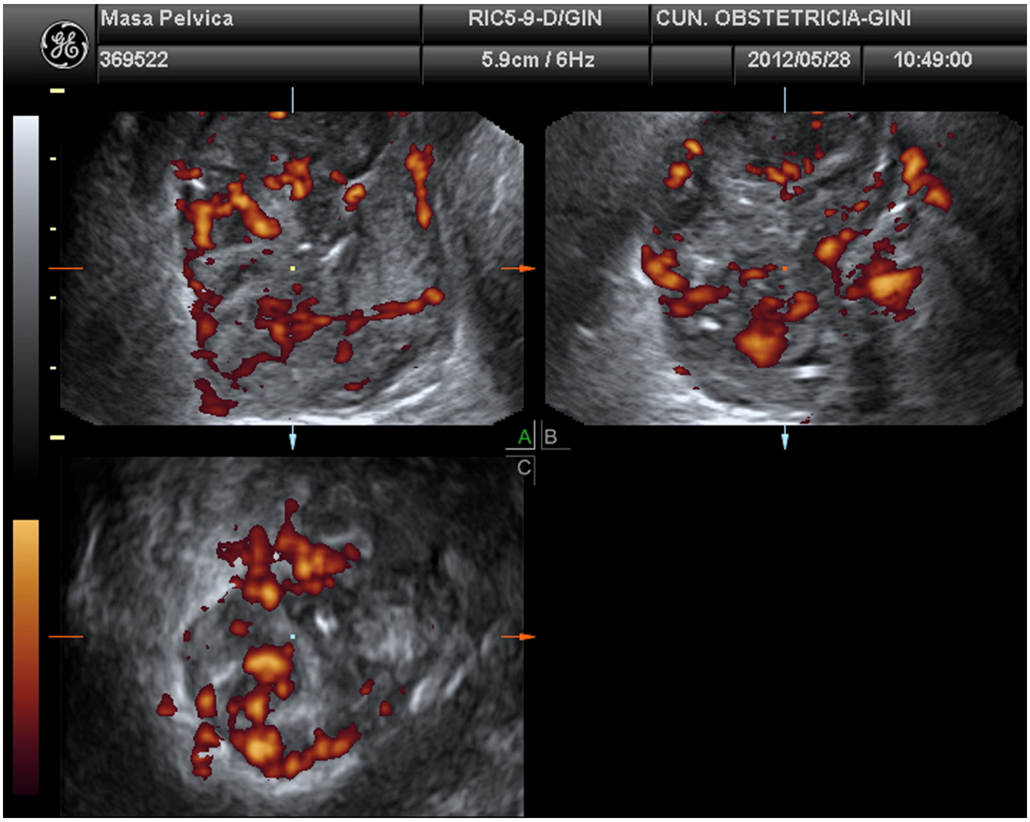

Fig. 2. Three-dimensional images as seen on multiplanar display from an irregular solid adnexal tumor with abundant vascularization.

Intra-observer agreement was good for all observers, with Kappa values ranging from 0.59 to 0.74 (Table 2).

Inter-observer agreement between examiners was moderate to good depending on the pair of examiners tested. Agreement was good when assessment was performed for pairs of expert examiners, but moderate when pairs involved non-expert examiners (Table 3).

4. Comments

In the present study we assessed whether 3D volumes can be used for assessing IOTA simple rules, and the observer agreement for IOTA simple rules for discriminating benign from malignant adnexal masses using stored 3D volumes. Our results show that these rules are reasonably reproducible among observers with different levels of experience.

In our opinion the IOTA simple rules are appealing since they constitute a simple user-friendly way to classify an adnexal mass as benign or malignant. According to reports from the IOTA group these rules could be applied in 76–77% of adnexal masses [5,12].

At least two additional studies have evaluated the use of the simple rules in adnexal masses. Fathallah et al. performed a prospective study using the simple rules in a series of 122 masses and they found that the simple rules were applicable in up to 89.3% of the masses [13]. More recently, Hartman et al. reported a prospective study in a series of 103 women and 110 adnexal tumors and they found that the simple rules could be applied in 82.7% of the tumors [14].

In all these studies reproducibility (either intra- or inter-observer) of the simple rules was not assessed. Our study provides some data about this issue. We observed that intra-observe reproducibility for classifying an adnexal mass as benign, malignant or inconclusive using the simple rules is good for all examiners independently of their level of experience. Inter-observer reproducibility was moderate to good among observers with different level of expertise.

Our study has some limitations, however, since our analysis was performed using stored 3D volumes instead of real-time ultrasound. Certainly, there is some evidence that analysis of 3D stored volumes of adnexal masses may render similar diagnostic performance to real-time ultrasound [9] and that assessment of tumor features is reproducible among different observers analyzing the same stored 3D volume [10,11]. We have to recognize, however, that for the specific assessment of simple rules the use of stored 3D volumes may be a limitation, since the presence or absence of some features may be limited, for example the presence

Table 1
Classification of adnexal masses by different examiners according to simple rules in a 3D volume set.

	Benign	Malignant	Inconclusive
Examiner A	49%	39%	13%
Examiner B	60%	33%	7%
Examiner C	36%	51%	13%
Examiner D	47%	34%	19%
Examiner E	42%	41%	17%

Table 2
Intra-observer reproducibility for observers with different levels of expertise.

	Kappa index[a]	Percentage of agreement[a]
Examiner A	0.71 (0.59–0.83)	83% (74–90%)
Examiner B	0.74 (0.63–0.85)	84% (75–91%)
Examiner C	0.69 (0.56–0.81)	83% (74–90%)
Examiner D	0.59 (0.47–0.73)	75% (65–83%)
Examiner E	0.69 (0.57–0.82)	82% (73–89%)

[a] 95% confidence intervals in parentheses.

Table 3
Inter-observer agreement between observers, expressed as kappa index and percentage of agreement.[a]

	Examiner B	Examiner C	Examiner D	Examiner E
Examiner A	0.58 (0.44–0.71) 74% (64–82%)	0.67 (0.55–0.79) 81% (72–88%)	0.48 (0.34–0.62) 68% (58–77%)	0.46 (0.32–0.60) 67% (57–76%)
Examiner B	NA	0.62 (0.49–0.74) 77% (67–85%)	0.51 (0.37–0.64) 69% (59–78%)	0.58 (0.45–0.71) 74% (64–82%)
Examiner C	NA	NA	0.47 (0.33–0.60) 68% (58–77%)	0.48 (0.34–0.61) 69% (59–78%)
Examiner D	NA	NA	NA	0.48 (0.34–0.61) 67% (57–76%)
Examiner E	NA	NA	NA	NA

NA, not available.
[a] 95% confidence interval for kappa index in parentheses.

of ascites or the number of papillary projections in large tumors, where the whole tumor cannot be included in one single 3D volume.

Additionally a selection bias may exist since the volumes were selected by one examiner: thus the results cannot be extrapolated to any 3D volume. Furthermore, the rate of inconclusive results was low in three out of the five examiners involved in the assessment as compared with data reported previously in prospective series using real-time ultrasound [12–14]. Finally, acquisition reliability was not assessed.

In spite of these limitations, we think that these rules are actually simple and easy to learn, and according to our data their reproducibility among examiners with different levels of experience is good when using stored 3D volumes.

References

[1] Sokalska A, Timmerman D, Testa AC, et al. Diagnostic accuracy of transvaginal ultrasound examination for assigning a specific diagnosis to adnexal masses. Ultrasound Obstet Gynecol 2009;34:462–70.
[2] Alcázar JL, Guerriero S, Laparte C, Ajossa S, Jurado M. Contribution of power Doppler blood flow mapping to gray-scale ultrasound for predicting malignancy of adnexal masses in symptomatic and asymptomatic women. Eur J Obstet Gynecol Reprod Biol 2011;155:99–105.
[3] Van Holsbeke C, Daemen A, Yazbek J, et al. Ultrasound experience substantially impacts on diagnostic performance and confidence when adnexal masses are classified using pattern recognition. Gynecol Obstet Invest 2010;69:160–8.
[4] Guerriero S, Alcazar JL, Pascual MA, et al. The diagnosis of ovarian cancer: is color Doppler imaging reproducible and accurate in examiners with different degrees of experience? J Womens Health 2011;20:273–7.
[5] Timmerman D, Testa AC, Bourne T, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. Ultrasound Obstet Gynecol 2008;31:681–90.
[6] Bromley B, Shipp TD, Benacerraf B. Assessment of the third-trimester fetus using 3-dimensional volumes: a pilot study. J Clin Ultrasound 2007;35:231–7.
[7] Rizzo G, Abuhamad AZ, Benacerraf BR, et al. Collaborative study on 3-dimensional sonography for the prenatal diagnosis of central nervous system defects. J Ultrasound Med 2011;30:1003–8.
[8] Alcázar JL, Iturra A, Sedda F, et al. Three-dimensional volume off-line analysis as compared to real-time ultrasound for assessing adnexal masses. Eur J Obstet Gynecol Reprod Biol 2012;161:92–5.
[9] Alcázar JL, García-Manero M, Galván R. Three-dimensional sonographic morphologic assessment of adnexal masses: a reproducibility study. J Ultrasound Med 2007;26:1007–11.
[10] Sladkevicius P, Valentin L. Intra-inter-observer agreement when describing adnexal masses using the International Ovarian Tumour Analysis (IOTA) terms and definitions: a study on three-dimensional (3D) ultrasound volumes. Ultrasound Obstet Gynecol 2013;41:318–27.
[11] Kundel HL, Polansky M. Measurement of observer agreement. Radiology 2003;228:303–8.
[12] Timmerman D, Ameye L, Fischerova D, et al. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. BMJ 2010;341:c6839.
[13] Fathallah K, Huchon C, Bats AS, et al. External validation of simple ultrasound rules of Timmerman on 122 ovarian tumors. Gynecol Obstet Fertil 2011;39:477–81.
[14] Hartman CA, Juliato CR, Sarian LO, et al. Ultrasound criteria and CA 125 as predictive variables of ovarian cancer in women with adnexal tumors. Ultrasound Obstet Gynecol 2012;40:360–6.