HAL
open science

O-RADS MRI score: analysis of misclassified cases in a prospective multicentric European cohort

I. Thomassin-Naggara, M. Belghitti, A. Milon, C. Abdel Wahab, E. Sadowski, A. Rockall

To cite this version:

I. Thomassin-Naggara, M. Belghitti, A. Milon, C. Abdel Wahab, E. Sadowski, et al.. O-RADS MRI score: analysis of misclassified cases in a prospective multicentric European cohort. European Radiology, 2021, 10.1007/s00330-021-08054-x . hal-03245758

HAL Id: hal-03245758

https://hal.sorbonne-universite.fr/hal-03245758v1

Submitted on 2 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

O-RADS MRI score: Analysis of misclassified cases in a multicentric European cohort

I Thomassin-Naggara [1 2], M Belghitti [3 4], A Milon [3 4], C Abdel Wahab [3 4], E Sadowski [5], A G Rockall [6 7], EURAD study group.

EURAD study group: I Thomassin-Naggara, E Poncelet, A Jalaguier-Coudray, A Guerra, L S Fournier, S Stojanovic, I Millet, N Bharwani, V Juhan, T M Cunha, G Masselli, C Balleyguier, C Malhaire, N Perrot, M Bazot, P Taourel, E Darai, A G Rockall

Affiliations

1 APHP, Sorbonne Université, Hôpital Tenon, Service de radiologie, 58 avenue Gambetta, 75020, Paris, France. isabelle.thomassin@aphp.fr.

2 Institut Universitaire de Cancérologie, Sorbonne Université, Hôpital Tenon, Service de radiologie, 75020, Paris, France. isabelle.thomassin@aphp.fr.

3 APHP, Sorbonne Université, Hôpital Tenon, Service de radiologie, 58 avenue Gambetta, 75020, Paris, France.

4 Institut Universitaire de Cancérologie, Sorbonne Université, Hôpital Tenon, Service de radiologie, 75020, Paris, France.

5 Departments of Radiology, Obstetrics and Gynecology, University of Wisconsin School of Medicine and Public Health, 600 Highland Ave, E3/372, Madison, WI, 53792-3252, USA.

6 Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK.

7 Department of Radiology, Imperial College Healthcare NHS Trust, London, UK.

ABSTRACT (250)

Objective: To retrospectively review the causes of categorization errors using O-RADS-MRI score and to determine the presumptive causes of these misclassifications.

Methods: EURAD database was retrospectively queried to identify misclassified lesions. In this cohort, 1194 evaluable patients with 1502 pelvic masses (277 malignant/1225 benign lesions) underwent standardized MRI to characterize adnexal masses with histology or 2 years follow-up as reference standard. An expert radiologist reviewed cases with two junior radiologists and lesions termed misclassified if malignant lesion was scored≤ 3, a benign lesion was scored≥4, the site of origin was incorrect, or a non-adnexal mass was incorrectly categorized as benign or malignant.

Results: There were 139/1502 (9.2%) misclassified masses in 116 women including 109 adnexal and 30 non-adnexal masses. False negative cases corresponded to 16 borderline or invasive malignant adnexal masses rated score≤3 (16/139,11.5%). False positive cases corresponded to 88 benign masses were rated score 4 (67/139,48.2%) or 5 (18/139,12.9%) or considered as suspicious non-adnexal lesions (3/139,2.2%). Misclassifications were only due to origin error in 12 adnexal masses (8 benign, 4 malignant) (8.6%,12/139) and 23 non-adnexal masses (18 benign, 5 malignant,16.5%,23/139) perceived respectively as non-adnexal and adnexal masses. Interpretive errors (n=104), failure to recognize technical insufficient exams (n=9) and perceptual errors (n=4) were found. Most interpretive were due to misinterpretation of solid tissue or incorrect assignment of mass origin. 84/139 cases were correctly reclassified by the readers with strict adherence to the score rules.

Conclusion: Most errors were due to misinterpretation of solid tissue or incorrect assignment of mass origin..

Keywords: Ovary, Magnetic resonance imaging, neoplasms

Key points:

1- Prospective assignment of O-RADS-MRI score resulted in misclassification of 9.25% of sonographically indeterminate pelvic masses
2- The greatest number of errors occurred in the O-RADS MR score 4 category (48.9%) (68/139).
3- Most errors were interpretive (74.8%) due to misinterpretation of solid tissue as defined by the lexicon or incorrect assignment of mass origin.
4- Pelvic inflammatory disease is a common source of misclassification (8.9%) (12/139)

Abbreviations:

ACR: American College of Radiology

ADNEX-MR score: ADNEXal Magnetic Resonance score

CCTIRS: Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé

DCE: Dynamic Contrast Enhanced

DW: Diffusion Weighted

EURAD: EURopean ADnexal

MRI: Magnetic Resonance Imaging

O-RADS: Ovarian Adnexal Reporting Data System

PID: Pelvic Inflammatory Disease

SIFEM: Société d'Imagerie de la Femme

STD: Standard Deviation

TIC: Time Intensity Curve

INTRODUCTION

Characterization and risk stratification of indeterminate adnexal masses is important to help determine if these women can undergo surveillance or should be referred for surgical evaluation. A validated scoring system was recently published named Ovarian Adnexal Reporting Data System (O-RADS) Magnetic Resonance Imaging (O-RADS-MRI) to standardize imaging reports and categorize the risk of malignant neoplasm as a triage test to decide whether surgery is appropriate and, if so, the extent of surgery required.

The accuracy of this score for risk stratification was validated in several single center studies under the name ADNEX-MR score, the original name of the score, and then in a large prospective European multicenter cohort, the EURAD study, which provided the evidence for the inaugural O-RADS-MRI score published in 2020 (1–7).

In the EURAD study, the score was assigned to each pelvic mass (figure 1) (8) and on a per patient basis, the score reached a sensitivity of 0.93 ($_{95\%}$CI, 0.89-0.96; 189 of 203 patients) and a specificity of 0.91 ($_{95\%}$CI, 0.89-0.93; 848 of 927 patients) with a good inter-rater agreement among both experienced and junior readers (κ=0.784; $_{95\%}$CI, 0.743-0824). O-RADS-MRI score combines morphological features, analysis of signal intensity on T2W, T1W and DW sequence and the analysis of time intensity curve (TIC) and is now recommended for use in French Guidelines on the management of patients with epithelial ovarian cancer (9). False positive and false negative cases are a crucial issue in this pathology where any delay of diagnosis may impact on mortality.

Until now, no studies already published have investigated the misclassification of O-RADS MR score. Diagnostic errors are usually predictable events with readily identifiable contributing factors, resulting in technical, perceptual and interpretive errors. Identifying contributing factors is one of the keys to developing interventions that reduce or mitigate diagnostic errors (10).

The objectives of this retrospective study were to identify and categorize the causes of errors and misinterpretations using the O-RADS-MRI score. Identifying the factors which pre-dispose to errors may allow effective strategies to reduce the frequency of false negative and false positive reads and thereby help to better categorize pelvic lesions.

MATERIAL AND METHODS

Population

The EURAD database was retrospectively queried to identify misclassified lesions by prospective (original) senior readers (>10 years' experience in gynecological imaging). The database is from the EURAD study which is a prospective, observational, multicenter study conducted between March 1st, 2013 and March 31st 2018, sponsored by SIFEM. According to French regulations at the time of study initiation, the study was approved by a national committee (CCTIRS, approval no. 13.090). In addition, the protocol was approved by the Ethics Committee of each participating site. All participating women provided written informed consent.

The EURAD database includes 1340 consenting women of whom 1194 were evaluable patients who underwent standardised MRI (Suppl Material 1) and the score (Figure 1) was applied to characterize adnexal masses, with subsequent histological diagnosis or 2-year follow-up as reference standard. There were 1130 women with at least one pelvic mass, and a total of 1502 pelvic masses in this population (277 malignant /1225 benign lesions), 1372 adnexal masses and 130 non-adnexal masses. (8) .

Review of cases

Open expert central review: one expert radiologist (ITN) (20 years'experience in gynecologic imaging) with two junior radiologists (one fellow (YG) and one resident (MB)) (5 and 2 years'experience in gynecologic imaging) analyzed in consensus all misclassified adnexal masses, based on the reference standard, to identify possible reasons why the lesions had been prospectively misinterpreted. Lesions were termed misclassified if borderline/malignant lesion was scored ≤ 3, a benign lesion was scored ≥ 4, the site of origin was incorrect, or a non-adnexal mass was incorrectly categorized as benign or malignant. The different causes of errors were classified as following:

When the original readers missed the specific abnormality that caused the error, there were two types of error:

- **Failure to recognize technical problem**: The expert review was also unable to identify the specific abnormality that caused the error. The different technical problem were poor quality images (missed solid tissue) that could not be interpreted, contrast not be given (time intensity curve uninterpretable), diffusion not interpretable

1) **Perceptual error**: The expert review identified the specific abnormality that caused the error, missed by original readers

When the original readers identified the abnormality but misinterpreted, the error was considered as an **interpretive (cognitive) error** and there were also two main types of error:

- The original reader misinterpreted a specific abnormality described in the O-RADS MR lexicon such as solid component misinterpreted as solid tissue, wrong curve analysis, signal intensity or peritoneal nodular thickening misinterpretation.

- The original reader misinterpreted the origin of the lesion due to anatomic location or surrounding anatomy that made the discernment truly challenging in some cases

When no reason was clearly identified, the radiologist recorded whether the lesion had any atypical features related to morphology, signal intensity or curve, or whether the lesion was complicated by adnexal torsion or the presence of collision tumor.

The expert review also classified the misclassified lesions according their morphological type (unilocular- no solid component, multilocular – no solid component- unilocular with solid component- multilocular with solid component and purely solid tumor) (11,12).

Blind review: Two radiologists (CW, AM) (7 years' experience in gynaecologic imaging) blind to the prospective classification and reference standard, reviewed the cases retrospectively in consensus to classify the lesion according O-RADS-MRI score, following training and with strict adherence to the defined score rules.

Statistical analysis

Reference standard was the final diagnosis established in EURAD study (7) . Prospective O-RADS-MRI analysis was considered as a misclassification according to the following definitions: Score 1 was only correct if there was no adnexal mass (all physiological findings are correctly allocated as score 1) or the lesion was reported as non-adnexal with correct origin, Score 2 and 3 were only correct if there was an adnexal mass which complied with score 2 or 3 rules and was reference standard benign (based on histopathology or 2 years follow up), Score 4 and 5 were only correct if there was an adnexal mass which complied with score 4 or 5 rules and was reference standard malignant (borderline or invasive tumor) (based on histopathology). Moreover, if a benign non-adnexal mass was rated score 1 but judged subjectively suspicious, the case was also considered as false positive. Finally, if a malignant non-adnexal mass was rated score 1 and subjectively judged non-suspicious, the case was considered as false negative.

The sizes of the misclassified lesions were assessed using mean values and standard deviation (STD) and associations with different morphological sub-types were evaluated. Chi square test was performed to compare the number of errors between 1.5T and 3T MR unit.

RESULTS

The final population included 139 pelvic masses (described in 116 women) misclassified according to O-RADS-MRI score among the 1502 masses (9.2%). The description of the population of 116 women is given in Table 1. The quality of MRI scans in this group was similar to the quality observed in the overall population of 1502 masses with 135/139 MR scan considered as good quality. Errors were more frequent at 3T than at 1.5T (p=0,009, OR=1,7). Misclassifications were found in 114 benign lesions, 10 borderline and 15 invasive tumors, including failure to recognize technically insufficient exams, interpretive and perceptual errors (Table 2).

Prospective assignment of O-RADS-MRI score compared to reference standard

O-RADS-MRI score misclassification mainly occurred when benign lesions were rated O-RADS-MRI score 4 (indeterminate) or score 5 (highly suspicious) (64.7%, 90/139) and in large tumors as the mean size of misclassified lesions according to O-RADS-MRI score was 70.9mm (min-max= 6–250mm). Interpretive errors were the most frequent type of error (74.9%, 104/139) (Table 2A). The different pathological subtypes misclassified and the different types of errors in each category of O-RADS-MRI score are presented in Table 2B and 3.

O-RADS-MRI score 1 (n=15): Twelve adnexal masses were prospectively misclassified as non-adnexal (8 benign adnexal masses classified as score 1 non-suspicious and 4 malignant tubo-ovarian masses rated as score 1 suspicious) (misinterpretation of origin). Ten (83.3%) of these 12 misclassified adnexal masses were larger than 5cm. In addition, 3 benign non-adnexal masses (2 myomas and one peritoneal hematoma) were considered as score 1 suspicious (i.e false positive) due to 2 misinterpretation of origin (uterine myomas misdiagnosed as digestive tumors) and one atypical morphology for peritoneal hematoma.

O-RADS-MRI score 2 (n=18)*:* Three malignant tumors (2 borderline tumors and one metastasis) were under-diagnosed (false negative). In the two borderline tumors, solid tissue was very small and missed by the original and central readers due to movement artfacts (failure to recognize technically insufficient exam) (n=1) or misinterpreted as dark dark (misinterpretation of O-RADS MR feature) (n=1). The metastasis came from invasive lobular carcinoma which are well known to be associated with dark T2 signal and low enhancement. In that case, DW signal was also negative possibly to a dark T2 though effect. This lesion was associated with a contralateral metastasis that was rated O-RADS-MRI 5 and not considered as false negative on a per patient basis in the princeps study. The 15 other misclassified O-RADS-MRI 2 lesions were benign non-adnexal lesions rated as benign adnexal masses (7 myomas, 3 peritoneal cyst, 3 peritoneal hematoma, and 2 others) (misinterpretation of origin).
O-RADS-MRI score 3 (n=16): Thirteen malignant tumors (8 borderline and 5 invasive tumors) were under-diagnosed (false negative). In this group, solid tissue was very small (3mm) and missed by both original and central readers due to movement artifacts and or partial volume in 5 cases (failure to recognize technically insufficient exam) (3 borderline and 2 tubal cancer) and only by original reader (1 mucinous borderline cystadenoma) (perceptual error). These 5 lesions were rated O-RADS-MRI 3 due to multilocularity. In other 5 cases, TIC was misinterpreted by original readers either because the plateau on TIC was missed (3 borderline serous) (misinterpretation of O-RADS MR feature) or because first sequence was already injected (One Granulosa cell tumor and one metastasis) in which one was correctly assessed by central readers (perceptual errors) and one was also missed (failure to recognize insufficient exam). (n=2). The 2 last false negative cases were due to a complication by adnexal torsion (borderline serous cystadenoma) and the presence of collision tumor (invasive mucinous cystadenocarcinoma with a struma ovarii). The 3 other misclassified O-RADS-MRI

3 lesions were benign non-adnexal lesions rated as benign adnexal masses (1 myoma, 1 peritoneal cyst, 1 hematoma) (misinterpretation of origin)

***O-RADS-MRI score 4 (n=68)*:** Sixty-seven benign pelvic masses were rated in this category (false positive). More than half of these cases were due to the description of a solid tissue which did not comply with the definition of solid tissue by the lexicon (50.7%, 34/67), including Rokitansky nodule (n=14), endosalpingial folds (n=8), normal ovarian parenchyma (n=4), thick smooth septa or wall (n=5) or small loculus (n=3) (figure 2) (misinterpretation of O-RADS MR feature). In 7 cases (10.7% 7/65), there was a misinterpretation of TIC quoted as intermediate-risk TIC (type 2) while no plateau existed (misinterpretation of O-RADS MR feature) and therefore should have been assigned low-risk curve, score 3. The other causes of errors were unfeasible TIC (i.e scored 4) due to movement artifacts by both central and original readers (n=2) (failure to recognize insufficient exams) or due to technical problems (injection) identified by central readers (n=2) (perceptual errors), wrong origin assignment with myomas rated as purely solid adnexal masses (n=3) (Figure 3), misinterpretation of T2 and DW signal which were dark-dark (n=3) (misinterpretation of O-RADS MR feature). Six further lesions were misclassified due to difficult interpretation in adnexal torsion (n=5) and the presence of collision tumor (n=1). No cause of misinterpretation was found in 10 cases, 8 of which were benign epithelial tumors enhancing with an intermediate-risk TIC (type 2) including 2 mucinous benign tumor, 1 benign ovarian fibroma with intermediate T2W signal and intermediate-risk TIC and 1 cystadenofibroma. The last case in this category was a Mullerian adenosarcoma which was rated O-RADS-MRI 4 as an adnexal mass (misinterpretation of origin).

***O-RADS-MRI score 5 (n=22)*:** Seventeen benign adnexal masses were rated 5 (false positive) including 12 cases due to the description of solid tissue that did not comply with definition of solid tissue (70.5%) (misinterpretation of O-RADS MR feature) which was confused with

Rokitansky nodule (n=6), endosalpingial folds (n=2), thick smooth septa (n=2), deep pelvic endometriotic nodule (n=1), or normal ovarian parenchyma (n=1). In this last case, a region of interest was placed on normal ovarian parenchyma generating a high-risk TIC resulting in an O-RADS-MRI 5 classification for a cystadenofibroma. The other cause was benign smooth thickening of peritoneum mis-interpretated as peritoneal carcinomatosis (n=3) (misinterpretation of O-RADS MR feature). In two further cases, no cause of misinterpretation was found: These two cases were one ovarian fibroma and one cystadenofibroma which displayed a very atypical enhancement with a high-risk TIC (type 3). In this category, 5 non-adnexal masses were misdiagnosed: 1 schwannoma (i.e false positive) and 4 malignant non-adnexal tumors (2 colorectal, 1 peritoneal and 1 urothelial cancer) (misinterpretation of the origin) which were misinterpreted with primitive ovarian carcinoma The different pathological subtypes misclassified in each morphological subtype score are presented in Table 4. The detailed analysis of these errors are presented in Supplemental Material 2.

Consensus retrospective blinded review

The blinded readers were asked to classify lesions with strict adherence to the score rules. A total of 84/139 (60,4%) O-RADS-MRI misclassifications were correctly reclassified including 7/15 (46.7%) O-RADS-MRI score 1, 7/18 (38.8%) O-RADS-MRI score 2, 6/16 (37.5%) O-RADS-MRI score 3, 50/68 (73.5%) O-RADS-MRI score 4, and 14/22 (63.6%) O-RADS-MRI score 5. These reclassifications included 64/88 (72.7%) false positive, 4/16 (25%) false negative and 16/35 (45.7%) errors exclusively due to origin. Then, O-RADS-MRI score reached the accuracy of 97.6% (1466/1502) including 12 false negative (7 borderline and 5 invasive), 24 false positive (benign lesions), with a sensitivity of 92% (255/277) and a specificity of 98% (1201/1225) (Table 3).

The 84/139 O-RADS-MRI misclassifications correctly reclassified were divided according the four categories of errors 1/ The blinded readers did not reclassified any failure to recognize technically insufficient exams 2/ The readers correctly reclassified 74% (77/104) of interpretive errors including 91.3% (42/46) due to solid component being incorrectly called solid tissue (the 4 residual misclassification were all benign lesions (1 serous and 1 mucinous cystadenomas, 1 mature cystic teratoma and 1 PID), 51.2% (21/41) of misinterpretations of origin, 90.9% (10/11) of curve type misinterpretation (only one serous borderline without plateau was not reclassified), 33.3% (1/3) of peritoneal nodular thickening misinterpretation reclassified O-RADS-MRI 3, and all signal intensity misinterpretations (all were dark-dark signal intensity) (3/3). 3/ The 4 perceptual errors were corrected by the blinded readers including 2 serous benign, 1 mucinous borderline and 1 metastasis

Finally, the blinded readers also correctly reclassified 22.7% (5/22) of lesions with atypical findings mainly based on the detection of associated adnexal torsion (50%, 3/6) correctly reclassified (residual one false negative and two false positive) and the detection of collision tumor with also 1/2 correctly reclassified. Only one myoma was correctly reclassified within the group of lesions atypical morphological features (n=8) and none of the 6 lesions with atypical curves could be reclassified.

DISCUSSION

Our study demonstrates that O-RADS-MRI score misclassification mainly occurs when benign lesions are rated as indeterminate (O-RADS-MRI score 4) or highly suspicious (O-RADS-MRI score 5) (64.7%,90/139) and in large tumors (mean size of misclassified lesions=71mm). Interpretive errors are the most frequent type of error (74.9%,104/139) with a particular difficulty to correctly recognize enhancing solid tissue or apply the term according to the strict definition (44.2%,46/104). The greatest number of errors occurred in the O-RADS-MRI score 4 category (48.9%,68/139). In the retrospective review, strict application of score rules allowed readers to correctly reclassify 84/139 lesions (60.4%) especially regarding false positive cases (72.7%,64/88).

The O-RADS-MRI score was recently developed to help the radiologist clearly communicate the risk of malignancy of adnexal masses to clinicians, based on MR imaging. In this study, we classified errors in 3 main types as already published in the literature (10): 1) Errors due to limitation of the technique (both original and central readers failed to recognize technically insufficient exams) were mainly represented by the missing of small solid tissue. 2) Perceptual errors were represented by lesion not seen by the original readers but seen by the expert review 3) Interpretive errors were the most frequent and corresponded to lesion seen but misinterpreted mainly due to errors in applying strict definitions of the lexicon. This last category was responsible formost O-RADS-MRI misclassifications due to false positive interpretation. This is probably in line with the history of this score; The original design of the score (named ADNEX-MR score) was published in 2013 (7) with the main objective to not miss any malignancy, ensuring early detection. Moreover, within the different publications and evolution of the score, the terms “solid tissue” and “solid component” were occasionally used interchangeably. A significant piece of work on homogenization of terms was conducted by the ACR with the recent publication of O-RADS lexicon (12) to fully distinguish the terms

“solid component” and “solid tissue”. Solid component refers to any non-fluid component of a lesion. There are two types of solid components: solid tissue and other solid components (not solid tissue). Solid tissue is defined as exhibiting post contrast enhancement and conforms to one of the following morphologies: papillary projections, mural nodules, irregular septations/walls and solid portion. Other solid components (not solid tissue) include smooth wall/septation, clot, debris, and fat within a lesion. These clarifications may help to avoid any misinterpretation regarding Rokitansky Protuberance which is a solid component but not defined as solid tissue (12). As reported in table 4, twenty mature cystic teratomas were misclassified including 19/20 rated O-RADS 4 or 5 due to the reader recording an intermediate or high-risk TIC in a Rokitansky nodule, even if the reader diagnosis was finally determined as benign teratoma. This element was already reported in a series published in 2013 (13) which underlined that different types of time intensity curve (low-risk 1, intermediate-risk 2 or high-risk 3) may be found in mature cystic teratoma and reflect respectively the presence of smooth muscular cells and fibrous, neuroglial or thyroid tissue. Another frequent cause of false positive was the misinterpretation between endosalpingial folds and papillary projections especially in the context of pelvic inflammatory disease (PID). PID is typically associated with inflammatory processes, such as thickening and marked enhancement of fallopian tube wall and endosalpingeal folds. In our study, 11/12 PID were over-rated (score 4 or 5) because TIC analysis was performed on endosalpingial folds or thick wall. This can be a difficult interpretation for radiologist who must be cautious and integrate other features such as tubular shape, premenopausal status, signs of inflammatory change in adjacent structures and the presence of pus within the lesion, which is easily recognized with DW signal.

According to the ACR O-RADS-MRI recommendations, women presenting with acute symptoms and an adnexal mass described at MRI need not be rated with O-RADS-MRI

score. However, PID is not always associated with acute pelvic pain in contrast with adnexal torsion which often has an acute clinical presentation. In EURAD study, the presence of acute symptoms was not an exclusion criterion. If women with acute symptoms were excluded from the study, the 4 errors of misclassification of adnexal torsion rated O-RADS-MRI 4 reported in this study would have been avoided.

Another type of error frequently reported is the misinterpretation of TIC. Errors in all cases were due to difficulty in recognizing a shoulder and plateau between low and intermediate-risk TIC (no confusion was found between high-risk TIC versus intermediate and low-risk). This misinterpretation was responsible for 7 false positives and 4 false negatives which all were correctly classified by the blind readers with strict adherence to the score rules. According to the lexicon, low-risk TIC (or curve type 1) consists in a curve without a shoulder and any plateau even, if the slope is acute while an intermediate-risk TIC (or curve type 2) describes a curve with a plateau but not steeper than that of outer myometrium.

MR imaging helped to correctly reclassify the origin of the presumed adnexal mass on ultrasonography with a sensitivity of 0.99 and a specificity of 0.78 in the princeps study (7). The analysis of the error of evaluation of the origin underlines that most of adnexal lesions were benign lesions and 10/12 were larger than 5cm. Misinterpretation of the origin may have a major potential impact in terms of morbidity and in terms of fertility in premenopausal patients as 17/41 patients of this study, in which misinterpretation of origin was present, underwent surgery without any symptom or history of infertility. The tumor size appears to impact on this error-type, as the mean size of misclassified lesions was higher than 7cm while the mean size of the masses among the princeps study (n=1502) was 55.6mm (+/-STD=41.7). Another cause of misclassifying the origin of the lesions is menopausal status. In this study, menopausal women represent 58,4% (24/41) of patients where the adnexal mass was

misclassified due to a wrong origin. Hormonal status is an understandable cause of origin misclassification error, as the normal ovaries may be difficult to identify due to atrophy in postmenopausal women.

Clearly identifying the type of errors is helpful for determining the focus of future educational and research initiatives. Interpretive errors represent the majority of errors (77.8%) and most of them (60.4%) were avoided when readers were fully trained and applied the score with strict adherence to the defined score rules. In this setting, the use of a lexicon with clear unambiguous descriptors is crucial. Moreover, the development of tools such as the calculator for practical training may help to follow the algorithm pathway.

This study has several limitations including a major one which is the retrospective design inherent to this type of analysis. The retrospective readers knew from the beginning that they should "find the classification error" Thus, accuracy calculated by taking into account this retrospective reading does not reflect the real diagnostic performance of MRI. We also analyzed only senior errors and possibly a larger study could be performed to identify the type of errors depending on the experience of the readers. Lastly, a single expert reader even with two juniors for analyzing misclassifications can introduce bias.

In conclusion, prospective assignment of O-RADS-MRI score resulted in misclassification of 9.25% of sonographically indeterminate pelvic masses. Most errors were interpretive due to misinterpretation of solid tissue as defined by the lexicon or incorrect assignment of mass origin. These should be the focus of further educational initiatives.

ACKNOWLEGMENT SECTION

The authors acknowledge the Société d'Imagerie de la femme (SIFEM) for promoting the study and giving a grant. Andrea Rockall acknowledges with respect to Imperial NIHR BRC and CRUK centres We thank Yoram Gueniche (YG) to participate to expert review

REFERENCES

1. Ruiz M, Labauge P, Louboutin A, Limot O, Fauconnier A, Huchon C. External validation of the MR imaging scoring system for the management of adnexal masses. Eur J Obstet Gynecol Reprod Biol. 2016 Oct;205:115–9.
2. Pereira PN, Sarian LO, Yoshida A, Araújo KG, Barros RHO, Baião AC, et al. Accuracy of the ADNEX MR scoring system based on a simplified MRI protocol for the assessment of adnexal masses. Diagn Interv Radiol Ank Turk. 2018 Apr;24(2):63–71.
3. Sasaguri K, Yamaguchi K, Nakazono T, Mizuguchi M, Aishima S, Yokoyama M, et al. External validation of ADNEX MR SCORING system: a single-centre retrospective study. Clin Radiol. 2019;74(2):131–9.
4. Pereira PN, Sarian LO, Yoshida A, Araújo KG, Silva ACB, de Oliveira Barros RH, et al. Improving the performance of IOTA simple rules: sonographic assessment of adnexal masses with resource-effective use of a magnetic resonance scoring (ADNEX MR scoring system). Abdom Radiol N Y. 2020 Oct;45(10):3218–29.
5. Basha MAA, Abdelrahman HM, Metwally MI, Alayouty NA, Mohey N, Zaitoun MMA, et al. Validity and Reproducibility of the ADNEX MR Scoring System in the Diagnosis of Sonographically Indeterminate Adnexal Masses. J Magn Reson Imaging JMRI. 2020 Jul 26;e27285.
6. Hottat NA, Van Pachterbeke C, Vanden-Houte K, Denolin V, Jani JC, Cannie MM. Magnetic resonance scoring system for the assessment of ovarian and adnexal masses: added value of diffusion-weighted imaging including the apparent diffusion coefficient map. Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol. 2020 May 21;
7. Thomassin-Naggara I, Poncelet E, Jalaguier-Coudray A, Guerra A, Fournier LS, Stojanovic S, et al. Ovarian-Adnexal Reporting Data System Magnetic Resonance Imaging (O-RADS MRI) Score for Risk Stratification of Sonographically Indeterminate Adnexal Masses. JAMA Netw Open [Internet]. 2020 Jan 24 [cited 2020 Oct 13];3(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6991280/
8. Thomassin-Naggara I, Aubert E, Rockall A, Jalaguier-Coudray A, Rouzier R, Daraï E, et al. Adnexal masses: development and preliminary validation of an MR imaging scoring system. Radiology. 2013 May;267(2):432–43.
9. Lavoue V, Huchon C, Akladios C, Alfonsi P, Bakrin N, Ballester M, et al. Management of epithelial cancer of the ovary, fallopian tube, and primary peritoneum. Long text of the Joint French Clinical Practice Guidelines issued by FRANCOGYN, CNGOF, SFOG, and GINECO-ARCAGY, and endorsed by INCa. Part 1: Diagnostic exploration and staging, surgery, perioperative care, and pathology. J Gynecol Obstet Hum Reprod. 2019 Jun;48(6):369–78.
10. Itri JN, Tappouni RR, McEachern RO, Pesch AJ, Patel SH. Fundamentals of Diagnostic Error in Imaging. Radiogr Rev Publ Radiol Soc N Am Inc. 2018 Oct;38(6):1845–65.
11. Andreotti RF, Timmerman D, Benacerraf BR, Bennett GL, Bourne T, Brown DL, et al. Ovarian-Adnexal Reporting Lexicon for Ultrasound: A White Paper of the ACR Ovarian-Adnexal Reporting and Data System Committee. J Am Coll Radiol JACR. 2018 Oct;15(10):1415–29.
12. Reinhold C, Rockall A, Sadowski EA, Siegelman ES, Maturen KE, Vargas HA, et al. Ovarian-Adnexal Reporting Lexicon for MRI: A White Paper of the ACR Ovarian-Adnexal Reporting and Data Systems MRI Committee. J Am Coll Radiol JACR. 2021 Jan 20;
13. Poncelet E, Delpierre C, Kerdraon O, Lucot J-P, Collinet P, Bazot M. Value of dynamic contrast-enhanced MRI for tissue characterization of ovarian teratomas: correlation with histopathology. Clin Radiol. 2013 Sep;68(9):909–16.

TABLES

Table 1: Population characteristics (n=116)

PERSONAL HISTORY	n	%
Menopausal	53	
History of pelvic surgery	34	
History of adnexal surgery	15	
History of infertility	10	
History of breast or ovarian cancer	7	
Known BRCA1/2 carriers	0	
CLINICAL PRESENTATION	**n**	**%**
Pelvic pain	37	
Vaginal bleeding	7	
Palpable mass or increasing abdominal volume	2	
Urinary symptoms	1	
Combination of previously mentioned symptoms	5	
None of these symptoms	64	
MR FINDINGS	**n**	**%**
Single lesion	85	
Two lesions	26	
Three lesions	5	
MANAGEMENT	**n**	**%**
Primary surgery	100	
Secondary surgery after initial follow-up	2	
Secondary surgery after primary chemotherapy	0	
24 months of clinical and/or imaging follow-up	14	
- Imaging follow-up		
• Disappearance	5	
• Decrease of the mass	1	
• Stability	4	
- Clinical follow-up (stability)	4	

Table 2A: Open review of the different type of errors and description of morphological subtypes

CAUSES	Benign masses (n=114)	Borderline tumors (n=10)	Invasive tumors (n=15)	Total (n=139)
Failure to recognize technically insufficient exams	**1.7% (2)**	**40% (4)**	**20% (3)**	**6.5% (9)**
Interpretive errors	**77.8% (91)**	**40% (4)**	**60% (9)**	**74.8% (104)**
Misinterpretation of the origin	28.1% (32)	-	60% (9)	29.5% (41)
Misinterpretation of O-RADS MR features	51.7% (59)	40% (4)	-	45.3% (63)
• Solid component misinterpreted as solid tissue ○ Rokitansky protuberance ○ Endosalpingial folds ○ Normal ovarian parenchyma ○ Thick smooth septa or wall ○ Small loculus ○ Deep endometriotic nodule	40.3% (46) 43.5 % (20) 21.7% (10) 10.9% (5) 13.0% (6) 8.7% (4) 2.2% (1)	-	-	33.1% (46)
• Curve analysis misinterpretation	6.1% (7)	30% (3)	-	7.2% (10)
• Signal intensity misinterpretation	2.6% (3)	10% (1)	-	2.9% (4)
• Peritoneal nodular thickening misinterpretation	2.6% (3)	-	-	2.2% (3)
Perceptual errors	**1.7% (2)**	**10% (1)**	**6.6% (1)**	**2.9% (4)**
No error found	**16.7% (19)**	**10% (1)**	**13.3% (2)**	**15.8% (22)**
Atypical morphology or signal features Atypical curve	6.1% (7) 5.3% (6)	-	6.7% (1)	5.8% (8) 4.3% (6)
Adnexal torsion	4.4% (5)	10% (1)	-	4.3% (6)
Collusion tumors	0.9% (1)	-	6.7% (1)	1.4% (2)

Table 2B: Open review of the different type of errors and description of morphological subtypes according to the O-RADS MR categories

CAUSES	Benign masses (n=114)	Borderline tumors (n=10)	Invasive tumors (n=15)	Total (n=139)
Failure to recognize technically insufficient exams	**2 (22.2%)**	**4 (44.4%)**	**3 (33.3%)**	**9**
O-RADS MR 1	**-**	**-**	**-**	
O-RADS MR 2	**-**	**1**	**-**	
O-RADS MR 3	**-**	**3**	**-**	
O-RADS MR 4	**2**	**-**	**3**	
O-RADS MR 5	**-**	**-**	**-**	
Misinterpretation of the origin	**32 (78%)**	**-**	**9 (21.9%)**	**41**
O-RADS MR 1	10	-	4	
O-RADS MR 2	15	-	-	
O-RADS MR 3	3	-	-	
O-RADS MR 4	3	-	1	
O-RADS MR 5	1	-	4	
Misinterpretation of O-RADS MR features	**59 (93.6%)**	**4 (6.3%)**	**-**	**63**
O-RADS MR 1	**-**	-	**-**	
O-RADS MR 2	**-**	1	**-**	
O-RADS MR 3	**-**	3	**-**	
O-RADS MR 4	44	-	**-**	
O-RADS MR 5	15	-	**-**	
Perceptual errors	**2 (50%)**	**1 (25%)**	**1 (25%)**	**4**
O-RADS MR 1	-	-	-	
O-RADS MR 2	-	-	-	
O-RADS MR 3	-	1	1	
O-RADS MR 4	-	-	-	
O-RADS MR 5	2	-	-	
No error found	**19 (86.3%)**	**1 (4.5%)**	**2 (9.1%)**	**22**
O-RADS MR 1	1	-	-	
O-RADS MR 2	-	-	1	
O-RADS MR 3	-	1	1	
O-RADS MR 4	16	-	-	
O-RADS MR 5	2	-	-	

Table 3: Description of type of lesions incorrectly classified by O-RADS-MRI score

	Prospective misclassification	Type of error	Type of Final diagnosis	Nb lesions prospectively rated	Nb lesions correctly reclassified
EURAD Study cohort **1502 pelvic masses**	**O-RADS 1** 10.4% (n= 15/115)	False positive (rated O-RADS 1 suspicious) (n=3)	Myoma	2	1
			Peritoneal hematoma	1	1
		Origin errors only* (n=12)	Cystadenocarcinoma	3	2
			Benign serous cystadenoma	3	1
			Metastasis	1	0
			Endometrioma	1	0
			Functional cyst	1	0
			Mature cystic teratoma	1	1
			Paraovarian cyst	1	0
			Pelvic inflammatory disease	1	1
	O-RADS 2 2.3% (n=18/783)	False negative (n=3)	Metastasis (ILC)	1	0
			Borderline tumors (serous)	2	0
		Origin error only (n=15)	Benign non-adnexal lesions	15	7
	O-RADS 3 5.9% (n=16/272)	False negative (n=13)	Borderline tumors - Serous - Mucinous	8 5 3	3 3 0
			Tubal cancer	2	0
			Mucinous cystadenocarcinoma	1	0
			Granulosa cell tumor	1	0
			Metastasis	1	1
		Origin error only (n=3)	Benign non-adnexal lesions	3	2
	0-RADS 4 46.6% (n=68/146)	False positive (n=67)	Epithelial benign tumors - Serous - Mucinous - Cystadenofibroma - Brenner	25 15 5 4 1	12 7 2 3 0
			Mature cystic teratoma	14	13
			Pelvic inflammatory disease	8	7
			Ovarian fibromas	8	6
			Adnexal torsion	4	3
			Uterine leiomyomas	3	3
			Endometriomas	2	2
			Functional cysts	3	3
		Origin error only (n=1)	Malignant non-adnexal tumor	1	1
	O-RADS 5 9.1% (n=22/186)	False positive (n=18)	Mature cystic teratoma	5	5
			Epithelial benign tumors - Cystadenofibroma - 1 Brenner - 1 serous	4 2 1 1	3 1 1 1
			Functional cysts	3	1
			Pelvic inflammatory disease	3	3
			Struma ovarii	1	1
			Ovarian fibroma	1	0
			Schwannoma	1	0
		Origin error only (n=4)	Malignant non-adnexal tumors	4	1

* all malignant tumors were rated O-RADS 1 suspicious and all benign lesions rated O-RADS 1 non -suspicious)

Table 4: Description of the different misclassifications depending on the morphological subtypes and final diagnosis of the pelvic mass

MORPHOLOGICAL SUBTYPES	Benign masses (n=114)	Borderline tumors (n=10)	Invasive tumors (n=15)	Total (n=139)
Unilocular – no solid component	10.5% (12)	0	13.3% (2)	10.1% (14)
Multilocular – no solid component	4.4% (5)	40% (4)	6.7% (1)	7.2% (10)
Unilocular- Solid component	35.1% (40)	40% (4)	0	31.7% (44)
Multilocular – Solid component	30.7% (35)	10% (1)	4.3 % (6)	30.2% (42)
Purely solid	19.3% (22)	1	4.3 % (6)	20.9% (29)

FIGURES

Figure 1: O-RADS MR scoring system : per patient analysis (n=1194) * *Data form Thomassin-Naggara et . JAMA Network open 2020*

Figure 2: Solid component versus Solid tissue. All cases are presented with AxT2W sequence and AxT1W after gadolinium

A and B - Endosalpingial folds that mimics papillary projections (Pelvic inflammatory disease with left pyosalpinx)

C and D- Debris that mimics thickened irregular septa (No enhancement= No solid tissue) (Mature cystic teratoma)

E and F- Small loculus that mimics mural nodule (Only wall enhancement) (Mature cystic teratoma)

Figure 3: Misinterpretation of origin

A, B, C, D - This non-adnexal mass, a subserous leiomyoma, was rated as an ovarian suspicious mass i.e O-RADS-MRI 4 and was considered as false positive. This patient had no symptoms or history of infertility and underwent surgery based on MR report.

E-This non-adnexal mass, a urothelial cancer was rated as an ovarian suspicious mass i.e O-RADS-MRI 5

Supplemental Material 1: MR Protocol

In this cohort, each patient underwent a routine pelvic MRI (1.5T or 3T), including morphological sequences (T2, T1 with and without fat suppression and T1 after gadolinium injection) and functional sequences (perfusion and diffusion-weighted sequences). If at the time of inclusion and MRI scan, the adnexal mass had disappeared on T2 and T1 weighted sequences, functional sequences and gadolinium injection were not mandatory. Each center used functional sequences with the following criteria: DCE MRI sequence: 3D isotropic, delay between the beginning of the sequence and injection: 1 minute, total duration after injection: 3min, slice thickness 3mm no gap (1 slice on 2 with the exact same location and field of view as the axial T2), spatial resolution: 3mm, temporal resolution < 15s, box size: 15cm. DWI sequence: exactly the same slice thickness as the T2, with a high value of b1000-b1200 (to ensure that urine in the bladder is dark).

Quality criteria for DCE MR Sequence

- The **3D isotropic voxels,** in order to allow reformatting in any plane with a maximal slice thickness of 3mm.
- **Axial plane**, in line with the T2 and DWI for ease of cross correlation of the SI characteristics.
- Minimum time resolution of **15 seconds per acquisition**
- **Subtraction** will allow the suppression of any pre-contrast high T1 signal intensity, whether the DCE is acquired with or without fat saturation.
- **The acquisition should start prior to contrast injection (**30 seconds prior to contrast injection) and should continue for at least 3mn.
- Curve analysis technique:
    - Ensure the time intensity curve is evaluated using **percentage or relative enhancement** (not absolute enhancement)
    - Identify the **most rapid/highest level enhancement** in the adnexal mass for selection of ROI for solid tissue enhancement curve. C**olor coded maps** facilitate the identification of the most suspicious spot within the solid portion of the lesion
    - Place ROI on the outer myometrium.
    - **Curve type 1 or low-risk curve**: Gradual increase in signal intensity of solid tissue without a well-defined shoulder.
    - **Curve type 2 or intermediate risk curve**: A moderate initial increase in the signal intensity of solid tissue relative to that of myometrium, followed by a plateau.
    - **Curve type 3 or high-risk curve**: An initial increase in the signal intensity of solid tissue that is steeper than that of myometrium.

Supplemental material 2: Prospective analysis by MRI morphology compared to reference standard.

Unilocular – no solid component: Fourteen pelvic masses prospectively misclassified were perceived as unilocular without solid component including 2 invasive tumors and 12 benign masses. The 2 invasive tumors were 2 false negative cases corresponding to tubal cancers classified as benign because the solid tissue was very small (<3mm) and was missed. The 12 benign masses included (1) 5 benign adnexal masses rated as non-adnexal masses (perceived as peritoneal cyst (n=3), tailgut cyst (n=1) or peritoneal hematoma (n=1)) and 4 non-

adnexal masses perceived as adnexal mass. Mean size of these lesions was 70.4mm. (2) Three benign lesions were false positive, over-rated either as O-RADS-MRI score 5 due to benign peritoneal thickening or as O-RADS-MRI 4 because a small solid tissue (5mm) was present in a cystadenofibroma but movement artifact prevented meaningful TIC analysis.

Multilocular – no solid component*:* Ten pelvic masses prospectively misclassified were perceived as multilocular without solid component including 1 invasive tumor, 4 borderline tumors and 5 benign masses. All the malignant tumors (borderline and invasive) were under rated O-RADS-MRI 3 (i.e false negative cases): The invasive mucinous cystadenocarcinoma was colluded with struma ovarii rendering the diagnosis to be difficult and the 4 borderline tumors (3 mucinous and 1 serous) contained very small papillary projections which were missed. The 5 benign masses included 3 adnexal lesions (one case pelvic inflammatory disease and one associated with adnexal torsion rated O-RADS MR score 4, one large paratubal cyst (92mm) misdiagnosed as peritoneal cyst and two peritoneal cysts confused with adnexal masses.

The mean size of cyst without solid component misclassified according to O-RADS-MRI score was 87.2mm (min-max= 20–250mm)

Unilocular cyst with solid component: Forty-four pelvic masses prospectively misclassified were perceived as unilocular with solid component including 4 borderline tumors and 40 confirmed benign lesions (90.9%). The 4 underdiagnosed serous borderline tumors were rated O-RADS-MRI score 2 (n=2) because small papillary projections were prospectively not identified or O-RADS-MRI score 3 (n=2) due to partial volume effect on DCE MR resulting in a difficult to interpret TIC which was incorrectly assigned low risk curve. In the group of 40 benign lesions, 28 adnexal masses (70%) were misclassified because the reader recorded solid tissue which did not comply with the lexicon definition, including 15 mature cystic teratoma, 5 pelvic inflammatory disease, 4 epithelial tumors confirmed at pathology and 3 luteal cyst, 1 endometrioma. The further 12 misclassified benign masses included 5 serous benign cystadenoma, 2 hematoma,1 pelvic inflammatory disease, 1 cystadenofibroma, 1 functional cyst and 1 very atypical uterine leiomyoma and 1 peritoneal cyst (figure 3A).

Multilocular cyst with solid component: Forty-two pelvic masses prospectively misclassified were perceived as multilocular with solid component including 6 invasive tumors (14.2%), 1 borderline (2,4%) and 35 benign lesions (83.3%). The invasive tumors included 4 invasive non-adnexal tumors (2 colorectal, 1 peritoneal and 1 urothelial cancer (figure 3B) rated O-RADS-MRI score 5, 1 invasive ovarian tumors (one endometrioid cystadenocarcinoma misinterpreted as a colorectal cancer and 1 Granulosa cell tumors rated O-RADS-MRI score 3 with a technical problem during injection and a very low enhancement of the myometrium). The

borderline tumor was a serous cystadenoma with moderate enhancement interpretated as low risk TIC while there was a plateau. In the group of 35 benign lesions, there were 18/35 cases (51.4%) where the reader recorded solid tissue which did not comply with the lexicon including 7 epithelial tumors with thick smooth septa (n=4), 4 mature cystic teratoma, 3 pelvic inflammatory disease, 1 luteal cyst, 1 endometrioma, 1 struma ovarii, and 1 extra adnexal mass (uterine leiomyomas). The further 17 benign adnexal lesions included 10 epithelial benign cystadenomas (6 serous, 2 mucinous, 1 Brenner, 1 cystadenofibroma), 2 pelvic inflammatory disease, 1 ovarian fibroma, 1 adnexal torsion and 3 benign non-adnexal masses (1 schwannoma, 1 hematoma and 1 myoma)

The mean size of cyst with solid component misclassified according to O-RADS-MRI score was 65.1mm (min-max= 6–230mm)

Purely solid mass: Twenty-nine pelvic masses prospectively misclassified were perceived as purely solid including 6 invasive tumors, 1 borderline tumor and 22 benign lesions. For the 6 invasive tumors, there were errors of origin in 4 cases (2 cystadenocarcinoma, 1 metastasis, and 1 Mullerian adenosarcoma), and 2 false negative cases which were one metastasis from invasive lobular carcinoma (enhancing according a low risk TIC), one metastasis from colorectal origin where the initial time-point of the DCE images were already enhanced and TIC analysis misinterpreted as low risk curve. The borderline tumor was a serous cystadenoma which was colluded with ovarian fibroma and complicated of torsion. The 22 benign lesions included 9 myomas, 8 ovarian fibromas, 1 cystadenofibroma, 2 adnexal torsions, and 2 other diagnosis.

The mean size of purely solid masses misclassified according to O-RADS MR score was 74.9mm (min-max= 20–190mm)