EJGO European Journal of Gynaecological Oncology

# Interobserver variability and positive predictive value for ultrasonographic BI-RADS categories requiring pathohistological evaluation

**A. Dobrosavljevic[1], Z. Milosevic[2,5], S. Plesinac[3,5], A. Dmitrović[3], A. Jankovic[4], M. Nadrljanski[2], S. Rakic[1,5], V. Pazin[1,5], S. Jankovic Raznatovic[1,5], A. Jurisic[1,5]**

[1] *Clinic of Obstetrics and Gynecology" Narodni Front", Belgrade;* [2] *Institute for Oncology and Radiology of Serbia, Belgrade*
[3] *Clinic of Obstetrics and Gynecology of the Clinical Center of Serbia, Belgrade;* [4] *Euromedik General Hospital, Belgrade*
[5] *Medical Faculty, University of Belgrade, Belgrade (Serbia)*

**Summary**

*Objective:* The objective of this study was an analysis of interobserver variability and positive predictive value (PPV) for BI-RADS categories requiring pathohistological evaluation: 4A, 4B, 4C, and 5. *Material and Methods:* Interobserver variability for each of descriptors as well as PPV for final BI-RADS categories requiring pathohistological evaluation was measured in a retrospective study which included 30 ultrasonographic reports, with pathohistological verification, randomly selected from ultrasonographic reports from Institute for Oncology and Radiology of Serbia where about 1,100 breast cancers are verified every year. Ten observers, seven gynecologists, and three radiologists, independently rated each ultrasonographic report according to the fourth edition of BI-RADS atlas. Interobserver variability was measured with $k$ coefficient. *Results:* There was most conformity for a category of orientation ($k = 0.79$). Substantial degree of conformity was also present for both boundary ($k = 0.71$) and shape ($k = 0.65$) categories. Moderate degree of conformity was achieved for posterior features ($k = 0.54$) and margins ($k = 0.41$) descriptors, while there was poor conformity in echogenicity ($k = 0.38$). In case of a final score, common conformity for all BI-RADS 4A, 4B, 4C, and 5 categories was ($k = 0.51$); it was the greatest for category 5 ($k = 0.50$), and it was less for categories 4C ($k = 0.37$), 4B ($k = 0.32$), and 4A ($k = 0.29$). *Conclusions:* Interobserver conformity for ultrasonographic descriptors and final evaluation of BI-RADS 4A, 4B, 4C, and 5 categories is good. PPV implies that not only division into categories 4 and 5, but also classification into categories 4 and subcategories 4A, 4B, and 4C are justified and clinically applicable.

*Key words:* BI-RADS; Interobserver variability; PPV.

## Introduction

Diagnostics of breast changes includes clinical and imaging evaluation. Most commonly used imaging methods are mammography (M), ultrasonography (US), and magnetic resonance imaging (MRI). Various modalities of examination and their combinations are applied depending on the age of the patient. Mammography and ultrasonography are used in women 40 years old and older. Ultrasonography of breasts has been proved as useful in evaluation of lesions identified by either mammography or clinical examination, not only in terms of distinguishing cystic and solid, but to distinguish benign and malignant [1-4]. Ultrasonography has been used independently in patients younger than 40 years of age, and in case that it is necessary, patients are referred to further diagnostic procedures [5].

Mammography has been proved as an effective method in detection palpable as well as clinically occult breast cancers. However, a small number of even palpable cancers cannot be identified by mammography. False negative mammographic report rate is 4 - 34% [6-10]. In case of mammography, a failure to detect or characterize a change is a result of several factors: breast density, characteristics of lesion related to size, localization (unavailable to mammographic projections, like changes located deeply in glandular tissue immediately before prepectoral fat tissue or in subareoal region), absence of desmoplastic reaction, poor radiology technique in terms of inadequate positioning, compression or inappropriate exposition, as well as failure in interpretation of mammogram [11, 12].

If mammographic finding is ambiguous, the next step in evaluation is ultrasonography examination. Additional ultrasonography examination may identify early breast cancer without its spreading into regional lymph nodes and which has not been identified by mammography [13-17], and its accuracy is increased in case of glandular, dense breasts [13]. Berg *et al.* reported that combination of ultrasonography and mammography had higher diagnostic yield

Revised manuscript accepted for publication July 3, 2014



*7847050 Canada Inc.*
*www.irog.net*

being 11.8 per 1,000 women screened than only mammography being 7.6 per 1,000 women screened. Therefore, the supplemental yield was 4.2 per 1,000 women screened. There is similarity to rates of cancers identified only by ultrasonography and which are 2.7 to 4.6 cancers per 1,000 women screened in other series [13-17].

Ultrasound BI-RADS classification has been described for the first time in the forth edition of BI-RADS atlas, and it has been used for less time than this classification has been applied in mammography. Therefore, the objective of this paper is retrospective evaluation of interobserver variability in description of breast lesion between sonographers and PPV for the categories involving pathohistological analysis (BI-RADS 4A, 4B, 4C, and 5).


Figure 1. — Ultrasonography finding of a solid tumor change categorized as BI-RADS 5 by all observers. Pathohistological evaluation confirmed a malignant change.

## Materials and Methods

Ten physicians, seven gynecologists, and three radiologists analyzed ultrasonographic images of BI-RADS 4 and 5 categories in a retrospective study. This study involved a group of randomly selected 30 women who had ultrasonography breast examination in the Clinic for Radiotherapy and Radiology Diagnostics, being a part of the Institute for Oncology and Radiology of Serbia where about 1,100 breast cancers are verified every year. All these women, after having clinical, mammography, and ultrasonography breast examination, also had surgical biopsy of breasts with pathohistological verification of change within the breast during 2011.

A change detected by ultrasonography, was rated according to BI-RADS classification as BI-RADS 4A (low suspicious of malignancy), BI-RADS 4B (intermediate suspicious of malignancy), BI-RADS 4C (moderate suspicious of malignancy), and BI-RADS 5 (highly suspicious of malignancy), being the final assessment of malignancy risk and the reason for setting indication for pathohistologic analysis of the change.

Ultrasonography of the breast was performed by 5-12 MHz linear probe and 5-9 MHz probe. Ultrasonography of breasts was performed through standard approach, with woman being in supination and lateral decubitus position, with examination of regional lymph nodes and use of power Doppler color signalization.

Several features, such as shape, margins, long axis orientation, inner echo texture, boundary, and acoustic transmission form basis for US mass interpretation according to the fourth edition of BI-RADS lexicon comprising criteria for ultrasonographic evaluation. Benign structures are oval, with circumscribed margins having not more than two or three slight lobulations, with long axis parallel to skin, predominantly hypoechoic uniform internal echo texture, abrupt interface with surrounding tissue, and with no elements implying malignancy. Malignancy criteria comprise irregular shape, microlobulation, indistinctiveness, angular structure, speculated margins, nonparallel orientation, hypoechoic appearance or heterogeneous echo texture, and posterior acoustic shading [19-22]. For further diagnostic procedures all available data were taken into account. Each case was assigned into one of final BI-RADS category according to ACR definitions.

Pathohistologic (PH) analysis involved ex tempore check of samples gained through excision biopsy, followed by standard analysis of samples stained by hematoxylin eosin.

The Cohen $k$ statistic was used to asses inter-reader agreement for all variables. The PPV values were determined by using data from the assessments of all readers. Based on $k$ coefficient, a degree of conformity is assigned into the following categories: 0-0.20 low; 0.21 - 0.40 fair; 0.41 - 0.60 moderate; 0.61 - 0.80 substantial, and 0.81 - 1 nearly perfect.

## Results

Considering sonographic descriptors in the present study, the authors registered the greatest, substantial interobserver conformity for orientation ($k = 0.79$) being defined as either parallel or vertical. Substantial conformity was also registered in evaluation of boundary described as abrupt or with presence of echogenic halo ($k = 0.71$). Evaluation of lesion shape was also within statistically substantial conformity ($k = 0.65$), and it was the greatest when it was irregular shape ($k = 0.61$), then oval ($k = 0.41$), lobular ($k = 0.31$), and it was the slightest for circular shape ($k = 0.21$).

There was moderate conformity for descriptor posterior features ($k = 0.54$), and it was the greatest for attenuation feature ($k = 0.41$), then amplification ($k = 0.39$), and absence of features ($k = 0.37$). In case of alternating posterior features, there was the slightest degree of conformity ($k = 0.19$) and this can be explained by the fact that even three observers did not register this kind of posterior features. There was similar, moderate degree of conformity in the evaluation of contour ($k = 0.41$) and as follows: the greatest in case of well-defined ($k = 0.71$), while in case of poorly defined, spicular, and microlobular ($k = 0.39$, 0.31, and 0.25, respectively). There was the least conformity for angular contour ($k = 0.19$).

Fair degree of conformity was registered in echogenicity ($k = 0.38$). For lesion characterized as hypoechogenic there was the greatest degree of conformity ($k = 0.39$), than for complex ($k = 0.29$), and slightly less for hyperechogenic ($k = 0.19$), and isoechogenic characterization ($k = 0.09$).

Table 1. — *Interobserver variability in descriptors according to US BI-RADS lexicon.*

| Descriptors and final evaluation | $k$ values in our study | $k$ values in a study by Lazarus *et al.* [24] | $k$ values in a study by Park *et al.* [25] |
|---|---|---|---|
| Shape | 0.65 | 0.66 | 0.42 |
| Orientation | 0.79 | 0.61 | 0.61 |
| Margin | 0.41 | 0.4 | 0.32 |
| Boundary | 0.71 | 0.69 | 0.55 |
| Echogenicity | 0.38 | 0.29 | 0.36 |
| Posterior features | 0.54 | 0.4 | 0.53 |
| Final category | 0.51 4A, 4B, 4C, and 5 | 0.28 For all categories | 0.49 For all categories |


Figure 2. — This image shows a malignant change where there is no conformity regarding description of a contour although all observes categorized it as BI-RADS 5. Three observers described it as microlobular, i.e. spicular, while two observers consider it as angular, and two as non-defined contours.

In case of the final rate of lesion, on the whole, when every BI-RADS category was considered (4A, 4B, 4C, 5), there was moderate conformity ($k$ = 0.50). There was the greatest conformity for a category being highly suspicious of malignant lesion, BI-RADS 5 ($k$ = 0.51), Figure 1, while in case of 4C, 4B, and 4A categories there was slight degree of conformity ($k$ = 0.37, 0.32, and 0.29, respectively).

Pathohistological reports were available for every observed change. Samples were obtained by surgical biopsy. Combination of BI-RADS rates provided by all observers yielded PPV for observed categories are as follows: category 5 (94%), category 4c (79%), category 4b (38%). and category 4a (9%).

## Discussion

A mammographic BI-RADS lexicon has been clinically used for more than a decade through four successive editions. History of an ultrasound BI-RADS lexicon is shorter, for it was published in the fourth edition of BI-RADS atlas for the first time, in 2003 [23] and therefore there are less data on use of the ultrasound and BI-RADS nomenclature in practice compared to the use of mammography. Most commonly used criteria for evaluation of nature of a breast change implied Stavros criteria which characterize certainly benign lesion as the one fulfilling the following conditions: first of all, absence of the following characteristics: spiculated, microlobular or angular contours, vertical orientation, ductal distribution, branch type of lesion, hypoechogenicity, posterior attenuation of ultrasonographic beam, and calcifications. The second criterion is presence in at least one of the mentioned features: homogeneous hyperechogenic lesion, oval or macrolobular shape with thin hyperechogenic margin [3]. By using Stavros criteria, ultrasonographers could evaluate the type of lesion. What was missing was uniform, standardized descriptive terminology which resulted in inconsistency of diagnosis in relation to ultrasonographic description of a change. Introduction of standardized BI-RADS terminology into ultrasonographic diagnostic aimed to reduce interobserver variability during characterizing of a lesion as well as better communication between physicians dealing with breast pathology. According to the data from the literature, the greatest objectiveness and conformity in evaluation among various physicians who perform examination exists for the criterion of tumor change orientation (horizontal, i.e. parallel to pectoral fascia, intrinsic to benign changes, or vertical, i.e. perpendicular, typical for malignant changes) and shape (irregular for malignant lesion, while oval, macrolobular are more common features of benign changes) [24]. Results of the present study conform to the results of other authors in this matter. There was the most conformity for a category of orientation ($k$ = 0.79). Assessment whether it is either parallel of vertical change is generally easy and it explains slight degree of interobserver variability (Table 1). The next one by statistical significance is a degree of conformity relating to a descriptor boundary ($k$ = 0.71) which could also be explained in a similar way, namely with a fact that there are only two characterizations: abrupt, i.e. present hyperechogenic halo. This result also conforms to results of other observers. [24, 25]. In a study conducted by Lazarus *et al.* the highest degree of conformity between observers as compared to other descriptors was obtained for this descriptor ($k$ = 0.69) [24].

There is the least conformity in evaluation of the margin of the change (circumscribed margin usually in benign changes, not circumscribed or poorly defined usually in malignant changes), and its echogenicity (benign changes

Table 2. — *Interobserver variability in the final evaluation of BI-RADS categories.*

| BI-RADS category | K value in this study | K values in a study by Lazarus *et al.* |
|---|---|---|
| 4A | $k = 0.29$ | $k = 0.14$ |
| 4B | $k = 0.32$ | $k = 0.16$ |
| 4C | $k = 0.37$ | $k = 0.26$ |
| 5 | $k = 0.51$ | $k = 0.56$ |

are either anechoic or hyperechoic while hypoechoic and isoechoic changes, and complex lesions are noted in both kinds of changes, but more suspicious of malignant changes), while the posterior phenomena are somewhere between these two extremities [24, 25]. The present study also revealed similar results: there was the least conformity in echogenicity ($k = 0.38$), while in case of evaluation of margins there was moderate level of conformity, but on the border with slight (k=0.41). Lazarus et al. also obtained similar results, while in a study conducted by Park *et al.* [25] there was the least conformity for descriptor margin, and slightly higher for echogenicity. However, conformity for both descriptors still belongs to a zone of fair conformity of pronounced k values which can be seen in Table 1. Since evaluation of margins is one of the key features of the lesion, particularly when we have to make a decision if it should be subjected to biopsy, this determinant may have significant effect on the final BI-RADS score. There is probably less conformity regarding a fact that there are often several features of the margin of the same lesion which can make accurate description difficult (Figure 2).

When it is about the final score of categories requiring pathohistological evaluation, our results reveal a moderate degree of conformity considering all categories of that type (k=0.50).

Ultrasonographically, BI-RADS 5 category implies irregular, star-like lesion, heterogeneously hypoechogenic, with hyperechogenic margins, interrupted ligaments, posterior attenuation of ultrasonographic beam and thick skin. According to the literature, this is a typical ultrasound view of malignant tumors, with carcinoma specificity of even by 98% [26]. In our study, a conformity degree for this category was the highest (k=0.51) and this also conforms to the results of other authors: Park et al. (k=0.51) and Lazarus et al. (k=0.56) [24, 25].

BI-RADS 4 category represents a quite heterogeneous group of changes which is due to great differences among pathohistological reports divided into 3 categories to obtain as homogenous subcategories as possible in terms of prediction of malignancy. Probability of malignancy in BI-RADS 4 category ranges between 2% and 95% depending on subcategories where the change has been assigned to BI-RADS 4A, 4B or 4C [20]. Therefore, it is logical regarding heterogeneity of lesion that a degree of conformity in these subcategories is less than in category 5. This degree becomes less from a category being highly suspicious of malignancy 4C (k=0.37) to categories 4B (k=0.32) and 4A (k=0.29) in our study as well in a study conducted by Lazarus et al. (Table 2) [24]. However, PPV for subcategories 4A, 4B and 4C implies that stratification into subcategories depending on probability of malignant lesion is useful and it can have a role in a therapy approach.

Finally, total conformity for all categories is slightly higher in our study as compared to already mentioned studies, but it should be taken into account that this study evaluated only interobserver variability for categories requiring pathohistological evaluation, while studies conducted by Lazarus et al. and Park et al. comprised all BI-RADS categories [24, 25].

This study also has certain limitations. As in many other studies, a specific training related to implementation of BI-RADS categories was not conducted in this study, and it would probably reduce subjective factor in evaluation of descriptors as well as in the final score, particularly when it is about subcategories 4A, 4B, and 4C. Secondly, there is a relatively small number of ultrasonography reports, but it should be taken into account that they were randomly taken from an institution where about 1,100 breast cancers are verified every year. The next limitation is related to the fact that used images were static and they were evaluated most often from one projection.

## Conclusions

Interobserver conformity for ultrasonographic descriptors and final evaluation of BI-RADS 4A, 4B, 4C, and 5 categories is good. This study has also validated correlation between BI-RADS 4 and 5 categories and pathohistological reports. The 4A report is nearly always related to benign changes and all other investigated categories with malignant (PPV for BI-RADS 5 is 94 %, for BI-RADS 4C 79%, BI-RAD 4B 38%, and 4A 9%). Therefore, it can be concluded that division into subcategories 4A, 4B, and 4C is useful and it can have a role in therapy approaches.

## References

[1] Raza S., Chikarmane S.A., Neilsen S.S., Zorn L.M., Birdwell R.L.: "BI-RADS 3, 4, and 5 lesions: value of US in management--follow-up and outcome". *Radiology,* 2008, 248, 773. doi: 10.1148/radiol.2483071786. Epub 2008 Jul 22.

[2] Rahbar G., Sie A.C., Hansen G.C., Prince J.S., Melany M.L., Reynolds H.E., *et al.*: "Benign versus malignant solid breast masses: US differentiation". Radiology 1999, *213*, 889.

[3] Stavros A.T., Thickman D., Rapp C.L., Dennis M.A.,Parker S.H., Sisney G.A.: "Solid breast nodules: use of sonography to distinguish between benign and malignant lesions". *Radiology*, 1995, *196*, 123.

[4] Skaane P., Engedal K.: "Analysis of sonographic feature in the differentiation of fibroadenoma and invasive ductal carcinoma". *AJR Am. J. Roentgenol.*, 1998, *170*, 109.

[5] Milosević Z., Karapandzić V.P., Jovićević A., Vuković V., Vucićević J., Pavlović T., Borojević N.: "Diagnosis of breast cancer in women age 40 and younger: mammography and breast ultrasound". *Acta Chir Iugosl.*, 2009, *56*, 77. [Article in Serbian]

[6] Bird R.E., Wallace T.W., Yankaskas B.C.: "Analysis of cancers missed at screening mammography". *Radiology*, 1992, *184*, 613.

[7] Goergen S.K., Evans J., Cohen G.P., MacMillan J.H.: "Characteristics of breast carcinomas missed by screening radiologists". *Radiology*, 1997, *204*, 131.

[8] Mann B.D., Giuliano A.E., Bassett L.W., Barber M.S., Hallauer W., Morton D.L.: "Delayed diagnosis of breast cancer as a result of normal mammograms". *Arch. Surg.*, 1983, *118*, 23.

[9] Wallis M.G., Walsh M.T., Lee J.R.: "A review of false negative mammography in a symptomatic population". *Clin. Radiol.*, 1991, *44*, 13.

[10] Cahill C.J., Boulter P.S., Gibbs N.M., Price J.L.: "Features of mammographically negative breast tumors". *Br. J. Surg.,* 1981, *68*, 882.

[11] Huynh P.T., Jarolimek A.M., Daye S.: "The false-negative mammogram". *Radiographics*, 1998, *18*, 1137.

[12] Siegal E.C., Angelakis E.J., Hartman A.: "Can peer review contribute to earlier detection of breast cancer? A quality initiative to learn from false-negative mammograms". *Breast J.*, 2008, *14*, 330.

[13] Buchberger W., Niehoff A., Obrist P., DeKoekkoek-Doll P., Dunser M.: "Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography". *Semin. Ultrasound CT MR*, 2000, *21*, 325.

[14] Crystal P., Strano S.D., Shcharynski S., Koretz M.J.: "Using sonography to screen women with mammographically dense breasts". AJR *Am. J. Roentgenol.*, 2003, *181*, 177.

[15] Gordon P.B., Goldenberg S.L.: "Malignant breast masses detected only by ultrasound. A retrospective review". *Cancer*, 1995, *76*, 626.

[16] Kaplan S.S.: "Clinical utility of bilateral whole-breast US in the evaluation of women with dense breast tissue". *Radiology*, 2001, *221*, 641.

[17] Leconte I, Feger C, Galant C, Berlière M., Berg B.V., D'Hoore W., Maldague B.: "Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density". *AJR Am. J. Roentgenol.*, 2003, *180*, 1675.

[18] Berg W.A., Blume J.D., Cormack J.B., Mendelson E.B., Lehrer D., Böhm-Vélez M., *et al.*: "Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer". *JAMA*, 2008, *299*, 2151. doi: 10.1001/jama.299. 18.2151.

[19] Kerlikowske K., Grady D., Rubin S.M., Sandrock C., Ernster V.L.: "Efficacy of screening mammography: a meta-analysis". *JAMA*, 1995, *273*, 149.

[20] D'Orsi C.J., Bassett L.W., Berg W.A., *et al.*: "BI-RADS: mammography". 4th ed. *In:* D'Orsi C.J., Mendelson E.B., Ikeda D.M., *et al.,* (eds). *Breast imaging reporting and data system: ACR BI-RADS – breast imaging atlas*. Reston (VA): American College of Radiology, 2003.

[21] Mendelson E.B., Baum J.K., Berg W.A., Merritt C.R., Rubin E.: "Ultrasound". *In: Breast imaging reporting and data system (BI-RADS).* 4th ed. Reston, Va: American College of Radiology, 2003.

[22] Ikeda D.M., Hylton M.N., Kuhl C.K., *et al.*: "Magnetic resonance imaging". 1st ed. *In:* D'Orsi C.J., Mendelson E.B., Ikeda D.M., *et al.,* (eds). *Breast imaging reporting and data system: ACR BI-RADS – breast imaging atlas*. Reston (VA): American College of Radiology, 2003.

[23] Levy L., Suissa M., Chiche J.F., Teman G., Martin B.: "BIRADS ultrasonography". *Eur. J. Radiol.*, 2007, *61*, 202.

[24] Lazarus E, .Mainiero M.B., Schepps B., Koelliker S.L., Livingston L.S.: "BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value". *Radiology*, 2006, *239*, 385. Epub 2006 Mar 28.

[25] Park C.S., Lee J.H., Yim H.W., Kang B.J., Kim H.S., Jung J.I., *et al.*: Observer agreement using the ACR Breast Imaging Reporting and Data System (BI-RADS)-ultrasound, First Edition (2003)". *Korean J. Radiol.*, 2007, *8*, 397.

[26] Pichler E.: "Ultrazvučni atlas dojke: diferencijalna dijagnoza i interventne tehnike". Zagreb: Skolska knjiga, 2005.

Address reprint requests to:
A. DOBROSAVLJEVIC, M.D.
Makisma Gorkog, 50
11000 Belgrade (Serbia)
e-mail: dobrosavljevical@gmail.com