Accuracy of sonographic findings in breast cancer: correlation between BI-RADS® categories and histological findings*

Acurácia dos achados ultrassonográficos do câncer de mama: correlação da classificação BI-RADS® e achados histológicos

José Hermes Ribas do Nascimento¹, Vinicius Duval da Silva², Antonio Carlos Maciel³

Abstract

OBJECTIVE: The main purpose of the present study is to evaluate the accuracy of ultrasonography (BI-RADS) in the diagnosis of breast cancer whereas the additional specific objectives are to describe the frequency of different sonographic findings and evaluating interobserver agreement. MATERIALS AND METHODS: Images of 110 patients who had been referred for biopsy with previous diagnosis of breast nodules were independently reviewed by two specialists according to the BI-RADS classification. Histological findings were utilized as a gold-standard. The accuracy of findings was determined. The chi-squared test for categorical variables was utilized in the analysis of the differences resulting from the groups comparison, and the interobserver agreement was calculated with kappa (κ) statistics. RESULTS: Among 110 breast masses evaluated by ultrasonography, 76 (69%) were benign and 34 (30.9%) were malignant. According to the radiologists, the sensitivity ranged from 70.5% to 82.3%, negative predictive value, from 81.1% to 87.5%, positive predictive value, from 42.1% to 45.1%, specificity from 56.58% to 55.2%, and accuracy from 60.9% to 63.6%. The global interobserver agreement was considered as moderate (κ = 0.50). CONCLUSION: The fourth edition of BI-RADS provides radiologists with an accurate clinical decision support system for the diagnosis and management of breast disease.

Keywords: Breast cancer; Ultrasonography; BI-RADS; Anatomopathological; Accuracy.

INTRODUCTION

Ultrasonography as an adjuvant to clinical examination and mammography is considered as the most effective method for the diagnosis of breast lesions (1). This is due to the technological development of ultrasound devices such as real-time transducers with high digital frequency of 7.5 MHz, 10 MHz and 13 MHz, and harmonic imaging. These multi-frequency transducers provide high-resolution images, depth penetration and a high number of scanning lines (2,3).

Although breast ultrasonography has been historically utilized for differentiating solid from liquid lesions, there is an increasing interest in the utilization of this method for differentiating malignant from benign masses. Additionally, ultrasonography has become a valuable tool in the characterization of nodules found at mammog-
raphy, thus avoiding unnecessary biopsies and eliminating the necessity of follow-up mammography\(^{4,5}\). The breast ultrasonography sensitivity has been reported as superior to that of mammography\(^{7,8}\) in premenopausal women and, recently, sono-
graphic screening has also been recom-
\(^{9}\) \mbox{mended for evaluation of dense breasts.} Studies have demonstrated the usefulness of ultrasonography for detection of clini-
cally and mammographically occult, non-
palpable breast carcinomas.\(^{10,11}\)

The Breast Imaging Reporting and Data
System (BI-RADS\(^{5}\)) lexicon for ultra-
songraphy was developed by the Amer-
ican College of Radiology (ACR) aiming at
increasing the clinical efficacy of the method and at standardizing the reports
organization and wording. There is a spe-
cific vocabulary for describing each lesion, and, at the end of the report, the lesion is
classified into categories ranging from 0 to 6 according to the findings suspicion de-
gree based on the positive predictive value
(PPV) of the study for breast cancer\(^{12-14}\).

The BI-RADS lexicon includes a sono-
diagnostic description of breast nodules or masses considering contours, orientation, margins, lesions limits, internal echoes pattern, characterization of posterior acoustic shadowing, borders and abnormalities in adjacent tissues. At the end of the de-
scription, the lesion is assigned to a BI-
RADS category\(^{13,15}\).

The present study primarily proposes an
evaluation of the sonographic BI-RADS
classification accuracy for differentiating benign lesions from malignant masses. The
secondary objectives were the description of the frequency of different sonographic
findings and the evaluation of the interob-
server agreement.

MATERIALS AND METHODS

Two physicians specialized in breast im-
ageing diagnosis independently reviewed
\(^{\text{studies of 110 patients referred to a clinic in the Northwestern region of the Rio
Grande do Sul state (Brazil) for core biopsy. Previously, all of them had been
sonographically diagnosed with breast
nodules or masses classified into BI-RADS
categories 3, 4 or 5. Each specialist, with
more than ten years of professional expe-
ience, course of residency in radiology, specialist title and/or course of specialization
in mammography by Colégio Brasileiro de Radiologia e Diagnóstico por Imag-
egem (CBR), blindly reviewed the sono-
graphic studies, utilizing the BI-RADS ter-
minology, evaluation and recommendations
included in the most recent lexicon for echography. Later, the reviewed studies
were compared with the anatomopa-
thological results.

The ultrasonography studies were performed with a high-resolution Sonoline G50 (Siemens Medical Solutions; Berlin,
Germany) equipment with 7.5 MHz and 10
MHz linear array transducers.

The accuracy of the BI-RADS classifi-
cation in ultrasonography was evaluated by
calculating sensitivity, specificity, PPV and
negative predictive value (NPV) for each of
the described characteristics, and in the
differentiation between benign and malig-
nant lesions. Histological findings were
utilized as standard criteria.

The interobserver agreement for the fi-
nal categories and separately for each cat-
\(^{\text{egory was analyzed by the kappa test (k)} \)
and the differences between groups were
analyzed through the chi-square test for
categorical variables.

The BI-RADS lexicon for ultrasonog-
raphy considers the following terms for de-
scribing breast nodules: shape, margins,
\(^{\text{orientation of the nodule in relation to the skin axis, lesion borders, internal echoes
\text{pattern, posterior acoustic characteristics
\text{and alterations in adjacent tissues.}}}
\)

After their description in compliance
with the BI-RADS criteria, all the lesions
were classified into categories 3, 4 or 5
(Chart 1).

Category 3 included the well-defined le-
sions, ovoid or rounded in shape (contour),

\begin{center}
\begin{tabular}{|c|}
\hline
\textbf{Chart 1} Final clinical conduct according to BI-RADS classification\(^{12}\). \\
\hline
\textbf{Incomplete evaluation} \\
Category 0 (zero): requires additional imaging evaluation \\
\textbf{Complete evaluation} \\
Category 1: negative \\
Category 2: negative findings \\
Category 3: probably benign findings – suggesting short interval follow-up \\
Category 4: suggestive of abnormality – biopsy should be considered (indeterminate) \\
Category 5: highly suggestive of malignancy – an appropriate conduct should be adopted \\
Category 6: biopsy-proved malignant disease \\
\hline
\end{tabular}
\end{center}
Sonographic accuracy in breast cancer

**Table 1** Distribution of false- and true-positive results and false- and true-negative results based on pathological and discriminate diagnoses – observer A.

<table>
<thead>
<tr>
<th>Test result</th>
<th>Positive disease</th>
<th>Negative disease</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>T+ (category 4, 5)</td>
<td>24 (42.11) TP</td>
<td>33 (57.89) FP</td>
<td>57 (100)</td>
</tr>
<tr>
<td>T– (category 3)</td>
<td>10 (18.87) FN</td>
<td>43 (81.13) TN</td>
<td>53 (100)</td>
</tr>
<tr>
<td>Total</td>
<td>34 (30.9)</td>
<td>76 (69.1)</td>
<td>110 (100)</td>
</tr>
</tbody>
</table>

Parameters

- Sensitivity: TP/(TP + FN) 70.59 (with disease and positive test)
- Specificity: TN/(TN + FP) 56.58 (without disease and negative test)
- Positive predictive value: TP/(TP + FP) 42.1
- Negative predictive value: TN/(TN + FN) 81.1
- Accuracy: (TP + TN)/Total 60.9

T+, positive test (lesion rated as category 4 or 5); T–, negative test (lesion rated as category 3); TP, true-positive; FP, false-positive; TN, true-negative; FN, false-negative.

**Table 2** Distribution of false- and true-positive results and false- and true-negative results based on pathological and discriminate diagnoses – observer B.

<table>
<thead>
<tr>
<th>Test result</th>
<th>Positive disease</th>
<th>Negative disease</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>T+ (category 4, 5)</td>
<td>28 (45.16) TP</td>
<td>34 (54.84) FP</td>
<td>62 (100)</td>
</tr>
<tr>
<td>T– (category 3)</td>
<td>6 (12.5) FN</td>
<td>42 (87.5) TN</td>
<td>48 (100)</td>
</tr>
<tr>
<td>Total</td>
<td>34 (30.9)</td>
<td>76 (69.1)</td>
<td>110 (100)</td>
</tr>
</tbody>
</table>

Parameters

- Sensitivity: TP/(TP + FN) 82.3 (with disease and positive test)
- Specificity: TN/(TN + FP) 55.2 (without disease and negative test)
- Positive predictive value: TP/(TP + FP) 45.1
- Negative predictive value: TN/(TN + FN) 87.5
- Accuracy: (TP + TN)/Total 63.6

T+, positive test (lesion rated as category 4 or 5); T–, negative test (lesion rated as category 3); TP, true-positive; FP, false-positive; TN, true-negative; FN, false-negative.

**Table 3** Distribution of margins and relationship with the BI-RADS classification for ultrasonography – observer A.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Category 3</th>
<th>Category 4</th>
<th>Category 5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Benign</td>
<td>Malignant</td>
<td>Benign</td>
<td>Malignant</td>
</tr>
<tr>
<td>Circumscribed — n (%)</td>
<td>37 (84.1)</td>
<td>7 (15.9)</td>
<td>17 (85.0)</td>
<td>3 (15.0)</td>
</tr>
<tr>
<td>Noncircumscribed — n (%)</td>
<td>6 (66.7)</td>
<td>3 (33.3)</td>
<td>12 (63.2)</td>
<td>7 (36.8)</td>
</tr>
<tr>
<td>Total</td>
<td>43</td>
<td>10</td>
<td>29</td>
<td>10</td>
</tr>
</tbody>
</table>

Radiol Bras. 2009 Jul/Ago;42(4):235–240 237
NPV for circumscribed margins was of 80.3%. Sensitivity was 64.7%, and specificity, 64.5%.

c) Evaluation of internal echoes patterns

Internal echoes patterns were observed as follows: hypoechoic in 85 cases, isoechoic in two, hyperechoic in two, and complex in 14 cases. All the hyperechoic lesions were benign, and 71.4% of lesions with complex pattern were benign. Among the 85 (77.3%) hypoechoic nodules, 26 were malignant, with NPV of 30.6%.

According to the observer B, the internal echoes pattern were hypoechoic in 79 cases, isoechoic in two, hyperechoic in five, and complex in 17 cases. Particularly, the complex pattern represented 15 (92.8%) of the nodules classified as categories 4 and 5. All the hyperechoic lesions were benign, and 58.8% of lesions with complex pattern were benign. Among the 79 (71.8%) hypoechoic nodules, 22 were malignant, with NPV of 27.8%.

d) Evaluation of nodule orientation in relation to the skin axis

For both observers, the mean size of the lesions parallel to the skin axis was 14.2 ± 9.9 mm, and 9.4 ± 4.5 mm in the lesions with vertical orientation, with prevalence in the upper breast quadrants.

Parallel orientation in relation to the skin axis was present in 101 cases (28 malignant, and 73 benign lesions) for a NPV of 72.3%. Antiparallel orientation was present in seven cases (four malignant and three benign lesions) for a PPV of 72.3%.

e) Description of the posterior acoustic characteristic

According to the observer A, the absence of posterior acoustic characteristic presented a NPV of 58.6%. Out of the 110 lesions, 34 presented posterior acoustic enhancement and, among them, 29 were benign, with NPV of 85.3%; and posterior acoustic shadowing was described in 11 cases, of which four were malignant, with PPV of 36.4%.

According the observer B, 31 (28.1%) of cases presented abrupt interface, and 66 (60%) of the 110 cases demonstrated echogenic halo. Echogenic halo was described in 27 cases classified as category 3, with NPV of 72.3%. Abrupt interface presented NPV of 68.4%.

According to the observer B, 31 (28.1%) of cases presented abrupt interface, and 66 (60%) of the 110 cases demonstrated echogenic halo. Echogenic halo was described in 24 cases classified as category 3, with NPV of 72.1%. Abrupt interface presented NPV of 61.2%.

f) Description of lesions borders

According to the observer A, 38 (34.5%) of 110 cases presented abrupt interface and 65 (59.0%) demonstrated echogenic halo. Echogenic halo was described in 27 cases classified as category 3, with NPV of 72.3%. Abrupt interface presented NPV of 68.4%.

According to the observer B, 31 (28.1%) of cases presented abrupt interface, and 66 (60%) of the 110 cases demonstrated echogenic halo. Echogenic halo was described in 24 cases classified as category 3, with NPV of 72.1%. Abrupt interface presented NPV of 61.2%.

g) Adjacent tissues appearance

According the observer A, 96 of the 110 described masses did not present alteration in adjacent tissues and, among them 52 were classified as category 3, with NPV of 76.6%. Out of 24 malignant lesions, 11 (45%) presented alteration in adjacent tissues, with PPV of 45%. Skin thickening was not observed in any case.

According the observer B, 96 of the 110 described masses did not present alteration in adjacent tissues and, among them 45 were classified as category 3, with NPV of 72.6%. Out of 34 malignant lesions, 26 (76%) presented alteration in adjacent tissues, with PPV of 76%. Skin thickening was not observed in any case.

h) Interobserver agreement (Table 5)

As far as the sonographic description is concerned, a moderate interobserver agreement was observed in the evaluation of nodules orientation (κ = 0.52), that was described as parallel or antiparallel in relation to the skin axis.

A moderate agreement was observed in the evaluation of the lesions contour (κ = 0.50). Low interobserver agreement (κ = 0.29) was observed in the evaluation of the lesion borders.

Moderate agreement was also observed in the evaluation of the lesion margins (κ = 0.53) and in the description of internal echoes pattern (κ = 0.56).

The different terms utilized for describing posterior acoustic characteristic has also determined a moderate interobserver agreement (κ = 0.51).

A moderate interobserver agreement (κ = 0.51) was also observed in the evaluation of adjacent tissues, especially in cases where no alteration was found.

The κ value, for unified categories 4 and 5 was 0.36. The prevalence of breast cancer in the present study was of 34 (30.9%).

**DISCUSSION**

The BI-RADS classification for mammography was the first attempt to standard-
In the present study, the sonographic accuracy ranged from 60.9% to 63.6% in the differentiation between benign and malignant lesions with the utilization of the BI-RADS. The PPV for the category 3 ranged from 81.1% to 87.5% between the observers, with a PPV ranging between 42.1% and 45.1%, similarly to the studies developed by Costantini et al. (15) and Roveda Jr et al. (20), who have demonstrated a PPV ranging between, respectively, 92.3% and 70.58% for category 3.

Thus, the utilization of the category 3, as probably benign, is a tool utilized by radiologists to avoid unnecessary biopsies, considering that the risk for malignancy of lesions described in this category corresponds to less than 2% (12). If an increase in the lesions dimensions were observed in the follow-up, there would be a trend towards changing to BI-RADS category 4, so that the biopsy could be appropriate.

The classification of breast nodules into category 4 presents the same clinical impact and meaning as those described for category 5, since in both cases biopsy would be indicated. In the present study, the PPV for categories 4 and 5 was 0, respectively, 45.2% and 42.2%, similarly to the study developed by Roveda Jr et al. (20), with a 50% PPV in category 4.

The analysis of the sonographic characteristics associated with the classification into categories 4 and 5 demonstrated that lesions with proved malignancy were frequently associated with hypoechoicinity, irregular contours, noncircumscribed margins, and antiparallel orientation in relation to the skin axis, although many of the benign nodules classified as BI-RADS categories 4 and 5 were hypoechoic, even being associated with circumscribed margins and parallel orientation in relation to the skin axis.

It could be observed that with the presence of three of the following findings, such as posterior acoustic shadowing, irregular contours, noncircumscribed margins, hypoechoic halo and antiparallel orientation in relation to the skin axis, the lesions were normally classified into categories 4 and 5, in accordance with the findings described by Chen et al. (22).

Masses demonstrating more than three characteristics suggestive of malignancy were classified into category 5 by both observers.

Nodule margins represented a relevant criterion in the differentiation between benign and malignant lesions, with a NPV ranging between 82.4% and 80.3% for circumscribed margins, not very different from the findings reported by Calas et al. (23), who had observed a NPV of 97% for circumscribed lesions. In the present study, the PPV for noncircumscribed margins ranged between 52.4% and 44.9%, differently from the NPV described by Calas et al. (23), corresponding to 70.4%.

Rounded contour (shape) was associated with high NPV that ranged between 83.3% and 76.1%; and irregular contour, with high PPV, that ranged between 84% and 65% for both observers.

Hypoechoic halo demonstrated a lower PPV than irregular contour and noncircumscribed margins. Echogenic halos presented a NPV ranging between 72.3% and 72.1%, and abrupt interface presented a NPV ranging from 68.4% to 61.2% between the observers.

The posterior acoustic characteristic is a result of the sound attenuation. The posterior acoustic enhancement presents a NPV between 81% and 85%, for both observers. Posterior acoustic shadowing presented a low PPV, ranging between 15% and 35% in the BI-RADS categories 4 and 5. Although posterior acoustic shadowing is a sonographic characteristic of malignant lesions (15), this finding was not confirmed in the present study, being also observed in benign lesions. Fine bilateral shadowing was considered as a sign of benign lesion.

In the analysis of the interobserver variability, a moderate agreement was observed for the evaluation of the lesions orientation in relation to the skin axis (κ = 0.52), evaluation of contours (κ = 0.50), margins (κ = 0.53), posterior acoustic characteristic (κ = 0.51), internal echoes pattern (κ = 0.56) and evaluation of adjacent tissues (κ = 0.51); and low interobserver agreement was obtained for the evaluation of the lesions borders (κ = 0.29) (Table 5).

Radiol Bras. 2009 Jul/Ago;42(4):235–240
CONCLUSION

The sonographic evaluation of breasts utilizing the BI-RADS classification is an accurate method, with the interobserver variability ranging between 60.9% and 63.3% in the differentiation of malignant from benign lesions. The most frequent sonographic findings of neoplasms were irregular nodules with noncircumscribed margins and antiparallel orientation. In the present study, complex or hypoechoic internal echoes pattern, the hypoechoic borders of the lesions, and posterior acoustic shadowing presented low PPV. The overall interobserver variability was moderate.

It is believed that the practice, the systematic follow-up periods, the double-reading technique, and BI-RADS training courses for physicians should be implemented to improve even more the accuracy in the diagnosis of breast diseases, thus reducing the number of unnecessary and expensive invasive procedures.

REFERENCES