Introduction

Breast cancer is the most common neoplasia in women and the second most common cancer type worldwide. Each year breast cancer accounts for about 22% of new cancer cases in women with an annual incidence estimated in 2008 in Brazil at 49,400 new cases, with an estimated risk of 51 cases out of 100,000 women, representing the main cause of death by cancer in women [1]. Unfortunately, it is not possible to predict who will develop breast cancer [2].

Category 4 of BI-RADS® includes those lesions that have a 3 to 94% chance of malignancy. Due to this great variation, the last edition of BI-RADS® kept stratification in 4A as low suspicion; 4B as intermediary suspicion and 4C as moderate suspicion. The guide chapter of the new BI-RADS® contained in the recent update of the American College of Radiology quotes some examples of how to conduct this stratification. What we notice in practice is the subdivision done subjectively or oftentimes not guided by breast echography. There is also a paucity of data in literature on the probabilities of cancer for each subdivision of category 4 of BI-RADS®. The lesions classified in this category require a cytological or histological sample as a next step in the investigation [3].

Despite scoring breast ultrasound findings for malignancy based on the criteria used for BI-RADS breast ultrasound having a high accuracy, comparable to that obtained by BI-RADS for mammography [4], classification into subdivision 4A, 4B, and 4C is poorly reproducible [5]. According to most of the radiologists who have specialized in breast imaging and medical doctors specialized in breast diseases, the BI-RADS system requires further refining [2]. We believe that refining BIRADS 4 stratification could help us guide the management with respect to the real need for surgical removal of the breast nodule.

It becomes necessary to have well-defined objective criteria for the subdivision of category 4 such that the biopsy indications can be optimized. We believe that 4A stratification is frequently an overestimation of BIRADS 3 and a short follow-up every 4–6 months would be enough to clarify the...
question helping to define benign alteration. Therefore, the purpose of this study is to evaluate the accuracy in the diagnosis of suspicious breast nodules by echographic study based on the last edition of the BIRADS® system for ultrasound, with subdivision of category 4 and inter-observer agreement with the use of the new lexicon.

Methods

A prospective study of diagnostic test was conducted by selecting patients for screening and diagnostic breast ultrasound, followed by mammography, from January to December 2007 in the Ultrasound & Radiology Services of Hospital Regina e Mãe de Deus Center, the first two in the city of Novo Hamburgo and the last in the city of Porto Alegre and by the Post-Graduate Pathology Program of UFCSPA, located in the state of Rio Grande do Sul- Brazil, after signing the informed consent form. The project was approved by the institutional review board of UFSCPA.

Of the patients sent for echography with or without tracking or diagnostic mammography, those who showed nodules in the echographic study were selected. The ultrasound exams were done by a group of radiologist doctors specialized in diagnostic imaging of the breast (n=6). The cases selected by the group of radiologist doctors specialized in the diagnostic imaging of the breast were reclassified by two doctors, also specialized, in a blind study in which the classification was redone and later compared. None of doctors had prior knowledge of any of the cases selected. The nodules were photographed in two incidences, analyzing the morphological characteristics, and grouping was done according to established criteria. A revision of the initial cases was carried out and two views selected for study were chosen by two of them (MTP and AAB) because of the scientific interest of these researchers. The radiologists’ group used only one ultrasound system, a 2005 Sonix Op device by Ultrasonix Medical Corporation Ultrasound Diagnostic, including a linear transducer with 10 to 14-MHz of frequency, with tracking in radial and transversal planes and the nomenclature used was recommended by the BI-RADS of 2013 in which the lexicon was extended to ultrasound.

A total of 70 patients were included, who showed solid nodules identifiable by ultrasound, with suspicious characteristics and were classified according to the BIRADS 4A, 4B, 4C, and 5 categories. We have not included BIRADS 3 because this category does not require anathomopathology diagnosis to rule out cancer.

After conducting the ultrasound, the mammograms were evaluated by observing the breast tissue type and associated changes. The following were excluded: patients with previous breast surgery, nodules in BI-RADS category 3, and other echographic findings that could not be characterized as nodules (e.g.: areas, complex cysts), patients in whom no anatomopathological results were obtained, and patients with metastatic disease for hidden cancer research.

These lesions were classified prospectively in the ultrasound, according to the morphological criteria of Sickles and Stavros, in the categories 4A, 4B, 4C, and 5 [6,7], taking into account the heterogeneity of the breast cancer characteristics, being biopsy becoming necessary in the presence of a suspicious finding.

The gold standard considered was the result of the anatomopathological exam, obtained in the nodules 4A by core biopsy and 4B, 4C, and 5 by core biopsy and surgical excision. If on analyzing the biopsy results, they were not as expected for the category suspected, then it would be re-discussed with the doctor and a new procedure, usually surgical excision, would be conducted to confirm the results.

Despite some examples of how to stratify BIRADS system having been supplied by BI-RADS (2013), after the recent update done by ACR, it does not determine objective criteria to subdivide subcategory 4 and most of the time we cannot find stratification directed to breast ultrasound. In our article we propose the only criteria more objective to each BIRADS 4 stratification for ultrasound as follows.

We defined that the nodules classified in category 4A should contain the criteria of nodules 3 (circumscribed margins and oval shape) and hypoechoic echogenicity. However, one of the margins has three microlobulations.

Despite some examples of how to stratify BIRADS system having been supplied by BI-RADS (2013), after the recent update done by ACR, it does not determine objective criteria to subdivide subcategory 4 and most of the time we cannot find stratification directed to breast ultrasound. In our article we propose the only criteria more objective to each BIRADS 4 stratification for ultrasound as follows.

We defined that the nodules classified in category 4A should contain the criteria of nodules 3 (circumscribed margins, homogenous echogenicity, posterior acoustic reinforcement, and greater longitudinal axis), being that if some of them were altered, for example heterogeneous echogenicity, it would denote a suspicious characteristic (Figure 1). In addition, nodules of recent appearance in women over 40 years and modifications in size of the nodules previously classified in category 3 remained in this category.

In the BI-RADS 4B category, we included nodules that had microlobulate margins associated with change in echogenicity (Figure 2).

Category 4C was defined as base in indistinct or angulated margins, marked hypoechogenicity and central acoustic shadowing posterior (Figure 3).

Figure 1: BIRADS 4A contains the criteria of nodules 3 (circumscribed margins and oval shape) in addition to heterogeneous echogenicity, which denote a suspicious characteristic.

Figure 2: BIRADS 4B contains the criteria of benign lesions: oval shape and hipoechoic echogenicity. However, one of the margins has three microlobulations.
In category 5, the nodules with suspicious characteristics were listed as well as those with 4 C, or spiculate margins, but those besides this were associated with mammographic changes with distortion of the breast tissue, speculations in the margins or suspicious microcalcifications, or associated with secondary signs, such as nipple retraction, thickening of the skin, or lymph node enlargement (Figure 4).

Statistical analysis was conducted by evaluating the accuracy, through calculations of sensitivity, specificity, positive and negative predictive value, and obtaining the area under the ROC (Receiver-Operator Characteristic) curve of the BI-RADS® characteristics described.

Furthermore, the prevalence was obtained in its respective BI-RADS ranges directed toward the breast echography exam and comparison of the agreement rate among different breast radiology experts by the Kappa method with p< 0.05 considered significant.

**Results**

In this preliminary study, 70 patients were selected. For each BI-RADS sub-category 4 and category 5, the number of patients is included as follows: 4A (n=20); 4B (n=7); 4C (n=16); 5 (n=27). 40 tested positive for malignancy were classified as invasive ductal carcinoma (n=35), invasive lobular cancer (n=3), adenocarcinoma (n=1) and in situ ductal cancer (n=1).

Considering the various classification ranges for category 4, there was no malignancy in any of the patients classified with category 4A nodules and in (2/7) 28.6%, (12/16) 75.0% and (26/27) 96.3% for categories 4B, 4C and 5, respectively.

Regarding BI-RADS accuracy of its descriptors, the positive predictive value (PPV) of irregular, oval, and lobulate shape are 80.8%; 11.1%, and 0%, respectively. Spiculate, indistinct, angular, microlobulate, and circumscribed margins had PPV of 92.8%; 80.8%; 60%; 27.3% and 0%, respectively. Regarding echogenicity, we divided descriptor in hypoechoic, heterogeneous, and isoechoic echogenicity and we observed PPV 62%, 45.4%, and 37.5%, respectively. Presence of posterior acoustic shadow was associated with PPV of 100%.

The agreement rates between two imaging experts resulted in a weighted Kappa of 0.90 (0.837 to 0.955), which denotes excellent agreement between different researchers when objective criteria are meticulously applied.

Figure 5 shows the area under the curve of nodules classified into BI-RADS categories. Area under the curve is 0.939 (CI 95% 0.85 to 0.98) p=0.0001.

Table 1 shows in detail the values of sensitivity, specificity, and predictive positive and negative values for each category of BI-RADS. As shown, 4C is the best accuracy point, with sensitivity of 92.5% (CI 95% 79.6 to 98.3) and specificity of 83.3% (CI 95% 65.3 to 94.3). Sonographic characteristics in breast nodules assigned to the BI-RADS categories 4A, 4B, 4C, and 5 are shown in Table 2.

**Discussion**

Our research included 70 patients who were women with suspicious breast nodules scheduled for echograph study for breast cancer without previous breast surgery in whom anathomopathology had been performed; 40 of them had malignancies. This study can be considered like an audit, because we can correctly access what has been done and how accurate the BI-RADS 4 stratification is.

Ultrasound has become in recent decades, especially as of 1980, an important diagnostic tool in its concomitant use with mammography and increasing diagnostic accuracy in dense breasts [8-16].
Among the main disadvantages of the ultrasound exam is the fact that it is operator-dependent, difficult to reproduce, and depends on the type of device used; thus only an experienced doctor would be able to locate and determine the characteristics of each lesion [2,17,18].

Accuracy of the breast imaging exams can be also affected by countless factors, such as technical aspects, differences related to the characteristics of the study population, age of patient, radiologist doctor’s experience, use of double-reading techniques or computer programs, as well as variability in interpretations of the radiologist doctor in the use of BI-RADS® [14,19].

In 1995, Stravos et al. [7] related suspicious morphological findings to breast cancer, classifying the lesions into three sequences: malignant and benign, depending on the presence of a suspicious finding or absence of this finding and, indeterminate, when not included in this criterion. The central point of this study was the demonstration of excellent rates of accuracy, with sensitivity of 98.4% and a negative predictive value of 99.5% for malignancy. This study is considered one of the bases for the creation of the BI-RADS lexicon for breast ultrasound. Countless other works have been conducted [3,20,21] among others. Studies have also been conducted to evaluate the predictive value of ultrasound in breast lesions, as is the case of the work by Pande et al. [22], conducted in 2003, in which sensitivity and specificity were 95% and 94.1% in the evaluation of palpable nodules and the study by Park et al., 2007 [23] that evaluated the inter-observer variability in the characterization of breast abnormalities making use of the new US lexicon with Kappa rates calculated for each variable found with a final Kappa rate of 0.65.

Today, there is a certain consensus on the morphological findings that involve higher level of suspicion of breast lesions [19].

Our study demonstrates that irregular shape, speculate, indistinct margins and presence of posterior acoustic shadow, features that could be evaluated by means of ultrasound examination were associated with PPV of 80.8% to 100%.

The Breast Imaging Report and Data System (BI-RADS®) of classification of the American College of Radiology (ACR) was introduced in 1993 with the aim of standardizing mammographic reports and guiding the mastology doctor with regard to the probability of a certain lesion being malignant, aiding in the most objective conduct of this investigation. In its fourth edition in 2003, there was an updating of terms for characterization of lesions (lexicon), including in addition to mammography, breast ultrasound and breast magnetic resonance, that was kept, in its fifth edition stratification in 4A as low suspicion; 4B as intermediary suspicion and 4C as moderate suspicion [3]. The American College of Radiology had done the original grading of BI-RADS. Regarding its accuracy, it has been demonstrated to be accurate, but with great variability (3 to 94%) while stratification of BI-RADS subcategory 4 were not performed. In addition, there are no conclusive data with regard to the probability of breast cancer for each BI-RADS 4 stratification. In practice, we can identify that the BI-RADS 4 subdivision is performed in a subjective manner or not related to ultrasound exam.

The importance of standardizing the correct evaluation of breast lesions is widely recognized, resulting in quality control programs of the ACR especially in Brazil to establish the use of BI-RADS®. BI-RADS of ultrasound was a radiological adaptation, using descriptors of the ACR especially in Brazil to establish the use of BI-RADS®. BI-RADS of ultrasound was a radiological adaptation, using descriptors of the ACR especially in Brazil to establish the use of BI-RADS® (lexicon), including in addition to mammography, breast ultrasound and breast magnetic resonance, that was kept, in its fifth edition stratification in 4A as low suspicion; 4B as intermediary suspicion and 4C as moderate suspicion [3]. The American College of Radiology had done the original grading of BI-RADS. Regarding its accuracy, it has been demonstrated to be accurate, but with great variability (3 to 94%) while stratification of BI-RADS subcategory 4 were not performed. In addition, there are no conclusive data with regard to the probability of breast cancer for each BI-RADS 4 stratification. In practice, we can identify that the BI-RADS 4 subdivision is performed in a subjective manner or not related to ultrasound exam.

Aiming to determine the performance of ultrasound variables and trying a greater uniformity in the evaluation and morphological classification of breast nodules, in the subdivision of categories 4A, 4B, 4C and 5, we classified the suspicious nodules into BI-RADS subcategory 4, comparing the observations of two radiologists experienced in breast cancer diagnosis in a double-blind study, and we assigned key clues to help us to define the best BI-RADS 4 subdivision for each case.

In our study, the agreement rates between two experts in breast imaging in the final categories resulted in a weighted Kappa of 0.939, denoting excellent agreement between different researchers when objective criteria were meticulously applied. It is worth mentioning that the two Radiologist Doctors, Breast Diagnostic Imaging experts, are actually working only 10 hours/a week in the same institution in addition to working 30 hours/a week in different institutions, as we need to take into account that when working in the same department, very often they tend to have the same opinions.

### Table 1: Diagnostic test for each BIRADS category.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Sensitivity</th>
<th>95% CI</th>
<th>Specificity</th>
<th>95% CI</th>
<th>+LR</th>
<th>-LR</th>
<th>+PV</th>
<th>-PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;=4A</td>
<td>100</td>
<td>91.1–100.0</td>
<td>0</td>
<td>0.0–11.7</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>&gt;4A</td>
<td>100</td>
<td>91.1–100.0</td>
<td>0</td>
<td>66.67</td>
<td>47.2–82.7</td>
<td>3</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>&gt;4B *</td>
<td>92.5</td>
<td>79.6–98.3</td>
<td>0</td>
<td>33.33</td>
<td>65.3–94.3</td>
<td>5</td>
<td>0</td>
<td>0.09</td>
</tr>
<tr>
<td>&gt;4C</td>
<td>60</td>
<td>43.3–75.1</td>
<td>0</td>
<td>96.67</td>
<td>82.7–99.4</td>
<td>18</td>
<td>0.41</td>
<td>66.7</td>
</tr>
<tr>
<td>&gt;5</td>
<td>0</td>
<td>0.0–8.9</td>
<td>100</td>
<td>88.3–100.0</td>
<td>1</td>
<td>90</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*4C is the best accuracy point

### Table 2: Sonographic characteristics in breast nodules assigned to the BI-RADS category 4A, 4B, 4C, and 5.

<table>
<thead>
<tr>
<th>Sonographic characteristics</th>
<th>BIRADS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4A</td>
</tr>
<tr>
<td>Shape</td>
<td></td>
</tr>
<tr>
<td>Irregular</td>
<td>2</td>
</tr>
<tr>
<td>Oval</td>
<td>13</td>
</tr>
<tr>
<td>Lobulated</td>
<td>5</td>
</tr>
<tr>
<td>Margins</td>
<td></td>
</tr>
<tr>
<td>Angular</td>
<td>0</td>
</tr>
<tr>
<td>Circumscribed</td>
<td>14</td>
</tr>
<tr>
<td>Spiculate</td>
<td>1</td>
</tr>
<tr>
<td>Indistinct</td>
<td>1</td>
</tr>
<tr>
<td>Microlobulate</td>
<td>4</td>
</tr>
<tr>
<td>Echogenicity</td>
<td></td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>4</td>
</tr>
<tr>
<td>Hypoechoic</td>
<td>11</td>
</tr>
<tr>
<td>Isoechoic</td>
<td>5</td>
</tr>
<tr>
<td>Posterior Acoustic shadow</td>
<td>0</td>
</tr>
</tbody>
</table>
Several studies were predominantly conducted evaluating the agreement between the radiologists when referring to morphological criteria for nodules and the final categories of BI-RADS. In 2006, in a study by Lazarus et al. [24], which evaluated the inter-observer agreement in the use of descriptors of lesions in mammography and ultrasound according to the fourth edition of BI-RADS in 4A, 4B and 4C, the Kappa index was found to be 0.28. In the US descriptors, the orientation had a Kappa of 0.61, with shape 0.66 and limit 0.69. In the margin, the Kappa was 0.4 and echogenicity was 0.29. Berg et al., in 2006 [25], also in an inter-observer agreement study in the detection of lesions and their categorization, obtained Kappa indices of 0.62 for shape, 0.67 for margins, and 0.51 for final evaluation of solid lesions. In 2007, Park et al. [23] also used descriptors of the fourth edition of BI-RADS for US, with inter-observer agreement indices (Kappa) for lesions smaller than 0.61 and intra-observer smaller than 0.73. Recently, Abdullah et al. [5] demonstrated that agreement for subdivisions 4a, 4b and 4c was fair.

Despite preliminary results, and we do recognize this weakness, our study is important because we have a base to improve and to analyze the indicators of BIRADS 4 stratification evaluated by ultrasound. Medical Doctors, who work with breast lesions, frequently recognize that BIRADS can facilitate their work, but also recognize that BIRADS system requires refining of its indicators. In spite of the limited sample of cases, the results observed here are promising and supply important information regarding the positive predictive value of each indicator. Further studies with larger sample of cases are necessary to confirm these findings.

As expected, BIRADS subdivision 4C has the highest accuracy for diagnosis of malignancy because features associated with 4C are closest to those associated with a BI-RADS grading 5. Although the findings are not pathognomonic, in our study we adopted more objective criteria for subcategory 4 and category 5 of nodules in US that guided our classification.

Based on our findings, we conclude that category 4C has the highest accuracy for diagnosis of malignancy, with sensitivity of 92.5% and specificity of 83.3%. The agreement rates between two imaging experts resulted in a weighted Kappa of 0.89, denoting good agreement between different radiologists when objective criteria were meticulously applied. This study demonstrates that irregular shape, indistinct margins, and presence of posterior acoustic shadow, features that can be evaluated by means of ultrasound examination were associated with PPV of 80.8% to 100%.

Acknowledgment

This study received financial support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We also thank Dr. Antonio Hartmann and Dr. Paulo Zen for suggestions regarding this article.

References
