

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA: CIÊNCIAS MÉDICAS**

ANDREA TEIXEIRA CADAVAL GONÇALVES

**VARIABILIDADE INTER E INTRA-OBSERVADOR DO SISTEMA BI-RADS®
INCLUIDO A SUBDIVISÃO DA CATEGORIA 4 PARA MICROCALCIFICAÇÕES
MAMÁRIAS ENCAMINHADAS PARA AGULHAMENTO PRÉ-OPERATÓRIO**

Porto Alegre

2006

ANDREA TEIXEIRA CADAVAL GONÇALVES

**VARIABILIDADE INTER E INTRA-OBSERVADOR DO SISTEMA BI-RADS®
INCLUIDO A SUBDIVISÃO DA CATEGORIA 4 PARA MICROCALCIFICAÇÕES
MAMÁRIAS ENCAMINHADAS PARA AGULHAMENTO PRÉ-OPERATÓRIO**

Dissertação de Mestrado apresentada
ao Programa de Pós-Graduação em Ciências
Médicas da Faculdade de Medicina da
Universidade Federal do Rio Grande do Sul

Orientadora: Profa. Dra. Maria Isabel Albano Edelweiss

Porto Alegre

2006

Para o meu grande amor e esposo Sandro, pelo seu estímulo, compreensão e genialidade, que me orientam nesta e em todas as etapas importantes da minha vida.

Para minha filha Lara, cuja presença me inspira e amplia minha visão do futuro.

AGRADECIMENTOS

Ao meu querido amigo e colega Dr. Álvaro Borba que com seu dinamismo e competência tornou possível a realização deste trabalho.

À professora Luciana Neves Nunes pelo seu criterioso auxílio e interesse na análise estatística.

Às minhas amigas e colegas Dra. Luciane Scherer e Dra. Cláudia Demarchi pela participação indispensável neste projeto.

A toda a equipe técnica e administrativa do Centro de Ecografia e Radiologia pela imensa contribuição.

Aos meus colegas da Serdil, Mãe de Deus Center e Centro de Ecografia e Radiologia que, com seu apoio, me deram condições para a conclusão deste projeto.

Ao querido Dr. Dakir Duarte responsável pela base do meu aprendizado e pelo meu interesse em mamografia.

À Profa. Dra. Sandra Costa Fuchs, pela sua competência e amizade e por ter me direcionado em momentos cruciais deste curso.

À minha amiga Ana Cláudia Tonelli de Oliveira, pelo seu estímulo e ajuda constante.

À minha professora orientadora Dra. Maria Isabel Edelweiss, que confiou na minha capacidade, pelo seu acolhimento e incentivo em todos os momentos.

À minha família pelo amor e compreensão incondicionais.

RESUMO

Objetivo: Determinar a variabilidade interobservador e intra-observador da terminologia da quarta edição do *Breast Imaging Reporting and Data System* (BI-RADS®) na avaliação de microcalcificações mamárias, incluindo a subdivisão da categoria 4. O objetivo secundário foi avaliar a acurácia das categorias finais do BI-RADS®.

Material e Método: Quatro médicos radiologistas especialistas em mamografia interpretaram, retrospectivamente, 109 mamografias de 89 pacientes encaminhadas para demarcação pré-operatória de microcalcificações. Foi utilizada a terminologia da quarta edição do BI-RADS®. Após seis meses foi realizada nova interpretação. O diagnóstico final baseou-se no resultado anatomopatológico da peça cirúrgica. A concordância entre os médicos para categorias finais e separadamente para cada categoria, morfologia e tipo de distribuição foi calculada através da estatística *kappa*. A acurácia foi determinada através da razão de verossimilhança (*likelihood ratio* - LR) e do valor preditivo positivo (VPP) para cada categoria do BI-RADS®.

Resultados: Foram estudadas 47 lesões malignas (46%), 29 carcinomas ductais *in situ* (CDIS) e 18 carcinomas ductais invasores (CDI), e 55 benignas (54%). Observou-se moderada concordância geral interobservador ($k = 0,49$ na primeira e segunda leitura) e intra-observador ($k = 0,57$) para todas as categorias finais. Na morfologia e na distribuição das microcalcificações houve baixa a moderada concordância geral (valores médios de k entre 0,27 a 0,49). As maiores inconsistências foram no uso das categorias 3, 4a e 4b e na morfologia amorfa. O VPP nas duas leituras para a subdivisão da categoria 4 foi 4a: 25% e 35%; 4b: 42% e 41%; e 4c: 60% e 74%. A acurácia foi menor nas categoriais 4a (LR, 0,39 e 0,63) e 4b (LR, 0,84 e 0,80). As categorias 4c (LR, 1,74 e 3,32) e 5 (LR, > 16,5 e > 20) estiveram associadas com alto risco para malignidade.

Conclusão: A subdivisão da categoria 4 e a terminologia da quarta edição do BI-RADS® demonstraram baixa a moderada concordância geral inter e intra-observador. Obteve-se boa estratificação da probabilidade de malignidade entre os casos de microcalcificações avaliados, com melhor acurácia nas categorias 4c e 5. Deve-se atentar para limitações na concordância das categorias 3, 4a e 4b e da morfologia amorfa e na acurácia nas categorias 4a e 4b para casos de microcalcificações com indicação de biópsia.

ABSTRACT

Purpose: To evaluate interobserver and intraobserver agreement by using the BI-RADS® (fourth edition) terminology and the final assessment including subcategorization of category 4 for breast microcalcifications. We also assess the accuracy of the BI-RADS® final categories.

Materials and Methods: Four breast radiologists retrospectively interpreted 109 mammograms showing microcalcifications in 89 patients who underwent needle-localization open breast biopsy according with the BI-RADS® lexicon (fourth edition). All cases were re-examined and classified again after six months by the same team. The final diagnosis was based upon the pathologic result of the surgical specimen. The reader agreement for final assessment category and for each category, morphology and distribution descriptions was measure by *kappa* statistics. The accuracy was determined with the likelihood ratio (LR) and the positive predictive value (PPV) for final assessment of each BI-RADS® category.

Results: Pathological examination revealed 47 malignant lesions (46%), 29 ductal carcinomas in situ (DCSI) and 18 invasive carcinomas (IDC) and 55 benign lesions (54%). Moderate agreement was achieved in interobserver ($k = 0.49$ in the first and in the second readings) and intraobserver ($k = 0.57$) of the final assessment of all categories. For microcalcifications morphology, the mean k was 0.49 for intraobserver and 0.27 and 0.34 for interobserver. For the distribution, the mean k was 0.48 for intraobserver and 0.42 and 0.39 for interobserver. We found some inconsistencies in assessment of final categories 3, 4a and 4b and amorphous morphology. PPV in the two readings sessions for subcategorization of category 4 was 4a: 25% and 35%; 4b: 42% and 41%; and 4c: 60% and 74%. The accuracy was lower in categories 4a (LR, 0.39 and 0.63) and 4b (LR, 0.84 and 0.80). The categories 4c (LR, 1.74 and 3.32) and 5 (LR, $> 16,5$ and > 20) were associated with high risk for malignancy.

Conclusion: The terminology of BI-RADS® fourth edition including subcategorization of category 4 was associated to an overall fair to moderate agreement. We found good stratification of the likelihood of malignancy with better accuracy for categories 4c and 5. Attention for inconsistencies in assessment of final BI-RADS® categories 3, 4a and 4b and amorphous or indistinct morphology and lower accuracy in categories 4a and 4b it is important concern for microcalcifications.

SUMÁRIO

INTRODUÇÃO.....	11
1. REVISÃO DA LITERATURA	13
1.1 Câncer de Mama e Rastreamento e Diagnóstico Mamográfico.....	13
1.2 Padronização de Laudos Mamográficos.....	16
1.3. Microcalcificações Mamárias	19
1.4. Estudos de Concordância em Laudos Mamográficos.....	24
2. OBJETIVOS	29
2.1. Objetivo Primário	29
2.2. Objetivo Secundário.....	29
3. BIBLIOGRAFIA.....	30
ARTIGO EM INGLÊS	35
ARTIGO EM PORTUGUÊS	54

LISTA DE FIGURAS

Figura 1 - Incidência mamográfica em perfil verdadeiro demonstrando o fio da demarcação pré-operatória por agulhamento	15
Figura 2 - Incidência mamográfica ampliada da peça cirúrgica, demonstrando a presença de microcalcificações.....	16
Figura 3: Desenho adaptado de Le Gal: nomenclatura proposta em 1984 para descrição das características morfológicas de microcalcificações	20
Figura 4 - Mamografia ampliada demonstrando microcalcificações puntiformes (<0,5mm), de distribuição segmentar	22
Figura 5: Mamografia ampliada demonstrando microcalcificações amorfas, de distribuição linear	14
Figure 1A to 1F (artigo em inglês) - Spot-magnification mammograms show microcalcifications	45
Fig. 1A a 1F (artigo em português) - Mamografias com incidências ampliadas de microcalcificações mamárias da amostra estudada	64

LISTA DE TABELAS

Tabela 1 - BI-RADS® 4ª edição, categorias finais	17
Tabela 2 - BI-RADS® 4ª edição, terminologia atual utilizada para descrição das microcalcificações	21
Tabela 3: Valores de Kappa e grau de concordância.....	25
Tabela 1 (artigos)	
Final BI-RADS® assessment categories (fourth edition).....	39
BI-RADS® 4ª edição, categorias finais	58
Tabela 2 (artigos)	
Interobserver and intraobserver variability in assigning final BI-RADS® assessment category	42
Variabilidade inter e intra-observador das categorias finais do BI-RADS® (quarta edição)	61
Tabela 3 (artigos)	
Interobserver and intraobserver variability for each final BI-RADS® categories. ...	42
Variabilidade interobservador entre os quatro médicos simultaneamente e intra-observador em cada categoria final do BI-RADS® (quarta edição)	62

Tabela 4 (artigos)

Interobserver and intraobserver variability in the morphology description of microcalcifications - BI-RADS® fourth edition terminology..... 44

Variabilidade interobservador entre os quatro médicos simultaneamente e intra-observador de acordo com a morfologia das microcalcificações (terminologia da quarta edição BI-RADS®) 63

Tabela 5 (artigos)

Interobserver and intraobserver variability in the distribution description of microcalcifications - BI-RADS® fourth edition terminology..... 46

Variabilidade interobservador entre os quatro médicos simultaneamente e intra-observador de acordo com a distribuição das microcalcificações (terminologia da quarta edição BI-RADS®) 65

Tabela 6 (artigos)

LR and PPVs of the consensus opinion for final BI-RADS® assessment category (fourth edition) in the first and second readings..... 47

Razão de verossimilhança (likelihood ratio - LR) e VPP para câncer de mama de acordo com as categorias do BI-RADS® – interpretação de consenso na primeira e na segunda leitura..... 66

LISTA DE ABREVIATURAS E SIGLAS

BI-RADS®	-	Breast Imaging Report and Data System
MQSA	-	Mammography Quality Standards Act
CADS	-	Computer-Aided Detection Systems
ACR	-	Colégio Americano de Radiologia
CBR	-	Colégio Brasileiro de Radiologia
VPP	-	Valor Preditivo Positivo
LR	-	Likelihood Ratio
k	-	Valores de <i>Kappa</i>
SPSS	-	Statistic Package for Social Science
CDI	-	Carcinoma Ductal Invasor
CDIS	-	Carcionoma Ductal <i>In Situ</i>

INTRODUÇÃO

O câncer de mama é a principal causa de morte por câncer e o tipo mais freqüente entre as mulheres (23% de todas as neoplasias), sendo considerado também o câncer mais prevalente no mundo¹. O rastreamento mamográfico é recomendado pela maioria das diretrizes mundiais a partir dos 40 anos de idade, embora alguns estudos tenham demonstrado que o seu maior impacto na redução da mortalidade (20%-35%) ocorre em mulheres de 50 a 69 anos^{6,7,8,9,10,11,12,13,14}. Sabe-se que a acurácia da mamografia pode ser afetada por inúmeros fatores, entre eles a redução da concordância intra e interobservador nos laudos mamográficos^{4,5,6,7,25}.

As microcalcificações mamárias constituem um dos achados mamográficos relacionados com estágios iniciais de câncer de mama, com uma taxa aproximada de malignidade de 22% a 37% em lesões referidas para biópsia. Cerca de 30% a 50% dos cânceres de mama não-palpáveis apresentam-se apenas como microcalcificações^{41,42,43,44,45,16}. As microcalcificações mamárias também são apontadas como um dos achados mamográficos com maior variabilidade na sua descrição entre os médicos radiologistas^{41,42,43,51,57}.

Na tentativa de padronizar a descrição das lesões e a recomendação de condutas em laudos mamográficos, o Colégio Americano de Radiologia (ACR) e o Colégio Brasileiro de Radiologia (CBR) recomendam a utilização do *Breast Imaging Reporting and Data System* (BI-RADS®)^{35,40}. Este sistema utiliza termos específicos para descrição da morfologia e distribuição das microcalcificações identificadas na mamografia e sugere que no final do laudo seja emitida uma classificação de acordo com o grau de suspeição de câncer de mama. Na última edição do BI-RADS®^{35,40} ocorreram significativas alterações, incluindo uma subdivisão opcional da categoria 4 em baixa suspeita (categoria 4a), intermediária suspeita (categoria 4b) e moderada

suspeita (categoria 4c). Uma das vantagens da subdivisão seria a possibilidade de melhor estratificação das probabilidades de câncer de acordo com as categorias, com otimização das indicações de biópsia, o que poderia resultar aumento da acurácia da mamografia. Entretanto, antes e após a quarta edição do BI-RADS®, vários trabalhos têm demonstrado considerável variabilidade na caracterização e classificação final das lesões pelos médicos radiologistas, em especial quando se tratam de microcalcificações, e observa-se que muitas das modificações sugeridas nesta última edição carecem de dados na literatura que as apoiem^{6,7,51,57,58,59,61,63, 64}.

1. REVISÃO DA LITERATURA:

1.1. Rastreamento e Diagnóstico Mamográfico do Câncer de Mama

O câncer de mama é a neoplasia mais freqüente no sexo feminino e a principal causa de morte por câncer nas mulheres, com 1,15 milhões de novos casos e 411 mil mortes estimados no ano de 2002, apontado também como o tipo de câncer mais prevalente no mundo¹. De acordo com o Instituto Nacional do Câncer (INCA), o número de casos novos de câncer de mama esperados para o Brasil em 2005 foi de 49.470, com um risco estimado de 53 casos a cada 100 mil mulheres. Na região Sudeste, o câncer de mama é o mais incidente, entre as mulheres, com um risco estimado de 73 casos novos por 100 mil. Sem considerar os tumores de pele não-melanoma, este tipo de câncer, também é o mais freqüente nas mulheres das regiões Sul (71/100.000), Centro-Oeste (38/100.000) e Nordeste (27/100.000)².

O prognóstico do câncer de mama é relativamente bom se diagnosticado nos estágios iniciais. Estima-se que a sobrevida média geral cumulativa, após cinco anos, seja de 65% nos países desenvolvidos e de 56% para os países em desenvolvimento³. Na população mundial, a sobrevida média após cinco anos é de 61%¹. No Brasil, as taxas de mortalidade por câncer de mama continuam elevadas, muito provavelmente, porque a doença ainda seja diagnosticada em estádios avançados⁴.

Ensaio clínicos randomizados têm demonstrado que a mamografia de rastreamento pode reduzir a mortalidade, sendo o maior impacto na faixa etária de 50 a 69 anos, com uma redução da mortalidade em torno de 25% (20%-35%) neste grupo^{5,6,7,8,9,10,11,12,13,14}. Embora estes achados tenham sido criticados em metanálise recentemente publicada¹⁵, a maioria das diretrizes mundiais, indica a mamografia para mulheres acima de 40 anos^{5,15,16,17,18}. Além da mamografia de rastreamento, existe a mamografia diagnóstica, utilizada em mulheres sintomáticas ou com um

exame anormal, com o objetivo de caracterizar melhor a alteração encontrada e indicar o próximo passo da investigação. De acordo com Elmore e cols.⁵ uma em dez mulheres com alteração palpável ou com mamografia anormal realmente terá câncer de mama. Os resultados de sete programas de rastreamento com base comunitária nos Estados Unidos, revelaram sensibilidade de 75% e especificidade de 92,3% da mamografia para o diagnóstico de câncer de mama^{5,17}. Porém em mamas densas e em pacientes com menos de 50 anos a sensibilidade é mais baixa, entre 63% e 69%¹⁷. Até o momento a mamografia digital e a mamografia convencional, não apresentaram diferença significativa para acurácia¹⁸.

A acurácia da mamografia pode ser afetada por inúmeros fatores, entre eles, aspectos técnicos, diferenças relacionadas às características da população em estudo, idade da paciente, experiência do médico radiologista e utilização de técnicas de dupla leitura ou de programas de computador dedicados para uma primeira leitura mamográfica (*Computer-Aided Detection Systems - CADs*) utilizados em mamografia digital ou imagem digitalizada^{19,20,21}. Vários pesquisadores têm demonstrado que existe variabilidade nas interpretações de mamografias entre os médicos radiologistas, alterando a acurácia da mamografia, a efetividade dos programas de rastreamento e do diagnóstico do câncer de mama^{4,22,23,24,25}.

Após a identificação de uma lesão mamária por um método de imagem podem ser realizadas a punção ou biópsia percutânea ou a biópsia excisional. As opções disponíveis para técnica percutânea são: punção aspirativa com agulha fina (PAAF) para avaliação citológica, biópsia com pistola automática (*core biopsy*) ou mamotomia (biópsia direcional assistida a vácuo) para avaliação histológica. Para a biópsia excisional, usualmente realiza-se a demarcação pré-cirúrgica, para nódulos não-palpáveis preferencialmente com orientação ecográfica e para microcalcificações com orientação mamográfica (figura 1 e 2) guiada pelo método biplanar ou pela estereotaxia^{26,27}. A demarcação pré-cirúrgica por técnica de agulhamento apresenta uma sensibilidade de em torno de 99%^{28,29}. Johnson e cols.³⁰ demonstraram redução de custos e tempo quando a demarcação pré-cirúrgica por agulhamento é realizada já como etapa inicial na investigação de microcalcificações suspeitas.

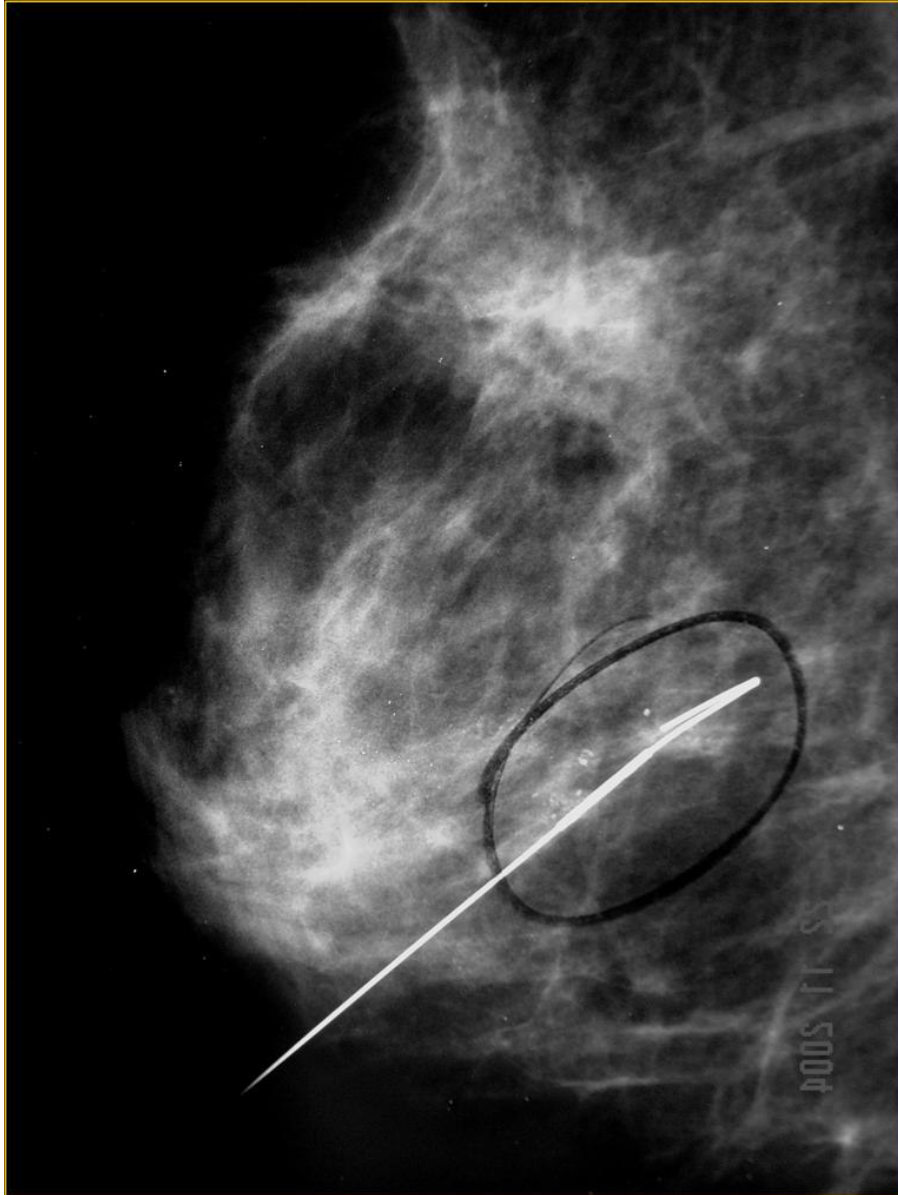


Figura 1: Incidência mamográfica em perfil verdadeiro demonstrando o fio da demarcação pré-operatória por agulhamento de microcalcificações mamárias.

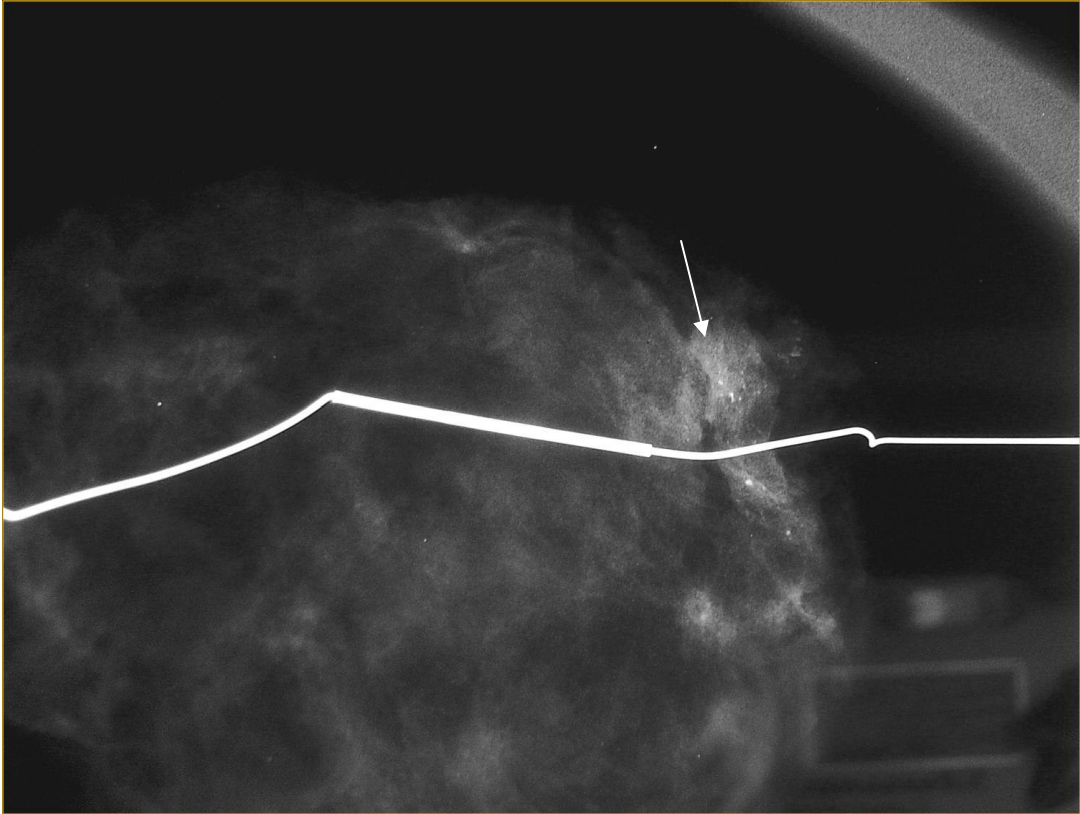


Figura 2: Incidência mamográfica ampliada da peça cirúrgica, demonstrando a presença de microcalcificações (seta).

As recomendações atuais preconizam um valor preditivo positivo (VPP) de 25% a 40% para câncer de mama entre as lesões que são encaminhadas para biópsia^{31,32}. Entretanto, um recente estudo³³ demonstrou que há uma tendência crescente nas indicações de biópsias mamárias, com aumento do número de resultados benignos e sem acréscimo do número de casos de câncer diagnosticados. Estes resultados são atribuídos em grande parte ao aumento na detecção e nas indicações de biópsias em casos de microcalcificações^{32,33}.

1.2. Padronização de Laudos Mamográficos

O programa de controle de qualidade da mamografia norte-americano, *Mammography Quality Standards Act* (MQSA)³⁴, preconiza o uso do *Breast Imaging Report and Data System* (BI-RADS®) do Colégio Americano de Radiologia (ACR)³⁵. O sistema de classificação BI-RADS® foi introduzido em 1993³⁶, com o objetivo de

padronizar os laudos de mamografia e orientar o médico mastologista quanto à chance de determinada lesão ser maligna, ajudando a conduzir a investigação. O BI-RADS® é um trabalho de órgãos Norte Americanos, entre eles, membros de vários Departamentos do Instituto Nacional do Câncer, de Centros de Controle e Prevenção da Patologia Mamária, da Administração de Alimentos e Drogas, da Associação Médica Americana, do Colégio Americano de Radiologia, do Colégio Americano de Cirurgiões e do Colégio Americano de Patologistas^{37,38}. Existe um vocabulário específico, para a descrição de cada lesão, e no final do laudo é emitida uma classificação (tabela 1), de acordo com o grau de suspeição, baseadas no VPP de cada categoria para câncer de mama. Na sua quarta edição ilustrada em 2003³⁸, inclui além da mamografia, a ecografia mamária e a ressonância magnética mamária. Definições ampliadas e novas terminologias para a caracterização das lesões (léxico) são introduzidas nesta versão. Propõem-se a adição da categoria 6 para pacientes com resultado já confirmado de câncer de mama e recomenda-se uma subdivisão opcional da categoria 4 em 4a, baixa suspeita de malignidade, 4b, intermediária suspeita de malignidade e 4c, moderada suspeita de malignidade. No momento muitas das modificações propostas não possuem dados na literatura que as apoiem, entretanto o comitê responsável julgou que as inclusões realizadas seriam necessárias para tornarem o léxico um instrumento prático que possibilite uma base unificada para pesquisa no diagnóstico por imagem da mama³⁸.

Tabela 1: BI-RADS® 4ª edição, categorias finais^{38,40}.

Categoria	Definições
1	Negativo
2	Achado(s) Benigno(s)
3	Achado Provavelmente Benigno
0	Necessária Avaliação Adicional
4a	Baixa Suspeita de Malignidade
4b	Intermediária Suspeita de Malignidade
4c	Moderada Suspeita de Malignidade
5	Altamente Sugestivo de Malignidade

No Brasil o primeiro consenso para padronização de laudos mamográficos foi em 1998, por iniciativa do Colégio Brasileiro de Radiologia (CBR), com a participação da Sociedade Brasileira de Mastologia e da Federação Brasileira das Sociedades de Ginecologia e Obstetrícia^{39,39}. A proposta seguiu as diretrizes internacionais, com a adoção do BI-RADS®. A tradução para a língua portuguesa do BI-RADS® foi disponibilizada apenas na sua quarta edição ilustrada, sendo realizada em 2005⁴⁰, um trabalho conjunto do CBR, em especial da Comissão Nacional de Mamografia do CBR, e do ACR, considerado o melhor método a ser adotado na confecção de laudos mamográficos, com o objetivo de tornar-se uma linguagem universal de fácil entendimento. Entretanto, sabe-se que apesar da sistematização dos achados, pode haver interpretações subjetivas, não havendo menção explícita sobre uma forma exata para a categorização final das lesões^{35,39,40}.

Uma mamografia é considerada negativa para câncer de mama, nas categorias 1, 2 e 3 do BI-RADS® e positiva nas categorias restantes. Na categoria 1 não há nenhum achado digno de nota, os seios são simétricos sem calcificações, nódulos, assimetrias, distorções focais ou outras alterações. Na categoria 2 são descritos achados definitivamente benignos e na 3, achados que apresentem menos de 2% de chance de malignidade, recomendando-se reavaliação em 6 meses, para verificar estabilidade da lesão. A categoria 0 representa um estudo incompleto, sendo solicitado um exame de imagem complementar ou mesmo comparação com exame prévio^{35,40}. Na categoria 4 encontram-se lesões com uma chance de malignidade de 3 a 94% e na 5, com mais de 95% de acordo com BI-RADS® e a conduta indicada nestas categorias é a solicitação de investigação citológica ou histológica^{38,40}. A subdivisão opcional da categoria 4 foi recomendada na tentativa de melhor estratificar a probabilidade de malignidade, para que futuramente as indicações de biópsia possam ser otimizadas, aumentando, desta forma, a acurácia da mamografia^{38,40}.

Os VPPs apresentados no parágrafo acima para as categorias 3, 4 e 5 estão baseados em inúmeros trabalhos de importante magnitude citados e extensamente revisados pelo comitê responsável pela elaboração do BI-RADS®³⁸. Entretanto, até o momento, não existem dados consistentes na literatura em relação às probabilidades de câncer para cada subdivisão da categoria 4 do BI-RADS®.

1.3. Microcalcificações Mamárias

As microcalcificações mamárias representam um dos achados com maior variabilidade para descrição e categorização entre médicos radiologistas^{41,42,43,51,10}. O diagnóstico diferencial das microcalcificações mamárias inclui lesões benignas não proliferativas, lesões benignas proliferativas com ou sem atipias, esteatonecrose, aterosclerose, alterações cutâneas, carcinoma *in situ* e carcinoma invasivo. O índice geral de malignidade entre microcalcificações mamárias é entre 22% a 37% e cerca de 30% a 50% dos carcinomas mamários não palpáveis apresentam como único achado, microcalcificações^{41,42,43,44,45,46}.

Desde a descrição inicial da associação de câncer de mama e microcalcificações mamárias por Leborgne em 1951⁴⁷ vários aspectos relacionados à descrição de sua morfologia e distribuição têm sido discutidos e atualizados. Em 1984 Le Gal e cols.⁴⁸ avaliaram microcalcificações identificadas pela mamografia correlacionando com achados histológicos, criando terminologias para sua descrição detalhada. De acordo com a classificação de Le Gal⁴⁸ (figura 1), apenas o tipo 1, anelar é considerado definitivamente benigno. Alguns destes conceitos e terminologias foram adotados pelo BI-RADS® e, em cada nova edição, modificados e ampliados (tabela 2)^{35,36,37,38}. A última edição do BI-RADS® traz alterações na descrição das microcalcificações, alguns conceitos novos são discutidos no capítulo guia, como a adição do termo “heterogêneas grosseiras” e a orientação da utilização da morfologia “pleomórficas finas” para descrever microcalcificações menores que 0,5 mm. A separação entre calcificações “redondas” e “puntiformes” é desaconselhada, a diferença relacionar-se-ia ao tamanho, as menores de 0,5 mm definidas como puntiformes, entretanto ambas estariam no mesmo grupo morfológico. O grau de suspeita das microcalcificações de acordo com a morfologia, nesta quarta edição do BI-RADS®, foi separado como mostra a tabela 2. Observa-se que há a indicação de dois grupos de microcalcificações suspeitas de acordo com a morfologia, as de intermediária suspeita (amorfas ou indistintas e heterogêneas grosseiras) e de alta probabilidade de malignidade (pleomórficas finas e finas lineares ou finas lineares ramificadas). Os conceitos relacionados aos padrões de distribuição das microcalcificações permanecem os mesmos (tabela 2)³⁸.

A maioria das calcificações identificadas na mamografia são as descritas na tabela 2 como tipicamente benignas⁴⁹. As puntiformes (redondas <0,5 cm) têm sido associadas a menos de 2% de malignidade⁵⁰, podendo ser classificadas como provavelmente benignas dependendo de sua distribuição. Já as finas lineares ou finas lineares ramificadas são consideradas altamente suspeitas, especialmente em distribuição segmentar ou linear, em algumas séries^{51,52} associadas com lesões malignas em 81% a 92% dos casos. De acordo com Liberman e cols.⁵¹ as pleomórficas finas em cerca de 41% estão associadas a malignidade. As microcalcificações amorfas ou indistintas, nesta edição indicadas como morfologia de intermediária suspeita³⁸, em dois importantes estudos^{41,51}, apresentaram uma taxa de malignidade em torno de 20% a 26%, especialmente associadas com casos de CDIS e, na maioria dos casos malignos em distribuição segmentar e linear.

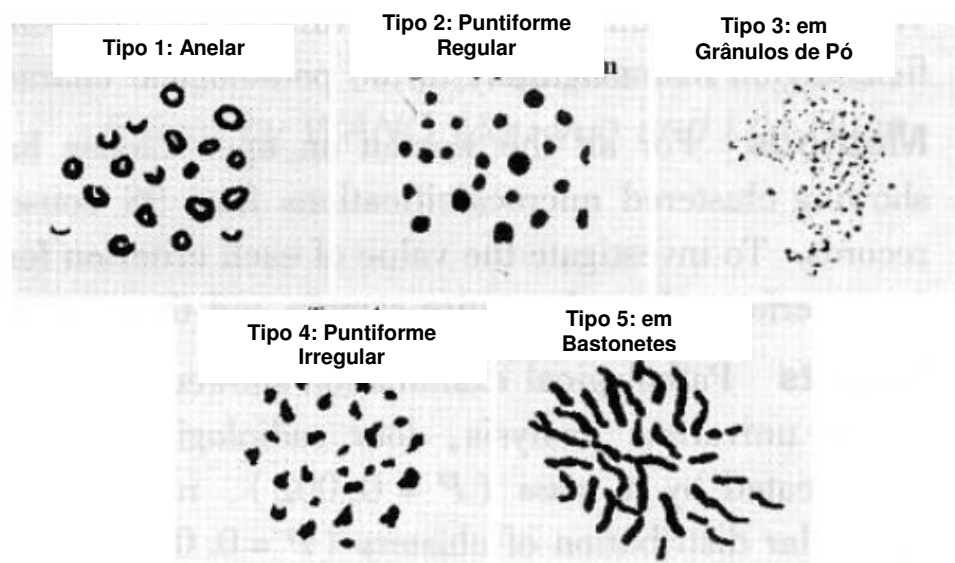


Figura 3: Desenho adaptado de Le Gal⁴⁸: nomenclatura proposta em 1984 para descrição das características morfológicas de microcalcificações.

Tabela 2: BI-RADS® 4ª edição, terminologia atual utilizada para descrição das microcalcificações^{35, 40}.

BI-RADS® 4ª EDIÇÃO	Caracterização mamográfica utilizada nas microcalcificações
Morfologia Tipicamente Benigna	Vasculares
	Grosseiras/semelhantes a “pipoca”
	Na pele
	Com centro radiotransparente
	Em “leite de cálcio”
	Bastonetes longos
	Em “casca de ovo” ou em “anel”
	De fios de sutura
	Distróficas
	Redondas/puntiformes
Morfologia de Intermediária Suspeita	Amorfas
	Heterogêneas grosseiras
Morfologia Altamente Suspeita	Pleomórficas finas
	Finas lineares ou finas lineares ramificadas
Distribuição	Em grupo
	Linear
	Segmentar
	Regional
	Difusa



Figura 4: Mamografia ampliada demonstrando microcalcificações puntiformes (<0,5mm), de distribuição segmentar.

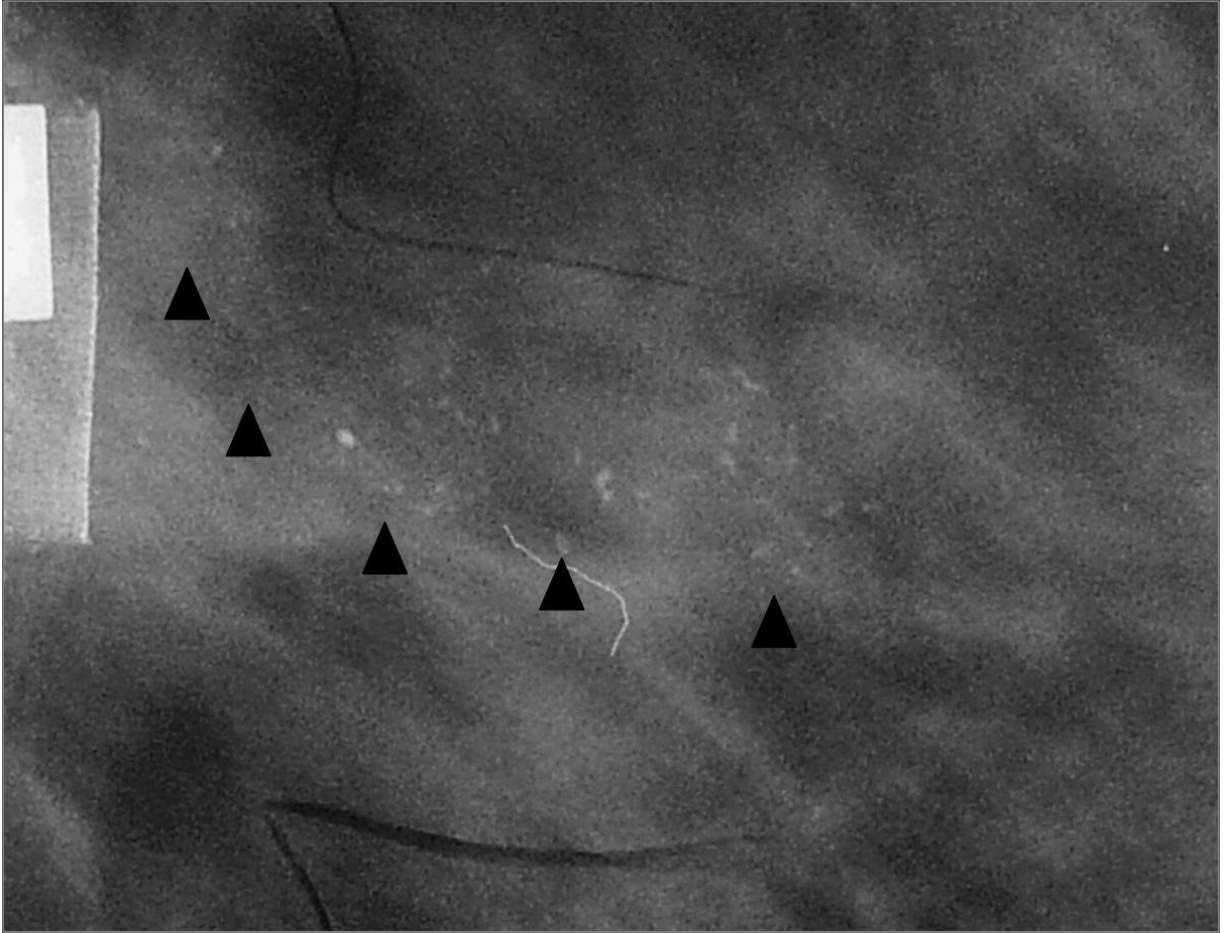


Figura 5: Mamografia ampliada demonstrando microcalcificações amorfas, de distribuição linear.

Portanto, devemos caracterizar as microcalcificações de acordo com sua morfologia, levando em conta a sua distribuição, para então classificá-las nas categorias do BI-RADS® de 2 a 5. Na categoria 4, como já exposto, estão aquelas lesões que têm uma probabilidade de malignidade que pode variar de 3 a 94%^{38,40}. Devido a esta grande variação, optou-se pela subdivisão desta categoria em 4a, baixa suspeita, 4b, intermediária suspeita e 4c, moderada suspeita³⁸. Então, como podemos observar sugere-se três subdivisões para a categoria 4, sendo possível que haja subjetividade na escolha entre as categorias 4a, 4b e 4c, já que há dois grupos de morfologias suspeitas. Por isso é fundamental uma análise conjunta da morfologia predominante e do padrão de distribuição das microcalcificações para chegarmos categorização final. Devemos ainda lembrar que a estabilidade não é um critério de benignidade totalmente confiável especificamente para estas lesões⁵³, o que torna questionável a utilização da categoria 3 do BI-RADS®, que recomenda

controle em 6 meses. No capítulo guia fica enfatizado que se trata de terminologia nova, destinada à orientação, tendo como principal objetivo uma melhor estratificação da probabilidade de câncer, facilitando pesquisas e auditorias internas de cada serviço de imagem e dos médicos radiologistas individualmente, não implicando em padrões exigidos para a prática³⁸.

1.4. Estudos de Concordância em Laudos Mamográficos

Devemos estar cientes no momento que avaliamos os estudos de concordância que, quanto maior o número de observadores e quanto maior o número de variáveis comparadas, mais baixos serão os níveis de concordância ou reprodutibilidade⁵⁴. As diferenças nos grupos de comparação são na maioria dos estudos analisadas pela estatística *Kappa* de Cohen ponderado (weighted) e não-ponderado ou generalizado (unweighted)^{55,56}. Perfeita concordância é indicada por um valor *Kappa* de 1,0 e ausência de concordância por *Kappa* de 0. Estudos realizados a priori^{54,19}, têm sugerido valores menores ou iguais a 0,20 como mínima concordância; 0,21-0,40, baixa concordância; 0,41-0,60 moderada concordância; 0,61-0,80 importante concordância e 0,81-1,0, quase perfeita concordância. Pode-se utilizar *Kappa bi-rater* (comparação entre pares de leitores) e *kappa multirater* (comparação simultânea de mais de dois leitores). É muito importante a descrição detalhada do tipo de análise estatística realizada e das limitações de cada uma delas¹⁹. O *kappa* ponderado é especialmente interessante quando se comparam categorias que não indicariam mudança de conduta médica com outras que indicariam. Por exemplo, uma variabilidade na escolha entre as categorias 4a, 4b, 4c e 5, não mudaria a indicação de biópsia, já uma mudança da categoria 2 para a 4a, implica modificação da conduta médica com relevância clínica¹⁹. No *Kappa* ponderado atribui-se diferentes pesos para estas situações, assim no exemplo acima, a primeira divergência pesaria menos que a segunda para o cálculo do *kappa*. Na avaliação de trabalhos que utilizam o *kappa* ponderado é indispensável que haja uma explicação clara de como foram atribuídos os pesos, podendo existir limitação na comparação entre artigos que utilizem estratégias de ponderação muito diferentes.

Tabela 3: Valores de *Kappa* e grau de concordância¹⁹.

Valores de <i>kappa</i>	<i>Graus de Concordância</i>
0,21-0,40	Baixa Concordância
0,41-0,60	Moderada Concordância
0,61-0,80	Importante Concordância
0,81-1,0	Quase Perfeita Concordância

Alguns importantes estudos^{57,58} avaliaram a concordância do BI-RADS®, antes da sua quarta edição, na recomendação de condutas de acordo com os achados mamográficos, encontrando significativa inconsistência. Um estudo multicêntrico⁵⁹ realizado pelo *Breast Cancer Surveillance Consortium* comparou a concordância das categorias finais e da recomendação de condutas para rastreamento mamográfico e para mamografia diagnóstica, antes e após as regras de estabelecidas pela MQSA. Este estudo demonstrou que de 1996 a 2001 ocorreu melhora da concordância nos laudos mamográficos, possivelmente relacionada às publicações do BI-RADS® (quarta edição não incluída nesta análise), as regras da MQSA para o controle de qualidade e a utilização de sistemas de laudos informatizados.

Em 1998 Kerlikowiske e cols.⁶ utilizando a análise de par de leitores e *kappa* ponderado, demonstraram uma concordância na impressão final de laudos mamográficos com o uso do BI-RADS® de 78% ($k = 0,58$) entre dois observadores e de 86% ($k = 0,73$) intra-observador na avaliação de 2.616 mamografias. O tipo e a morfologia das calcificações apresentaram baixa concordância, com $k = 0,33$ e $0,46$, respectivamente. De 73% a 78% dos casos com câncer foram considerados anormais e foi descrita uma concordância geral moderada, porém, nesta investigação, não foi incluída a categoria 3 (provavelmente benigno), que sabemos ser uma das mais geradoras de discordância entre os médicos mamografistas⁵⁹.

Berg e cols.⁷ com uso do *kappa* ponderado entre múltiplos leitores (cinco médicos radiologistas e 103 casos), apresentaram coeficientes de concordância menores, com maior variabilidade na impressão final interobservador ($k = 0,37$) e intra-observador ($k = 0,60$). Neste estudo a variável que apresentou pior

concordância interobservador foi a forma dos nódulos ou massas ($k = 0,28$), seguido pela morfologia das microcalcificações ($k = 0,36$). A morfologia amorfa apresentou $k = 0,25$. Este mesmo autor, em 2002²⁵, através do *kappa* generalizado, avalia a concordância no uso do BI-RADS® antes e após treinamento específico, observando uma concordância baixa a moderada na classificação final (antes do treinamento $k = 0,31$ e após $k = 0,45$), na caracterização da distribuição ($k = 0,35$ e $k = 0,36$) e morfologia ($k = 0,36$ e $k = 0,44$) das microcalcificações. A concordância geral entre um subgrupo de três médicos experientes foi maior ($k = 0,57$). Dos achados mamográficos avaliados, observou-se que a distribuição das microcalcificações não apresentou melhora da concordância após treinamento específico.

A utilização de CADS especificamente para casos de microcalcificações em um trabalho publicado em 2001⁶⁰ reduziu a variabilidade interobservador, com uma melhora do coeficiente *Kappa* de 0,22 (múltiplos leitores, sem ponderação). Entretanto, a maioria destes programas não está disponível em nosso meio devido ao alto custo, sendo a maioria deles aplicável apenas em sistemas de mamografia digital ou em imagens digitalizadas.

Em 2004 um trabalho⁶¹ realizado com dados de um estudo prévio COBRA - *study* (core biopsy after radiological localization)⁶² avaliou 533 lesões não-palpáveis caracterizadas por microcalcificações. Dois médicos radiologistas realizaram a primeira leitura de todos os casos e após 6 meses, segunda leitura de uma amostra de 100 destes casos. As microcalcificações foram classificadas em seis morfologias (linear, grosseiras, granulares, granulares finas contáveis, granulares muito finas incontáveis, “cottonwool” e grosseiras), em distribuição difusa, em grupo ou segmentar e em categorias de 3 a 5 do BI-RADS®. A análise estatística foi através do *kappa bi-rater*. Neste estudo encontrou-se maior inconsistência nas categorias 4 e 3, com uma moderada concordância na classificação final ($k = 0,54$) e boa concordância na categoria 5 ($k = 0,73$). Este trabalho não apresenta dados da variabilidade interobservador e não utiliza a subdivisão da categoria 4 do BI-RADS®.

Como será discutido abaixo, a quarta edição do BI-RADS®, até a presente data, foi avaliada por um pequeno número de trabalhos recentemente publicados com diferentes metodologias, observamos que alguns optam pela subdivisão da categoria 4 e outros não.

S. Ciatto e cols.⁶³ avaliaram a variabilidade interobservador e intra-observador, sendo a segunda leitura realizada com um intervalo médio de duas semanas, entre doze médicos radiologistas para as categorias finais do BI-RADS® quarta edição. Neste estudo foram incluídas lesões caracterizadas por nódulos e/ou microcalcificações. Na avaliação estatística obtiveram-se os seguintes valores de *kappa* generalizado interobservador (todos os leitores contra todos) para cada categoria: 0,44 para a categoria 2; 0,12 para a 3; 0,25 para a categoria 4 sem a subdivisão; 0,08 para a 4a; 0,07 para a 4b; 0,10 para a 4c; 0,41 para a 5 e 0,20 para todas as categorias combinadas. A variabilidade intra-observador apresentou melhor concordância, como era de se esperar, já que o intervalo entre as duas leituras foi em média apenas duas semanas, com *kappas* que variaram de 0,24 a 0,90 para todas as categorias, com melhora da concordância para os médicos com mais anos de experiência em mamografia. O viés de memória (intervalo de apenas 2 semanas entre as leituras) é um fator que deve ser fortemente considerado neste estudo. O grande número de médicos comparados neste trabalho também pode ser referido como uma explicação para os menores valores de *kappa* na avaliação interobservador por categorias em comparação com outros trabalhos.

Cosar e cols.¹⁶ na avaliação de 83 casos de microcalcificações utilizando a terminologia da quarta edição do BI-RADS®, encontrou uma moderada concordância na descrição da morfologia entre três diferentes observadores (comparados em pares), com *kappa* médio não ponderado de 0,42 (0,33 a 0,51) e a maior variabilidade na morfologia “amorfas” ($k = 0,33$) e na distribuição “regional” ($k=0,16$). Neste trabalho não se optou pela subdivisão da categoria 4, as microcalcificações foram classificadas em intermediária e altamente suspeitas pela morfologia e em 3, 4 ou 5 pelas categorias finais.

Lazarus e cols.⁶⁴ em maio de 2006 avaliou a variabilidade interobservador da quarta edição do BI-RADS®, incluindo a subdivisão da categoria 4, em 94 lesões encaminhadas para biópsia, identificadas na mamografia ou na ecografia mamária. Entre os casos avaliados, 32 eram de microcalcificações. Observou-se baixa concordância na categorização final das lesões (2, 3, 4a, 4b, 4c e 5), com um coeficiente *Kappa* generalizado de 0,28, sendo a maior concordância encontrada nas lesões altamente suspeitas (categoria 5) e a menor nas categorias 4a ($k = 0,14$) e 4b ($k = 0,16$). A morfologia das microcalcificações apresentou *kappa* de 0,32 e a

distribuição de 0,50. Não foi realizada análise da variabilidade intra-observador. Os cinco médicos que interpretaram os exames pertenciam ao mesmo serviço e possuíam qualificação profissional semelhante. Os VPPs apresentados neste estudo para a categoria 4 foram: 6% na 4a, 15% na 4b e 53% na 4c. Os autores referem que uma das possibilidades para explicar a baixa concordância observada é o maior número de categorias, incluindo a subdivisão, mas concluem que o BI-RADS® quarta edição mostrou-se um bom instrumento para a estratificação do grau de suspeita das lesões.

Numa análise conjunta destes dados, observamos que certamente existem inconsistências que devem ser identificadas e questionadas para que a sistematização dos laudos de métodos diagnósticos por imagem da mama torne-se realmente uma ferramenta útil na melhora do diagnóstico do câncer de mama. A adequação da terminologia do BI-RADS® na sua quarta edição, que traz substantivas mudanças em especial na descrição e classificação de microcalcificações, constitui-se uma importante questão de pesquisa.

2. OBJETIVOS:

2.1. Objetivo Primário:

O objetivo deste trabalho é determinar a variabilidade interobservador e intra-observador da terminologia da quarta edição do BI-RADS® na avaliação de microcalcificações mamárias, incluindo a subdivisão da categoria 4, em pacientes encaminhadas para agulhamento pré-cirúrgico.

2.2. Objetivo Secundário:

Como objetivo secundário será avaliado também a acurácia da utilização da quarta edição do BI-RADS®, com a estratificação das probabilidades de câncer de mama para cada uma das categorias utilizadas neste grupo de pacientes.

3. REFERÊNCIAS BIBLIOGRÁFICAS:

1. Parkin MD, Bray F, Ferlay J, Pisani P. Global Cancer Statistics, 2002. *CA Cancer J Clin* 2005;55:74-108.
2. Instituto Nacional do Câncer. Estimativa 2005 Incidência de Câncer no Brasil. <http://www.inca.gov.br/estimativa/2005/versaofinal.pdf> (acessado em 06/abr/2005).
3. Ward E, Jemal A, Cokkinides V, et al. Cancer disparities by race/ethnicity and socioeconomic status. *CA Cancer J Clin* 2004; 54:78 –93.
4. Curi Hallal, AL, Gotlieb SLD, Latorre MRDO. Evolução da mortalidade por neoplasias malignas no Rio Grande do Sul, 1979-1995. *Rev. Bras. Epidemiol* 2001; 4:169-177.
5. Elmore JG, Armstrong K, Lehman CD, Fletcher SW. Screening for breast cancer. *JAMA* 2005; 293:1245-1256.
6. Shapiro S, Venet W, Strax P, Venet L. Periodic Screening for Breast Cancer: The Health Insurance Plan Project and Its Sequelae, 1963-1986. Baltimore, Md:Johns Hopkins University Press; 1988.
7. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* 2002;359:909-919.
8. Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmo Mammographic Screening Program. *J Natl Cancer Inst Monogr* 1997:63-67.
9. Tabar L, Fagerberg G, Chen HH, et al. Efficacy of breast cancer screening by age: new results from the Swedish Two-County Trial. *Cancer* 1995;75:2507-2517.
10. Frisell J, Lidbrink E. The Stockholm Mammographic Screening Trial: risks and benefits in age group 40-49 years. *J Natl Cancer Inst Monogr* 1997:49-51.
11. Bjurstam N, Bjorneld L, Duffy SW, et al. The Gothenburg Breast Cancer Screening Trial: preliminary results on breast cancer mortality for women aged 39-49. *J Natl Cancer Inst Monogr* 1997:53-55.
12. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. *J Natl Cancer Inst* 2000;92:1490-1499.

13. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up: a randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med* 2002;137:305-315.
14. Alexander FE, Anderson TJ, Brown HK, et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet* 1999;353:1903-1908.
15. Olsen O, Gøtzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet* 2001;358:1340-2.
16. Fletcher SW, Elmore JG. Clinical practice: mammographic screening for breast cancer. *N Engl J Med* 2003;348:1672-1680.
17. Kerlikowske K, Smith-Bindman R, Ljung BM, Grady D. Evaluation of abnormal mammography results and palpable breast abnormalities. *Ann Intern Med* 2003;139:274-284.
18. Pisano ED, Gatsonis C, Hendrick E et al. Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening. *N Engl J Med* 2005;353:1773-83.
19. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. *N Engl J Med* 1995; 331:1493–1499.
20. Warren Burhenne LJ, Wood S, D'Orsi C, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology* 2000; 215:554–562.
21. Baker JA, Lo JY, DeLong DM, Floyd CE. Computer-aided detection in screening mammography: variability in cues. *Radiology* 2004; 233:411–417.
22. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists: findings from a national sample. *Arch Intern Med* 1996; 156:209–213.
23. Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. *J Natl Cancer Inst* 1998;90:1801–1809.
24. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: Inter- and Intraobserver Variability in Feature Analysis and Final Assessment. *AJR* 2000;174:1769-77.
25. Berg WA, D'Orsi CJ, Jackson VP et al. Does Training in the Breast Imaging Reporting and Data System (BI-RADS) Improve Biopsy Recommendations or Feature Analysis Agreement with Experienced Breast Imagers at Mammography? *Radiology* 2002; 224:871–880.
26. Parker S, Stavros A, Dennis M. Needle biopsy techniques. *Radiol Clin North Am* 1995; 33:1171–1186.

27. Liberman L. Clinical management issues in percutaneous core breast biopsy. *Radiol Clin North Am* 2000; 38:791–807.
28. Verkooijen HM, Peeters PH, Pijnappel RM, Koot VC, Schipper ME, Rinkes IH. Diagnostic accuracy of needle-localized open breast biopsy for impalpable breast disease. *Br J Surg* 2000;87:344-7.
29. Markopoulos C et al. Open surgical biopsy for nonpalpable breast lesions detected on screening mammography. *Eur J Gynaecol Oncol* 2005;26:311-4.
30. Johnson JM, Dalton RR, Landercasper J, Travelli R, Lambert PJ. Image-guided or needle-localized open biopsy of mammographic malignant-appearing microcalcifications? *nJ Am Coll Surg* 1998; 187:604–609.
31. Brown ML, Houn F, Sickles EA, Kessler LG. Screening mammography in community practice: positive predictive value of abnormal findings and yield of follow-up diagnostic procedures. *AJR Am J Roentgenol* 1995;165:1373–1377.
32. Burnside ES, Rubin DL, Fine JP, Shachter RD, Sisney GA, Leung WK. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. *Radiology* 2006 Sep;240:666-73.
33. Gur D, Wallace LP, Klym AH, et al. Trends in recall, biopsy, and positive biopsy rates for screening mammography in an academic practice. *Radiology* 2005;235:396–401.
34. Monsees BS. The Mammography Quality Standards Act: an overview of the regulations and guidance. *Radiol Clin North Am* 2000;38:759–772.
35. American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS), Breast Imaging Atlas 4th ed. Reston, VA: American College of Radiology, 2003.
36. American College of Radiology. Breast imaging reporting and data system (BI-RADS). Reston, VA: American College of Radiology, 1993.
37. American College of Radiology. Breast Imaging Reporting and Data System, 2nd ed. Reston, VA: American College of Radiology, 1995.
38. American College of Radiology. Illustrated Breast Imaging Reporting and Data System (BI-RADS), 3rd ed. Reston, VA: American College of Radiology, 1998.
39. Colégio Brasileiro de Radiologia. Programa para Treinamento em Mamografia. Colégio Brasileiro de Radiologia 2004.
40. Colégio Brasileiro de Radiologia. Sistema de Laudos e Registro de Dados de Imagem da Mama. Colégio Brasileiro de Radiologia, 2005.
41. Berg WA, Arnoldus CL, Teferra E, Bhargavan M. Biopsy of Amorphous Breast Calcifications: Pathologic Outcome and Yield at Stereotactic Biopsy. *Radiology* 2001; 221:495–503.

42. Egan RL, McSweeney MB, Sewell CW. Intramammary calcifications without an associated mass in benign and malignant diseases. *Radiology* 1980;137:1–7.
43. Roses DF, Harris MN, Gorstein F, Gumpert SL. Biopsy for microcalcification detected by mammography. *Surgery* 1980; 87:248–252.
44. Hall FM, Storella JM, Silverstone DZ, Wyshak G. Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography. *Radiology* 1988; 167:353–358.
45. Hall JA, Murphy DC, Hall BR, Hall KA. Mammographic abnormalities and the detection of carcinoma of the breast. *AmJ Obstet Gynecol* 1993; 168:1677–1680; discussion 1680–1682.
46. Cosar ZS, Cetin M, Tepe TK, Cetin R, Zarali AC. Concordance of mammographic classifications of microcalcifications in breast cancer diagnosis: Utility of the Breast Imaging Reporting and Data System (fourth edition). *Clin Imaging* 2005; 29:389-95.
47. Leborgne R. Diagnosis of Tumors of the Breast by Simple Roentgenography: Calcifications in Carcinoma. *AJR* 1951;65:1-11.
48. Le Gal M, Chavanne G, Pellier D. Valeur diagnostique des microcalcifications groupees decouvertes par mammographies: a propos de 277 cas avec verification histologique et sans tumeur du sein palpable. *Bull Cancer (Paris)* 1984;71:57-64.
49. Sickles EA. Breast calcifications: mammographic evaluation. *Radiology* 1986; 160:289–293.
50. Sickles EA. Periodic mammographic follow-up of probably benign lesions: results in 3,184 consecutive cases. *Radiology* 1991; 179:463–468.
51. Liberman L, Abramson A, Squires FB, Glassman JR, Morris EA, Dersshaw DD. The Breast Imaging Report Breast Imaging Reporting and Data System: Positive Predictive Value of Mammographic Features and Final Assessment Categories. *AJR* 1998; 171: 35-40.
52. Knutzen AM, Gisvold JJ. Likelihood of malignant disease for various categories of mammographically detected, nonpalpable breast lesions. *Mayo Clin Proc* 1993; 68:454–460.
53. Lev-Toaff AS, Feig SA, Saitas VL, Finkel GC, Schwartz GF. Stability of malignant breast microcalcifications. *Radiology* 1994;192:153-156.
54. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43:543-9.
55. Crewson PE. Reader Agreement Studies. *AJR* 2005; 185:1391-7.
56. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.

57. Baker JA, Kornguth PJ, Floyd CE. Breast Imaging Reporting and Data System Standardized Mammography Lexicon: Observer Variability in Lesion Description. *AJR Am J Roentgenol* 1996;166:773–8.
58. Taplin SH, Ichikawa LE, Kerlikowske K, et al. Concordance of breast imaging reporting and data system assessments and management recommendations in screening mammography. *Radiology* 2002;222(2):529–535.
59. Geller BM, Ichikawa LE, Buist DS et al. Improving the concordance of mammography assessment and management recommendations. *Radiology* 2006;241:67-75.
60. Jiang Y, Nishkawa RM, Schmidt RA, Toledano AY, Doi K. Potential of Computer-aided Diagnosis to Reduce Variability in Radiologists' Interpretations of Mammograms Depicting Microcalcifications. *Radiology* 2001; 220:787-793.
61. Pijnappel, M, Peeters HM, Hendriks HCL, Mali MP. Reproducibility of mammographic classifications for nonpalpable suspect lesions with microcalcifications. *The British Journal of Radiology* 2004; 77:312-314.
62. Verkooijen HM. Diagnostic accuracy of stereotactic largecore needle biopsy for nonpalpable breast disease: results of a multicenter prospective study with 95% surgical confirmation. *Int J Cancer* 2002;99:853–9.
63. Ciatto S, Houssami N, Apruzzese A et al. Reader variability in reporting breast imaging according to BI-RADS assessment categories (the Florence experience). *Breast* 2006;15(1):44-51.
64. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. *Radiology* 2006; 239:385-91.

Artigo em Inglês:

Breast Microcalcifications: Variability Intraobserver and Interobserver in Mammographic Interpretations Using the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS®) Fourth Edition

Gonçalves ATC¹, Borba AA², Scherer L², Demarchi C², Edelweiss MIA³.

1. Radiologist, currently taking MSc at Postgraduate Course in Medical Science of School of Medicine, Federal University of Rio Grande do Sul (UFRGS), Brazil.

2. Radiologist, M.D. of Mammography Service of Centro de Radiologia de Novo Hamburgo and Mãe de Deus Center Hospital, Brazil.

3. Pathologist PhD, Associated Professor in the Department of Pathology of Medical School of UFRGS, Brazil.

Correspondence to: Andrea T Cadaval Gonçalves

Av Soledade, 569 – 6th floor, CEP 904703440

Bairro Três Figueiras. Porto Alegre-RS. Brazil

andreadaval@hotmail.com

Key words - BI-RADS® – microcalcifications – mammography

ABSTRACT

Purpose: To evaluate interobserver and intraobserver agreement by using the BI-RADS® (fourth edition) terminology and the final assessment including subcategorization of category 4 for breast microcalcifications. We also assess the accuracy of the BI-RADS® final categories.

Materials and Methods: Four breast radiologists retrospectively interpreted 109 mammograms showing microcalcifications in 89 patients who underwent needle-localization open breast biopsy according with the BI-RADS® lexicon (fourth edition). All cases were re-examined and classified again after six months by the same team. The final diagnosis was based upon the pathologic result of the surgical specimen. The reader agreement for final assessment category and for each category, morphology and distribution descriptions was measure by *kappa* statistics. The accuracy was determined with the likelihood ratio (LR) and the positive predictive value (PPV) for final assessment of each BI-RADS® category.

Results: Pathological examination revealed 47 malignant lesions (46%), 29 ductal carcinomas in situ (DCSI) and 18 invasive carcinomas (IDC) and 55 benign lesions (54%). Moderate agreement was achieved in interobserver ($k = 0.49$ in the first and in the second readings) and intraobserver ($k = 0.57$) of the final assessment of all categories. For microcalcifications morphology, the mean k was 0.49 for intraobserver and 0.27 and 0.34 for interobserver. For the distribution, the mean k was 0.48 for intraobserver and 0.42 and 0.39 for interobserver. We found some inconsistencies in assessment of final categories 3, 4a and 4b and amorphous morphology. PPV in the two readings sessions for subcategorization of category 4 was 4a: 25% and 35%; 4b: 42% and 41%; and 4c: 60% and 74%. The accuracy was lower in categories 4a (LR, 0.39 and 0.63) and 4b (LR, 0.84 and 0,80). The categories 4c (LR, 1.78 and 3.32) and 5 (LR, > 16,5 and > 20) were associated with high risk for malignancy.

Conclusion: The terminology of BI-RADS® fourth edition including subcategorization of category 4 was associated to an overall fair to moderate agreement. We found good stratification of the likelihood of malignancy with better

accuracy for categories 4c and 5. Attention for inconsistencies in assessment of final BI-RADS® categories 3, 4a and 4b and amorphous or indistinct morphology and lower accuracy in categories 4a and 4b it is important concern for microcalcifications.

INTRODUCTION

The breast cancer is still one of the leading causes of cancer death among women and is the most prevalent cancer in the world¹. The breast cancer screening with mammography is recommended by most experts groups to begin at 40 years of age² although researches demonstrated strongest evidences supporting the usefulness of mammographic screening in mortality reduction (20-35%) for women aged 50 to 69 years³. Microcalcifications constitute one of the findings in early stage of breast cancer, in situ carcinoma, invasive carcinoma and is also associated with considerable variability in description^{4,5,6,7,8,9,10}. The overall rate of malignancy among calcifications that are referred for biopsy ranges from 22% to 37% and between 30% and 50% of nonpalpable breast cancers presenting themselves as microcalcifications alone^{11,12,13,14,15,16}. The accuracy of mammography can be reduced by intra and interobserver variability among radiologists^{4,6}.

In an attempt to standardize the assessment and reporting of breast lesions identified on mammograms the American College of Radiology recommend the Breast Imaging Reporting and Data System (BI-RADS®)¹⁷. This system includes terms for describing the morphology and distribution of calcifications and classifies the radiologist's level of suspicion about the mammographic abnormality. In the BI-RADS® last edition it as been suggested the use of optional division of Category 4 into three subcategorizations: lesions into those with a low (category 4a), intermediate (category 4b), or moderate (category 4c) likelihood of malignancy⁷. This is helpful in stratifying the likelihood of malignancy in lesions recommended for biopsy by making auditing easier, but it can be correlated with fair agreement in interpretation of mammograms, especially for microcalcifications¹⁸.

The purpose of this study was to evaluate interobserver and intraobserver agreement by using the BI-RADS® (fourth edition) terminology to characterize and define the final assessment in mammograms including subcategorization of category 4 in breast microcalcifications. We also assess the accuracy of the BI-RADS® final categories.

MATERIAL AND METHODS:

Sample

Four radiologists with subspecialty expertise in breast imaging retrospectively evaluated 102 consecutive mammograms showing calcifications in 89 women (mean age, 51 years; age range, 36-79 years) who underwent needle-localization breast biopsy between January 1, 2002 and 31 December, 2004 in a private radiology practice. We included all cases with available good-quality original mammograms and definitive pathologic results. Mammographic magnification views were obtained for all cases.

Interpretation

All radiologists came from the same group working together in two different private medical institutions, they have been interpreting breast images for 5 to 12 years and all have the Radiology and Diagnostic Imaging Specialty Diploma by the Brazilian College of Radiology as qualified interpreting physicians. All cases were read independently and randomly without the pathologic results. Identification and clinical history of the patients and prior mammograms was not provided. After a six month period all cases were re-examined and classified again by the same team. No previous formal training in the BI-RADS® lexicon was provided because the four radiologists were familiar with the terminology of the fourth edition of the BI-RADS® and subcategorization.

The cases was interpreted by using the terminology for microcalcifications of the fourth edition of the BI-RADS® standard lexicons and assigned a final assessment category, including the subcategories of BI-RADS® category 4. The calcifications could be described according to predominant morphology in: vascular, coarse or popcorn like, rod like, round or punctuate, lucent center, ring or eggshell, milk of calcium, suture, dystrophic, coarse heterogeneous, fine pleomorphic and fine-linear branching. According with distribution in: grouped or clustered, linear, segmental, regional, diffuse or scattered. The final assessment BI-RADS® categories was 2, 3, 4a, 4b, 4c and 5 (table 1)¹⁷.

Table 1: Final BI-RADS® assessment categories (fourth edition).

Category	Definition
2	Benign
3	Probably Benign
4a	Low Suspicion of Malignancy
4b	Intermediate Suspicion of Malignancy
4c	Moderate Suspicion of Malignancy
5	Highly Suggestive of Malignancy

For the consensus opinion we considered data from the assessments of all readers, using the most frequently reported assessment category for each case, on the first and on the second reading session separately.

The final diagnosis was based upon the pathologic result of the surgical specimen (needle localized breast biopsy) and grouped in: benign lesions and malignant lesions, ductal carcinoma in situ (DCIS) and invasive ductal carcinoma (IDC). We did not have any case of lobular carcinoma. Pathologic diagnosis confirmation and intraoperative radiography of the surgical specimen were available for all cases.

Statistical Analysis

The interobserver and intraobserver variability was calculated by comparing the results of morphology, distribution descriptions and the final assessment category of the first and second reading sessions for each case using the Cohen's *kappa* statistics. To access the final assessment category interobserver and intraobserver variability we used weighted *kappa* between pairs of readers¹⁹. We also analyzed separately the subgroups benign, IDC and DCIS. Weights were established after data collection using the Fleiss strategy²⁰. For BI-RADS® category 2 we established value 1, for BI-RADS® category 3 we established value 2, for BI-RADS® category 4a we established value 3, and successively. The difference between the values was used to estimate the weight in each cell. The Fleiss strategy for assigning weights for the weighted *kappa* was done using the formula:

$$w_{ij} = 1 - \frac{(i_w - j_w)^2}{(k-1)^2}$$

where i_w and j_w are the values for the readers i and j ; k is the number of rating categories. The weighted *kappa* was calculated using the observed and expected proportions multiplied by the given weight.

To access interobserver consistency of each final assessment category, each description of morphology and of distribution we compared each radiologist's report against the others on the first and on the second reading sessions using multirater generalized Cohen's *kappa*. Prior reports have suggested the following level of agreement between observers for the indicated *kappa* values: 0.20 slight agreement; 0.21–0.40 fair agreement; 0.41–0.60 moderate agreement; 0.61–0.80 substantial agreement; and 0.81–1.00 almost perfect agreement between observers²¹.

The likelihood ratio (LR) and positive predictive values (PPV) was used to access the accuracy for each category for the consensus opinion. The data were analyzed using the SPSS v13.0 and the MS Excel. This study was approved by the ethical committee of Federal University of Rio Grande do Sul.

RESULTS:

Pathological examination revealed 47 malignant lesions (46%), 29 DCIS and 18 DIC and 55 benign lesions (54%).

Final Assessment Categories

The interobserver agreement of final assessment of all categories in the first and second reading and intraobserver agreement according to BI-RADS® lexicon (fourth edition) are presented in table 2. Moderate agreement was achieved in interobserver (mean weighted *kappa* = 0.49 in the first and in the second readings) and intraobserver (mean weighted *kappa* = 0.57) of the final assessment of all categories. Consistency was higher according with years of practice.

In evaluating the subgroups (benign, IDC and DCIS) between all pairs of readers, agreement was moderate. The mean weighted *kappa* values for benign were 0.54 (interobserver first reading), 0.53 (interobserver second reading) and 0.65 (intraobserver) on the final assessment. The weighted *kappa* values for IDC were 0.41 (interobserver first reading), 0.45 (interobserver second reading) and 0.60 (intraobserver) on the final assessment. And for DCIS the weighted *kappa* values were 0.65 (interobserver first reading), 0.60 (interobserver second reading) e 0.73 (intraobserver) on the final assessment.

Table 2: Interobserver and intraobserver variability in assigning final BI-RADS® assessment category*.

Observers **	INTEROBSERVER		INTRAOBSERVER	
	First reading Weighted <i>Kappa</i> *	Second reading Weighted <i>Kappa</i> *	Reader	Weighted <i>Kappa</i> *
Reader 1 x Reader 2	0.63	0.51	Reader 1	0.72
Reader 1 x Reader 3	0.53	0.47	Reader 2	0.53
Reader 1 x Reader 4	0.40	0.49	Reader 3	0.61
Reader 2 x Reader 3	0.56	0.47	Reader 4	0.42
Reader 2 x Reader 4	0.48	0.37	-----	-----
Reader 3 x Reader 4	0.33	0.46	-----	-----
<i>Mean Weighted Kappa</i>	0.49	0.49	-----	0.57

* BI-RADS® categories 2-5.

** Reader 1: 12 years of experience; reader 3: 9 years; reader 2: 6 years; reader 4: 5 years

Each BI-RADS® final categories were analyzed separately in the first and in the second reading (table 3). We found slight agreement in category 3 on the first reading, 4a and 4b on the first and second reading. For intraobserver analysis we found fair agreement in category 4b. When we compared the four observers for the subgroup analysis the highest inconsistencies were category 3 for benign (*kappa* values 0.11 - 0.33), 4a for benign and IDC, (*kappas* values 0.07 - 0.30) e 4b for DCIS (*kappas* values 0.01 - 0.15).

Table 3: Interobserver and intraobserver variability for each final BI-RADS® categories.

Final assessment category BI-RADS®	INTEROBSERVER		INTRAOBSERVE R
	First reading <i>Kappa</i> value*	Second reading <i>Kappa</i> value*	<i>Kappa</i> value*
2	0.39	0.49	0.51
3	0.20	0.25	0.35
4a	0.10	0.14	0.23
4b	0.19	0.10	0.17
4c	0.25	0.24	0.27
5	0.63	0.39	0.52
Mean <i>Kappa</i>	0.29	0.25	0.34

* Generalized *kappa* value.

Calcification Morphology

For microcalcifications morphology, the overall agreement was moderate for intraobserver analysis and fair to moderate when we compared the four observers at the same time (table 4). The most frequent disagreement observed was in amorphous (*kappa* values 0.11 - 0.24) and fine pleomorphic (*kappas* values 0.30 - 0.42) morphologies. The overall generalized *kappa* values were similar when evaluating the subgroups benign, IDC and DCIS (data not shown).

Repartition of different microcalcifications morphologies in our series of 18 cases of IDC, using consensus opinion on the first and the second reading session, was: fine pleomorphic: 13 (72%), fine-linear branching: 3 (17%) and coarse heterogeneous: 2 (11%). The microcalcifications associated with 29 cases of DCIS were classified as fine pleomorphic: 12 (42%); amorphous or indistinct: 7 (24%); fine-linear branching: 4 (14%); coarse heterogeneous: 3 (10%); dystrophic: 2 (7%); milk of calcium: 1 (3%). The amorphous or indistinct, dystrophic and milk of calcium morphologies were used for false negatives cases of DCIS. The morphologies amorphous or indistinct and fine pleomorphic used in the false positive cases, in individual and consensus opinion of two readings.

Table 4: Interobserver and intraobserver variability in the morphology description of microcalcifications, BI-RADS® fourth edition terminology.

BI-RADS® Descriptor For Morphology	INTEROBSERVER		INTRAOBSERVER
	First reading <i>Kappa</i> value*	Second reading <i>Kappa</i> value*	<i>Kappa</i> value*
Fine Pleomorphic	0.30	0.32	0.42
Fine-linear branching	0.21	0.16	0.56
Coarse heterogeneous	0.22	0.52	0.68
Amorphous	0.11	0.24	0.20
Dystrophic	0.31	0.41	0.61
Punctate/round	0.31	0.32	0.39
Milk of calcium	0.41	0.41	0.59
Lucent center	-	-	-
Coarse or popcornlike	0.66	0.66	1.00
Rod like	-	-	-
Vascular	-	-	-
Mean <i>Kappa</i>	0.27	0.34	0.49

* Generalized *kappa* value.

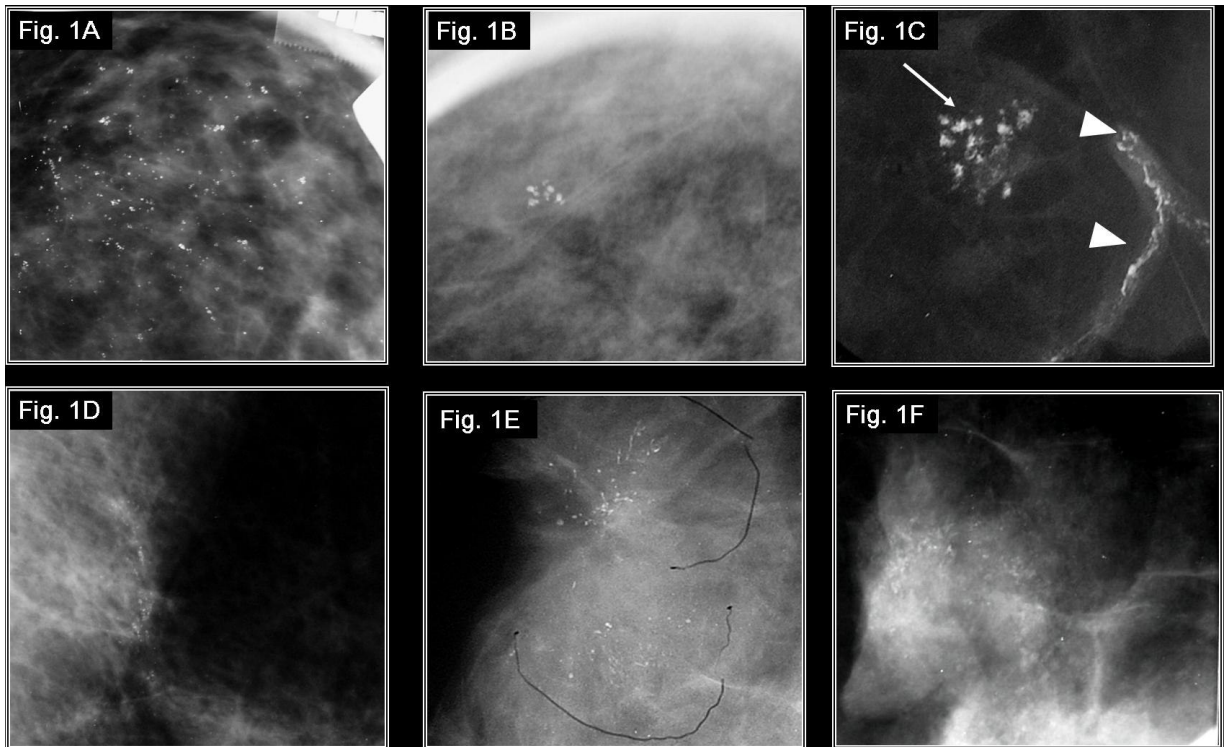


Figure 1A to 1F: Spot-magnification mammograms show microcalcifications. Fig.1A: 54-year-old woman with round microcalcifications in a regional distribution - BI-RADS® category 2 (benign). Fig 1B: 49 year-old woman with a cluster of dystrophic microcalcifications - BI-RADS® category 3 (benign). Fig 1C: 62 year-old woman, arrow: a cluster of coarse heterogeneous microcalcifications - BI-RADS® category 4b (benign); head arrows: vascular calcifications. Fig 1D: 51 year-old woman with fine pleomorphic microcalcifications, in a linear distribution - BI-RADS® category 4c (IDC with DCIS areas). Fig. 1E: 49 year-old woman with fine linear branching microcalcifications in a segmental distribution - BI-RADS® category 4c (IDC). Fig 1F: 46 year-old woman, with amorphous microcalcifications, in a regional distribution - BI-RADS® category 4a (benign).

Calcification Distribution

There was fair to moderate agreement in the distribution of microcalcifications among the four observers (table 5). The diffuse or scattered distribution was not used for any observer for any case. The results were similar when evaluating the subgroups benign, IDC and DCIS (data not shown). The greatest agreement was found for grouped or clustered.

Table 5: Interobserver and intraobserver variability in the distribution description of microcalcifications - BI-RADS® fourth edition terminology.

BI-RADS® Descriptor For Distribution	INTEROBSERVER		INTRAOBSERVER
	First reading <i>Kappa value*</i>	Second reading <i>Kappa value*</i>	<i>Kappa value*</i>
Grouped or clustered	0.62	0.48	0.62
Linear	0.38	0.50	0.45
Segmental	0.32	0.28	0.44
Regional	0.37	0.29	0.42
Diffuse or scattered	-	-	-
Mean <i>Kappa</i>	0.42	0.39	0.48

* Generalized *kappa* value.

Accuracy Evaluation

The accuracy measures of the consensus opinion for final assessment categories in the first and the second reading are showed in table 6. When we evaluated the malignancy, we observed the majority of 18 IDC was classified in category 5. The repartition of final assessment for IDC was: category 5 (11 cases in the first reading and 7 in the second), category 4c (4 in the first reading and 7 in the second), 4b (2 in the first reading and 3 in the second) e 4a just one case in the first and in the second readings. For the 29 DCIS the repartition of final assessment was: category 5 (3 cases in the first reading and 4 in the second), 4c (11 in the first reading and 10 in the second), 4b (8 in the first reading and 8 in the second) e 4a (3 in the first reading and e 6 in the second). We had four false negatives in the first reading and just one in the second, all were DCIS and was classified as category 3.

Table 6: LR and PPVs of the consensus opinion for final BI-RADS® assessment category (fourth edition) in the first and second readings.

Category	First reading		Second reading	
	LR	PPV	LR	PPV
2	<0,01	----	<0,01	----
3	0.36	----	0.08	----
4a	0.39	25%	0.63	35%
4b	0.84	42%	0.80	41%
4c	1.78	60%	3.32	74%
5	> 16.5	93%	> 20,0	100%

DISCUSSION:

Since the initial description of microcalcifications in breast cancer by Leborgne in 1951²² the morphological and the distribution aspects of microcalcifications in mammograms have been a source of discussion. In 1984, Le Gal et al. evaluated microcalcifications discovered by mammography with histological verification according to their morphology²³. Latter, in 1993, American College of Radiology (ACR) developed Breast Imaging Reporting and Data System (BI-RADS®) to standardize mammographic reporting²⁴. The last edition, the illustrated BI-RADS®, Fourth Edition¹⁷ was translated to Portuguese in 2005²⁵ and has many substantive changes in terminology in descriptions of the mammographic appearance of microcalcifications. In this study we use all this recommendations, including the optional subcategorization of category 4.

Our results were similar to previous studies^{7,10} using the BI-RADS® without the subcategorization with an overall fair to moderate intraobserver and interobserver agreement for the description of microcalcifications. An important aspect in agreement studies is the raising fractionation of reporting categories, number of descriptions and of observers implying a decrease in consistency¹⁹. In our study a weighted *kappa* was used to assess the final assessment categories agreement and a generalized *kappa* to each category, morphology and distribution. The analysis between pairs of readers using the weighted *kappa* for final assessment categories

(mean $kappa$ = 0.49 and 0.57 interobserver and intraobserver respectively) has shown more frequent changes between near categories (for example 3 for 4a and 4a for 4b) than changes with major weighted. Furthermore we found improving agreement for final assessment categories with increasing number of years in practice (table 2), in concordance with other studies^{4,6}.

Cosar et al¹⁶ investigated 83 cases of microcalcifications using the fourth edition BI-RADS® terminology, reporting similar results, in which three radiologist was compared in pairs with an overall fair to moderate agreement (mean $kappa$ = 0.49). For morphology the mean $kappa$ value was 0.42 (0.33 – 0.66), the least agreement was for the description of microcalcifications as amorphous. Those investigators did not use the subcategorization of category 4 lesions. Among 35 malignancies, there were 9 (11%) in situ carcinomas. In our study we had 29 DCIS (62%). Ductal carcinoma in situ represents a heterogeneous group of preinvasive lesions that exhibit a wide variety of architectural patterns²⁶. Necrosis in DCIS lesions is often associated with dystrophic calcifications that have a fine pleomorphic branching-type pattern found on mammography²⁷. It has been demonstrated inconsistent correlation between histological types of DCIS and predominant mammographic appearance, with significant overlap^{28,29,30}.

The BI-RADS® recommend two levels of suspicions by morphology of microcalcifications: intermediate (amorphous or indistinct and coarse heterogeneous) and higher probability of malignancy (fine pleomorphic and fine branching or casting). There is no precise recommendation that clearly determine how to perform the subcategorization of category 4 in the BI-RADS®, although there are some exemplifications. We need to consider the morphology and distribution of microcalcifications to choose the final category and this can be associated with inherent variability especially for the subcategorization of category 4. The use of category 3 for microcalcifications is not completely clear because stability is not a reliable criterion for benignity³¹. In our study the category 3 was used for amorphous or indistinct microcalcifications in two false negatives cases (DCIS). The larger source of variation appeared among radiologists in descriptions of amorphous or indistinct morphology in categories 3, 4a and 4b in our study. The amorphous or indistinct microcalcifications in two different studies^{11,32} were associated with malignancy in 20%-26%, in the majority of cases DCIS and in segmental or linear

distribution. It is described variable recommendations for follow-up (BI-RADS® category 3) or biopsy for this morphology type^{8,11}. Perhaps a more refined explanation for categorization of microcalcifications especially for 3, 4a and 4b categorizations and a differential concern about the amorphous morphology could reduce the inconsistencies for lesions near the threshold for biopsy⁸.

In the Guidance Chapter of BI-RADS®^{17,25} the phrase “coarse heterogeneous” was added and the “fine pleomorphic” was recommended to be used to describe calcifications smaller than 0.5 mm. In our results we found improving agreement in the second reading especially for “coarse heterogeneous” morphology. This can be associated with the learning curve of this new terminology in the practice in six months.

S. Ciatto et al³³ compared 12 radiologists for interobserver and intraobserver agreement using the BI-RADS® fourth edition. They included lesions characterized by nodules and microcalcifications. The *kappa* values for interobserver analysis was 0.44 for category 2, 0.12 for 3, 0.25 for all category 4 (0.08 for 4a, 0.07 for 4b, 0.10 for 4c), 0.41 for 5 and 0.20 for all the categories combined. The intraobserver agreement was better (*kappa* 0.24 – 0.90) but the first and the second readings sessions had just two weeks of mean interval. The larger number of readers and some small differences in the *kappa* statistic analysis are limitations to perform a direct comparison with our data. In our study with four radiologists we also found more interobserver than intraobserver variability in the final assessment of each category, in morphologies and in distribution descriptions, however with overall generalized *kappa* values a little higher.

A limitation of our study is that we only evaluated lesions referred for biopsy. This could explain some of the inconsistencies and the number of false positives found. Mammographic magnification views were previous available for all cases. This sample can be useful for study in diagnostic mammography but do not represent women seen in clinical practice.

A recent study¹⁸ evaluated interobserver variability among five radiologists using terminology of the fourth edition of the BI-RADS® to categorize lesions on sonograms and mammograms, with 32 cases of microcalcifications. There were fair

agreement for overall BI-RADS® category reported ($kappa = 0.28$) including the subcategorization. They found a PPV of category 4 between 6% to 53% and larger number of false positives and inconsistencies in categories 4a and 4b, in concordance with our results. Our PPV for lesions in category 4 was 25%-30% to 60%-74% using the subcategorization of final assessment. There was a trend toward improved accuracy in the second reading. The accuracy was lower in categories 4a (LR, 0.39 and 0.63) and 4b (LR, 0.84 and 0,80). The categories 4c (LR, 1.78 and 3.32) and 5 (LR, > 16.5 and > 20) were associated with high risk for malignancy. Therefore we demonstrated a gradual increasing in the likelihood of malignancy with a good stratification using the subcategories 4a, 4b and 4c and the category 5 of BI-RADS®. The current practice standards recommend that the PPV of breast biopsy should be between 25%–40%^{34,35}. Although a recent study³⁶ advocate that the tendency toward an increasing number of biopsies in breast image practice may increase only the number of benign biopsy results, rather than the number of cancer diagnosis. This result have been attributed to increasing detection and biopsy of microcalcifications^{34,35}

The evaluation of applicability of BI-RADS® in different languages and contexts it is very important and the efforts toward to improvement in each new edition are remarkable. Our study strengthens the current knowledge showing an overall fair to moderate agreement in the use of the BI-RADS® fourth edition terminology and the subcategorization of category 4, good stratification of the likelihood of malignancy with better accuracy for categories 4c and 5. We found some inconsistencies in assessment of final BI-RADS® categories 3, 4a and 4b and amorphous or indistinct morphology and lower accuracy in categories 4a and 4b. Until further investigations helping to refine the lexicon it is very important consider these aspects for a clinical decision based in a mammogram report in a patients with microcalcifications near the threshold for biopsy.

REFERENCES

- ¹ Parkin MD, Bray F, Ferlay J, Pisani P. Global Cancer Statistics, 2002. *CA Cancer J Clin* 2005;55:74-108.
- ² Elmore JG, Armstrong K, Lehman CD, Fletcher SW. Screening for breast cancer. *JAMA* 2005; 293:1245-1256.
- ³ Fletcher SW, Elmore JG. Clinical practice: mammographic screening for breast cancer. *N Engl J Med*. 2003; 348:1672-1680.
- ⁴ Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. *N Engl J Med* 1995; 331:1493–1499.
- ⁵ Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists: findings from a national sample. *Arch Intern Med* 1996; 156:209–213.
- ⁶ Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. *J Natl Cancer Inst* 1998;90:1801–1809.
- ⁷ Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: Inter- and Intraobserver Variability in Feature Analysis and Final Assessment. *AJR* 2000;174:1769-77.
- ⁸ Berg WA, D'Orsi CJ, Jackson VP et al. Does Training in the Breast Imaging Reporting and Data System (BI-RADS®) Improve Biopsy Recommendations or Feature Analysis Agreement with Experienced Breast Imagers at Mammography? *Radiology* 2002; 224:871–880.
- ⁹ Taplin SH, Ichikawa LE, Kerlikowske K, et al. Concordance of breast imaging reporting and data system assessments and management recommendations in screening mammography. *Radiology* 2002;222(2):529–535.
- ¹⁰ Baker JA, Kornguth PJ, Floyd CE. Breast Imaging Reporting and Data System Standardized Mammography Lexicon: Observer Variability in Lesion Description. *AJR Am J Roentgenol* 1996;166:773–8.
- ¹¹ Berg WA, Arnoldus CL, Teferra E, Bhargavan M. Biopsy of Amorphous Breast Calcifications: Pathologic Outcome and Yield at Stereotactic Biopsy. *Radiology* 2001; 221:495–503.
- ¹² Egan RL, McSweeney MB, Sewell CW. Intramammary calcifications without an associated mass in benign and malignant diseases. *Radiology* 1980;137:1–7.
- ¹³ Roses DF, Harris MN, Gorstein F, Gumport SL. Biopsy for microcalcification detected by mammography. *Surgery* 1980;87:248–252.

- ¹⁴ Hall FM, Storella JM, Silverstone DZ, Wyshak G. Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography. *Radiology* 1988; 167:353–358.
- ¹⁵ Hall JA, Murphy DC, Hall BR, Hall KA. Mammographic abnormalities and the detection of carcinoma of the breast. *AmJ Obstet Gynecol* 1993; 168:1677–1680; discussion 1680–1682.
- ¹⁶ Cosar ZS, Cetin M, Tepe TK, Cetin R, Zarali AC. Concordance of mammographic classifications of microcalcifications in breast cancer diagnosis: Utility of the Breast Imaging Reporting and Data System (fourth edition). *Clin Imaging*. 2005; 29:389-95.
- ¹⁷ American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS), Breast Imaging Atlas 4th ed. Reston, VA: American College of Radiology, 2003.
- ¹⁸ Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS® lexicon for US and mammography: interobserver variability and positive predictive value. *Radiology*. 2006; 239:385-91.
- ¹⁹ Crewson PE. Reader Agreement Studies. *AJR* 2005; 185:1391-7.
- ²⁰ Fleiss JL. Statistical methods for rates and proportions, 2nd ed. New York, NY: Wiley, 1981.
- ²¹ Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- ²² Leborgne R. Diagnosis of Tumors of the Breast by Simple Roentgenography: Calcifications in Carcinoma. *AJR* 1951;65:1-11.
- ²³ Le Gal M, Chavanne G, Pellier D. Valeur diagnostique des microcalcifications groupees decouvertes par mammographies: a propos de 277 cas avec verification histologique et sans tumeur du sein palpable. *Bull Cancer (Paris)* 1984;71:57-64.
- ²⁴ American College of Radiology. Breast imaging reporting and data system (BI-RADS®). Reston, VA: American College of Radiology, 1993.
- ²⁵ Colégio Brasileiro de Radiologia. Sistema de Laudos e Registro de Dados de Imagem da Mama. Colégio Brasileiro de Radiologia, 2005.
- ²⁶ Arpino G, Laucirica R, Elledge RM. Premalignant and In Situ Breast Disease: Biology and Clinical Implications. *Ann Intern Med* 2005;143:446-457.
- ²⁷ Fisher ER, Costantino J, Fisher B, Palekar AS, Redmond C, Mamounas E. Pathologic findings from the National Surgical Adjuvant Breast Project (NSABP) Protocol B-17. Intraductal carcinoma (ductal carcinoma in situ). The National Surgical Adjuvant Breast and Bowel Project Collaborating Investigators. *Cancer* 1995;75:1310-9.

- ²⁸ Stomper PC, Connolly JL. Ductal carcinoma in situ of the breast: correlation between mammographic calcification and tumour subtype. *AJR* 1992;159:483-5.
- ²⁹ Yeh ED, Slanetz PJ, Oyama T, Giardino AA, Koerner FC. Mammographic appearance of ductal carcinoma in-situ (DCIS) does not reliably predict histologic subtype. *Radiology* 1998;209(P):271.
- ³⁰ Dinkel Hp, Gassel Am, Tschammler A. Is the appearance of microcalcifications on mammography useful in predicting histological grade of malignancy in ductal cancer in situ? *The British Journal of Radiology* 2000;73:938-944.
- ³¹ Lev-Toaff AS, Feig SA, Saitas VL, Finkel GC, Schwartz GF. Stability of malignant breast microcalcifications. *Radiology*. 1994;192:153-156.
- ³² Liberman L, Abramson A, Squires FB, Glassman JR, Morris EA, Dersshaw DD. The Breast Imaging Report Breast Imaging Reporting and Data System: Positive Predictive Value of Mammographic Features and Final Assessment Categories. *AJR* 1998; 171: 35-40.
- ³³ Ciatto S, Houssami N, Apruzzese A et al. Reader variability in reporting breast imaging according to BI-RADS assessment categories (the Florence experience). *Breast*. 2006 Feb;15(1):44-51.
- ³⁴ Brown ML, Houn F, Sickles EA, Kessler LG. Screening mammography in community practice: positive predictive value of abnormal findings and yield of follow-up diagnostic procedures. *AJR Am J Roentgenol* 1995;165:1373–1377.
- ³⁵ Burnside ES, Rubin DL, Fine JP, Shachter RD, Sisney GA, Leung WK. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. *Radiology* 2006 Sep;240:666-73.
- ³⁶ Gur D, Wallace LP, Klym AH, et al. Trends in recall, biopsy, and positive biopsy rates for screening mammography in an academic practice. *Radiology* 2005;235:396–401.

Artigo em Português:

VARIABILIDADE INTER E INTRA-OBSERVADOR DO SISTEMA BI-RADS® INCLUIDO A SUBDIVISÃO DA CATEGORIA 4 PARA MICROCALCIFICAÇÕES MAMÁRIAS ENCAMINHADAS PARA DEMARCAÇÃO PRÉ-OPERATÓRIA POR AGULHAMENTO

Gonçalves ATC¹, Borba AA², Scherer L², Demarchi C², Edelweiss MIA³.

1. Médica Radiologista, mestranda do Programa de Pós-Graduação em Ciências Médicas da UFRGS, Brasil.
2. Médicos Radiologistas do Serviço de Mamografia do Centro de Ecografia e Radiologia de Novo Hamburgo e do Mãe de Deus Center, Brasil.
3. Professor Adjunto do Departamento de Patologia da Faculdade de Medicina da UFRGS, Brasil.

Correspondência para: Andrea T Cadaval Gonçalves

Av Soledade, 569 – 6º andar, CEP 904703440

Bairro Três Figueiras. Porto Alegre-RS. Brasil

andreadaval@hotmail.com

Palavras chave - BI-RADS® - microcalcificações - mamografia

RESUMO:

Objetivo: Determinar a variabilidade interobservador e intra-observador da terminologia da quarta edição do *Breast Imaging Reporting and Data System* (BI-RADS®) na avaliação de microcalcificações mamárias, incluindo a subdivisão da categoria 4. O objetivo secundário foi avaliar a acurácia das categorias finais do BI-RADS®.

Material e Método: Quatro médicos radiologistas especialistas em mamografia interpretaram, retrospectivamente, 109 mamografias de 89 pacientes encaminhadas para demarcação pré-operatória de microcalcificações. Foi utilizada a terminologia da quarta edição do BI-RADS®. Após seis meses foi realizada nova interpretação. O diagnóstico final baseou-se no resultado anatomopatológico da peça cirúrgica. A concordância entre os médicos para categorias finais e separadamente para cada categoria, morfologia e tipo de distribuição foi calculada através da estatística *kappa*. A acurácia foi determinada através da razão de verossimilhança (*likelihood ratio* - LR) e do valor preditivo positivo (VPP) para cada categoria do BI-RADS®.

Resultados: Foram estudadas 47 lesões malignas (46%), 29 carcinomas ductais *in situ* (CDIS) e 18 carcinomas ductais invasores (CDI), e 55 benignas (54%). Observou-se moderada concordância geral interobservador ($k = 0,49$ na primeira e segunda leitura) e intra-observador ($k = 0,57$) para todas as categorias finais. Na morfologia e na distribuição das microcalcificações houve baixa a moderada concordância geral (valores médios de k entre 0,27 a 0,49). As maiores inconsistências foram no uso das categorias 3, 4a e 4b e na morfologia amorfa. O VPP nas duas leituras para a subdivisão da categoria 4 foi 4a: 25% e 35%; 4b: 42% e 41%; e 4c: 60% e 74%. A acurácia foi menor nas categoriais 4a (LR, 0,39 e 0,63) e 4b (LR, 0,84 e 0,80). As categorias 4c (LR, 1,78 e 3,32) e 5 (LR, > 16,5 e > 20) estiveram associadas com alto risco para malignidade.

Conclusão: A subdivisão da categoria 4 e a terminologia da quarta edição do BI-RADS® demonstraram baixa a moderada concordância geral inter e intra-observador. Obteve-se boa estratificação da probabilidade de malignidade entre os casos de microcalcificações avaliados, com melhor acurácia nas categorias 4c e 5.

Deve-se atentar para limitações na concordância das categorias 3, 4a e 4b e da morfologia amorfa e na acurácia nas categorias 4a e 4b para casos de microcalcificações com indicação de biópsia.

INTRODUÇÃO:

O câncer de mama é a principal causa de morte por câncer e o tipo mais freqüente entre as mulheres, sendo ainda atualmente considerado também o câncer mais prevalente no mundo¹. O rastreamento mamográfico é recomendado pela maioria das diretrizes mundiais a partir dos 40 anos de idade, embora alguns estudos tenham demonstrado que o seu maior impacto na redução da mortalidade (20%-35%) ocorre em mulheres de 50 a 69 anos². As microcalcificações mamárias constituem um dos achados mamográficos relacionados com estágios iniciais de câncer de mama, com uma taxa aproximada de malignidade de 22% a 37% em lesões referidas para biópsia^{3,4,5,6,7,8,9}. Cerca de 30% a 50% dos cânceres de mama não-palpáveis apresentam-se apenas como microcalcificações^{10,11,12,13,14,15}. As microcalcificações também são apontadas como um dos achados mamográficos com maior variabilidade na sua descrição entre os médicos radiologistas^{10,11,12,16,17}. Sabe-se que a acurácia da mamografia pode ser afetada por inúmeros fatores, entre eles a redução da concordância entre os médicos radiologistas nos laudos mamográficos^{4,6}.

Na tentativa de padronizar a descrição das lesões e a recomendação de condutas em laudos mamográficos, o Colégio Americano de Radiologia (ACR) e o Colégio Brasileiro de Radiologia (CBR) recomendam a utilização do *Breast Imaging Reporting and Data System* (BI-RADS®)^{18,19}. Este sistema utiliza termos específicos para descrição da morfologia e distribuição das microcalcificações identificadas na mamografia e sugere o laudo contenha uma classificação final de acordo com o grau de suspeição de cada lesão. Na última edição do BI-RADS®^{18,19} ocorreram significativas alterações, incluindo subdivisão opcional da categoria 4 em baixa suspeita (categoria 4a), intermediária suspeita (categoria 4b) e moderada suspeita (categoria 4c). Uma das vantagens da subdivisão seria a possibilidade de melhor estratificação das probabilidades de câncer de acordo com as categorias, com

otimização das indicações de biópsia, o que poderia resultar em aumento da acurácia da mamografia. Entretanto, antes e após a quarta edição do BI-RADS®, vários trabalhos têm demonstrado considerável variabilidade na caracterização e classificação final das lesões pelos médicos radiologistas, em especial quando se tratam de microcalcificações^{6,7,7,8,10,16,20,33}.

O objetivo deste trabalho é determinar a variabilidade interobservador e intra-observador da quarta edição do BI-RADS® na avaliação de microcalcificações mamárias, incluindo a subdivisão da categoria 4, em pacientes encaminhadas para agulhamento pré-cirúrgico. Como objetivo secundário também foi avaliada a acurácia das classificações finais.

Material e Métodos:

Amostra

Quatro médicos radiologistas, com treinamento específico em mamografia, avaliaram retrospectivamente 102 casos consecutivos de mamografias contendo microcalcificações. A amostra constitui-se de 89 mulheres (média de idade 51 anos; variando de 36 a 79 anos) encaminhadas para demarcação pré-operatória por agulhamento guiado por mamografia, em um serviço privado de radiologia, no período de janeiro de 2002 a dezembro de 2004. Foram incluídos os casos que dispunham de filmes mamográficos de boa qualidade e de resultados anatomopatológicos conclusivos. Todos os casos dispunham de incidências ampliadas das lesões de interesse.

Interpretação:

Os médicos observadores trabalham no mesmo grupo no setor de mamografia de duas clínicas privadas de radiologia, possuindo de 5 a 12 anos de experiência em mamografia, com título de especialização em diagnóstico por imagem do CBR. A avaliação dos casos ocorreu de forma independente e randômica, sem o conhecimento dos diagnósticos anatomopatológicos. Os médicos

eram cegos para dados de identificação da paciente e história clínica, não sendo disponibilizados exames prévios para comparação. Após um período de seis meses todos os casos foram novamente interpretados e classificados da mesma forma pelos quatro médicos. Nenhum treinamento prévio específico no léxico do BI-RADS® foi fornecido antes das duas leituras, porém todos os médicos já vinham utilizando este sistema de laudos desde sua primeira edição.

Os casos foram interpretados utilizando-se a terminologia e as categorias finais da quarta edição do BI-RADS®, incluindo a subdivisão da categoria 4. As calcificações poderiam ser descritas de acordo com a morfologia predominante em: vasculares, em pipoca/grosseiras, centro radiotransparente, leite de cálcio, bastões longos, distróficas, redondas/puntiformes, amorfas ou indistintas, heterogêneas grosseiras, pleomórficas finas, finas lineares ou finas lineares ramificadas. De acordo com sua distribuição: em grupo, linear, segmentar, regional e difusa ou esparsa. As categorias finas do BI-RADS® poderiam ser 2, 3, 4a, 4b, 4c e 5 (tabela 1)^{18,19}.

Tabela 1: BI-RADS® 4ª edição, categorias finais^{18,19}.

Categoria	Definições
1	Negativo
2	Achado(s) Benigno(s)
3	Achado Provavelmente Benigno
0	Necessário Avaliação Adicional
4a	Baixa Suspeita de Malignidade
4b	Intermediária Suspeita de Malignidade
4c	Moderada Suspeita de Malignidade
5	Altamente Sugestivo de Malignidade

A interpretação de consenso foi determinada pelo resultado mais freqüente na interpretação dos quatro médicos em cada leitura. Em casos de empate estabeleceu-se a categoria mais elevada.

O diagnóstico final foi dado pelo estudo anatomopatológico da peça cirúrgica (biópsia aberta com demarcação pré-operatória por agulhamento) e dividido em

lesões benignas ou lesões malignas: carcinoma ductal *in situ* (CDIS) e carcinoma ductal invasor (CDI). Não houve nenhum carcinoma lobular. Todos os casos tiveram radiografias intra-operatórias das peças cirúrgicas, comprovando a presença de microcalcificações, e confirmação do diagnóstico anatomopatológico.

Análise Estatística

A variabilidade interobservador e intra-observador foi calculada através da estatística *Kappa* de Cohen²¹. Para a análise da variabilidade interobservador e intra-observador da classificação final, comparou-se os resultados da primeira e da segunda leitura entre os pares de leitores, utilizando-se o *kappa* ponderado, sendo os pesos calculados após a coleta dos dados, utilizando-se algoritmos simples citados na literatura¹⁹. Desta mesma forma foi realizada análise de subgrupos (benignos, CDI e CDIS). Usando o esquema de ponderação de Fleiss²² foi atribuído valor 1 para a categoria 2, valor 2 para categoria 3, valor 3 para categoria 4a e assim sucessivamente. A diferença entre os valores foi usada para estimar o peso de cada célula da tabela. A atribuição dos pesos para o cálculo do *kappa* ponderado foi realizada através da estratégia de Fleiss, seguindo a seguinte fórmula:

$$w_{ij} = 1 - \frac{(i_w - j_w)^2}{(k-1)^2}$$

onde i_w e j_w são os valores atribuídos pelos médicos i e j , e k é o número de categorias possíveis de classificação. O *kappa* então é calculado a partir das proporções observadas e esperadas multiplicadas pelos pesos atribuídos.

Para análise de cada categoria final separadamente, da morfologia e da distribuição compararam-se os quatro leitores simultaneamente, utilizando-se o *kappa* generalizado. Perfeita concordância é indicada por um valor *Kappa* de 1,0 e ausência de concordância por *Kappa* de 0¹⁹. Estudos realizados a priori, têm sugerido valores menores ou iguais a 0,20 como mínima concordância; 0,21-0,40, baixa concordância; 0,41-0,60 moderada concordância; 0,61-0,80 importante concordância e 0,81-1,0, quase perfeita concordância²³.

Os parâmetros de acurácia foram razão de verossimilhança (*likelihood ratio - LR*) e valor preditivo positivo (VPP) para cada categoria final, conforme interpretação de consenso na primeira e na segunda leitura separadamente. Os dados foram analisados utilizando-se os programas SPSS 13.0 e MS Excel. Este estudo foi aprovado pelo comitê de Ética da Universidade Federal do Rio Grande do Sul.

Resultados:

Entre os 102 casos de microcalcificações avaliados, de 89 pacientes, os exames anatomopatológicos revelaram 47 casos malignos (46%), destes 29 CDIS e 18 CDI, e 55 casos benignos (54%).

Categorias Finais:

Na tabela 2 são demonstrados os valores de *kappa* ponderado por pares de médicos na primeira e na segunda leitura, observando-se moderada concordância na variabilidade interobservador (*kappa* médio de 0,49 na primeira e na segunda leitura). Na análise intra-observador, também se obteve moderada concordância (*kappa* médio de 0,57), observando-se melhora dos valores de *kappa* à medida que aumentava o número de anos de experiência do médico radiologista.

Foi realizada análise de subgrupos (benignos, CDI e CDIS) também entre os pares de leitores, também obtendo-se moderada concordância. Os casos benignos apresentaram *kappas* ponderados de 0,54 (interobservador primeira leitura), 0,53 (interobservador segunda leitura) e 0,65 (intra-observador) nas categorias finais. Para os casos de CDI, *kappas* ponderados de 0,41 (interobservador primeira leitura), 0,45 (interobservador segunda leitura) e 0,60 (intra-observador) nas categorias finais. Para os casos de CDIS, *kappas* ponderados de 0,65 (interobservador primeira leitura), 0,60 (segunda leitura) e 0,73 (intra-observador) nas categorias finais.

Tabela 2: Variabilidade inter e intra-observador das categorias finais do BI-RADS® (quarta edição).

Observadores**	INTEROBSERVADOR		INTRA-OBSERVADOR	
	Primeira leitura Kappa ponderado*	Segunda leitura Kappa ponderado*	Observador	<i>Kappa</i> ponderado*
Médico 1 x Médico 2	0,63	0,51	Médico 1	0,72
Médico 1 x Médico 3	0,53	0,47	Médico 2	0,53
Médico 1 x Médico 4	0,40	0,49	Médico 3	0,61
Médico 2 x Médico 3	0,56	0,47	Médico 4	0,42
Médico 2 x Médico 4	0,48	0,37	-----	-----
Médico 3 x Médico 4	0,33	0,46	-----	-----
<i>Kappa</i> médio	0,49	0,49	-----	0,57

* *Kappa* ponderado, pesos calculados após a coleta de dados.

** Médico1: 12 anos de experiência; médico 3: 9 anos; médico 2: 6 anos; médico 4: 5 anos.

As categorias finais foram analisadas separadamente, encontrando-se mínima concordância nas categorias 3 na primeira leitura, 4a e 4b na primeira leitura e na segunda leitura (tabela 3). Na avaliação intra-observador, a categoria 4b apresentou mínima concordância. As maiores inconsistências, na comparação dos quatro médicos simultaneamente nos subgrupos, foram as categorias 3 nos casos benignos (*Kappas* generalizados 0,11 a 0,33), 4a nos benignos e nos CDI, (*kappas* generalizados 0,07 a 0,30) e 4b nos CDIS (*kappas* generalizados 0,01 a 0,15).

Tabela 3: Variabilidade interobservador entre os quatro médicos simultaneamente e intra-observador em cada categoria final do BI-RADS® (quarta edição).

Categorias do BI- RADS® 4 ^a edição	INTEROBSERVADOR		INTRA- OBSERVADOR
	Primeira leitura	Segunda leitura	Kappa*
	<i>Kappa*</i>	<i>Kappa*</i>	
2	0,39	0,49	0,51
3	0,20	0,25	0,35
4a	0,10	0,14	0,23
4b	0,19	0,10	0,17
4c	0,25	0,24	0,27
5	0,63	0,39	0,52
Combinadas	0,29	0,25	0,34

* *Kappa* generalizado.

Morfologia das Microcalcificações

Os quatro médicos radiologistas avaliados simultaneamente apresentaram baixa a moderada concordância geral (*kappa* médio interobservador 0,27 na primeira leitura, 0,34 na segunda leitura e intra-observador 0,49) na descrição da morfologia das microcalcificações (tabela 4). As morfologias que apresentaram pior concordância foram: amorfas (*kappas* generalizados de 0,11 a 0,24) e pleomórficas finas (*kappas* generalizados de 0,30 a 0,42). Observaram-se semelhantes valores de *kappa* na análise dos subgrupos benignos, CDI e CDIS (dados não expostos).

Considerando-se a interpretação de consenso, dentre os 18 casos de CDI, 13 (72%) foram descritos como pleomórficas finas, 3 (17%) como finas lineares ou finas lineares ramificadas e 2 (11%) como heterogêneas grosseiras. Dos 29 casos de CDIS, 12 (42%) foram interpretados como pleomórficas finas, 7 (24%) como amorfas ou indistintas, 4 (14%) como finas lineares ou finas lineares ramificadas, 3 (10%) como heterogêneas grosseiras, 2 (7%) como distróficas e 1 (3%) como leite de cálcio. As morfologias amorfas ou indistintas, distróficas e leite de cálcio foram as descritas nos casos de falsos negativos de CDIS. E as morfologias amorfa ou

indistinta e pleomórficas finas nos casos de falsos positivos, tanto nas descrições individuais como na interpretação de consenso das duas leituras.

Tabela 4: Variabilidade interobservador entre os quatro médicos simultaneamente e intra-observador de acordo com a morfologia das microcalcificações (BI-RADS® quarta edição).

Morfologias BI-RADS® 4ª edição	INTEROBSERVADOR		INTRA- OBSERVADOR
	Primeira leitura <i>Kappa</i> *	Segunda leitura <i>Kappa</i> *	Médias dos médicos <i>Kappa</i> *
Pleomórficas Finas	0,30	0,32	0,42
Heterogêneas Grosseiras	0,22	0,52	0,68
Finas Lineares ou Finas Lineares Ramificadas	0,21	0,16	0,56
Amorfas	0,11	0,24	0,20
Centro Radiotransparente Em Pipoca/Grosseiras	-	-	-
Redondas/Puntiformes Regulares	0,31	0,32	0,39
Vasculares	-	-	-
Distróficas	0,31	0,41	0,61
Leite de Cálcio	0,41	0,41	0,59
Bastões Longos	-	-	-
<i>Kappa</i> Médio	0,27	0,34	0,49

* *Kappa* generalizado.

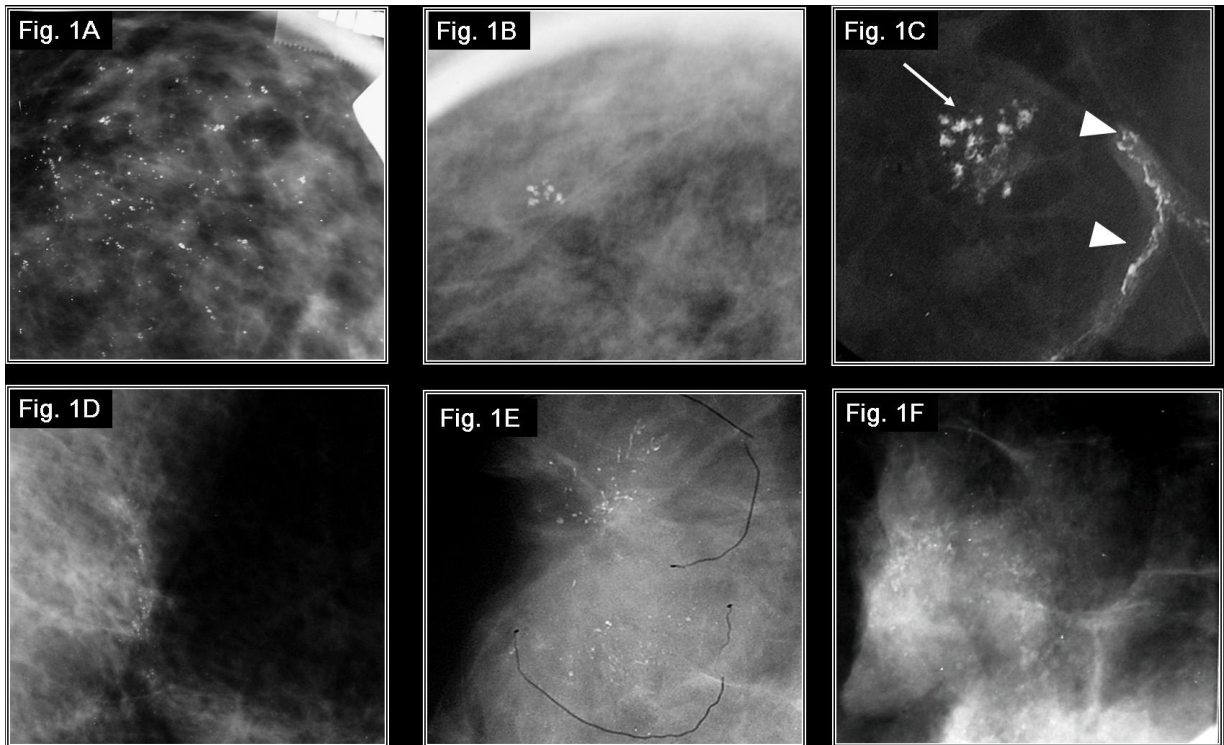


Fig. 1A a 1F: Mamografias com incidências ampliadas de microcalcificações mamárias da amostra estudada. Fig.1A: Paciente de 54 anos, microcalcificações redondas, distribuição regional, BI-RADS® 2, processo benigno. Fig 1B: 49 anos, grupo de microcalcificações distróficas, BI-RADS® 3, processo benigno. Fig 1C: 62 anos, seta: grupo de microcalcificações heterogêneas, BI-RADS® 4b, processo benigno, pontas de setas: calcificações vasculares. Fig 1D: 51 anos, microcalcificações pleomórficas finas de distribuição linear, BI-RADS® 4c, CDI com áreas de CDIS. Fig 1E: 49 anos, microcalcificações finas lineares, distribuição segmentar, BI-RADS® 4c, CDI. Fig 1F: 46 anos, microcalcificações amorfas, distribuição regional, BI-RADS® 4a, processo benigno.

Distribuição das Microcalcificações

Na tabela 5 podemos observar também baixa a moderada concordância intra-observador na descrição da distribuição das microcalcificações, comparando-se os quatro médicos simultaneamente. A distribuição “difusa ou esparsa” não foi utilizada para nenhum caso. A melhor concordância na distribuição foi na descrição “em grupo”. Observaram-se semelhantes valores de *kappa* na análise dos subgrupos benignos, CDI e CDIS (dados não mostrados),

Tabela 5: Variabilidade interobservador entre os quatro médicos simultaneamente e intra-observador de acordo com a distribuição das microcalcificações (BI-RADS® quarta edição).

Distribuição BI-RADS 4ª edição	INTEROBSERVADOR		INTRA- OBSERVADOR
	Primeira leitura <i>Kappa*</i>	Segunda leitura <i>Kappa*</i>	<i>Kappa*</i>
Em grupo	0,62	0,48	0,62
Linear	0,38	0,50	0,45
Segmentar	0,32	0,28	0,44
Regional	0,37	0,29	0,42
Difusa ou esparsa	-	-	-
<i>Kappa</i> médio	0,42	0,39	0,48

* *Kappa* generalizado.

Avaliação da Acurácia

Os parâmetros de acurácia da interpretação de consenso para as duas leituras são demonstrados na tabela 6. Avaliando-se os casos de malignidade, observou-se que dos 18 CDI, a maioria concentrou-se na categoria 5. Obtendo-se a seguinte distribuição: categoria 5 (11 casos na primeira leitura e 7 na segunda), categoria 4c (4 casos na primeira leitura e 7 na segunda), 4b (2 na primeira leitura e 3 na segunda) e 4a apenas um na primeira e um na segunda leitura. Nos 29 CDIS observou-se a seguinte distribuição: categoria 5 (3 casos na primeira leitura e 4 na segunda), 4c (11 casos na primeira e 10 na segunda), 4b (8 casos na primeira leitura e 8 na segunda) e 4a (3 casos na primeira leitura e 6 na segunda). Na primeira leitura ocorreram quatro falsos negativos e na segunda, apenas um, todos haviam sido classificados na categoria 3 e tratavam-se de CDIS.

Tabela 6: Razão de verossimilhança (*likelihood ratio* - LR) e VPP para câncer de mama de acordo com as categorias do BI-RADS® – interpretação de consenso na primeira e na segunda leitura.

Categorias BI-RADS® 4ª edição	Primeira leitura		Segunda leitura	
	LR	VPP	LR	VPP
2	<0,01	----	<0,01	----
3	0,36	----	0,08	----
4a	0,39	25%	0,63	35%
4b	0,84	42%	0,80	41%
4c	1,78	60%	3,32	74%
5	> 16,5	93%	> 20,0	100%

Discussão:

Desde a associação das microcalcificações com câncer de mama reportada inicialmente por Leborgne em 1951²⁴, vários aspectos relacionados à descrição mamográfica de morfologia, distribuição e categorização têm sido discutidos. Em 1984, Le Gal e cols. através da análise da correlação histológica das microcalcificações identificadas na mamografia, recomendaram uma forma bastante didática na descrição morfológica²⁵. Em 1993, o ACR desenvolveu o *Breast Imaging Reporting and Data System* (BI-RADS®) com o principal objetivo de padronizar o vocabulário e as recomendações dos laudos mamográficos²⁶. A sua última edição, o BI-RADS® ilustrado quarta edição¹⁷, traduzido para a língua portuguesa em 2005¹⁹, inclui algumas modificações na terminologia para caracterização das microcalcificações mamárias. Além disso, sugere a subdivisão opcional da categoria 4. Neste estudo procuramos seguir essas recomendações.

Nossos resultados foram semelhantes aos de estudos prévios^{7,10} que utilizaram outras edições do BI-RADS® sem a subdivisão da categoria 4 e apresentaram valores médios de *kappa* 0,36 a 0,57, também indicando baixa a moderada concordância geral. É importante ressaltar que quanto maior o número de variáveis comparadas e de observadores, mais baixos serão os níveis de reprodutibilidade dos testes de concordância. Observamos que alguns estudos utilizam estratégias de ponderação para o cálculo do *Kappa* e outros não, o que

também pode trazer limitações na comparação entre os seus resultados¹⁹. Neste estudo, optou-se pelo uso do *kappa* ponderado para a análise geral da variabilidade entre as categorias finais e do *kappa* generalizado para avaliação individual de cada categoria, morfologia e padrão de distribuição. Dessa forma, utilizando a ponderação do *Kappa* na avaliação de pares de leitores para categoria final (*kappa* médio de 0,49 interobservador e 0,57, intra-observador) observamos que mudança sutis de categorias, de 3 para 4a e 4a para 4b por exemplo, foram mais freqüentes do que mudanças com maior peso. Também de acordo com outros estudos^{3,6}, houve melhora da concordância à medida que aumentou o número de anos de experiência do médico radiologista, o médico, com maior número de anos de prática em mamografia apresentou importante concordância na categorização final (*kappa* 0,70).

Cosar e cols.¹⁶ na avaliação de 83 casos de microcalcificações utilizando a terminologia da quarta edição do BI-RADS®, semelhante aos nossos resultados, encontrou baixa a moderada concordância geral (*kappa* médio de 0,49) entre três diferentes observadores (comparados em pares). Na descrição da morfologia o *kappa* foi de 0,42 (0,33 a 0,66), com maior variabilidade também na morfologia “amorfas”. Neste trabalho não se optou pela subdivisão da categoria 4, as microcalcificações foram classificadas em tipicamente benignas, intermediária e altamente suspeitas e entre os casos de câncer apenas 9 (10,8%) tratavam-se de CDIS. No nosso estudo 62% dos casos de câncer foram CDIS. A maior variabilidade encontrada nos casos de CDIS também foi na descrição da morfologia. Sabemos que o carcinoma ductal *in situ* representa um grupo heterogêneo de lesões pré-invasivas que exhibe uma variedade de padrões histológicos²⁷. A necrose encontrada nos casos de CDIS é frequentemente associada à microcalcificações distróficas, que na mamografia podem apresentar-se como pleomórficas finas e finas lineares ramificadas²⁸. Entretanto, tem sido demonstrada variável correlação entre a histologia dos tipos de CDIS e os padrões mamográficos, com significativas sobreposições^{29,30,31}.

Os critérios para a subdivisão da categoria 4 não são explicitamente mencionados no BI-RADS®. Indica-se a separação do grau de suspeita das microcalcificações em dois grupos de morfologia definidas, as de intermediária suspeita (amorfas ou indistintas e heterogêneas grosseiras) e as de alta

probabilidade de malignidade (pleomórficas finas e finas lineares ou finas lineares ramificadas). Juntamente com a morfologia deve-se considerar também a avaliação do padrão de distribuição, podendo haver substancial subjetividade na escolha entre as categorias finais. Observamos que não há uma tendência no BI-RADS® na recomendação da utilização da categoria 3 (achado provavelmente benigno, que recomenda seguimento em 6 meses) para microcalcificações, pois sabe-se estabilidade não é um critério de benignidade totalmente confiável para este tipo de lesão³². A utilização da categoria 3 para microcalcificações descritas como amorfas ou indistintas foi observada neste estudo em dois casos falsos negativos (casos de CDIS) e as principais discordâncias entre os observadores nas duas leituras foram na morfologia amorfas ou indistintas nas categorias 3, 4a e 4b. Este tipo de morfologia, considerada de intermediária suspeita, em dois importantes estudos^{10,17}, apresentou uma taxa de malignidade em torno de 20% a 26%, especialmente associadas com casos de CDIS, na maioria dos casos em distribuição segmentar e linear. Entretanto, observa-se que há variabilidade na classificação final das microcalcificações amorfas ou indistintas de acordo com a literatura¹⁰ e para muitas destas lesões é indicado seguimento mamográfico (BI-RADS® categoria 3)^{7,10}. Questionamos se uma explicação mais objetiva da maneira de realizar a subdivisão em especial nas categorias de 3 a 4b, com uma abordagem diferenciada da morfologia amorfa ou indistinta no texto do BI-RADS® poderia reduzir estas inconsistências, que resultam na problemática da subjetividade na indicação de biópsias observada atualmente⁷.

No capítulo guia do BI-RADS®¹⁷, observa-se a adição do termo “heterogêneas grosseiras” e a orientação para a utilização da morfologia “pleomórficas finas” na descrição microcalcificações menores que 0,5 mm. A concordância observada no tipo morfológico “heterogênea grosseira”, apresentou melhora na segunda leitura deste estudo, podendo estar relacionado à incorporação desta terminologia nova, com o aprendizado no período de seis meses.

S. Ciatto e cols.³³ avaliaram a variabilidade interobservador e intra-observador (sendo a segunda leitura realizada com um intervalo médio de 2 semanas) entre 12 médicos radiologistas para as categorias finais do BI-RADS® quarta edição. Neste estudo foram incluídas lesões caracterizadas por nódulos e/ou microcalcificações. Na avaliação estatística obteve-se valores de *kappa* generalizado interobservador de

0,44 para a categoria 2; 0,12 para a 3; 0,25 para a categoria 4 como um todo; 0,08 para a 4a; 0,07 para a 4b; 0,10 4c; 0,41 para a categoria 5 e 0,20 para todas as categorias combinadas. A concordância intra-observador apresentou maiores valores *kappa* (0,24 a 0,90) para todas as categorias, mas o intervalo entre as duas leituras foi em média apenas duas semanas. O viés de memória é um fator que deve ser fortemente considerado neste estudo. O maior número de observadores e algumas diferenças na utilização da estatística *kappa* são fatores que limitam a comparação deste estudo com os nossos resultados. Observamos que no nosso estudo, com quatro médicos comparados, as categorias finais, as morfologias e os padrões de distribuição analisados individualmente também apresentaram maior variabilidade interobservador que intra-observador, entretanto com valores *kappa* generalizado mais elevados.

Uma limitação que deve ser considerada neste trabalho, bem como de grande parte de trabalhos semelhantes aqui já citados, é a seleção da amostra. Todos os casos avaliados possuíam uma indicação prévia de biópsia. Isto também pode estar relacionado ao número de falsos positivos e a algumas das inconsistências na classificação final encontradas. Todos os radiologistas dispunham de incidências ampliadas das microcalcificações de interesse. Este tipo de amostra poderia ser representativo de mamografia diagnóstica, mas não reflete completamente a prática clínica.

Lazarus e cols.¹⁸, na análise por cinco médicos de 94 casos de lesões referidas para biópsia (32 casos de microcalcificações), identificadas na ecografia e na mamografia, também utilizando a terminologia da quarta edição do BI-RADS®, obteve baixa concordância na subdivisão da categoria 4 (*kappa* = 0,28). Na avaliação da acurácia, este autor, encontrou um VPP que variou de 6% a 53%, com o maior número de falsos positivos e menor concordância nas categorias 4a e 4b, similar aos nossos resultados. No nosso estudo, o VPP da categoria 4 com a subdivisão ficou entre 25%-30% a 60%-74%. A acurácia foi menor nas categoriais 4a (LR, 0,39 e 0,63) e 4b (LR, 0,84 e 0,80). Parece ter havido uma tendência a melhora da acurácia da segunda leitura. A classificação sugerida pelo BI-RADS® nas subdivisões 4c (LR 1,78 e 3,32, na primeira e na segunda leitura, respectivamente) e 5 (LR > 16,5 e > 20, na primeira e na segunda leitura, respectivamente) estiveram associadas com alto risco para câncer de mama.

Portanto, observou-se um aumento gradativo da probabilidade de malignidade da categoria 4a à 5, com uma boa estratificação em comparação com os dados existentes na literatura até o momento. As recomendações atuais preconizam um VPP de 25% a 40% para câncer de mama entre as lesões que são encaminhadas para biópsia^{34,35}. Entretanto, um recente estudo³⁶ demonstrou que há uma tendência crescente nas indicações de biópsias mamárias, com aumento do número de resultados benignos e sem acréscimo do número de casos de câncer diagnosticados. Estes resultados são atribuídos em grande parte ao aumento na detecção e no número de biópsias em casos de microcalcificações^{35,36}.

A avaliação da aplicabilidade da ferramenta BI-RADS® em diferentes contextos e línguas é muito importante e os esforços realizados a cada nova edição são extremamente elogiáveis. Nosso trabalho expande o conhecimento atual, demonstrando que a terminologia da quarta edição do BI-RADS®, incluindo a subdivisão da categoria 4, apresentou baixa a moderada concordância geral intra-observador e interobservador, obtendo-se boa estratificação da probabilidade de malignidade entre os casos de microcalcificações avaliados, com maior acurácia nas categorias 4c e 5. Identificamos maior variabilidade no uso das categorias 3, 4a e 4b e da morfologia amorfa e menor acurácia nas categorias 4a e 4b nos casos de microcalcificações avaliados. Enquanto aguardamos os efeitos de novos aprimoramentos do léxico do BI-RADS® a observação desses aspectos pode ser útil tanto na prática radiológica, durante a interpretação, quanto na prática clínica, durante a tomada de decisões baseadas em laudo mamográfico de pacientes com microcalcificações mamárias.

Bibliografia:

- ¹ Parkin MD, Bray F, Ferlay J, Pisani P. Global Cancer Statistics, 2002. *CA Cancer J Clin* 2005;55:74-108.
- ² Fletcher SW, Elmore JG. Clinical practice: mammographic screening for breast cancer. *N Engl J Med*. 2003; 348:1672-1680.
- ³ Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. *N Engl J Med* 1995; 331:1493–1499.
- ⁴ Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists: findings from a national sample. *Arch Intern Med* 1996; 156:209–213.
- ⁵ Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. *J Natl Cancer Inst* 1998;90:1801–1809.
- ⁶ Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: Inter- and Intraobserver Variability in Feature Analysis and Final Assessment. *AJR* 2000;174:1769-77.
- ⁷ Berg WA, D'Orsi CJ, Jackson VP et al. Does Training in the Breast Imaging Reporting and Data System (BI-RADS®) Improve Biopsy Recommendations or Feature Analysis Agreement with Experienced Breast Imagers at Mammography? *Radiology* 2002; 224:871–880.
- ⁸ Taplin SH, Ichikawa LE, Kerlikowske K, et al. Concordance of breast imaging reporting and data system assessments and management recommendations in screening mammography. *Radiology* 2002;222(2):529–535.
- ⁹ Baker JA, Kornguth PJ, Floyd CE. Breast Imaging Reporting and Data System Standardized Mammography Lexicon: Observer Variability in Lesion Description. *AJR Am J Roentgenol* 1996;166:773–8.
- ¹⁰ Berg WA, Arnoldus CL, Teferra E, Bhargavan M. Biopsy of Amorphous Breast Calcifications: Pathologic Outcome and Yield at Stereotactic Biopsy. *Radiology* 2001; 221:495–503.
- ¹¹ Egan RL, McSweeney MB, Sewell CW. Intramammary calcifications without an associated mass in benign and malignant diseases. *Radiology* 1980; 137:1–7.
- ¹² Roses DF, Harris MN, Gorstein F, Gumport SL. Biopsy for microcalcification detected by mammography. *Surgery* 1980; 87:248–252.
- ¹³ Hall FM, Storella JM, Silverstone DZ, Wyshak G. Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography. *Radiology* 1988; 167:353–358.

- ¹⁴ Hall JA, Murphy DC, Hall BR, Hall KA. Mammographic abnormalities and the detection of carcinoma of the breast. *AmJ Obstet Gynecol* 1993; 168:1677–1680; discussion 1680–1682.
- ¹⁵ Cosar ZS, Cetin M, Tepe TK, Cetin R, Zarali AC. Concordance of mammographic classifications of microcalcifications in breast cancer diagnosis: Utility of the Breast Imaging Reporting and Data System (fourth edition). *Clin Imaging*. 2005; 29:389-95.
- ¹⁶ Berg WA, Arnoldus CL, Teferra E, Bhargavan M. Biopsy of Amorphous Breast Calcifications: Pathologic Outcome and Yield at Stereotactic Biopsy. *Radiology* 2001; 221:495–503.
- ¹⁷ Liberman L, Abramson A, Squires FB, Glassman JR, Morris EA, Dersshaw DD. The Breast Imaging Report Breast Imaging Reporting and Data System: Positive Predictive Value of Mammographic Features and Final Assessment Categories. *AJR* 1998; 171: 35-40.
- ¹⁸ American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS), Breast Imaging Atlas 4th ed. Reston, VA: American College of Radiology, 2003.
- ¹⁹ Colégio Brasileiro de Radiologia. Sistema de Laudos e Registro de Dados de Imagem da Mama. Colégio Brasileiro de Radiologia, 2005.
- ²⁰ Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS® lexicon for US and mammography: interobserver variability and positive predictive value. *Radiology* 2006; 239:385-91.
- ²¹ Crewson PE. Reader Agreement Studies. *AJR* 2005; 185:1391-7.
- ²² Fleiss JL. Statistical methods for rates and proportions, 2nd ed. New York, NY: Wiley, 1981.
- ²³ Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- ²⁴ Leborgne R. Diagnosis of Tumors of the Breast by Simple Roentgenography: Calcifications in Carcinoma. *AJR* 1951;65:1-11.
- ²⁵ Le Gal M, Chavanne G, Pellier D. Valeur diagnostique des microcalcifications groupees decouvertes par mammographies: a propos de 277 cas avec verification histologique et sans tumeur du sein palpable. *Bull Cancer (Paris)* 1984;71:57-64.
- ²⁶ American College of Radiology. Breast imaging reporting and data system (BI-RADS®). Reston, VA: American College of Radiology, 1993.
- ²⁷ Arpino G, Laucirica R, Elledge RM. Premalignant and In Situ Breast Disease: Biology and Clinical Implications. *Ann Intern Med* 2005;143:446-457.

- ²⁸ Fisher ER, Costantino J, Fisher B, Palekar AS, Redmond C, Mamounas E. Pathologic findings from the National Surgical Adjuvant Breast Project (NSABP) Protocol B-17. Intraductal carcinoma (ductal carcinoma in situ). The National Surgical Adjuvant Breast and Bowel Project Collaborating Investigators. *Cancer* 1995;75:1310-9.
- ²⁹ Stomper PC, Connolly JL. Ductal carcinoma in situ of the breast: correlation between mammographic calcification and tumour subtype. *AJR* 1992;159:483-5.
- ³⁰ Yeh ED, Slanetz PJ, Oyama T, Giardino AA, Koerner FC. Mammographic appearance of ductal carcinoma in-situ (DCIS) does not reliably predict histologic subtype. *Radiology* 1998;209:271.
- ³¹ Dinkel Hp, Gassel Am, Tschammler A. Is the appearance of microcalcifications on mammography useful in predicting histological grade of malignancy in ductal cancer in situ? *The British Journal of Radiology* 2000;73:938-944.
- ³² Lev-Toaff AS, Feig SA, Saitas VL, Finkel GC, Schwartz GF. Stability of malignant breast microcalcifications. *Radiology*. 1994;192:153-156.
- ³³ Ciatto S, Houssami N, Apruzzese A et al. Reader variability in reporting breast imaging according to BI-RADS assessment categories (the Florence experience). *Breast* 2006 Feb;15(1):44-51.
- ³⁴ Brown ML, Houn F, Sickles EA, Kessler LG. Screening mammography in community practice: positive predictive value of abnormal findings and yield of follow-up diagnostic procedures. *AJR Am J Roentgenol* 1995;165:1373–1377.
- ³⁵ Burnside ES, Rubin DL, Fine JP, Shachter RD, Sisney GA, Leung WK. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. *Radiology* 2006 Sep;240:666-73.
- ³⁶ Gur D, Wallace LP, Klym AH, et al. Trends in recall, biopsy, and positive biopsy rates for screening mammography in an academic practice. *Radiology* 2005;235:396–401.