

ESTUDO INTRA-OBSERVADOR DA INTERPRETAÇÃO ULTRA-SONOGRÁFICA DE LESÕES DE MAMA SEGUNDO BI-RADS®

M. J. G. Calas*, R.M.V.R. Almeida** e W. C. A. Pereira**

*CDPI – Centro de Diagnóstico por Imagem, Rio de Janeiro, Brasil

**Programa de Engenharia Biomédica – COPPE/UFRJ, Rio de Janeiro, Brasil

e-mail: mjpgcalas@terra.com.br wagner@peb.ufrj.br

Abstract: The standardization of criteria to interpretation and diagnosis of medical images is a desired but not simple task. The *American College of Radiology (ACR)* has proposed in recent years the BI-RADS® system, which is a set of rules to characterize the morphometric aspects of breast masses. This paper presents an initial study of intraobserver performance when applying the BI-RADS® system to 40 breast ultrasound images (19 benign, 21 malign). The result with 3 specialists was divided in 3 groups: total agreement (8 cases), partial agreement (11 cases) and no agreement (21). The consequences for the patient of each level of agreement (including the Kappa index) are discussed.

Key words: ultra-som, BI-RADS, mama.

Introdução

Na busca do diagnóstico precoce do câncer de mama, a associação de métodos diagnósticos tem sido utilizada com êxito. A ultra-sonografia mamária (USM) tem grande participação como exame complementar à mamografia e à clínica, consolidando-se como um importante exame de imagem de doenças mamárias. Além de poder detectar algumas lesões suspeitas iniciais não vistas à mamografia, ajuda na diferenciação de nódulos císticos e sólidos, como também nódulos sólidos benignos e malignos [1, 2].

A classificação BI-RADS® para a USM, criada pelo Colégio Americano de Radiologia (ACR) [3], propõe uma escala de categorias de avaliação para o risco de malignidade. No Brasil, os ultra-sonografistas vêm aderindo a este sistema.

Para que uma característica ecográfica isolada seja aplicável na diferenciação de uma lesão, esta deve ter um alto índice de concordância inter-observador. Para tal, por ser a ultra-sonografia um método operador-dependente, torna-se indispensável a qualificação do profissional que, aliada à experiência, permite uma avaliação acurada das características morfológicas e a classificação da imagem [4, 5, 6].

Segundo Baker e Soo [5], um modelo de classificação e diferenciação de lesões benignas e malignas deve apresentar: (1) especificidade moderada, permitindo redução no número de biópsias benignas; (2) alto valor preditivo negativo (VPN), permitindo o

seguimento das lesões caracterizadas como benignas; (3) melhorar a precisão da utilização associada a mamografia; (4) apresentar uma baixa variação inter e intra-observador e (5) ser conveniente para uso clínico rotineiro, podendo ser reproduzido em diferentes equipamentos e em diferentes populações de pacientes.

O objetivo deste trabalho é realizar um estudo intra-observador de avaliação de imagens ecográficas mamárias, utilizando a classificação BI-RADS®.

Material e Método

O banco de dados contém 40 imagens ecográficas, com registro nos eixos longitudinal e transversal, e diagnóstico histopatológico. As imagens foram digitalizadas (*Scanner* Gotec 5760) e gravadas em CD.

O sistema BI-RADS® é composto por sete classes: Categoria 0 (Incompleta), Categoria 1 (Negativa), Categoria 2 (Achado benigno), Categoria 3 (Provavelmente benigno), Categoria 4 (Anormalidade suspeita), Categoria 5 (Altamente sugestivo de malignidade) e Categoria 6 (Malignidade já conhecida).

Para avaliação dos observadores foram escolhidas apenas as categorias 2, 3, 4 e 5, por serem os casos que possuíam diagnóstico histopatológico.

Participaram deste estudo três profissionais (A, B e C), que atuam em radiologia mamária em diferentes instituições, com experiência de 4, 13 e 23 anos, respectivamente.

Cada observador recebeu um CD com as imagens, e com um protocolo de critérios ecográficos BI-RADS®. Após três meses da primeira leitura, os observadores receberam novamente as imagens. Uma terceira análise ocorreu após seis meses da segunda leitura. Os observadores não sabiam que as 40 imagens eram as mesmas a cada rodada; não tiveram acesso ao prontuário das pacientes, nem ao exame histopatológico ou aos resultados das suas análises prévias.

Os protocolos preenchidos pelos especialistas foram classificados em três grupos, segundo o nível de concordância das categorias BI-RADS® para cada imagem: (a) Concordância Total: imagens para as quais cada especialista indicou a mesma categoria BI-RADS® nos três estudos; (b) Concordância Parcial: imagens para as quais as categorias atribuídas pelos especialistas divergem, porém a conduta médica é a mesma. Ou seja, foram agrupadas as categorias 4 e 5 (prosseguir a

investigação) e as categorias 2 e 3 (controle USM periódico); (c) Discordância: imagens cujas categorias atribuídas pelos profissionais não coincidiram e a conduta médica posterior também é diversa.

Uma análise estatística intra-observador foi feita com o índice Kappa modificado (K) [7] que permite avaliar a concordância simultânea entre vários observadores.

Para os testes diagnósticos, as imagens de categorias 2 e 3 foram consideradas como caso benigno e as de categorias 4 e 5, como caso maligno.

Resultados

Das 40 lesões, 19 eram benignas e 21 malignas (Tabela 1). O número de casos de concordância e discordância intra-observador pode ser visto na Tabela 2. Na Tabela 3 estão os valores de Kappa calculados para cada observador, para as categorias em separado e em grupo categorias (2+3) e (4+5).

Tabela 1: Histopatologia das 40 imagens.

Histologia	Nº de casos
Cisto	4
Alterações fibrocísticas	1
Hematoma	1
Fibroadenoma	13
*CDI Grau I	8
*CDI Grau II	7
*CDI Grau III	3
Carcinoma Lobular Infiltrante	1
Carcinoma Metástase Linfonodo	1
Carcinoma Medular	1

*CDI – carcinoma ductal infiltrante

Tabela 2: Nº de casos de concordância total, parcial e discordância para cada especialista, nas três rodadas.

Observadores	total	Parcial	Discordância
A	15	17	8
B	19	17	4
C	22	11	7

Tabela 3: Índice Kappa modificado para cada observador. Kappa* indica classes agrupadas (2+3) e (4+5).

Observadores	Kappa	Kappa*
A	0,37	0,73
B	0,51	0,87
C	0,59	0,76

Observador A - Dos oito casos de discordância, seis eram de carcinomas, onde em pelo menos uma rodada, houve indicação errada de controle, sendo o caso de carcinoma medular o mais grave, pois foi sugerido o controle por duas vezes. Para duas lesões benignas, houve desnecessária indicação de biópsia, sendo o caso de hematoma indicado duas vezes.

Observador B - Todos os quatro casos de discordância eram carcinomas, havendo indicação errada de controle semestral para todos. No caso de carcinoma medular, a indicação ocorreu por duas vezes.

Observador C - Todas as sete discordâncias eram lesões benignas, havendo indicação desnecessária de biópsia em todas, com dupla indicação em dois casos de fibroadenoma.

Na tabela 4 estão a média dos três estudos, por cada observador de Sensibilidade, Especificidade, Valor preditivo positivo (VPP), Valor preditivo negativo (VPN) e Acurácia, dados em valores percentuais.

Tabela 4: Parâmetros de avaliação percentual por observador.

Obs.	Sens	Espec	VPN	VPP	Acurácia
A	79,3	94,7	80,9	94,8	86,6
B	87,2	94,7	87,6	94,7	90,8
C	95,2	78,9	93,6	87,1	87,5

Discussão

A ultra-sonografia é um exame altamente dependente do operador e do equipamento, conduzindo a variações de interpretação da imagem e de recomendações de conduta. Para realizar um procedimento com qualidade, o examinador deve conhecer bem as características do equipamento, a técnica, a interpretação do exame, a história da paciente, e as patologias mamárias [8, 9].

No presente estudo, os observadores consideraram como limitações importantes a falta de correlação com história clínica e mamografia e o fato das imagens serem estáticas (sem tempo-real). Apesar do processo de digitalização das fotos originais ter causado perda de qualidade, nenhum observador excluiu nenhuma imagem por este motivo.

Com relação a discordância no caso do Observador A, houve indicação errada de controle em seis dos oito casos, sendo um de carcinoma medular. Este tipo de câncer possui contorno parcialmente circunscrito e discreto reforço acústico posterior, facilmente confundível com um cisto complicado (cisto espesso). Para duas lesões benignas, houve desnecessária indicação de biópsia, sendo o caso de hematoma, explicável pois (sem correlação clínica) apresenta imagem similar a cisto complexo, que tem categoria 4.

Nos casos de discordância do Observador B, todos eram carcinomas, havendo indicação errada de controle semestral para todos (duas vezes para o carcinoma medular, conforme explicado para o Observador A).

No Observador C, nas sete discordâncias, houve indicação desnecessária de biópsia, com dupla indicação em dois casos de fibroadenomas, porém sem prejudicar a paciente com atraso de diagnóstico.

Baker e Soo [5], analisando 152 lesões, encontraram interpretação divergente em 23 casos (15,1%). Foram consideradas divergências quando havia mudança de conduta para a paciente. Estas diferenças se devem a quatro casos falso-negativos (dois casos de câncer classificados como benignos e dois casos de lesões não vistas na USM), e 14 casos falso-positivos (cinco casos benignos dados como suspeitos e nove casos de “falsas imagens” ou seja, artefatos). Houve ainda discordância em três casos descritos como cistos e que a punção

confirmou serem sólidos e dois casos onde houve discrepância entre os achados ecográficos e os mamográficos.

No estudo de Rahbar e cols. [10], a análise inter-observador mostrou concordância dos três observadores para os critérios de benignidade, inclusive em um caso de classificação incorreta de carcinoma medular, semelhante ao que ocorreu no presente trabalho pelos observadores A e B.

No trabalho de Arger e cols. [4], a concordância entre os observadores ocorreu em 39 casos (55,7%). Das 42 massas categoria 5, houve discordância em 11 (26%). Destas 11, seis casos ficaram entre categorias 4 e 5 (sem alteração de conduta) e em cinco casos, três observadores classificaram como categoria 3 (mudança de conduta), no entanto eram 3 casos benignos. Das 20 massas categoria 4, em 15 (75%) houve discordância. Destas 15, duas lesões foram classificadas como 3 e se tratavam de dois carcinomas. Das 8 massas classificadas como categoria 3, cinco (62%) foram discordantes. Nenhuma das massas apresentou resultado de malignidade nesta categoria, não havendo teoricamente alteração na conduta clínica. No entanto, os autores justificam um valor preditivo positivo alto (10,6%) para esta categoria (diferente dos 2% sugerido pelo sistema BI-RADS®) em função de serem selecionadas pacientes para prosseguir com biópsia, o número pequeno de casos, os observadores não terem acesso a mamografia e o estudo não incluir categoria 1 (mama normal).

Na Tabela 5 estão os resultados dos principais estudos inter e intra-observadores encontrados na literatura, que usam o índice Kappa como parâmetro.

Tabela 5: Resultados da literatura segundo o Kappa.

Autores ano/ ref.	Número casos	Número observador	Kappa
Skaane 1997 [11]	200	4	0,48 (0,37-0,61) Inter 0,51 (0,41-0,60) Intra 0,66
Baker 1999 [5]	60	5	Inter 0,72-0,75 0,71 – 0,83 - lesões > 7mm 0,37 – 0,74 - lesões < 7mm
Bosch 2003 [12]	60	3	0,28
Del Frate 2006 [13]	256	3	0,52
Lazarus 2006 [14]	94	5	Inter 0,62 Intra – 0,61-0,80
Berg 2006 [15]	10	11	Intra A- 0,37 B- 0,51 C- 0,59
Lee 2008 [16]	150	4	
Este trabalho	40	3	

O trabalho de Skaane e cols. [12] apresentou resultados inter-observador na mamografia, USM e em ambos os métodos combinados, concluindo que a variação inter-observador é um sério problema no estudo da imagiologia mamária. O índice Kappa foi menor na interpretação da USM (0,48) e maior para os métodos combinados. Estes resultados confirmam o ponto de vista de que a USM deve ser realizada com o conhecimento prévio dos achados da mamografia.

A presente pesquisa apresenta um número pequeno de casos (40 casos) em comparação com os artigos descritos na Tabela 5 (variando entre 60 e 256 casos). No entanto não existe na literatura estudo com 3 observadores, que realizaram avaliação intra-observador por três vezes. Apenas nos estudos de Baker e cols. [5], com cinco observadores, um deles realizou uma segunda análise seis meses depois; e Lee e cols. [16], onde os quatro radiologistas realizaram o estudo duas vezes, num intervalo de oito semanas. O trabalho de Berg e cols. [15] apresenta 11 radiologistas, porém apenas 10 pacientes foram avaliadas em tempo real.

Lazarus e cols. [14] determinaram utilizando o sistema BI-RADS® para a determinação do VPP das categorias 4a, 4b e 4c da mamografia e da ultrasonografia, encontraram uma concordância “modesta” ($k = 0,28$). A maior concordância ocorreu para lesões classificadas como categoria 5.

Neste trabalho, para classes BI-RADS® separadas obteve-se uma concordância “modesta” para o observador A e “moderada” para os observadores B e C. No entanto, valores de Kappa obtidos são inferiores aos trabalhos de estudo intra-observador da literatura [5, 16]. Para classes BI-RADS® agrupadas (2+3) e (4+5), obteve-se um “acordo substancial” para todos os 3 observadores.

Na comparação da literatura dos testes diagnósticos para USM, Shimamoto e cols. [18] avaliaram 54 lesões (30 benignas e 24 malignas), sendo os observadores divididos em grupo *senior* e *junior*. A sensibilidade variou de 50 a 62,5% no grupo *junior* e de 62,5 a 83,3% no *senior*. A acurácia variou entre 53,7% e 61,1% no grupo *junior* e entre 64,8 e 72,2 % no *senior*. Os seus resultados sugerem que a concordância inter-observador é mais dependente da dificuldade de cada caso do que a experiência do operador.

Na literatura, os testes diagnósticos variam: sensibilidade entre 79,3 e 99,8%, especificidade de 32,9 a 96,8%, VPP de 25,3 e 85,5%, VPN de 89,4 a 96,6% e uma acurácia de 53,8 a 97% [17, 19-22].

Na casuística do presente trabalho, os observadores com maior tempo de experiência apresentaram os melhores resultados de testes diagnósticos.

Os testes sensíveis são utilizados quando há necessidade de se realizar o rastreamento de uma doença em grupos populacionais (p.ex., a mamografia). Os testes específicos são utilizados para confirmar um diagnóstico, já que um teste muito específico raramente resultará positivo na ausência da doença. Com relação ao estudo dos valores preditivos, deve-se considerar a prevalência de câncer na casuística estudada, sabendo-se que maiores prevalências tendem a aumentar os VPP. Em resumo, é mais útil para o especialista um resultado negativo de um teste muito específico, com menor probabilidade de falsos positivos.

Entretanto a diagnose é essencialmente baseada na avaliação subjetiva dos achados. Assim, no intuito de evitar ambigüidade em observações subjetivas e dependência do operador em exames por ultra-som, vários tipos de sistemas de auxílio diagnóstico por

computador (CAD-*computer-aided diagnosis*) têm sido propostos [22]. A idéia é que avaliações quantitativas ajudem a diminuir as discordâncias entre observadores.

Apesar de todos os trabalhos da literatura utilizarem equipamentos de alta resolução e profissionais treinados, as divergências observadas são basicamente devidas às diferentes metodologias de cada estudo.

Conclusão

O presente estudo intra-observador apresentou uma concordância modesta para o observador com menos tempo de experiência, e moderada para os observadores com maior tempo de experiência.

A casuística ainda é limitada e um estudo mais amplo deverá ser realizado, porém já foi possível identificar pontos críticos que merecem atenção, p.ex., a importância da formação do profissional, o treinamento na utilização do sistema BI-RADS® e conseqüentemente uma avaliação crítica da indicação desnecessária de biópsias.

Referências

- [1] Calas M.J.G., Castro R., Manoel V.R., Pasqualette H.A., Soares-Pereira P.M. (2002) "Proposta de Normatização dos Laudos de Ultra-Sonografia Mamária" *Femina*, v.30, n.2, p.103-110.
- [2] Paulinelli R.R., Calas MJG, Freitas Junior, R.(2007) "BI-RADS e ultra-sonografia mamária – uma análise crítica" *Femina*, v.35, n.9, p.537-608.
- [3] American College of Radiology. Illustrated Breast Imaging Reporting and Data System Atlas (BI-RADS®) (2003): Ultrasound. In: 4th ed, Reston, V.A.
- [4] Arger P.H., Sehgal C.M., Conant E.F., Zuckerman J., Rowling S.E., Patton J.A. (2001) "Interreader variability and predictive value of US descriptions of solid breast masses: pilot study", *Acad Radiol* v.8, n.4, p.335-342.
- [5] Baker J.A., Kornguth P.J., Soo M.S., Walsh R., Mengoni P. (1999) "Sonography of solid breast lesions: observer variability of lesion description and assessment", *Am J Roentgl*, v.172, n.6, p.1621-1625.
- [6] Levy L., Suissa M., Chiche J.F., Teman G., Martin B. (2007) "BIRADS ultrasonography", *European Journal of Radiology*, v. 61, n.2, p.202-211.
- [7] Sim J, Wright CC. (2005) "The Kappa statistic in reliability studies: use, interpretation, and sample size requirements", *Physical Therapy*, v.85, n.3, p.257-268.
- [8] Baez E, Strathmann K, Vetter M, Madjar H, Hackelöer B-J. (2005) "Likelihood of malignancy in breast lesions characterized by ultrasound with a combined diagnostic score", *Ultrasound in Med & Biol*, v.31, n.2, p.179-184.
- [9] Berg W.A., Blume J.D., Cormack J.B., Mendelson E.B. (2006) "Operator dependence of physician-performed whole-breast US: lesion detection and characterization", *Radiology*, v. 241, n.2, p.355-365.
- [10] Rahbar G., Sie A.C., Hansen G.C., et al. (1999) "Benign versus malignant solid breast masses: US differentiation", *Radiology* v. 213, n.3, p.889-894.
- [11] Skaane P., Engedal K., Skjennald A. (1997) "Interobserver variation in the interpretation of breast imaging. Comparison of mammography, ultrasonography, and both combined in the interpretation of palpable noncalcified breast masses", *Acta Radiologica*, v. 38, p.497-502.
- [12] Bosch A.M., Kessels A.G.H., Beets G.L., et al. (2003) "Interexamination variation of whole breast ultrasound", *The Brit. J. of Radiol.*, v.75, p.328-331.
- [13] Del Frate C., Bestagno A., Cerniati R., Soldano F., Isola M., Puglisi F., Bazzocchi M. (2006) "Sonographic criteria for differentiation of benign and malignant solid breast lesions: size is of value", *Radiol Med*, v.111, n.6, p.783-796.
- [14] Lazarus E., Mainiero M.B., Schepps B., Koelliker S.L., Livingston L.S. (2006) "BI-RADS Lexicon for US and Mammography: interobserver variability and positive predictive value", *Radiology*, v.239, n. 2, p.385-91.
- [15] Berg W.A., Blume J.D., Cormack J.B., Mendelson E.B., Madsen E.L. (2006) "Lesion detection and characterization in a Breast US Phantom: results of the ACRIN 6666 investigators", *Radiology*, v. 239, n.3, p.693-702.
- [16] Lee H-J, Kim E-K, Kim M.J. et al. (2007) "Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound", *European Journal of Radiology*, v.65,n.2, p.293-298.
- [17] Hong A.S., Rosen E.L., Soo M.S., Baker J.A. (2005), "BI-RADS for sonography: positive and negative predictive values of sonographic features", *American Journal of Radiology*, v. 184, p.1260-1265.
- [18] Shimamoto K., Sawaki A., Ikede M., et al. (1998) "Interobserver agreement in sonographic diagnosis of breast tumors", *Eur. J. Ultrasound*, v.8, n.1, p.25-31.
- [19] Costantini M., Belli P., Lombardi R., Franceschini G., Mulè A., Bonomo L. (2006) "Characterization of solid breast masses. Use of sonographic Breast Imaging Reporting and Data System Lexicon" *J Ultrasound Med*, v. 25, n.5, p 649-659.
- [20] Costantini M., Belli P., Ierardi C., Franceschini G., La Torre G., Bonomo L. (2007) "Solid breast mass characterisation: use of the sonographic BI-RADS classification" *Radiol Med*, v.112, n.6, p. 877-894.
- [21] Paulinelli R.R., Freitas-Junior R., Moreira M.A., et al. (2005) "Risk of malignancy in solid breast nodules according to their sonographic features", *J Ultrasound Med*, v.24, p.635-641.
- [22] Drukker K, Giger M.L., Vyborny C.J, Mendelson E.B. (2004) "Computerized Detection and Classification of Cancer on Breast Ultrasound" *Acad Radiol*, v.11, n.5, p.526-535.