

Breast Imaging Reporting and Data System Lexicon for US: Interobserver Agreement for Assessment of Breast Masses

1. Nouf Abdullah, MD,
2. Benoît Mesurolle, MD,
3. Mona El-Khoury, MD and
4. Ellen Kao, MD

± Author Affiliations

1. *'From the Department of Radiology, Cedar Breast Clinic, McGill University Health Center, Royal Victoria Hospital, 687 Pine Ave W, Montreal, QC, Canada H3A 1A1.*

1. **Author contributions:** Guarantor of integrity of entire study, B.M.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, N.A., B.M.; clinical studies, N.A., B.M., M.E.; statistical analysis, N.A.; and manuscript editing, all authors

Abstract

Purpose: To retrospectively evaluate the interobserver agreement of radiologists who used the Breast Imaging Reporting and Data System (BI-RADS) lexicon to characterize and categorize ultrasonographic (US) features of breast masses.

Materials and Methods: No institutional review board approval or patient consent was required. Five breast radiologists retrospectively independently evaluated 267 breast masses (113 benign and 154 malignant masses in 267 patients) by using the BI-RADS US lexicon. Reviewers were blinded to mammographic images, medical history, and pathologic findings. Interobserver agreement was assessed with the Aickin revised κ statistic.

Results: Interobserver agreement varied from fair for evaluation of mass margins ($\kappa = 0.36$) to moderate for evaluation of lesion boundary ($\kappa = 0.48$), echo pattern ($\kappa = 0.58$), and posterior acoustic features ($\kappa = 0.47$) to substantial for evaluation of mass orientation ($\kappa = 0.70$) and shape ($\kappa = 0.64$). For small (≤ 0.7 cm; $n = 49$) or malignant ($n = 154$) masses, low concordance was noted for margin descriptors ($\kappa = 0.30$ and 0.28 , respectively) and BI-RADS category ($\kappa = 0.21$ and 0.26 , respectively). Overall, only fair agreement was obtained for BI-RADS category ($\kappa = 0.30$). Agreement for subdivisions 4a, 4b, and 4c of BI-RADS category 4 was fair ($\kappa = 0.33$), fair ($\kappa = 0.32$), and poor ($\kappa = 0.17$), respectively.

Conclusion: Reproducibility of US BI-RADS terminology is good except for margin evaluation. A trend toward lower concordance was noted for the evaluation of small

masses and malignant lesions. Classification into subdivisions 4a, 4b, and 4c was poorly reproducible.