Published in final edited form as:

J Am Coll Radiol. 2009 December ; 6(12): 851–860. doi:10.1016/j.jacr.2009.07.023.

The ACR BI-RADS® Experience: Learning From History

Elizabeth S. Burnside, MD, MPH, MS^a, Edward A. Sickles, MD^b, Lawrence W. Bassett, MD^c, Daniel L. Rubin, MD, MS^d, Carol H. Lee, MD^e, Debra M. Ikeda, MD^d, Ellen B. Mendelson, MD^f, Pamela A. Wilcox^g, Priscilla F. Butler^g, and Carl J. D'Orsi, MD^h

^a Department of Radiology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin

^b Department of Radiology, University of California, San Francisco, Medical Center, San Francisco, California

^c David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California

^d Department of Radiology, Stanford University School of Medicine, Stanford, California

^e Department of Radiology, Memorial Sloan-Kettering Cancer Center, New York, New York

^f Department of Radiology, Northwestern Memorial Hospital, Chicago, Illinois

^g Department of Quality and Safety, American College of Radiology, Reston, Virginia

^h Breast Imaging Center, Department of Radiology, Emory University Hospital, Atlanta, Georgia

Abstract

The Breast Imaging Reporting and Data System® (BI-RADS®) initiative, instituted by the ACR, was begun in the late 1980s to address a lack of standardization and uniformity in mammography practice reporting. An important component of the BI-RADS initiative is the lexicon, a dictionary of descriptors of specific imaging features. The BI-RADS lexicon has always been data driven, using descriptors that previously had been shown in the literature to be predictive of benign and malignant disease. Once established, the BI-RADS lexicon provided new opportunities for quality assurance, communication, research, and improved patient care. The history of this lexicon illustrates a series of challenges and instructive successes that provide a valuable guide for other groups that aspire to develop similar lexicons in the future.

Keywords

BI-RADS; breast cancer; breast imaging; diagnosis; informatics lexicon; mammography; screening; structured reporting

The Breast Imaging Reporting and Data System® (BI-RADS®), the first practice management system developed for imaging, is called a system because it contains several important components, including 1) a lexicon of descriptors, 2) a recommended reporting structure including final assessment categories with accompanying management recommendations, and 3) a framework for data collection and auditing. In this paper, we focus on the development of the lexicon within BI-RADS, but we also consider important aspects of the system as a whole when indicated.

A HISTORICAL PERSPECTIVE

As mammography utilization increased in the 1980s, wide variability of practices, including disparate quality and inconsistent radiation doses, were cited as substantial problems [1–3]. Organizations such as the American Medical Association also asserted that mammography reports too often contained unintelligible descriptions and ambiguous recommendations [4]. In response, the ACR convened a committee of radiologists, medical physicists, and a US Food and Drug Administration (FDA) representative to develop a voluntary mammography accreditation program in 1986 [5]. The ACR recognized that meaningful descriptors of findings and the precise communication of recommendations in mammography reports were important parts of a quality assurance program. Thus, a separate ACR committee also was charged with drafting guidelines on mammography reporting and management under the title of the Breast Imaging Reporting and Data System [6].

Many well-respected groups, including the American Medical Association, the National Cancer Institute, the Centers for Disease Control and Prevention, the FDA, the American College of Surgeons, and the College of American Pathologists, participated in this development initiative to establish a broad base of support [6]. The inclusion of diverse stakeholders in the development process helped promote consensus and facilitated acceptance.

The first version of BI-RADS included recommendations for the conduct of mammographic imaging, an overall structure for mammography reports, final assessment categories with management recommendations, and a mammography lexicon. The introduction of the document outlined the recommended method to provide efficient and cost-effective mammography: screening mammography, which can be performed without an interpreting physician in attendance (batch reading) to optimize efficiency; and diagnostic mammography, which should be performed with “direct supervision” so that examinations can be tailored to individual patients. Direct supervision is defined as a physician’s being present and immediately available to furnish assistance and direction throughout the performance of the procedure. Direct supervision may also be accomplished “via tele mammography as long as the interpreting physician is immediately available” [7]. These guidelines clarified the optimal practice of mammography services for radiologists, referring physicians, and patients.

The original BI-RADS document also described the overall structure of the breast imaging report, which included a summary of breast density, a description of significant findings (using appropriate descriptors as well as size and location), and a final assessment and management section. The inclusion of a statement describing the general breast tissue type arose from evidence in the literature establishing that increased breast density is accompanied by decreased sensitivity [8–11]. Subsequently, evidence has mounted that increased breast density also is associated with increased breast cancer risk [12–14]. Although these hypotheses are still active areas of research, the inclusion of 4 categories describing breast density (ranging from the almost entirely fatty breast to the extremely dense breast) in the standard mammography report is designed to improve the communication of predicted mammographic performance and breast cancer risk.

The descriptors in the BI-RADS lexicon were selected on the basis of their ability to discriminate between benign and malignant findings as determined by well-designed reader studies. These studies suggested that feature lists containing predictive descriptors could improve diagnostic accuracy. In the first study, investigators developed a decision aid consisting of predictive terms and scales used to assess xero-radiography and reported increased decision-making accuracy in a small reader study [15]. These early features were

then adapted to screen-film mammography and studied in a larger reader study of 150 cases and 6 radiologists [16,17]. These initial studies helped refine the feature lists and led to the development of computer programs to improve both sensitivity and specificity. Furthermore, careful cataloging of descriptors significantly improved performance as case difficulty increased and brought generalist performance to the level of the mammography specialist [16,17]. The studies also guided the development of descriptors in the original BI-RADS lexicon. Once the terms were chosen, they were precisely defined to eliminate ambiguity. The final list of terms was intended to be evidence based and predictive and to foster the clear and accurate communication of mammographic findings.

The BI-RADS Committee went beyond advocating the use of clear and standardized terms and recommended that mammography be “decision oriented.” During the early days of mammography, the American Medical Association specifically complained that mammography interpretation was often indecisive and confusing [4]. In response, the BI-RADS Committee recommended that final impressions be summarized by choosing only one among several standardized final assessment categories at the end of a report, each of which included a matched, also standardized management recommendation. These categories currently are as follows: category 1: negative; category 2: benign finding(s); category 3: probably benign finding—initial short-interval follow-up suggested; category 4: suspicious abnormality—biopsy should be considered; category 5: highly suggestive of malignancy—appropriate action should be taken; and category 6: known biopsy-proven malignancy—appropriate action should be taken. An incomplete category was also provided, category 0: need additional imaging evaluation and/or prior mammograms for comparison. These categories also took an evidence-based approach. The “probably benign” category was based on literature demonstrating that follow-up rather than biopsy is safe and effective management for a clearly defined subset of findings that are very likely benign [18]. The final assessment of “highly suggestive for malignancy” was included at the request of representatives of the American College of Surgeons. Thus, the “highly suggestive” category implied a “classic” finding for malignancy, which enabled women living in underserved areas (without expertise in image-guided diagnosis) to be scheduled for operative diagnosis, using frozen section, and immediate surgical management.

ENVIRONMENTAL INFLUENCES

Concurrent with the development of the BI-RADS lexicon, the mammography community was recognizing the importance of automating clinical practice for a variety of mission critical reasons. First, as a population-based initiative, screening mammography involves the management of a substantial amount of data. Second, the conduct of medical audits to ensure optimal performance and fulfill compliance requirements is onerous if performed manually. Third, optimal tracking for the accurate communication of results requires reliable and reproducible data storage and retrieval. Automated, computer-based management and tracking of mammography results addressed these challenges more efficiently than conventional manual methods [19,20].

The explosive increase in data was not unique to breast imaging practices during the late 1980s. The entire biomedical infrastructure was straining under the increases in data inherent in clinical care and research. There were other, more general efforts to standardize biomedical terminology that paralleled the construction of the BI-RADS mammography lexicon. For example, the Unified Medical Language System (UMLS) was designed and constructed by the National Library of Medicine in 1986 to enhance access to medical data and the scientific literature. The UMLS provides the infrastructure to collect and link controlled vocabularies to facilitate the development of computer systems that understand, retrieve, and classify biomedical literature. The National Library of Medicine itself uses

components of UMLS for its PubMed system. The UMLS, a powerful tool to enable computers to communicate about biomedical data generally, parallels BI-RADS, which provides the same capabilities in the domain of breast imaging.

THE EVOLUTION OF BI-RADS

BI-RADS in general and the lexicon specifically were not intended to be static [6]. After the initial creation of BI-RADS in 1993 [21], 3 more editions were created in 1995 [22], 1998 [23], and 2003 [24]. However, the path to a successfully adopted breast imaging lexicon was not always smooth. Controversy involving the BI-RADS lexicon arose in 1994 with the publication of an editorial in which the author contested that “expertise is the heart of the problem, not terminology. BI-RADS, with its emphasis on words and definitions, is barking up the wrong tree” [25]. The authors of the BI-RADS lexicon responded in a follow-up editorial that addressed many concerns that were being voiced by the community. In particular, the response clarified that the BI-RADS lexicon was intended to be a tool that radiologists would use to communicate with clinicians to convey concise and orderly descriptions of findings in understandable, standardized language, which in turn contributes to an orderly thought process and logical assessments and recommendations. Furthermore, BI-RADS was designed to encourage improvements in expertise because it provided standardized recommendations that could be used for performance tracking: “Without standardized terms to describe important features ... there is no means of obtaining objective data to improve. Indeed, this format is important for all reports we generate, not only mammography” [26]. This thoughtful and productive debate strengthened support for the lexicon.

Each BI-RADS revision added components that were important for clarification, management, or quality assurance. The third edition of BI-RADS incorporated an atlas that provided an artist’s renderings of examples of each descriptor. The fourth edition made several changes in lexicon terminology [27]. To decrease confusion between terminology for overall breast density and the descriptor *density* (referring to a noncalcified finding seen on only one of the two standard mammographic views), the authors adopted the descriptor term *asymmetry* in place of *density*. To parallel this change in terminology, *asymmetric breast tissue* was renamed *global asymmetry* (a nonmass finding seen on at least two views that occupies at least a quadrant), and *focal asymmetric density* was renamed *focal asymmetry* (a nonmass finding seen on at least two views that occupies less than a quadrant). Furthermore, a study demonstrating that amorphous microcalcifications carried a 20% risk for malignancy [28] prompted the BI-RADS Committee to subcategorize suspicious microcalcification descriptors into “intermediate risk” (including the amorphous descriptor) and “higher probability of malignancy.” Additional published data related to microcalcification descriptors demonstrated that the *pleomorphic* descriptor was not stratifying risk beyond the overall risk for suspicious microcalcifications [29]. In response, the BI-RADS Committee divided “pleomorphic” microcalcifications into two more specific categories: “coarse heterogeneous” and “fine pleomorphic.” This distinction was subsequently shown to effectively stratify the probability of malignancy among these types of calcifications [30]. To further help with risk stratification, the fourth edition provided the option to subdivide BI-RADS assessment category 4 into 4A (low suspicion for malignancy), 4B (intermediate suspicion for malignancy), and 4C (moderate concern but not classic for malignancy). These subdivisions provide assessment and reporting options designed to help both physicians and patients understand likely biopsy findings and probable follow-up recommendations [27,31].

The development of a breast ultrasound lexicon, BI-RADS–Ultrasound, first published as part of the fourth edition of BI-RADS, demonstrated similar themes. In 1998, the ACR

received a grant from the Office on Women's Health of the US Department of Health and Human Services to support protocol development for research in breast ultrasound (contract 282-97-0076, Federal Technology Transfer Program to Advance Novel Breast Imaging Technologies, US Public Health Service, Office on Women's Health, Department of Health and Human Services). An ACR expert working group of national and international breast imagers with special interest in breast ultrasound met in Maine to design research projects that might advance use of ultrasound in conjunction with mammography and other imaging modalities. These potential studies included 1) the identification of criteria to differentiate benign from malignant solid masses, 2) ultrasound for breast cancer screening, and 3) using ultrasound to guide diagnostic interventions and as a therapeutic agent (high-frequency ultrasound). Until that time, ultrasound was used primarily for cyst-solid differentiation, despite FDA premarket approval of Advanced Technology Laboratories' (Bothell, Washington; now Philips Medical Systems, Andover, Massachusetts) "high-definition imaging" in 1996. This approval was based on an international, multicenter study involving ultrasound evaluation of nearly 1,000 breast lesions (published as a monograph by Advanced Technology Laboratories rather than in the peer-reviewed literature), which indicated that ultrasound improved specificity for masses found to be indeterminate on mammography and physical examination. This study asserted the need for additional research in defining diagnostic criteria for classifying solid breast masses on ultrasound.

The ACR expert working group proposed a standardized lexicon, similar to BI-RADS for mammography, to provide a foundation for research characterizing solid masses for risk stratification and evidence-based management. For example, this strictly defined lexicon could be used to determine benign and probably benign masses. Soon after the Maine meeting, a subcommittee of the ACR's BI-RADS Committee was formed to officially develop BI-RADS–Ultrasound. After several iterations, the consensus document was presented and tested at select subspecialty meetings, including the Society of Breast Imaging's biennial meeting in San Diego in 2001. The descriptors and assessment categories were validated by statistical analysis of interobserver consistency (κ), showing good agreement for most terms among the experienced and novice breast imaging participants [31].

BI-RADS–Ultrasound was predicated on high-quality images and real-time ultrasound observations and encourages the assessment of combined features. Using descriptors from several feature categories can balance the risk associated with all relevant features, but usually, the most suspicious feature will dominate the final assessment and recommendation [32]. Validation of the grouping of features is derived from univariate, bivariate, and multivariate analyses of features characterizing the lesions submitted as evidence for FDA premarket approval by Advanced Technology Laboratories for its high-definition imaging system. Groups of features were also used in the development of the mass classification algorithm proposed by Stavros et al [33]. For BI-RADS–Ultrasound, the three most important feature categories, taken together are shape, margin, and orientation, the last a feature unique to ultrasound.

BI-RADS–Ultrasound can also contribute to the investigation of emerging technologies. Computer-assisted diagnosis programs use segmentation and feature extraction to classify breast masses on the basis of features similar to BI-RADS–Ultrasound. Structured reporting, rapidly evolving and currently available in breast imaging reporting software packages initially designed for mammography, uses BI-RADS–Ultrasound to construct standardized reports and encourage consistent communication.

The development of BI-RADS–Ultrasound that began more than a decade ago will continue. New feature categories (eg, elasticity) will require standardization, evidence, and validation.

The continuing goals for BI-RADS of providing useful, comprehensive guides to breast imagers for analyzing, assessing, reporting, and managing breast lesions is especially critical for breast ultrasound, which has long been considered an operator-dependent modality.

The fourth edition of BI-RADS also incorporated breast MRI descriptors. Between the early 1970s and the late 1990s, contrast-enhanced breast MRI had shown great promise. Studies demonstrated near 100% sensitivity for detecting early invasive breast cancer, though these results were tempered by more modest specificity [34]. However, attempts to systematically evaluate the literature were stymied by nonuniform approaches to image acquisition and reporting. Variable magnet field strength, hardware, pulse sequences, and lesion characterization (including both morphologic and kinetic data) led to freedom of innovation but also impeded consensus development. A Web-based survey of the members of the Society of Breast Imaging conducted from September 2006 to January 2007 showed that poorly standardized breast MRI protocols were a serious problem [35]. Of 551 responding facilities, 84% indicated they never or rarely would interpret contrast-enhanced breast MRI examinations performed at other facilities because of protocol variability.

Recognizing a clear need to achieve consensus on MRI acquisition techniques and lesion terminology, the Public Health Service's Office on Women's Health funded the International Working Group for Breast MRI in 1997. This group's goal was to disseminate evidence-based consensus on the performance and interpretation of breast MRI. A subset of the international working group, the Lesion Diagnosis Working Group, composed of internationally recognized breast MRI investigators, was charged with developing a standardized breast MRI lexicon and reporting system [36]. This group later became the Subcommittee on MRI Lexicon Breast Cancer.

In 1998, the Lesion Diagnosis Working Group developed minimum reporting standards on MRI scanning techniques, region-of-interest kinetic curve acquisition, lesion architecture, and kinetic curve interpretation. These experts used the breast MRI literature to compile the most important descriptors for lesion diagnosis that would prompt specific patient management recommendations, such as biopsy. The morphologic descriptors were based on terms used in the BI-RADS mammography lexicon, when appropriate, to facilitate use and adoption in clinical practice. After the development of the preliminary breast MRI lexicon, this group performed several reader studies (funded by the National Cancer Institute, the Susan G. Komen Breast Cancer Foundation, and the ACR) to evaluate the reproducibility of these descriptors for the characterization of biopsy-proven MRI abnormalities [37]. Using the results of each study, portions of the lexicon were expanded and others eliminated in a stepwise, progressive manner.

Optimization and testing of the lexicon continued for a period of 6 years, resulting in BI-RADS–MRI, first published in the fourth edition of BI-RADS in 2003 [24]. The breast MRI lexicon is now widely used for breast MRI reporting, teaching, research, and communication. However, the work of the Subcommittee on MRI Lexicon Breast Cancer, researchers, and practitioners is far from over; the accuracy and reproducibility of the lexicon continue to be tested in the clinical and scientific arenas [38–41], and novel features and techniques are emerging rapidly. For example, the importance of background enhancement, the use of T2-weighted sequences, and bilateral scanning [42–44] prompted the committee to recommend the routine use of these techniques. Major advances in hardware, software, magnet field strength, and pulse sequence development, including parallel imaging, diffusion-weighted imaging, and MR spectroscopy, among other advances, promise to contribute to improved diagnostic capability as well as add complexity to the lexicon [45]. For example, the increased use of computer-aided diagnosis has led to major changes in breast MRI interpretation [35]. Furthermore, MRI-guided biopsy provides tissue

diagnosis for abnormalities seen only on MRI, enabling early breast cancer diagnosis as well as the collection and calculation of accurate performance metrics. In fact, the ACR's MRI Lexicon Committee is planning a new edition for 2010 that will clarify existing descriptors, eliminate descriptors that did not work in clinical practice, and add T2 weighting, background enhancement, and breast implant descriptions. Clearly, the BI-RADS–MRI is an evolving document that requires continued development to provide up-to-date, evidence-based standard terminology for its continued contribution to improving the accuracy of breast MRI. Technological progress, the multimodality evolution of breast imaging, and maintenance of the data-driven lexicon will serve as the foundation of the fifth edition of BI-RADS, which will include the second editions of BI-RADS–Ultrasound and BI-RADS–MRI.

There also have been spinoffs from BI-RADS, including the National Mammography Database (NMD) initiative [46]. The ACR originally launched the NMD in 1999 but subsequently put the initiative on hold in 2002 because of limited resources. However, in 2007, the ACR began the development of the National Radiology Data Registry, a data warehouse to collect quality improvement data across multiple modalities. The NMD is one of the registries within the National Radiology Data Registry. BI-RADS licensed software vendors are required to comply with NMD requirements, enabling the automatic upload of facility audit data directly to the ACR. During the second half of 2009, the ACR pilot-tested the NMD in early 2009 and opened NMD reporting to interested participants in the summer of 2009. The ACR will provide outcomes reports to participants to enable comparison with national benchmarks as well as practices similar in size, type, and region.

BI-RADS, MAMMOGRAPHY ACCREDITATION, AND THE MAMMOGRAPHY QUALITY STANDARDS ACT

The fact that the ACR Mammography Accreditation Program and the BI-RADS Committee were launched almost simultaneously (in 1987 and 1988, respectively) demonstrates that these initiatives were closely tied and interdependent. The Committee on Mammography Accreditation believed that once completed, compliance with BI-RADS reporting categories and the medical audit would be a critical element in improving the quality of mammography. BI-RADS was rapidly implemented by many experts in breast imaging, but both the ACR and the FDA's National Mammography Quality Assurance Advisory Committee believed that the greatest potential for improvement in interpretation lay with interpreting physicians who were not dedicated breast imagers. After the successful implementation of the ACR's Mammography Accreditation Program, the federal government followed suit by passing the Mammography Quality Standards Act of 1992 (MQSA), mandating accreditation and certification for all mammography facilities. The MQSA 1993 interim regulations [47] required that every mammography facility review outcome data from all mammography performed, including follow-up on the disposition of positive mammographic results and correlation of surgical biopsy results with mammographic reports. The 1999 MQSA final rules [48] mandated that every mammographic report include the language for a final assessment similar to those in BI-RADS. The MQSA also encouraged all interpreting physicians to review their performance compared with benchmarks established by the Agency for Health Care Policy and Research's [49] "Quality Determinants of Mammography," a document intended to provide guidance for the MQSA regulations.

The BI-RADS assessment categories have proved to be a unique resource for measuring and improving the quality of mammographic interpretation. In its 2005 report "Improving Breast Imaging Quality Standards," the Institute of Medicine [50] recognized that BI-RADS assessment provides an important tool for defining mammography positivity and negativity

to audit interpretive performance. In addition, the report stated that the audit requirements under MQSA are inadequate for measuring or improving the quality of interpretation and recommended that to achieve this improvement, an expanded BI-RADS-compatible audit should be required.

BI-RADS AND EDUCATION

The structure of BI-RADS lends itself to consistent and rational evaluation of mammographic findings and facilitates resident and breast imaging fellowship training. Both the ACR and the Society of Breast Imaging recommend that breast imaging education within residency and fellowship training should be designed to require the use of BI-RADS terminology, assessment categories, and management recommendations [51]. In fact, a 2003 survey indicated that 98% of radiology residents were taught to use BI-RADS in their mammographic reports [52]. Furthermore, the ACR's mammography self-assessment examinations, used by thousands of radiologists from 1993 to date, have exclusively used BI-RADS terminology, assessment categories, and management recommendations [53].

Because the BI-RADS lexicon descriptors are precisely defined and tied (through the literature) to breast cancer risk, proper term use leads logically to a final assessment and in turn to an appropriate management recommendation. The literature shows that training in BI-RADS can decrease variability and improve performance [54]. For example, scattered distribution and punctate morphology of calcifications are benign features and should not prompt a BI-RADS 4 or 5 final assessment or a recommendation for biopsy. However, interobserver variability remains a challenge [54,55], and the appropriate use of BI-RADS assessment categories and recommendations is not uniform [56–60].

BI-RADS AND COMMUNICATION

Although structured lexicons enable communication between humans and computers, they also facilitate communication between various physicians involved with complex clinical care. In particular, the BI-RADS final assessment categories and their accompanying management recommendations have become the standard by which physicians determine breast care on the basis of imaging. This standardization is unique among imaging reports and greatly aids in communication and facilitates the comprehension of imaging results by all members of the multidisciplinary breast care team: surgeons, pathologists, oncologists, radiation oncologists, and other health care providers. For example, the BI-RADS final assessment categories (eg, 4A, 4B, 4C, 5) are useful for communicating to pathologists the level of suspicion of lesions undergoing imaging-guided biopsy. This helps pathologists make accurate histologic diagnoses by encouraging imaging-pathology correlation. In addition, clarity of communication between pathologists and radiologists can promote the detection of possible sampling errors at percutaneous core biopsy and avoid a delay in cancer detection by prompting excisional biopsy.

BI-RADS AND RESEARCH

BI-RADS has also served to generate substantial scientific investigation that otherwise may not have been possible. A review of PubMed from 1985 to 2007 illustrates the possible effect that BI-RADS may have had on the mammography literature (Figure 1). As a baseline, papers catalogued in PubMed as addressing generic “mammography” or “mammography standards” demonstrate a steady increase in number within these years without apparent acceleration at any time point. However, publication rates for papers presenting mammographic performance data or observer studies began to outpace the more generic categories in the mid-1990s. Although it is difficult to assert a causal relationship between the establishment of BI-RADS and these types of performance and observer

studies, one can surmise that a standardized reporting system would facilitate such research. Reader studies that measure mammography performance, whether retrospective or prospective, require standardized imaging outcome assessments such as those provided by BI-RADS. Furthermore, the major North American peer-reviewed radiology journals unofficially but effectively require the use of BI-RADS terminology for the acceptance of papers related to breast imaging for publication, which in turn encourages the wide international readership to use BI-RADS. Currently, the ACR's *BI-RADS Atlas* has been translated into (or is in the process of being translated into) 8 languages: French, Spanish, Portuguese, Croatian, German, Russian, Mandarin Chinese, and Romanian.

BI-RADS provides a foundation on which to build a strong scientific tradition in breast imaging. The large, prospective ACRIN[®] trials in breast imaging (eg, the Digital Mammographic Imaging Screening Trial and the ultrasound screening trial) used interpretation guidelines that were modeled after BI-RADS (though perhaps not identical). In addition, decision support technologies using artificial intelligence algorithms such as artificial neural networks [61] and Bayesian networks [62] have been developed on the basis of standardized BI-RADS descriptors.

The National Cancer Institute–funded Breast Cancer Surveillance Consortium has used BI-RADS and its audit data elements as the basis for its multistate registry (including academic and private practices) since 1996. Participants in this registry include community-based mammography facilities as well as academic practices. The consortium sites have published 290 articles across a broad range of research topics in mammography and breast cancer.

DEVELOPING TECHNOLOGY AND FUTURE DIRECTIONS

Many directions promise to further advance the goals of BI-RADS: standardization and accuracy in breast imaging reporting. For example, technologies in medical informatics, including structured reporting (using specific data elements or formats that allow automated storage and indexing of report) and ontologies (the representation of a set of concepts and the relationships between those concepts), have the potential to further improve the application of the BI-RADS mammography, ultrasound, and MRI lexicons in clinical practice. Structured reporting is a currently available but rapidly evolving technology that provides reusable knowledge, such as templates or checklists, to the clinical reporting process to aid radiologists' consistency, accuracy, and completeness [63]. Structured reporting usually uses a point-and-click interface, prescribed lexical conventions, and a back-end database to record important variables and generate reports. For example, in breast imaging, these modules incorporate a clearly defined set of data-driven features to generate clinical reports and next management steps, mandating that a BI-RADS category be present on every report, as required by the FDA. Such systems record this information in the back-end database and allow the automated abstraction of important data for future use. However, although advantageous for standardization, mandating the use of acceptable terms can limit rich descriptive language that might be important in complex cases. Understandably, radiologists do not want to be constrained in their descriptions and thereby risk losing important subtleties of cases. So, structured reporting vendors invariably offer the ability to report free text, which increases the chance that BI-RADS descriptors will be erroneously or ambiguously applied.

Ontologies can be used to solve issues of ambiguity because they rigorously define relationships between terms within and outside of the lexicon. Such ontologies provide a mechanism for computers to understand concepts and the representation of these concepts in human language. For example “is-a” relationships would link more general terms to their more specific terms within a lexicon. “Associated” relationships would provide a conception

that there are specific terms sanctioned by BI-RADS that would be more appropriate than seemingly similar terms not sanctioned by BI-RADS.

Merging structured reporting and ontologies could provide powerful, reusable tools that could advance the accurate application of BI-RADS. For example, using relationships defined in an ontology could enable reporting systems to guide radiologists away from undesirable terms (eg, *stellate*) to more desirable BI-RADS terms (eg, *spiculated*) in real time in the clinic. Such functionality is not currently possible with lexicons alone. Furthermore, an ontology can clarify sources of confusion when a single descriptor may be used in different contexts but the specific context is ambiguous in free text. For example, in BI-RADS, the word *linear* may be used to describe 1) a microcalcification shape, 2) a microcalcification distribution, or 3) an MRI enhancement pattern. Although *linear* is not an official BI-RADS term on its own (*fine linear* and *linear distribution* are the precise terms recommended), it is widely used. When relationships are defined in an ontology and structured reporting is used on the basis of that ontology, it is possible to extract which “linear” is being referenced. As a final example, a radiologist may enter “Linear and heterogeneous microcalcifications are identified in the upper outer quadrant extending from the nipple into the axillary tail” in free text in a structured reporting module. Although this sentence is appropriately descriptive, the BI-RADS lexicon descriptors are not correct and therefore lack precise meaning to enable decision-driven management. If an ontology of BI-RADS terms were available, the structured reporting system would be able to suggest to the radiologist in real time, “By linear, do you mean fine-linear morphology or linear distribution?” as well as “By heterogeneous, are you referring to the morphology descriptor ‘coarse-heterogeneous’?” The radiologist would then know to improve the description and adhere to BI-RADS, and the appropriate adjustment would be “coarse-heterogeneous microcalcifications in a linear distribution are identified in the upper outer quadrant extending from the nipple into the axillary tail.” Structured reporting systems combined with ontologies have the potential to provide seamless and facile interfaces that “mandate” the use of standardized terms such as BI-RADS, preferably using point-and-click entry, while providing the freedom of expression that radiologists demand to generate precise and descriptive reports.

However, these solutions will not work if each vendor creates a different system to convey the concepts within BI-RADS; therefore, “harmonization” of BI-RADS with other lexicons has the potential to further codify the breast imaging lexicon in the context of other lexicons. *Harmonization* is a general term that can refer to defining the relationships of terms between lexicons, which does not imply that terms are “integrated” or the meanings of terms changed. Harmonization provides the opportunity to formalize an organizational structure that encourages communication among groups responsible for developing and curating lexicons. Harmonization, which encourages the use of uniform conventions but not necessarily the same terms, would facilitate global radiology informatics endeavors. For example, a harmonized collection of radiology lexicons provides the opportunity to develop general algorithms (eg, natural language processing tools) that could extract information from all types of radiology reports for data mining or quality assurance. The Radiological Society of North America has created RadLex (<http://www.rsna.org/radlex/>), a controlled terminology for radiology reporting, teaching, and research [64]. RadLex focuses on compiling a comprehensive set of terms for radiology and making the relationships among terms explicit in an ontology. Harmonization between BI-RADS and RadLex has the potential to benefit both systems. For example, the ontology in RadLex would provide reusable knowledge to standardize structured reporting interfaces and thereby encourage radiologists to use these systems to uniformly create BI-RADS-compliant reports. BI-RADS could help RadLex by providing a well-developed lexicon and a historical perspective regarding techniques that contributed to the development and evolution of a successful

lexicon in radiology. This harmonization process will take careful thought and planning and should be guided by experts in the field while informed by other stakeholders, as has been the BI-RADS tradition. Furthermore, RadLex harmonization will succeed only if established BI-RADS terminology and its principled, evidence-based evolution are rigorously preserved.

LESSONS LEARNED

The BI-RADS system has followed a logical and evidence-based path from its inception. The ACR committees have insisted on predictive features discovered through scientific investigation whenever possible. BI-RADS was developed from scientific data, expert guidance from leaders in breast imaging, and input from clinicians and other stakeholders. It was always meant to be a “living” document that changes as new data are acquired and more sophisticated patterns of breast care emerge. The BI-RADS lexicon can serve as an example of a highly successful standard terminology upon which lexicons in other domains can be modeled.

Acknowledgments

This study was supported by grant 1 K07 CA114181 from the National Cancer Institute of the National Institutes of Health, Bethesda, Maryland.

References

1. McLelland R. Mammography 1984: challenge to radiology. *AJR Am J Roentgenol.* 1984; 143:1–4. [PubMed: 6428201]
2. Galkin BM, Feig SA, Muir HD. The technical quality of mammography in centers participating in a regional breast cancer awareness program. *Radiographics.* 1988; 8:133–45. [PubMed: 3353530]
3. Conway BJ, McCrohan JL, Rueter FG, Suleiman OH. Mammography in the eighties. *Radiology.* 1990; 177:335–9. [PubMed: 2217765]
4. Scott, W. Establishing mammographic criteria for recommending surgical biopsy. Chicago, Ill: American Medical Association; 1989.
5. McLelland R, Hendrick RE, Zininger MD, Wilcox PA. The American College of Radiology Mammography Accreditation Program. *AJR Am J Roentgenol.* 1991; 157:473–9. [PubMed: 1872231]
6. D’Orsi CJ, Kopans DB. Mammography interpretation: the BI-RADS® method. *Am Fam Physician.* 1997; 55:1548–50. 52. [PubMed: 9105186]
7. American College of Radiology. Practice guidelines and technical standards. Reston, Va: American College of Radiology; 2008. ACR practice guideline for the performance of screening and diagnostic mammography; p. 525–34.
8. Fajardo LL, Hillman BJ, Frey C. Correlation between breast parenchymal patterns and mammographers’ certainty of diagnosis. *Invest Radiol.* 1988; 23:505–8. [PubMed: 3170137]
9. van Gils CH, Otten JD, Verbeek AL, Hendriks JH, Holland R. Effect of mammographic breast density on breast cancer screening performance: a study in Nijmegen, the Netherlands. *J Epidemiol Community Health.* 1998; 52:267–71. [PubMed: 9616416]
10. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst.* 2000; 92:1081–7. [PubMed: 10880551]
11. Mann BD, Giuliano AE, Bassett LW, Barber MS, Hallauer W, Morton DL. Delayed diagnosis of breast cancer as a result of normal mammograms. *Arch Surg.* 1983; 118:23–4. [PubMed: 6848070]
12. Hainline S, Myers L, McLelland R, Newell J, Grufferman S, Shingleton W. Mammographic patterns and risk of breast cancer. *AJR Am J Roentgenol.* 1978; 130:1157–8. [PubMed: 418657]
13. Carlile T, Kopecky KJ, Thompson DJ, et al. Breast cancer prediction and the Wolfe classification of mammograms. *JAMA.* 1985; 254:1050–3. [PubMed: 4021043]

14. Whitehead J, Carlile T, Kopecky KJ, et al. Wolfe mammographic parenchymal patterns. A study of the masking hypothesis of Egan and Mosteller. *Cancer*. 1985; 56:1280–6. [PubMed: 4027868]
15. Getty DJ, Pickett RM, D'Orsi CJ, Swets JA. Enhanced interpretation of diagnostic images. *Invest Radiol*. 1988; 23:240–52. [PubMed: 3372189]
16. Swets JA, Getty DJ, Pickett RM, D'Orsi CJ, Seltzer SE, McNeil BJ. Enhancing and evaluating diagnostic accuracy. *Med Decis Making*. 1991; 11:9–18. [PubMed: 2034078]
17. D'Orsi CJ, Getty DJ, Swets JA, Pickett RM, Seltzer SE, McNeil BJ. Reading and decision aids for improved accuracy and standardization of mammographic diagnosis. *Radiology*. 1992; 184:619–22. [PubMed: 1509042]
18. Sickles EA. Periodic mammographic follow-up of probably benign lesions: results in 3,184 consecutive cases. *Radiology*. 1991; 179:463–8. [PubMed: 2014293]
19. Sickles EA. The usefulness of computers in managing the operation of a mammography screening practice. *AJR Am J Roentgenol*. 1990; 155:755–61. [PubMed: 2119105]
20. Monticciolo DL, Sickles EA. Computerized follow-up of abnormalities detected at mammography screening. *AJR Am J Roentgenol*. 1990; 155:751–3. [PubMed: 2119104]
21. American College of Radiology. Breast Imaging Reporting and Data System® (BI-RADS®). Reston, Va: American College of Radiology; 1992.
22. American College of Radiology. Breast Imaging Reporting and Data System® (BI-RADS®). 2. Reston, Va: American College of Radiology; 1995.
23. American College of Radiology. Breast Imaging Reporting and Data System® (BI-RADS®). 3. Reston, Va: American College of Radiology; 1998.
24. American College of Radiology. Breast Imaging Reporting and Data System® (BI-RADS®). 4. Reston, Va: American College of Radiology; 2003.
25. Heilbrunn KS. The American College of Radiology's mammography lexicon: barking up the wrong tree? *AJR Am J Roentgenol*. 1994; 162:593–4. [PubMed: 8109502]
26. D'Orsi CJ, Kopans DB. The American College of Radiology's mammography lexicon: barking up the only tree. *AJR Am J Roentgenol*. 1994; 162:595. [PubMed: 8109503]
27. D'Orsi CJ, Newell MS. BI-RADS® decoded: detailed guidance on potentially confusing issues. *Radiol Clin North Am*. 2007; 45:751–63. [PubMed: 17888766]
28. 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. [PubMed: 11687695]
29. Liberman L, Abramson AF, Squires FB, Glassman JR, Morris EA, Dershaw DD. The Breast Imaging Reporting and Data System: positive predictive value of mammographic features and final assessment categories. *AJR Am J Roentgenol*. 1998; 171:35–40. [PubMed: 9648759]
30. Burnside ES, Ochsner JE, Fowler KJ, et al. Use of microcalcification descriptors in BI-RADS® 4th edition to stratify risk of malignancy. *Radiology*. 2007; 242:388–95. [PubMed: 17255409]
31. 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. [PubMed: 16569780]
32. Mendelson EB, Berg WA, Merritt CR. Toward a standardized breast ultrasound lexicon, BI-RADS: ultrasound. *Semin Roentgenol*. 2001; 36:217–25. [PubMed: 11475068]
33. Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiology*. 1995; 196:123–34. [PubMed: 7784555]
34. Stomper PC, Herman S, Klippenstein DL, et al. Suspect breast lesions: findings at dynamic gadolinium-enhanced MR imaging correlated with mammographic and pathologic features. *Radiology*. 1995; 197:387–95. [PubMed: 7480682]
35. Bassett LW, Dhaliwal SG, Eradat J, et al. National trends and practices in breast MRI. *AJR Am J Roentgenol*. 2008; 191:332–9. [PubMed: 18647898]
36. Ikeda DM. Progress report from the American College of Radiology Breast MR Imaging Lexicon Committee. *Magn Reson Imaging Clin North Am*. 2001; 9:295–302.

37. Ikeda DM, Hylton NM, Kinkel K, et al. Development, standardization, and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance imaging studies. *J Magn Reson Imaging*. 2001; 13:889–95. [PubMed: 11382949]
38. Macura KJ, Ouwerkerk R, Jacobs MA, Bluemke DA. Patterns of enhancement on breast MR images: interpretation and imaging pitfalls. *Radiographics*. 2006; 26:1719–34. [PubMed: 17102046]
39. Nie K, Chen JH, Yu HJ, Chu Y, Nalcioğlu O, Su MY. Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. *Acad Radiol*. 2008; 15:1513–25. [PubMed: 19000868]
40. Raza S, Vallejo M, Chikarmane SA, Birdwell RL. Pure ductal carcinoma in situ: a range of MRI features. *AJR Am J Roentgenol*. 2008; 191:689–99. [PubMed: 18716095]
41. Schnall MD, Blume J, Bluemke DA, et al. Diagnostic architectural and dynamic features at breast MR imaging: multicenter study. *Radiology*. 2006; 238:42–53. [PubMed: 16373758]
42. Jansen SA, Fan X, Karczmar GS, Abe H, Schmidt RA, Newstead GM. Differentiation between benign and malignant breast lesions detected by bilateral dynamic contrast-enhanced MRI: a sensitivity and specificity study. *Magn Reson Med*. 2008; 59:747–54. [PubMed: 18383287]
43. Kuhl C. The current status of breast MR imaging. Part I. Choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice. *Radiology*. 2007; 244:356–78. [PubMed: 17641361]
44. Kuhl CK. Current status of breast MR imaging. Part 2. Clinical applications. *Radiology*. 2007; 244:672–91. [PubMed: 17709824]
45. Yabuuchi H, Matsuo Y, Okafuji T, et al. Enhanced mass on contrast-enhanced breast MR imaging: lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images. *J Magn Reson Imaging*. 2008; 28:1157–65. [PubMed: 18972357]
46. Osuch JR, Anthony M, Bassett LW, et al. A proposal for a national mammography database: content, purpose, and value. *AJR Am J Roentgenol*. 1995; 164:1329–36. [PubMed: 7754870]
47. Mammography facilities: requirements for accrediting bodies and quality standards and certification requirements—interim rules. *Fed Reg*. 1993; 58:67558–72.
48. State certification of mammography facilities. *Fed Reg*. 2002; 67:5446–69.
49. Bassett, LW.; Hendrick, RE.; Bassford, TL.; Butler, PF.; Carter, D.; DeBor, M. Clinical Practice Guideline No. 13. AHCPR Publication No. 95-0632. Rockville, Md: Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services; 1994. Quality determinants of mammography.
50. Nass, SJ.; Ball, J. Improving breast imaging quality standards. Washington, DC: National Academy of Science; 2005.
51. Sickles EA, Philpotts LE, Parkinson BT, et al. American College of Radiology/Society of Breast Imaging curriculum for resident and fellow education in breast imaging. *J Am Coll Radiol*. 2006; 3:879–84. [PubMed: 17412188]
52. Bassett LW, Monsees BS, Smith RA, et al. Survey of radiology residents: breast imaging training and attitudes. *Radiology*. 2003; 227:862–9. [PubMed: 12728182]
53. Sickles EA. The American College of Radiology's Mammography Interpretive Skills Assessment (MISA) examination. *Semin Breast Dis*. 2003; 6:133–40.
54. 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; 224:871–80. [PubMed: 12202727]
55. Baker JA, Kornguth PJ, Floyd CE Jr. Breast imaging reporting and data system standardized mammography lexicon: observer variability in lesion description. *AJR Am J Roentgenol*. 1996; 166:773–8. [PubMed: 8610547]
56. Houssami N, Boyages J, Stuart K, Brennan M. Quality of breast imaging reports falls short of recommended standards. *Breast*. 2007; 16:271–9. [PubMed: 17270445]
57. Geller BM, Barlow WE, Ballard-Barbash R, et al. Use of the American College of Radiology BI-RADS® to report on the mammographic evaluation of women with signs and symptoms of breast disease. *Radiology*. 2002; 222:536–42. [PubMed: 11818625]

58. Lehman C, Holt S, Peacock S, White E, Urban N. Use of the American College of Radiology BI-RADS[®] guidelines by community radiologists: concordance of assessments and recommendations assigned to screening mammograms. *AJR Am J Roentgenol.* 2002; 179:15–20. [PubMed: 12076896]
59. 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; 222:529–35. [PubMed: 11818624]
60. D’Orsi CJ, Hall FM. BI-RADS[®] lexicon reemphasized. *AJR Am J Roentgenol.* 2006; 187:W557. [PubMed: 17056895]
61. Baker JA, Kornguth PJ, Lo JY, Williford ME, Floyd CE Jr. Breast cancer: prediction with artificial neural network based on BI-RADS[®] standardized lexicon. *Radiology.* 1995; 196:817–22. [PubMed: 7644649]
62. Burnside ES, Rubin DL, Shachter RD. Using a Bayesian network to predict the probability and type of breast cancer represented by microcalcifications on mammography. *Stud Health Technol Inform.* 2004; 107:13–7. [PubMed: 15360765]
63. Kahn CE Jr, Langlotz CP, Burnside ES, et al. Toward best practices in radiology reporting. *Radiology.* 2009; 252:852–6. [PubMed: 19717755]
64. Langlotz CP. RadLex: a new method for indexing online educational materials. *Radiographics.* 2006; 26:1595–7. [PubMed: 17102038]

Fig 1.

Publication rate calculated on the basis of the percentage of papers published in a given 2-year period as a percentage of the total in that category over the entire 22-year interval. The slopes of the lines represent the publication rates within the respective categories. Papers catalogued in PubMed demonstrate that observer and performance studies outpaced generic mammographic or mammography standards studies after the mid-1990s, when BI-RADS was increasingly adopted. The search was performed on May 14, 2009. Searches used Medical Subject Heading (MeSH) terms as follows: mammography = "Mammography[MeSH Terms]"; standards = "Mammography/standards[MAJR]"; observer studies = "Mammography[MeSH Terms] AND Observer Variation[MeSH Terms]"; and performance = "Mammography[MeSH Terms] AND Sensitivity and Specificity[MeSH Term]."