Published in final edited form as:

AJR Am J Roentgenol. 2016 February ; 206(2): 259–264. doi:10.2214/AJR.15.15396.

Breast Imaging in the Era of Big Data: Structured Reporting and Data Mining

Laurie R. Margolies¹, Gaurav Pandey², Eliot R. Horowitz³, and David S. Mendelson⁴

Laurie R. Margolies: Laurie.margolies@mountsinai.org

¹Department of Radiology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Pl, Box 1234, New York, NY 10029. Address correspondence to L.R. Margolies

²Department of Genetics and Genomic Science and Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY

³MongoDB, New York, NY

⁴Department of Radiology, Clinical Informatics, Radiology IT, Mount Sinai Health System, Icahn School of Medicine at Mount Sinai, New York, NY

Abstract

OBJECTIVE—The purpose of this article is to describe structured reporting and the development of large databases for use in data mining in breast imaging.

CONCLUSION—The results of millions of breast imaging examinations are reported with structured tools based on the BI-RADS lexicon. Much of these data are stored in accessible media. Robust computing power creates great opportunity for data scientists and breast imagers to collaborate to improve breast cancer detection and optimize screening algorithms. Data mining can create knowledge, but the questions asked and their complexity require extremely powerful and agile databases. New data technologies can facilitate outcomes research and precision medicine.

Keywords

breast cancer; breast imaging; data mining; precision medicine; structured reporting

Breast imaging has been the unparalleled innovative data leader in radiology since the introduction of BI-RADS more than 20 years ago [1]. Every day breast radiologists use components of imaging informatics, which is the radiology subspecialty that includes PACS, electronic medical records, structured reporting, computer-assisted diagnosis, natural language processing, archiving, radiation dosimetry, data mining, peer review, exchange of health information, and real-time education [2–4]. As breast imaging positions itself for the next 20 years of innovation in the era of big data, structured reporting and the ability to create and populate large databases that allow data mining are critical.

E. Horowitz is the chief technical officer and cofounder of MongoDB.

BI-RADS and Structured Reporting

BI-RADS, the first standardized vocabulary for imaging, was cutting edge in the late 1980s [5], and structured breast imaging reporting was innovative. Breast imaging was ideally suited to be the structured reporting leader; the single organ lends itself to a narrow lexicon, and reports are predominantly focused on cancer detection. The templates and descriptors are of limited number, and the high percentage of normal or unchanged examination findings allows reports to be generated in few clicks [6]. The BI-RADS mammography, breast ultrasound, and MRI lexicons encourage structured reporting, which facilitates simple and powerful analyses of data [7]. There have been many incarnations of breast structured reporting, beginning with the Breast Imaging Reporting and Database program created by the American College of Radiology (ACR) in the early 1990s [8]. The earliest structured reports were created by checking boxes on paper. Now structured reports can be created with voice recognition, point and click, or touch screen technology, and the BI-RADS mammography lexicon is in its 5th edition [9]. BI-RADS is now the backbone of structured reporting in 22 programs (Creech W, written communication, 2015). Because the systems must output data in a specific structure, the interpreting radiologists use guided paths with common terminology, concepts, and factors to report examination findings regardless of the reporting software or input method used.

By 2003, 93% of surveyed radiology residents reported always using BI-RADS, and only 1% never used it [10]. The latest edition of BI-RADS has sold approximately 2600 hard copies since February 2014 and 800 digital versions since September 2014 in addition to 20 institutional licenses. It has been translated into Spanish, and at least four other translations are in progress (Creech W, written communication, 2015). Data concerning utility of the BI-RADS lexicon show variable results. Although no improvement was found in a reinterpretation study [11], other studies have shown that sensitivity, specificity, and positive predictive value all improve to a statistically significant degree [12]. When the BI-RADS lexicon is taught, there is statistically significant improvement in agreement between new screening radiologists and experts [13], and use of the BI-RADS lexicon has been found to improve final assessment concordance [14]. Integrating clinical decision support tools with BI-RADS descriptors may improve radiologists' performance [15], and training in the use of the lexicon may be key to maximizing its benefits [16].

One important benefit of structured reporting is that it has a minimum expected content—a minimum best practice. A mammogram report cannot be completed, for example, without reporting of breast density. No longer subject to the vagaries of the individual radiologist, the referring physician or data extraction algorithm does not have to read through a narrative report with content in whatever order the radiologist deems appropriate. Because the report always has the same architecture, the recipient can always expect specific data elements and a consistent classification of the findings.

There are other benefits to structured reporting. Data such as family history, examination history, and procedure history can be electronically imported through an application programming interface. Vagueness and if-then type recommendations are limited, and errors are reduced [17–21]. Automated structured reports are time savers for many radiologists. For

others, however, cumbersome reporting systems increase reporting times, increase errors, and cause intense frustration [17, 22]. It is incumbent on radiology informatics specialists to ensure that software and associated interfaces improve rather than hinder radiologist performance.

Another benefit of structured reporting is that it facilitates automated alerts and notifications [2, 23, 24]. Dense breast tissue, for example, is an actionable finding in more than 20 states [25]. A breast density structured data field can trigger delivery of information and recommendations to the patient. Furthermore, the report may be presented in appropriately customized formats to the ordering clinician, a payer, and the patient [24, 26]. In parallel, critical alerting systems have been developed for entire radiology departments with automation of actionable result reporting. Examples include Powerscribe 360/Critical Results (Nuance) and ViSion (VisionSR).

Health care informatics has embraced standardization similar to BI-RADS. The Radiological Society of North America has established the Radlex vocabulary [27]. This validates the utility of BI-RADS; the importance of establishing consistency across practices is accepted. LI-RADS for the liver [28], Lung-RADS for lung CT screening [29], PI-RADS for the prostate [30], and other systems parallel BI-RADS and exist or are being developed.

Although structured reporting is recommended in the BI-RADS atlas, some radiologists do not use it [31]. Unstructured data sources require specialized algorithms, and it can be difficult to extract relevant information. It is difficult to compare unstructured dictations because structures and terms are highly inconsistent. Unstructured data require manual or natural language processing techniques to parse text, analyze the syntax, detect negatives, and detect acronyms and abbreviations [32]. Natural language processing solutions are evolving in which vendors such as M*Modal and VisionSR translate free text narrative into structured reports, but the important ability to enforce best practice conformance is lost. Hybrid solutions are evolving that may combine the best of both approaches.

Data to Big Data and Data Mining

Just as BI-RADS and the Mammography Quality Standards Act dramatically changed the breast imaging landscape, the intersection between high-quality breast imaging and big data is the current turning point. Many informaticists speak about the three Vs of big data: volume, velocity, and variety [33]. There is no specific number for volume, but one knows it when one sees it. The ability to integrate and quickly initiate (velocity) transactions on large volumes of disparate data types (variety) characterizes this era.

Data mining is the process of looking for patterns in data that were collected without specific hypotheses in mind. Most of the evidence that supports BI-RADS, however, is the result not of data mining but of guided analyses to validate a hypothesis. Determining whether palpable masses with benign imaging findings can be followed, for example, is a hypothesis that can be proved or disproved [34]. Data mining is most effective when researchers do not have such hypotheses; data mining is used to identify new hypotheses that can be validated in the same or other cohorts.

Data mining applied to breast imaging reports can improve current imaging reporting and help optimize imaging strategies for the future [35]. The concept of big data makes possible the integration of current, previous, and future breast examination data with a multitude of other data sources, such as electronic health records, genomic analysis, and lifestyle tracking. This integration can make future breakthroughs in risk modeling and disease detection possible [36]. To realize these breakthroughs, breast imagers and data engineers will need to collaborate. The ability to integrate the growing types and complexity of data sources is critical—this is big data for radiology. The key point is that we no longer need to predetermine which data to collect. Which data are relevant or desired are constantly evolving, and therefore data collection must be able to evolve as new questions are asked by future generations. Continuing to ask only the same family history and menarche questions limits the ability to develop true precision medicine. Choosing the right database architecture will allow us to address issues not imagined today. Because we will want to integrate ever-changing categories of risk factors, dataset nonuniformity will demand database products that are maximally flexible.

The conversion to modern database architecture allows mining of unstructured and seemingly unrelated data. Most would, for example, assume that grapefruit is a healthful food, and it likely is, but it may be a breast cancer risk factor [37]. Aluminum is commonly present in more than trace amounts in certain processed foods, medicines, vaccines, and some sources of drinking water; an extensive comprehensive database could be mined to determine whether aluminum exposure is related to breast cancer risk [38]. In essence, data mining can extract new information from sources as diverse as smart telephone tracking of exercise and diet to the electronic medical record and integration of the data with breast imaging information to identify previously unknown breast cancer risk factors and ultimately create highly refined models that predict an individual's breast cancer risk and optimize screening algorithms.

Data-mined structured reports analyzed with an inductive logic program, for example, generated millions of potential rules [35]. From 130 rules, two were selected as possible predictors of malignancy. One rule indicated that if a woman 65–70 years old had a normal mammogram, a second mammogram interpreted as probably benign, and a third read as highly suspicious (BI-RADS category 5) with a mass, the mass was always malignant. The probably benign readings were always wrong. From this knowledge, one can create a rule: a new mass in a woman 65–70 years old may not be assessed as BI-RADS 3; it must be biopsied unless unequivocally proven to be a benign mass such as a simple cyst [35]. Running this and other similar rules in real time could alert the reading radiologist to potential pitfalls and suggest that management be changed. Data mining can lead to real-time report analytics that improve patient care.

Other big data tools can assist radiologists as they interpret examinations. For example, when the BI-RADS and ACR National Mammography Database (NMD) descriptors are put into a logistic regression model and the radiologist is compared with the radiologist plus computer, the combination is superior [39]. Models such as these are capable of discerning radiologists' descriptions assessment and recommendation discordance [39]. If this approach works in real time, errors can be reduced. For example, if one describes a mass on an

ultrasound image as a complex mass and assesses it as a BI-RADS 2 benign finding, the computer would alert the radiologist that a mismatch likely exists; that is, the BI-RADS lexicon choice is likely inconsistent with the assessment. Similarly, if a big data model were correlating, for example, genetic data, BI-RADS descriptors and BI-RADS assessment real-time potential errors could be reduced. Knowing genetic information about a given patient may change the assessment of an image. For example, we know that the appearance of *BRCA*-related cancers often differs from that of non-*BRCA*-related cancers with more benign features [40–42]. Therefore, what might be called benign or probably benign in an individual of average risk might be a suspicious finding in a *BRCA*-positive patient. Adding 77 genetic variants to mammographic findings improves breast cancer diagnosis [43]. This is an example of real-time processing—enabled by new database technology—to combine the phenotypic finding and genetic knowledge to facilitate correct diagnoses by preventing medical errors and thereby improving patient care.

Current mammography databases include examination information and information about conventional risk factors, such as menarche and family history. When these databases were created, storing this much data was a herculean task [44]. Until recently, most mammography databases did not communicate or share data across organizations. Now, however, the breast imaging community can pool data in the NMD [45]. The NMD is limited—it collects data from structured fields and individual practice data to compare with national data for quality assurance and peer review. The NMD structure facilitates these goals, but its architecture does not allow flexibility for interaction with a multitude of other databases. These are clearly important goals, but a great deal more could be done. In the NMD, there is no opportunity, for example, to look for novel correlations beyond traditional risk factors. Images and pathology slides are not included. If practices could upload all the data they have, including structured and unstructured data, images of pathology slides, and breast images, and allow researchers access to deidentified HIPAA-compliant data, breast imaging could once again be the model that facilitates improvements in medical diagnosis. We should not limit ourselves to repeating what we have been doing; rather, we can and should explore how we can be innovative to determine what we are capable of doing.

The U.S. National Institutes of Health Clinical and Translational Science Award program is asking for broad collaborative efforts among researchers. Breast imaging can lead the way for precision medicine by being creative in its database architecture. The precision medicine initiative announced in January 2015 [46] asks for analysis of molecular, genomic, cellular, clinical, behavioral, physiologic, and environmental information [47]. Database architecture and data mining are crucial to the precision medicine initiative [48], which may ultimately lead to significantly improved models for predicting a woman's breast cancer risk, the best imaging strategy for optimal early detection, and lifestyle recommendations for prevention.

Imaging offers a variety of data sources in addition to report content. DICOM images themselves contain much metadata (e.g., imaging abnormalities, age, compression thickness, radiation dose). The wealth of information in images could be harnessed if the algorithms for automatically extracting information were more robust. A large warehouse of images and data and facts extracted from images can be used to apply new data-mining algorithms and to extract further knowledge [49, 50]. Many aspects of radiology are moving toward

quantitative imaging with definitive metrics delivered as part of the imaging examination. A combination of quantitative and textual (semantic) information is embedded in breast images (e.g., MRI regions of interest with kinetic data, breast density, mass margins, demographic information). Recent work has included, for example, automated density estimation, which addresses the issue of interobserver and intraobserver variability; the lexicon itself cannot eliminate this variability [51–55]. Some of this information can be exported directly from the image into the medical record [2, 26]; software developed by Volpara Solutions can be used to import a patient's quantitative breast density directly into a mammogram report. The Radiological Society of North America Annotation and Image Markup Project enables extraction and consumption of information by other applications [2, 23, 26]. It is complementary to the DICOM-structured report information.

Health care data have been collected and analyzed for many years, but the initial forays into data analytics focused on finance and operational issues. The richest data sources were the admission, discharge, and transfer and the finance and billing systems. Most of the data are stored in traditional relational databases in traditional data warehouses at many institutions. A health care enterprise data warehouse stores financial data, demographic data, and clinical data, such as pathology reports, radiology reports, and the electronic medical record. These data can be deidentified and made available to researchers. Once genomic information becomes part of the data warehouse, the amount of stored data increases dramatically. Adding individual lifestyle data from smart phones, including information on exercise, diet, alcohol intake, sleep patterns, and more, will likely become commonplace, and the complexity of the data to be analyzed will again increase greatly. As we move from treating illness to supporting wellness, this information becomes invaluable—assuming it is accessible for data mining. Personalized predictive precision medicine is our future, and the future can be now for breast imaging if we fully use modern database architecture.

The services that we routinely use on the Internet—social media, shopping, and financial applications—operate with a new generation of database technology that accommodates big data. Document storage databases, such as Hadoop, MarkLogic, and MongoDB, allow the quick integration and association of myriads of disparate data. These technical solutions are now migrating to the world of health care [56–58]. Discovery of associations between data elements can be expedited with this new generation of tools. For example, data mining to identify drug reactions can include, in addition to the medical literature, online self-reporting and electronic medical records [59]. Important associations that once would have been difficult, if not impossible, to imagine are now easier to substantiate with appropriate data-mining methods.

The Future

Finding new patterns such as unexpected correlations between environmental factors and breast cancer risk is facilitated when data are easier to find and data from multiple sources are easier to integrate. This approach works in fields like genomic medicine, in which data from genetics, genomics, and medical records have revealed patterns that facilitate identifying biomarkers for Parkinson and Alzheimer diseases [60]. Future researchers will be able to elucidate factors beyond *BRCA* and other mutations that increase breast cancer

risk. Data currently stored in silos can be exposed and integrated to show a 360° view of a patient that will ultimately benefit that individual. For example, we know that the breast cancer rate is increased among flight attendants [61, 62], but we do not know whether this is related to the flying hours, secondhand smoke exposure, or other factors. As we gain more data, we may be able to find the answer. Eventually, optimal cost-effective life-saving imaging paradigms for large cohorts of women with similar breast cancer risk profiles can be created.

Databases such as MongoDB and Mark-Logic have been used in many industries to create holistic views of customers to facilitate sophisticated analyses. Fields can be added or changed easily in these databases and the data mined to create knowledge; breast imaging can do the same. Modern document databases are designed to store a mix of structured and unstructured data on massive scales and can incorporate more than the BI-RADS and NMD mandated fields. The MetLife Wall, for example, collects structured and unstructured data from more than 70 different database systems to create a 360° view of more than 100 million clients [63, 64]. Medicine has lagged in interfacing numerous databases. The European Organization for Nuclear Research (CERN) has data from relational databases, document databases, blogs, wikis, and more that are aggregated and mined [65]. A user asks a question, and the system searches the data for results and merges them into an answer—essentially mining the data to create knowledge [66]. The WindyGrid in Chicago analyzes data in real time to react to and anticipate issues [67]. For example, data show that 7 days after a garbage complaint, there will likely be a rodent problem. The city can be proactive because of this knowledge [68]. Potentially, for the imaging community, having an immense document database will dramatically reduce the cost and time of creating and testing a hypothesis. Imagine what the breast imaging community could do if we had a system with the images, imaging reports, pathology images, pathology reports, and other data on our patients. Imagers would be in a better position to innovate because insightful questions could be asked and hypotheses generated and quickly tested against existing data. Breast imagers could have a database with volume, velocity, and variety.

If we grasp this opportunity, and do not continue to look only at the same risk factors that we have for decades, we can save patients' lives. Without change, we risk remaining beholden to those who overstate risk factor knowledge and who display unwarranted confidence in conclusions that lead to bad health care policy [69]. We remain at the mercy of those who create models, which may or may not be accurate [36]. It has been suggested, for example, that those younger than 50 years might not benefit from routine screening and that only those at high risk should undergo screening between the ages of 40 and 50 [70]. But 61% of women in their 40s who have screen-detected cancers have no family history of breast cancer [71]. More than 50% of those with breast cancer detected with MRI did not qualify for MRI according to American Cancer Society criteria [61]. These examples expose the dangers of illusion and incompleteness of knowledge; massive data can help the breast imaging community create evidence-based personalized early detection algorithms.

With currently available technology, one can see a quality loop emerging. ACR appropriateness guidelines are available in a computable format as ACR Select [2, 26] and are formally available as radiology order entry clinical decision support. The ordering

provider receives real-time feedback about whether an order fits the appropriateness guidelines and is given the opportunity to modify the order if it does not fit the criteria. The order is then executed and reported. All this information is available for data mining and outcomes research. We have been conducting trials and analyzing data for many years, and we now have robust tools for expediting this work. The iterative loop and associated outcomes research govern modifications in clinical decision support [2, 26], which is where the optimal breast imaging methods can be used and evolve.

When all of the potentially relevant data are amassed into the same database, breast imaging research can move quickly and fluidly. The future of breast imaging and associated data technologies can be incredibly robust with the deployment of flexible and agile databases. One challenge is to use database architecture that allows an unlimited number of structured data points. Others are having a single database that allows merging of structured and unstructured data sources and having every possible data point amassed and accessible in a safe, HIPAA-compliant, responsible manner. Security is its own specialty within imaging informatics [72].

In building a database, archiving data, and creating database applications, one must observe all security measures that are applied to data containing protected health information. Local and state laws and HIPAA apply to the database infrastructure. Issues include access and control policies, audit trails, multifactor authentication, data encryption, deidentification, and confidentiality. Authentication must include strong passwords and policies to govern creation and periodic change of passwords. Authentication may include biometrics [73]. If mobile devices are sources of data or are used to access data, there must be policies and security measures to accommodate loss or theft [74]. Security measures must be applied to all applications and data sources. Periodic vulnerability testing must be performed. These measures are intended to prevent damaging data breaches.

The medical community is only beginning to understand the immensity of what is possible. Breast imagers can be at the forefront of creating outcomes-based imaging guidelines by embracing new ways of collecting and interpreting data. New technologies that aggregate structured and unstructured data can facilitate outcomes research and personalized medicine. We have rich tools that can expose patterns and relationships never anticipated. The benefits of the power of data and of not leaving actionable information unused will add value to breast radiology and the totality of radiology.

Acknowledgments

G. Pandey is supported by NIH grants R01GM114434, U54OD020353, and 1U01HL107388 and an IBM faculty award.

References

1. American College of Radiology. Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology; 1992.
2. Mendelson DS, Rubin DL. Imaging informatics: essential tools for the delivery of imaging services. *Acad Radiol*. 2013; 20:1195–1212. [PubMed: 24029051]

3. Branstetter BF. Basics of imaging informatics. Part 1. Radiology. 2007; 243:656–667. [PubMed: 17431128]
4. Branstetter BF. Basics of imaging informatics. Part 2. Radiology. 2007; 244:78–84. [PubMed: 17581896]
5. Burnside ES, Sickles EA, Bassett LW, et al. The ACR BI-RADS experience: learning from history. J Am Coll Radiol. 2009; 6:851–860. [PubMed: 19945040]
6. Langlotz CP. Structured radiology reporting: are we there yet? Radiology. 2009; 253:23–25. [PubMed: 19789252]
7. Larson DB, Towbin AJ, Pryor RM, Donnelly LF. Improving consistency in radiology reporting through the use of department-wide standardized structured reporting. Radiology. 2013; 267:240–250. [PubMed: 23329657]
8. MagView website. [Accessed September 20, 2015] MagView: then and now. www.magview.com/company-history
9. D’Orsi, CJ.; Sickles, EA.; Mendelson, EB., et al. ACR BI-RADS atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology; 2013.
10. Bassett LW, Monsees BS, Smith RA, et al. Survey of radiology residents: breast imaging training and attitudes. Radiology. 2003; 227:862–869. [PubMed: 12728182]
11. 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. [PubMed: 9839520]
12. Lehman CD, Miller L, Rutter CM, Tsu V. Effect of training with the American College of Radiology Breast Imaging Reporting and Data System lexicon on mammographic interpretation skills in developing countries. Acad Radiol. 2001; 8:647–650. [PubMed: 11450966]
13. Timmers JM, van Doorne-Nagtegaal HJ, Verbeek AL, den Heeten GJ, Broeders MJ. A dedicated BI-RADS training programme: effect on the inter-observer variation among screening radiologists. Eur J Radiol. 2012; 81:2184–2188. [PubMed: 21899969]
14. 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. [PubMed: 12202727]
15. Benndorf M, Kotter E, Langer M, Herda C, Wu Y, Burnside ES. Development of an online, publicly accessible naive bayesian decision support tool for mammographic mass lesions based on the American College of Radiology (ACR) BI-RADS lexicon. Eur Radiol. 2015; 25:1768–1775. [PubMed: 25576230]
16. Antonio AL, Crespi CM. Predictors of interobserver agreement in breast imaging using the Breast Imaging Reporting and Data System. Breast Cancer Res Treat. 2010; 120:539–546. [PubMed: 20300960]
17. Johnson AJ, Chen MY, Swan JS, Applegate KE, Littenberg B. Cohort study of structured reporting compared with conventional dictation. Radiology. 2009; 253:74–80. [PubMed: 19709993]
18. Hobby JL, Tom BD, Todd C, Bearcroft PW, Dixon AK. Communication of doubt and certainty in radiological reports. Br J Radiol. 2000; 73:999–1001. [PubMed: 11064655]
19. Robinson PJ. Radiology’s Achilles’ heel: error and variation in the interpretation of the Röntgen image. Br J Radiol. 1997; 70:1085–1098. [PubMed: 9536897]
20. Hawkins CM, Hall S, Zhang B, Towbin AJ. Creation and implementation of department-wide structured reports: an analysis of the impact on error rate in radiology reports. J Digit Imaging. 2014; 27:581–587. [PubMed: 24859725]
21. Durack JC. The value proposition of structured reporting in interventional radiology. AJR. 2014; 203:734–738. [PubMed: 25247937]
22. Weiss DL, Langlotz CP. Structured reporting: patient care enhancement or productivity nightmare? Radiology. 2008; 249:739–747. [PubMed: 19011178]
23. Kahn CE, Langlotz CP, Burnside ES, et al. Toward best practices in radiology reporting. Radiology. 2009; 252:852–856. [PubMed: 19717755]
24. Reiner B. Radiology report innovation: the antidote to medical imaging commoditization. J Am Coll Radiol. 2012; 9:455–457. [PubMed: 22748782]

25. Are You Dense Advocacy website. [Accessed August 20, 2015] D.E.N.S.E. state efforts. areyoudenseadvocacy.org/dense
26. Kohli M, Dreyer KJ, Geis JR. Rethinking radiology informatics. *AJR*. 2015; 204:716–720. [PubMed: 25794061]
27. [Accessed August 20, 2015] Radiological Society of North America website. What is Radlex?. www.rsna.org/radlex.aspx
28. Mitchell DG, Bruix J, Sherman M, Sirlin CB. LI-RADS (Liver Imaging Reporting and Data System): summary, discussion, and consensus of the LI-RADS Management Working Group and future directions. *Hepatology*. 2015; 61:1056–1065. [PubMed: 25041904]
29. American College of Radiology website. [Accessed October 22, 2015] Lung CT Screening Reporting and Data System (Lung-RADS). www.acr.org/Quality-Safety/Resources/LungRADS
30. Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE. Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. *Radiology*. 2012; 265:59–69. [PubMed: 22723501]
31. Powell DK, Silberzweig JE. State of structured reporting in radiology, a survey. *Acad Radiol*. 2015; 22:226–233. [PubMed: 25442793]
32. Nassif, H.; Woods, R.; Burnside, E.; Ayvaci, M.; Shavlik, J.; Page, D. Information extraction for clinical data mining: a mammography case study; *Proc IEEE Int Conf Data Min*; 2009. p. 37-42.
33. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. *Health Aff (Millwood)*. 2014; 33:1115–1122. [PubMed: 25006136]
34. Graf O, Helbich TH, Fuchsjäger MH, et al. Follow-up of palpable circumscribed noncalcified solid breast masses at mammography and US: can biopsy be averted? *Radiology*. 2004; 233:850–856. [PubMed: 15486217]
35. Burnside ES, Davis J, Costa VS, et al. Knowledge discovery from structured mammography reports using inductive logic programming. *AMIA Annu Symp Proc*. 2005:96–100. [PubMed: 16779009]
36. Vilapriño E, Forné C, Carles M, et al. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. *PLoS One*. 2014; 9:e86858. [PubMed: 24498285]
37. Monroe KR, Stanczyk FZ, Besinque KH, Pike MC. The effect of grapefruit intake on endogenous serum estrogen levels in postmenopausal women. *Nutr Cancer*. 2013; 65:644–652. [PubMed: 23859031]
38. Willhite CC, Karyakina NA, Yokel RA, et al. Systematic review of potential health risks posed by pharmaceutical, occupational and consumer exposures to metallic and nanoscale aluminum, aluminum oxides, aluminum hydroxide and its soluble salts. *Crit Rev Toxicol*. 2014; 44(suppl 4): 1–80. [PubMed: 25233067]
39. Chhatwal J, Alagoz O, Lindstrom MJ, Kahn CE, Shaffer KA, Burnside ES. A logistic regression model based on the national mammography database format to aid breast cancer diagnosis. *AJR*. 2009; 192:1117–1127. [PubMed: 19304723]
40. Veltman J, Mann R, Kok T, et al. Breast tumor characteristics of *BRCA1* and *BRCA2* gene mutation carriers on MRI. *Eur Radiol*. 2008; 18:931–938. [PubMed: 18270717]
41. Kaas R, Kroger R, Peterse JL, Hart AA, Muller SH. The correlation of mammographic and histologic patterns of breast cancers in *BRCA1* gene mutation carriers, compared to age-matched sporadic controls. *Eur Radiol*. 2006; 16:2842–2848. [PubMed: 16924440]
42. Kaas R, Kroger R, Hendriks JH, et al. The significance of circumscribed malignant mammographic masses in the surveillance of *BRCA 1/2* gene mutation carriers. *Eur Radiol*. 2004; 14:1647–1653. [PubMed: 15083333]
43. Liu J, Page D, Peissig P, et al. New genetic variants improve personalized breast cancer diagnosis. *AMIA Jt Summits Transl Sci Proc*. 2014; 2014:83–89.
44. Osuch JR, Anthony M, Bassett LW, et al. A proposal for a national mammography database: content, purpose, and value. *AJR*. 1995; 164:1329–1334. discussion, 1335–1326. [PubMed: 7754870]
45. American College of Radiology. National mammography database. Reston, VA: American College of Radiology; 2001.

46. National Institutes of Health website. [Accessed August 1, 2015] About the Precision Medicine Initiative. www.nih.gov/precisionmedicine
47. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015; 372:793–795. [PubMed: 25635347]
48. Williams SM, Moore JH. Lumping versus splitting: the need for biological data mining in precision medicine. *BioData Min*. 2015; 8:16. [PubMed: 26085842]
49. Gundreddy RR, Tan M, Qiu Y, Cheng S, Liu H, Zheng B. Assessment of performance and reproducibility of applying a content-based image retrieval scheme for classification of breast lesions. *Med Phys*. 2015; 42:4241. [PubMed: 26133622]
50. Park SC, Sukthankar R, Mummert L, Satyanarayanan M, Zheng B. Optimization of reference library used in content-based medical image retrieval scheme. *Med Phys*. 2007; 34:4331–4339. [PubMed: 18072498]
51. Cheddad A, Czene K, Eriksson M, et al. Area and volumetric density estimation in processed full-field digital mammograms for risk assessment of breast cancer. *PLoS One*. 2014; 9:e110690. [PubMed: 25329322]
52. Nicholson BT, LoRusso AP, Smolkin M, Bovbjerg VE, Petroni GR, Harvey JA. Accuracy of assigned BI-RADS breast density category definitions. *Acad Radiol*. 2006; 13:1143–1149. [PubMed: 16935726]
53. Redondo A, Comas M, Macià F, et al. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. *Br J Radiol*. 2012; 85:1465–1470. [PubMed: 22993385]
54. Winkel RR, von Euler-Chelpin M, Nielsen M, et al. Inter-observer agreement according to three methods of evaluating mammographic density and parenchymal pattern in a case control study: impact on relative risk of breast cancer. *BMC Cancer*. 2015; 15:274. [PubMed: 25884160]
55. Bernardi D, Pellegrini M, Di Michele S, et al. Interobserver agreement in breast radiological density attribution according to BI-RADS quantitative classification. *Radiol Med (Torino)*. 2012; 117:519–528. [PubMed: 2228132]
56. Ercan, MZ.; Lane, M. An evaluation of NoSQL databases for EHR systems; Proceedings of the 25th Australasian Conference on Information Systems; 2014 Dec 8–10. aut.researchgateway.ac.nz/bitstream/handle/10292/8134/acs20140_submission_117.pdf?sequence=1&isAllowed=y
57. Schadt EE, Linderman MD, Sorenson J, Lee L, Nolan GP. Computational solutions to large-scale data management and analysis. *Nat Rev Genet*. 2010; 11:647–657. [PubMed: 20717155]
58. Schmitt, O.; Majchrzak, TA. Using document-based databases for medical information systems in unreliable environments. In: Rothkranz, L.; Ristvej, J.; Franco, Z., editors. Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management. Vancouver BC, Canada: Simon Fraser University; 2012.
59. Karimi S, Wang C, Metke-Jimenez A, Gaire R, Paris C. Text and data mining techniques in adverse drug reaction detection. *ACM Comput Surv*. 2015; 47:1–39.
60. Toga AW, Foster I, Kesselman C, et al. Big biomedical data as the key resource for discovery science. *J Am Med Inform Assoc*. 2015 Jul 21. [Epub ahead of print].
61. Hollingsworth AB, Stough RG. An alternative approach to selecting patients for high-risk screening with breast MRI. *Breast J*. 2014; 20:192–197. [PubMed: 24387050]
62. Tokumaru O, Haruki K, Bacal K, Katagiri T, Yamamoto T, Sakurai Y. Incidence of cancer among female flight attendants: a meta-analysis. *J Travel Med*. 2006; 13:127–132. [PubMed: 16706942]
63. Harris, D. [Accessed on October 3, 2015] The promise of better data has MetLife investing \$300M in new tech. Gigaom website. 2013 May 7. gigaom.com/2013/05/07/with-300m-earmarked-for-techinnovation-metlife-wants-to-remake-insurance
64. Henschen D. MetLife uses NoSQL for customer service breakthrough. *InformationWeek*. 2013 May 13. www.informationweek.com/software/information-management/metlife-uses-nosql-for-customerservice-breakthrough/d/d-id/1109919.
65. McKenna B. Cornell CERN project plumps for NoSQL DBMS. *Computer Weekly*. 2012 Oct 16. www.computerweekly.com/news/2240166469/Cornell-Cern-project-plumps-for-NoSQL-DBMS.
66. [Accessed September 29, 2015] MongoDB website. CERN CMS. 2015. www.mongodb.com/customers/cern-cms

67. Thornton S. Chicago's WindyGrid: taking situational awareness to a new level. Data Smart City Solutions. 2013 Jun 13. datasmart.ash.harvard.edu/news/article/chicagos-windygrid-taking-situational-awareness-to-a-new-level-259.
68. Murphy C. Chicago CIO pursues predictive analytics strategy. InformationWeek. 2013 Apr 19. www.informationweek.com/government/leadership/chicago-cio-pursues-predictive-analytics-strategy/d/d-id/1109623.
69. Mudd, P. The head game: high efficiency analytic decision-making and the art of solving complex problems quickly. New York, NY: Liveright; 2015.
70. U.S. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009; 151:716–726. [PubMed: 19920272]
71. Destounis SV, Arieno AL, Morgan RC, et al. Comparison of breast cancers diagnosed in screening patients in their 40s with and without family history of breast cancer in a community outpatient facility. AJR. 2014; 202:928–932. [PubMed: 24660726]
72. Mendelson DS, Erickson BJ, Choy G. Image sharing: evolving solutions in the age of interoperability. J Am Coll Radiol. 2014; 11:1260–1269. [PubMed: 25467903]
73. Schwartz J, Haarbrandt B, Fortmeier D, Haux R, Seidel C. Authentication systems for securing clinical documentation workflows: a systematic literature review. Methods Inf Med. 2014; 53:3–13. [PubMed: 24247789]
74. Hirschorn DS, Choudhri AF, Shih G, Kim W. Use of mobile devices for medical imaging. J Am Coll Radiol. 2014; 11:1277–1285. [PubMed: 25467905]