J Pathol Inform

Editor-in-Chief:
Anil V. Parwani , Pittsburgh, PA, USA
Liron Pantanowitz, Pittsburgh, PA, USA

OPEN ACCESS
HTML format

For entire Editorial Board visit : www.jpathinformatics.org/editorialboard.asp

Technical Note

Development of an electronic breast pathology database in a community health system

Heidi D. Nelson[1,2], Roshanthi Weerasinghe[1], Maritza Martel[1], Carlo Bifulco[1], Ted Assur[1], Joann G. Elmore[3], Donald L. Weaver[4]

[1]Providence Cancer Center, Providence Health and Services Oregon, Portland, Oregon; [2]Departments of Medical Informatics and Clinical Epidemiology and Medicine, Oregon Health and Science University, Portland, Oregon, [3]Department of Medicine, University of Washington School of Medicine, Seattle, Washington, [4]Department of Pathology, University of Vermont, Burlington, Vermont, USA

E-mail: *Heidi D. Nelson - nelsonh@ohsu.edu
*Corresponding author

Received: 03 October 2013 Accepted: 20 May 2014 Published: 30 July 2014

This article may be cited as:
Nelson HD, Weerasinghe R, Martel M, Bifulco C, Assur T, Elmore JG, *et al*. Development of an electronic breast pathology database in a community health system. J Pathol Inform 2014;5:26.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2014/5/1/26/137730

Abstract

Background: Health care systems rely on electronic patient data, yet access to breast tissue pathology results continues to depend on interpreting dictated free-text reports. **Objective:** The objective was to develop a method to electronically search and categorize pathologic diagnoses of patients' breast tissue specimens from dictated free-text pathology reports in a large health system for multiple users including clinicians. **Design:** A database integrating existing patient-level administrative and clinical information for breast cancer screening and diagnostic services and a web-based application for comprehensive searching of pathology reports were developed by a health system team led by pathologists. The Breast Pathology Assessment Tool and Hierarchy for Diagnosis (BPATH-Dx) provided search terms and guided electronic transcription of diagnoses from text fields on breast pathology clinical reports to standardized categories. **Approach:** Breast pathology encounters in the pathology database were matched with administrative data for 7332 women with breast tissue specimens obtained from an initial procedure in the health system from January 1, 2008 to December 31, 2011. Sequential queries of the pathology text based on BPATH-Dx categorized biopsies according to their worst pathological diagnosis, as is standard practice. Diagnoses ranged from invasive breast cancer (23.3%), carcinoma *in situ* (7.8%), atypical lesions (6.39%), proliferative lesions without atypia (27.9%), and nonproliferative lesions (34.7%), and were further classified into subcategories. A random sample of 5% of reports that were manually reviewed indicated 97.5% agreement. **Conclusions:** Sequential queries of free-text pathology reports guided by a standardized assessment tool in conjunction with a web-based search application provide an efficient and reproducible approach to accessing nonmalignant breast pathology diagnoses. This method advances the use of pathology data and electronic health records to improve health care quality, patient care, outcomes, and research.

Key words: Breast biopsy, breast pathology, electronic data systems

Access this article online

Website: www.jpathinformatics.org

DOI: 10.4103/2153-3539.137730

Quick Response Code:

INTRODUCTION

The majority of women in the United States will never develop breast cancer, however, all women are eligible for periodic mammography screening at age 40 or 50 and continuing every year or two for 25 years or more.[1-3] Over one episode of mammography screening, approximately 9-12/1000 women require breast biopsies because of suspicious radiographic lesions.[4] Breast biopsies are also required for women with physical findings, such as breast lumps or skin changes. The volume and complexity of breast imaging and biopsies have a major impact on health systems' services, delivery, and data systems.

Most breast biopsies do not result in an invasive breast cancer diagnosis, although several pathologic diagnoses are considered high-risk lesions, including various forms of atypical hyperplasia and carcinoma *in situ*. Future 10-year risks of breast cancer after biopsy have been estimated as 17-26% with atypical ductal hyperplasia (ADH), 21% with atypical lobular hyperplasia (ALH), and 24% with lobular carcinoma *in situ* (LCIS).[5] Prevention interventions and surveillance are recommended for women with these lesions, including more frequent mammography, additional imaging technologies, such as magnetic resonance imaging,[3,6,7] and risk-reducing medications.[8] These services are often underutilized because patients and their clinicians interpret benign results as normal and do not pursue personalized screening and prevention interventions.[9] How this affects clinical outcomes for individual patients and across different risk groups is not known.

In order to provide appropriate health care to patients and develop effective clinical services, health systems must be able to accurately identify, characterize, and track women with high-risk lesions. This is problematic because these diagnoses are not identified by International Classification of Disease (ICD) Terminology and are usually embedded within pathology reports of individual patients. Most existing health system data systems store dictated pathology reports as free-text documents and diagnoses are not easily extracted. As a result, women with high-risk lesions are not readily detected and opportunities to provide appropriate follow-up care and surveillance are missed.

The purpose of this project is to improve the access to breast pathology reports within a large community health system to support follow-up care and surveillance of patients. This project also tests the clinical applications of an assessment tool to map breast pathology diagnostic terms to clinically significant hierarchical categories that was developed for an ongoing study. We used this tool to develop a method that allows users to electronically search and categorize pathologic diagnoses of breast tissue specimens from dictated free-text pathology reports. This approach is unique because it was developed by a health system team led by pathologists and designed to provide direct access to patient data for multiple clinical and health system users. It also uses existing health system data that have been collected during the course of clinical care avoiding additional data coding and entry.

DEVELOPMENT

Health System Data Sources

This project was based at Providence Health and Services Oregon and used data from patients receiving care within the health system. Providence Health and Services Oregon is an integrated health system of eight community hospitals and affiliated outpatient facilities across the state that provides comprehensive care for breast cancer and related conditions, including screening, diagnosis, treatment, and survivorship care. Patients closely match the demographic and socioeconomic profiles of their communities. Breast tissue specimens are interpreted by 20 pathologists, including two subspecialists in breast pathology, in a centralized health system pathology department. As a pathology department policy, a second opinion is obtained for all new diagnoses of invasive carcinoma and ductal carcinoma *in situ* (DCIS).

The health system uses a master patient identifier for each patient accessing its extensive clinical network of hospitals and clinics. The master patient identifier and other unique patient identifiers are used to track patients across multiple encounters and over time. A breast care specific data mart developed by the health system integrates patient-level data from various internal sources, including administrative databases, electronic medical records, imaging data, and pathology data from the laboratory information system [Figure 1].[10] Data from these sources are extracted, transformed, and loaded to the health system data warehouses using a common data model, and a subset of data is extracted to create the breast health-specific data mart. Data are subsequently linked based on matching algorithms that group data for individual patients creating disease-specific data tables accessed by customized interactive queries. The database structure interfaces are continually updated with existing and new data sources as health system data sources change.

Data definitions and standards for the breast care data mart are based on the Breast Cancer Surveillance Consortium,[11] a national research collaborative sponsored by the National Cancer Institute; and the National Accreditation Program for Breast Centers.[12] The data mart primarily involves the collection and analysis of existing data that are obtained as part of routine patient care, but has the capacity to link to tumor registries, research data, and quality improvement initiatives. Data security and patient confidentiality are protected by existing health system safeguards and procedures, and

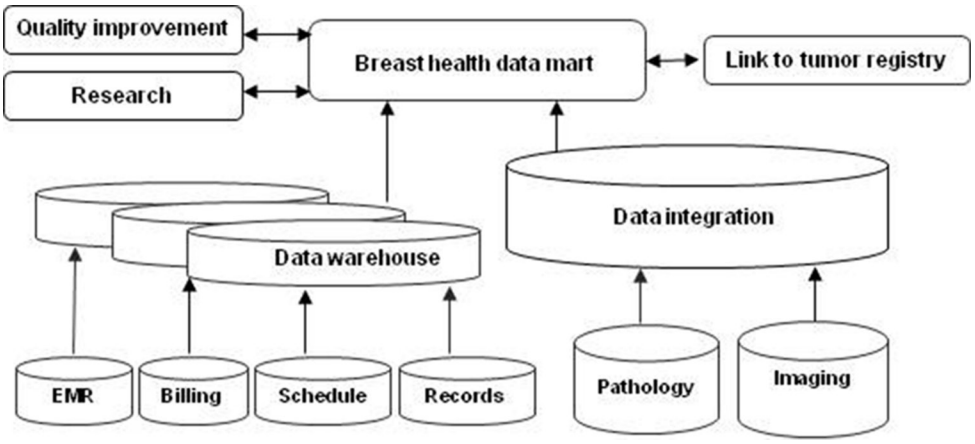

Figure 1: Health system data sources. Patient-level data from various internal sources are integrated in the breast health registry. EHR = Electronic health records

analysis of the data mart data has been approved by the health system's Institutional Review Board and Privacy Board.

Pathology Database Search Application

A health system team led by pathologists and including health system information technologists developed a web-based application for comprehensive electronic searches of anatomic pathology reports. The application receives and indexes nightly feeds of reports from the laboratory information system through a series of extraction, transformation, and loading processes applied to the data warehouse stores. This derived information is also added to an existing database of anatomical pathology reports in the health system. The application and its underlying database were designed and built on Microsoft platforms using an ASP.NET Model-View-Controller with an underlying MS-SQL 2008 database engine that support rapid and flexible solution development that can be leveraged beyond this single application.

The pathology search application uses Boolean and proximity operators for text retrieval, allowing users to interact with the search database in an intuitive and iterative manner. The application offers a simple free text search with date-range limiters, or a more advanced search for filtering on specific identifiers (including patient identifiers, providers, case numbers, etc.). Query logic includes the following search mechanisms:

- An asterisk (*) before the ending quotation mark for finding any words that start with that term. For example, "atypi*" will match with "atypia" and "atypical"
- AND: Such as "ductal hyperplasia" AND "breast*" will only match records where both terms are found somewhere in the document
- NEAR: Such as "ductal hyperplasia" NEAR "breast*" will only match records where both terms are found somewhere near each other in the document
- OR: Such as "right breast" OR "left breast" will match records where either term is found
- AND NOT: Such as "breast" AND NOT "melanoma" will match all records with "breast," but then exclude ones with "melanoma" from the final result.

A list of search results is returned in a simple and familiar interface with automated text highlighting of the requested search terms [Figure 2]. Selecting one of the results returns additional details of the case, including the entire anatomic pathology text [Figure 3]. The application uses internal network login credentials for authentication, and authorizes users into either a non-protected or protected mode where patient health information is made visible, or not, based on group privileges. This flexibility broadens the application's users to consulting clinicians and researchers. Results can be exported to comma-separated values format for additional reporting and manipulation in analytical software.

Breast Pathology Assessment Tool and Hierarchy for Diagnosis

The Breast Pathology Study (B-Path) is an ongoing project sponsored by the National Cancer Institute to evaluate the accuracy of pathologists' interpretations of breast tissue specimens.[13,14] The B-Path Study investigators, including three participating in this project, developed a standardized assessment tool and method to map breast pathology diagnostic terms to clinically significant hierarchical categories for the study (Breast Pathology Assessment Tool and Hierarchy for Diagnosis [BPATH- Dx]) [Figure 4]. The major diagnostic categories include invasive breast cancer, carcinoma *in situ* (DCIS and LCIS), atypical, proliferative lesion without atypia, and nonproliferative changes.

To evaluate the clinical applications of the BPATH-Dx form, investigators used it to transcribe diagnoses from text fields on breast pathology clinical reports at Providence to standardized BPATH-Dx categories. The

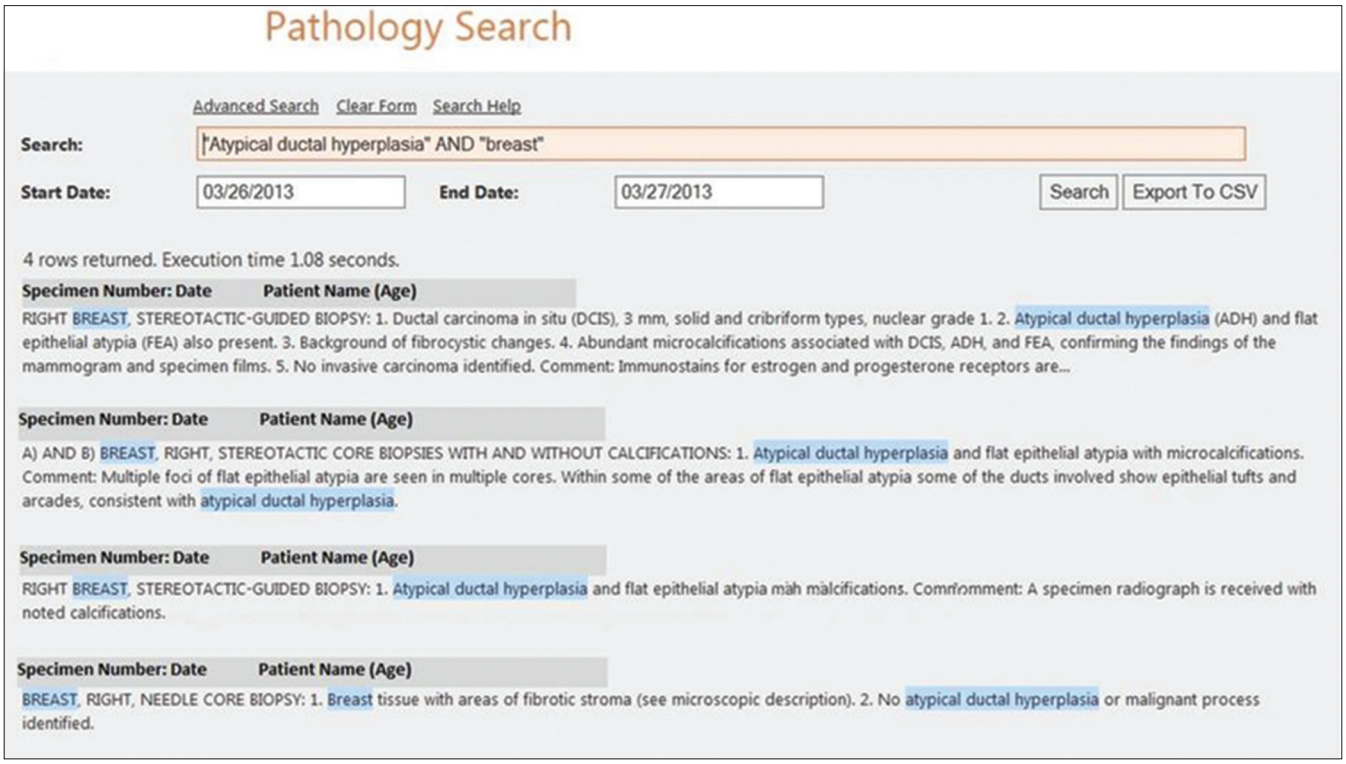

Figure 2: Screen shot of the search application

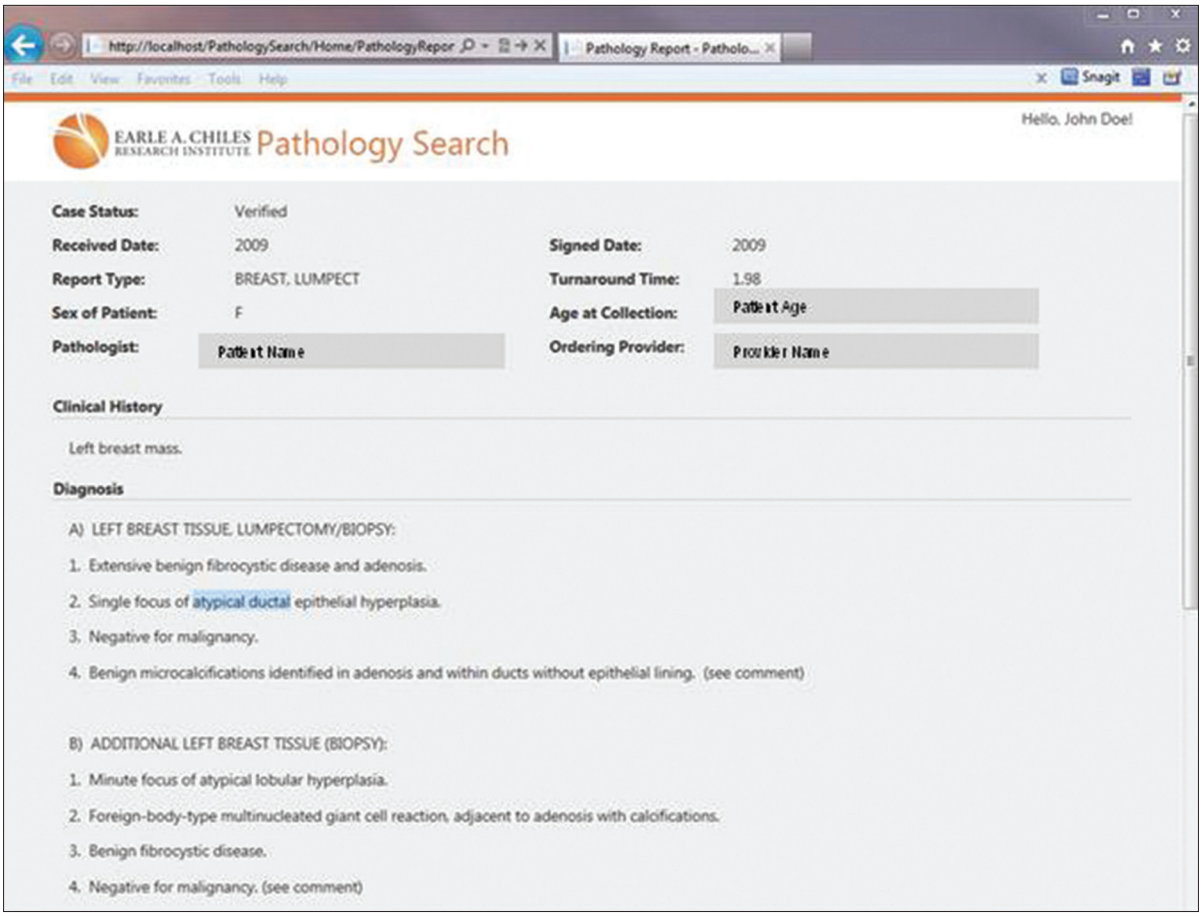

Figure 3: Screen shot of an individual patient's diagnosis from the pathology report

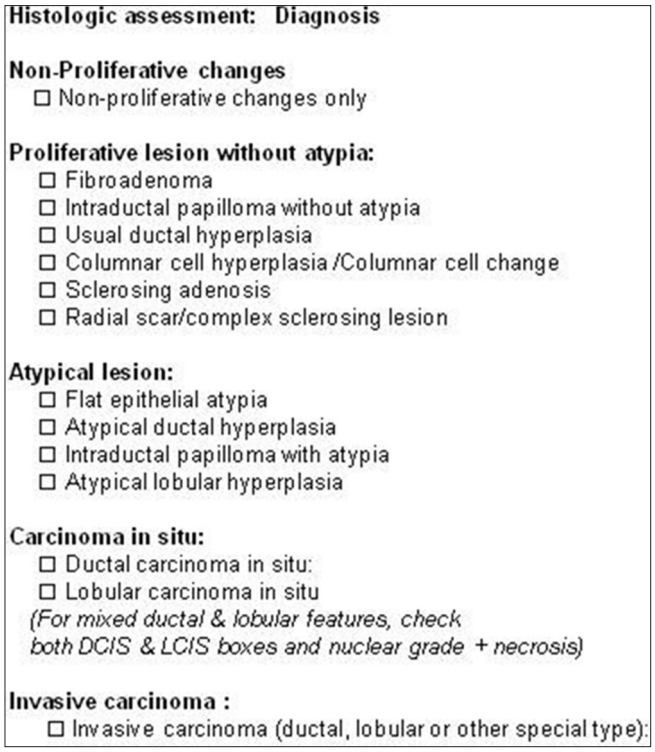

Figure 4: Breast Pathology Assessment Tool and Hierarchy for Diagnosis data collection form for the Breast Pathology Study (B-Path). In the B-Path Study, the data elements were presented in an electronic form as a web page with a series of pop-up windows. Flat epithelial atypia was grouped within the atypical lesion category on the form because the word atypia is in its name, but was coded and analyzed as a proliferative lesion without atypia because its associated risk for future carcinoma is low

health system breast pathologist determined a clinical hierarchy within the major diagnostic categories by ranking the diagnoses from the most to least serious in order to identify the worst pathologic diagnosis per patient, consistent with clinical practice [Table 1].

APPROACH

Selecting and Categorizing Cases

Breast tissue specimens in the health system were identified in the pathology database and matched with administrative data based on ICD-9-CM procedure codes (19102, 19103, 19120, and 19125) from January 1, 2008 to December 31, 2011. These data were entered into a MS-Access database. For purposes of this project, results were determined for individual women using specimens from an initial procedure occurring during a 1-year time interval. Approximately, 85% of initial procedures at Providence are core needle biopsies.

A research associate conducted manual sequential queries of the breast pathology reports using the pathology search application and entering the diagnostic categories from the BPATH-Dx form beginning with the worst pathologic category (i.e. invasive cancer) and progressing to benign

Table 1: Diagnostic coding hierarchy

Order	Diagnosis
1	Invasive carcinoma
2	DCIS
3	LCIS
4	ADH
5	ALH
6	Papilloma with atypia
7	FEA
8	Fibroadenoma
9	Papilloma without atypia
10	CCH
11	Sclerosing adenosis
12	Radial scar
13	UDH
14	Nonproliferative changes (including normal)
15	Other

DCIS: Ductal carcinoma *in situ,* LCIS: Lobular carcinoma *in situ*, ADH: Atypical ductal hyperplasia, ALH: Atypical lobular hyperplasia, FEA: Flat epithelial atypia, UDH: Usual ductal hyperplasia, CCH: Columnar cell hyperplasia

categories. If contradicting statements were found, such as "ADH present" and "no evidence of ADH," the report was manually reviewed. The final classification of a case depended on having a statement indicating its presence, no statements indicating its absence, and no diagnoses of higher severity. A health system breast pathologist was consulted to review cases for which a classification could not be determined by this approach. Results of this process were compared with a random sample of 5% of the pathology reports that were manually reviewed by physician investigators including a breast pathologist.

Implementation

Pathologic diagnoses of breast tissue specimens from 7332 women were identified from the pathology database and categorized using the search application and hierarchical classification approach. Search results indicated that pathologists used many different terms when reporting either the absence or presence of a diagnosis, and some used both types of statements in a single report. For example, for the diagnosis of ADH, pathologists used 20 different ways to describe its absence and 10 ways to describe its presence [Table 2]. The most common expressions indicating the absence of ADH were statements preceded by "no evidence of" and followed by "ADH," "atypia," "atypical feature or malignancy," or "atypical epithelial hyperplasia." The most common expressions used to indicate the presence of ADH were "ADH," "atypical duct hyperplasia,"and "atypical ductal epithelial hyperplasia." In our data, <1% of reports had contradictory statements that required manual review.

Results indicated 1709 (23.3%) women with invasive breast cancer, 491 (6.7%) with DCIS, 82 (1.1%) with LCIS , 459 (6.3%) with atypical lesions, 2044 (27.9%) with proliferative lesions without atypia, and 2547 (34.7%)

with nonproliferative lesions [Table 3]. These include four cases of apocrine atypia that were categorized separately within the atypical lesions group. A manual review of a random sample of 5% of reports (N = 359) indicated 97.5% diagnostic agreement, with discrepancies predominantly among reports that described borderline diagnoses or used unclear terminology.

Table 2: Pathologists' descriptions of ADH (*n*=852)

Search terms	*n* (%)
Absence of condition	
No evidence of ADH	174
No evidence of atypia	72
No evidence of atypical feature or malignancy	66
No evidence of atypical epithelial hyperplasia	56
Negative for atypia	26
No ADH	21
Negative for atypical	14
No evidence of dysplasia	10
No evidence of cytologic atypia	9
No evidence of epithelial atypia	6
No morphologic or immunohistochemical evidence of ADH	5
No evidence of atypical ductal epithelial	2
Other phrases "no, not identified, or negative"	17
Total cases	478 (56.1)
Presence of condition*	
ADH	292
Atypical duct hyperplasia	44
Atypical ductal epithelial hyperplasia	15
Atypical duct epithelial hyperplasia	9
Atypical intraductal proliferation	2
Epithelial hyperplasia with atypia	1
Atypia of ductal epithelium	1
Intraductal epithelial atypia	1
Atypical epithelial proliferation	1
Atypical ductal proliferation	1
Total cases	367 (43.1)
Statements requiring manual review	
Atypia mentioned in comments	6
Atypia mentioned in the clinical history	1
Total cases	7 (0.8)

*ADH was mentioned in the report, but for 31 cases it was not the most severe diagnosis. ADH: Atypical ductal hyperplasia

DISCUSSION

Our approach to accessing patients' breast pathology diagnoses with sequential queries of free-text pathology reports using a web-based search application and hierarchical diagnostic categories provides an efficient solution to capturing important clinical data. The distribution of diagnoses using our approach for 7332 breast biopsies at Providence is similar to a large national study of 26,748 breast biopsies.[15] Our approach uses existing patient information and a familiar interface and search strategy that does not require special coding or programming for each query. This allows multiple types of users' direct electronic access to pathology reports, accomplishing the major goal of our project.

While other systems for reporting pathology data have been developed, they could not be used to access the existing pathology reports in the health system. The Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT), a comprehensive clinical terminology system, requires entry of coded clinical information before data can be accessed.[16] The College of American Pathologists provides breast cancer reporting checklists and guidelines; however, these have not been implemented in clinical practice and are not searchable.[17] Furthermore, while other types of natural language processing (NLP) software are available commercially, their performance in tasks similar to ours and their other advantages are not clear.

Other studies using NLP to extract clinical information from free-text pathology reports indicate high accuracy, but also high complexity,[18,19] which can prohibit implementation and routine application. In a study of more than 76,000 breast pathology reports, NLP sensitivity and specificity were 99.1% and 96.5% when compared to expert human coders. However, many diagnostic terms were used, such as 124 different ways to describe invasive ductal carcinoma.[18] In another study

Table 3: Pathology results for 7332 women with breast tissue specimens from 2008 to 2011, *n* (%)

Age (years)	Invasive carcinoma	Carcinoma *in situ*		Atypical lesion			
		DCIS	LCIS	ADH	ALH	IP with atypia	Apocrine atypia
<40	55 (3.2)	11 (2.2)	2 (2.4)	15 (4.5)	7 (8.0)	2 (6.7)	0
40-49	229 (13.4)	94 (19.1)	19 (23.2)	102 (30.5)	23 (26.1)	1 (3.5)	1 (25.0)
50-59	418 (24.5)	147 (29.9)	40 (48.8)	106 (31.3)	28 (30.7)	7 (24.1)	1 (25.0)
60-69	476 (27.9)	128 (26.1)	11 (13.4)	75 (22.4)	18 (20.5)	10 (34.5)	2 (50.0)
70-79	297 (17.4)	69 (14.1)	8 (9.8)	24 (7.1)	10 (11.4)	6 (20.7)	0
≥80	234 (13.7)	42 (8.6)	2 (2.4)	14 (4.2)	3 (3.4)	4 (13.8)	0
Total	1709	491	82	336	89	30	4

DCIS: Ductal carcinoma *in situ*, LCIS: Lobular carcinoma *in situ*, ADH: Atypical ductal hyperplasia, ALH: Atypical lobular hyperplasia, IP: Intraductal papilloma

Table 3: Continued

Age (years)	Proliferative lesion without atypia							Nonproliferative
	FEA	Fibro-adenoma	IP no atypia	CCH/ CCC	Sclerosing adenosis	Radial scar	UDH	
<40	2 (2.8)	450 (35.1)	48 (17.8)	7 (11.7)	9 (11.2)	7 (12.7)	23 (10.2)	377 (14.8)
40-49	25 (34.7)	386 (30.1)	70 (25.9)	24 (40.0)	26 (32.5)	23 (41.8)	74 (32.9)	726 (28.5)
50-59	30 (41.7)	201 (15.7)	64 (23.7)	14 (23.3)	27 (33.8)	15 (27.3)	64 (28.4)	698 (27.4)
60-69	8 (11.1)	159 (12.4)	56 (20.7)	11 (18.3)	12 (15.0)	4 (7.3)	39 (17.3)	466 (18.3)
70-79	5 (6.9)	67 (5.2)	25 (9.3)	3 (5.0)	6 (7.5)	3 (5.5)	18 (8.0)	212 (8.3)
≥80	2 (2.8)	19 (1.5)	7 (2.6)	1 (1.7)	0	3 (5.5)	7 (3.1)	68 (2.7)
Total	72	1282	270	60	80	55	225	2547

CCH/CCC: Columnar cell hyperplasia/columnar cell change, FEA: Flat epithelial atypia, IP: Intraductal papilloma, UDH: Usual ductal hyperplasia

of NLP compared to a human coded gold standard, sensitivity was 90.6% and specificity 91.6%.[19] This study used a MedLEE NLP application that had to be modified with a preprocessor in order to address specific features of pathology reports that are difficult for NLP. While these studies indicate high levels of accuracy with NLP, the wide range of diagnostic terms and hierarchical nature of breast pathology diagnosis complicate its application for clinical uses.

Our approach also has limitations. For some reports, the search terms identified a diagnosis that was entered in the comments or history fields. While we were able to find these errors and manually resolve them when the report was explicitly contradictory [Table 2], some reports may have been less obvious. Furthermore, this project was not designed to evaluate the accuracy of the BPATH-Dx form with a gold standard;[20] although, we compared results with a manual review of a random sample of 5% of reports.

The inability of the current search engine to properly handle explicit negation in the diagnostic text (e.g. "no evidence of DCIS") required additional sorting steps that could have led to false-positive matches. Furthermore, the search terms sometimes led to multiple diagnoses when borderline diagnoses were described in the pathology report (e.g. "ADH bordering on low grade DCIS"). In addition, not all possible breast pathology diagnoses are currently included in the BPATH-Dx form and need to be added to address additional clinical conditions (e.g. phyllodes tumor, secondary neoplasms). The variable terminology for some lesions and difficulty categorizing them further complicate this task.

Our exploratory work highlights areas for improvement, including refinement of the diagnostic categories and hierarchy to include additional and borderline diagnoses. This effort could build on existing work, such as guidelines developed by the United Kingdom National Coordinating Committee for Breast Screening Pathology that uses numeric hierarchical diagnostic categories for core biopsies.[21] In addition, future steps to create procedures and algorithms for different kinds of users would help assure a uniform approach. Efforts to standardize diagnostic terms among health system pathologists could minimize the variability of dictated text phrases. These approaches to standardization for the interpretation of mammography examinations by radiologists resulted in the Breast Imaging Reporting and Data System (BI-RADS) classifications that are essential to current practice.[22] A similar effort for breast pathology would also be valuable.

Sequential human queries of free-text clinical breast pathology reports guided by standardized hierarchical diagnostic categories and using a web-based search application provides an efficient approach to accessing clinical diagnoses of patients. Our evaluation of a random sample of cases indicates that this approach is also reproducible. This informatics approach advances the use of electronic pathology data to improve health care quality, patient care, outcomes, and research.

HUMAN SUBJECTS PROTECTION

This project was performed in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects and was reviewed by Providence Health and Services IRB approval #11-045A renewal date: 04/08/2014.

ACKNOWLEGMENTS

This work was funded by the Providence Cancer Center and the Safeway Foundation. We thank Christopher Dubay, PhD for his suggestions for the manuscript.

REFERENCES

1. US Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med

2009;151:716-26, W-236.
2. National Cancer Institute. Screening mammograms: Questions and answers. Available from: http://www.cancer.gov/cancertopics/factsheet/Detection/screening-mammograms. [Last accessed on 2014 May 19].
3. Smith RA, Saslow D, Sawyer KA, Burke W, Costanza ME, Evans WP 3rd, *et al.* American Cancer Society guidelines for breast cancer screening: Update 2003. CA Cancer J Clin 2003;53:141-69.
4. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: An update for the U.S. Preventive Services Task Force. Ann Intern Med 2009;151:727-37, W237-42.
5. Coopey SB, Mazzola E, Buckley JM, Sharko J, Belli AK, Kim EM, *et al.* The role of chemoprevention in modifying the risk of breast cancer in women with atypical breast lesions. Breast Cancer Res Treat 2012;136:627-33.
6. D'Orsi CJ, Bassett LW, Berg WA. Follow-up and outcome monitoring. In: Breast Imaging Reporting and Data System: ACR BIRADS. 4th ed. Reston, VA: American College of Radiology; 2003. p. 229-51.
7. Rosenberg RD, Yankaskas BC, Abraham LA, Sickles EA, Lehman CD, Geller BM, *et al.* Performance benchmarks for screening mammography. Radiology 2006;241:55-66.
8. Moyer V, On behalf of the U.S. Preventive Services Task Force. Medications for risk reduction of primary breast cancer in women: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2013;159:698-708.
9. Waters EA, McNeel TS, Stevens WM, Freedman AN. Use of tamoxifen and raloxifene for breast cancer chemoprevention in 2010. Breast Cancer Res Treat 2012;134:875-80.
10. Nelson HD, Weerasinghe R. Actualizing personalized healthcare for women through connected data systems: Breast cancer screening and diagnosis. Glob Adv Health Med 2013;2:30-6.
11. Breast Cancer Surveillance Consortium. Available from: http://www.breastscreening.cancer.gov/. [Last accessed on 2013 Jun 26].
12. National Accreditation Program for Breast Centers. Available from: http://www.napbc-breast.org/. [Last accessed on 2013 Jun 26].
13. Oster NV, Carney PA, Allison KH, Weaver DL, Reisch LM, Longton G, *et al.* Development of a diagnostic test set to assess agreement in breast pathology: Practical application of the guidelines for reporting reliability and agreement studies (GRRAS). BMC Womens Health 2013;13:3.
14. Allison KH, Reisch LM, Carney PA, Weaver DL, Schnitt SJ, O'Malley FP, *et al.* Understanding diagnostic variability in breast pathology: Lessons learned from an expert consensus review panel. Histopathology 2014;[E pub 2014 Apr 2].
15. Weaver DL, Rosenberg RD, Barlow WE, Ichikawa L, Carney PA, Kerlikowske K, *et al.* Pathologic findings from the breast cancer surveillance consortium: Population-based outcomes in women undergoing biopsy after screening mammography. Cancer 2006;106:732-42.
16. SNOMED Clinical Terms (SNOMED CT). Available from: http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html. [Last accessed on 2014 May 19].
17. College of American Pathologists Reference Resources and Publications. Available from: http://www.cap.org/apps/cap.portal?_nfpb=true and _pageLabel=reference. [Last accessed on 2014 May 19].
18. Buckley JM, Coopey SB, Sharko J, Polubriaginof F, Drohan B, Belli AK, *et al.* The feasibility of using natural language processing to extract clinical information from breast pathology reports. J Pathol Inform 2012;3:23.
19. Xu H, Anderson K, Grann VR, Friedman C. Facilitating cancer research using natural language processing of pathology reports. Stud Health Technol Inform 2004;107:565-72.
20. Cheng LT, Zheng J, Savova GK, Erickson BJ. Discerning tumor status from unstructured MRI reports – Completeness of information in existing reports and utility of automated natural language processing. J Digit Imaging 2010;23:119-32.
21. Ellis IO, Humphreys S, Michell M, Pinder SE, Wells CA, Zakhour HD, *et al.* Best Practice No 179. Guidelines for breast needle core biopsy handling and reporting in breast screening assessment. J Clin Pathol 2004;57:897-902.
22. American College of Radiology. The American College of Radiology Breast Imaging Reporting and Data System (BI-RADS). 4th ed. Reston, VA: American College of Radiology; 2003.