

Published in final edited form as:

J Patient Saf. 2017 September; 13(3): 138–143. doi:10.1097/PTS.0000000000000127.

Using Natural Language Processing to Extract Abnormal Results from Cancer Screening Reports

Carlton Moore, MD, MS,

4023 Old Clinic Building, Campus Box 7110, Division of General Medicine and Clinical Epidemiology, Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC 27599-7110, Phone: 919-843-4978, Fax: 919-966-2274

Ashraf Farrag, MS, and

160 N. Medical Drive, Brinkhous-Bullitt, 2nd floor, The North Carolina Translational and Clinical Sciences Institute, University of North Carolina, Chapel Hill, NC 27599, Phone: 919-843-8992

Evan Ashkin, MD

590 Manning Drive, Department of Family Medicine, University of North Carolina School of Medicine, Chapel Hill, NC 27599, Phone: 919-966-4996

Abstract

OBJECTIVES—Numerous studies show that follow-up of abnormal cancer screening results, such as mammography and Papanicolaou (Pap) smears, are frequently not performed in a timely manner. A contributing factor is that abnormal results may go unrecognized because they are buried in free-text documents in electronic medical records (EMRs) and, as a result, patients are lost to follow-up. By identifying abnormal results from free-text reports in EMRs and generating alerts to clinicians, natural language processing (NLP) technology has the potential for improving patient care. The goal of the current study is to evaluate the performance of NLP software for extracting abnormal results from free-text mammography and Pap smear reports stored in an EMR.

METHODS—A sample of 421 and 500 free-text mammography and Pap reports; respectively, were manually reviewed by a physician and the results categorized for each report. We tested the performance of NLP to extract results from the reports. The two assessments ('gold' standard vs. NLP) were compared to determine the precision, recall and accuracy of NLP.

RESULTS—When NLP was compared to manual review for mammography reports, the results were as follows: precision = 98% (96–99%), recall = 100% (98–100%) and accuracy = 98% (96–99%). For Pap smear reports, the precision, recall and accuracy of NLP were all 100%.

CONCLUSION—Our study developed NLP models that accurately extract abnormal results from mammography and Pap smear reports. Future plans include using NLP technology to generate real-time alerts and reminders for providers to facilitate timely follow-up of abnormal results.

INTRODUCTION

Inadequate follow-up of abnormal test results represents a significant patient safety and malpractice concern. ^{1–3} In fact, the fastest growing area of malpractice litigation involves failures or delays in diagnosis. Of these, 25% are attributable to avoidable failures in the test follow-up system. ⁴ Accordingly, follow-up of outpatient test results has become a major priority for organizations and policy makers concerned with healthcare quality and patient safety; so much so, that the National Committee for Quality Assurance has included implementation of reliable systems to effectively track test results as a criteria for primary care practices to attain Patient-Centered Medical Home designation. ⁵ These safety concerns are echoed by physicians who perceive that follow-up of abnormal test results, as well as the systems used for follow-up, are sub-optimal. ^{6–9}

Test result follow-up is especially problematic for cancer screening, and inadequate test result follow-up has been implicated in missed and delayed cancer diagnoses. ^{3,10} For example, 28% of women don't receive timely follow-up of abnormal cervical cancer screening results and minority women are at especially high risk. ^{11,12} Major factors contributing to the lack of timely follow-up are sub-optimally designed work processes and the dearth of clinical decision support systems in the healthcare system. ^{13–15} In this environment, abnormal test results often fall through the cracks ⁸ and in the case of cancer screening test results, may lead to delayed diagnoses and increased morbidity and mortality.

Clinical Decision Support Systems (CDSSs) that generate alerts and reminders to clinicians improve guideline adherence to recommended care. ¹⁶ CDSSs use structured data elements in electronic medical records (EMRs) and conditional (if ..., then ...) logic to trigger alerts when pre-defined conditions are satisfied. Currently, unstructured clinical data, such as that found in free-text cancer screening reports (Pap smear and mammography), cannot be used in CDSSs.

Natural language processing (NLP) has the ability to extract results from free-text clinical reports stored in EMRs and convert the results into a structured format suitable for use by CDSSs. ^{17,18} The overall goal of our study is to develop and validate NLP that will accurately identify results in free-text mammography and Pap smear reports and convert them into a structured format suitable for CDSS.

METHODS

Study Setting and Population

The UNC Health Care System includes NC Memorial Hospital, a Neuroscience Hospital, the UNC Women's Hospital, the UNC Children's Hospital and an Ambulatory Care Center in which 154,224 unique patients were seen at UNC Ambulatory Centers for 973,122 visits. Overall, 10% of patients are Latino, 58% female, and 97,851 were between 19 and 59 years of age, while 34,862 were aged 60 or older.

Natural Language Processing (NLP) of Free-Text Clinical Documents

The NLP software used for the study employs dictionary-based named entity recognition ¹⁸ to identify the standard terminologies used to document mammography and Pap smear results in the free-text reports. All NLP models were developed using the IBM Content Analytics software package residing on a computer server at the Information Services Division of the University of North Carolina Hospital. The study was conducted at the University of North Carolina Healthcare System and was approved by the institutional review board (IRB number 11-1197). The project was funded by a National Institutes of Health grant 1UL1TR001111 through the North Carolina Translational and Clinical Sciences Institute.

Our goal is to attain NLP performance characteristics (precision, recall and accuracy) greater than 95%, with 95% confidence intervals of $\pm 1.5\%$ (range of 3%). For these performance specifications, we calculated that a minimum sample size of 400 reports, each for mammograms and Pap smears, was required.

Mammography Reports

Mammography results are reported as Breast Imaging Reporting and Data System (BIRADS) assessment categories ranging from 0 to 6, each with a specific recommendation for follow-up (table 1). ¹⁹ Figure 1 shows an example of a mammography report with the BIRADS result and follow-up recommendations highlighted in yellow. In the semi-structured mammography reports, results are always located after the word "IMPRESSION:" and the follow-up recommendations appear after a detailed description of the mammography findings. NLP algorithms were developed to detect the term "IMRESSION:" in the free-text mammography documents and to extract the BIRADS results that appear afterwards. In the corpus of mammography reports, there was some variability in how results were documented; for example, BIRADS 1 might appear as "BIRADS 1", "BIRADS 1", "BIRADS-1" or "BIRADS #1". Therefore algorithms were developed to correctly map all possible iterations of BIRADS 1 using conditional logic (if ..., then ... rules) in the NLP soft software.

We selected a random sample of 421 free-text mammography reports completed between January 2003 and January 2012 from the Carolina Data Warehouse for Health (CDW-H). The reports were manually reviewed by a general internist (CM) and the BIRADS result determined for each. We developed and tested the performance of an NLP model to extract the BIRADS results from the same set of reports. The two assessments ('gold' standard vs. NLP) were compared to determine the precision, recall and accuracy of NLP for extracting mammography results. Precision, recall and accuracy were calculated as follows:

 $\label{eq:precision} Precision = TP/(TP+FP), Recall = TP/(TP+FN), and Accuracy = (TP+TN)/Total; where TP=true-positive, FP=false-positive, TN=true-negative, and FN=false-negative.$

Papanicolaou (Pap) Smear Reports

Figure 2 shows a portion of a Pap smear report with results highlighted in yellow. Pap smear results are always located in free-text after the expression "RESULTS/INTERPRETATION"

and are reported using the Bethesda Classification²⁰ of cervical cytology (table 2). For example, results of the Pap report in figure 2 would map to "ASC-H" in table 2. NLP algorithms were developed that map the results identified in the free-text documents with the corresponding Bethesda Classification shown in table 2.

We used a random sample of 500 free-text Pap smear reports completed between January 2003 and January 2012 and stored in the CDW-H. Pap smear reports were manually reviewed by a general internist (CM) and the results (Bethesda Classification²¹) were determined for each report; this was used as the "gold" standard to compare with NLP results. The performance of NLP was assessed using Precision, Recall and Accuracy as described previously.

Finally, we determined inter-rater reliability for manual review of mammogram and Pap smear results by having a second reviewer (EA) independently abstract results from a 20% random sample of mammogram and Pap smear reports. We compared the results to those obtained by the 1st reviewer (CM). Inter-rater agreement was calculated using the kappa coefficient.

Follow-up of Abnormal Pap Spear and Mammogram Results

For each abnormal Pap smear and mammogram (BIRADS 4, 5, and 6) result identified using NLP, we reviewed the electronic medical record (EMR) to assess the amount of time to repeat testing or follow-up procedure (example; colposcopy or core biopsy). If none was found, we reviewed clinical notes to determine if the abnormal result was acknowledged and that medical management other than repeat testing was chosen or that follow-up was to be completed at an outside institution. All calculations were performed using Stata for Windows (Stata Corporation, College Station, Tx).

RESULTS

Mammography Reports

Manual review ('gold' standard) of the 421 mammography reports determined that 3.6%, 13.5%, 57.9%, 12.1%, 8.6%, 1.2%, and 3.1% of results were BIRADS 0 thru 6; respectively (table 1). Inter-rater agreement (kappa coefficient) between the two reviewers was 1.0. When we compared NLP against the 'gold' standard manual review, the results were as follows:

Precision = 98% (96–99%), Recall = 100% (98–100%) and Accuracy = 98% (96–99%). The only inaccuracies in the NLP model occurred with mammography reports that had initial BIRADS results of zero (0) with later addendums documenting result changes (example; changing BIRADS 0 to BIRADS 2) based on radiologists' reviews of previous mammography results. In these situations, the NLP model identified the BIRADS 0 as the final result and did not detect results in the addenda. This occurred in a total of 10 (2.4%) of the 421 reports and we have since modified the NLP model so that it now identifies results in report addenda.

Follow-up of Abnormal Mammograms

For abnormal mammogram results (BIRADS 4, 5 and 6) identified using NLP, we reviewed the medical record to determine time between incident abnormal results and any repeat testing or procedures (table 3). For BIRADS 4 mammogram results, 34 of 36 patients had follow-up (usually biopsy) within 2 months. Of the two patients not followed up within 2 months, one had fine needle aspiration of a left axillary mass (identified on the mammogram) approximately 8 months (251 days) after the incident BIRADS 4 result. Results of the fine needle aspiration showed metastatic adenocarcinoma consistent with breast primary. The second patient with BIRADS 4 and no follow-up within 2-months had biopsies of bilateral breast masses 79 days later with results showing benign pathology.

Four of 5 patients with BIRADS 5 had follow-up at 2 months (table 3). The one remaining patient had biopsies of 2 breast masses approximately 8 months (239 days) after the incident BIRADS 5 result and the pathology was consistent with invasive ductal carcinoma. All patients with BIRADS 6 results were followed up within 1 month.

Papanicolaou (Pap) Smear Reports

Results of the manual review ('gold' standard) of 500 Pap reports are shown in table 4. Interrater agreement between the two reviewers was 1.0. When we compared Pap results determined by NLP against the 'gold' standard, we obtained 100% each for Precision, Recall, and Accuracy for the NLP model.

Follow-up of Abnormal Pap Smear Results

For the abnormal Pap smear results identified using NLP, we reviewed the medical record in order to determine the time between the incident abnormal Pap result and any repeat testing or procedures (example; repeat Pap smear or colposcopy). For women with Pap smear results showing low-grade squamous intraepithelial lesion (LSIL), 37% had no documented follow-up testing at 18-months after their incident abnormal Pap (table 4). For the atypical squamous cells of undetermined significance (ASCUS) Pap abnormalities, 47% of women had no documented follow-up testing at 18-months.

DISCUSSION

Although mammography and Pap smears have well-established lexicons for results reporting, results of these diagnostic procedures are not routinely stored in clinical information systems as discrete query-able data elements, but are "buried" in free-text narrative reports. As a result, the full benefit of using standard lexicons to report results are not realized.^{22,23} A major barrier to structured reporting can be found in current adoption patterns in which breast imaging, cardiology, and gastroenterology are the most common use cases.²⁴ Diagnostic reports in these disciplines describe well-circumscribed anatomical sites that are subject to fairly narrow pathologies, thus making results reporting conducive to manageable structured templates. However, the spectrum of possible anatomical sites and potential pathologies increases significantly for other studies such as magnetic resonance imaging (MRI) scans, thus making structured results reporting potentially subject to inaccuracies and lack of completeness.²⁵ For example, a study by Johnson²⁵ found that

completeness and accuracy scores for MRI results declined significantly when physicians switched from a narrative reporting format to a structured format. However, other investigators have found improvements in satisfaction with diagnostic reports after adoption of structured reporting. ²⁶ Therefore, for some diagnostic studies, there may be trade-offs between result quality (completeness and accuracy) and structured formatting of results. Finally, some radiologists are concerned that structured data entry of results may adversely affect their ability to effectively interpret radiographic images because of sub-optimally designed human-computer interfaces. ²⁷ Future research is needed to investigate potential trade-offs of narrative versus structured results reporting on outcomes such as physician workflow, report quality, and timely follow-up of abnormal results.

In summary, much of the data generated by our healthcare system, including radiology reports, pathology reports, clinical notes and discharge summaries, will continue to be entered into EMRs as narrative text. Given the increasing need for data analytics to help drive evidence-based patient care, regulatory policies and NLP-enabled technologies will be needed to effectively leverage this unstructured data. Along these lines, one of the Stage-3 meaningful use regulations proposed by the Office of the National Coordinator for Health Information Technology is that EMRs have the ability to identify abnormal test results and notify ordering providers. ²⁸

Presumably, this should include abnormal results stored as both structured and unstructured data in EMRs. Our study developed NLP models that accurately identify abnormal results from mammography and Pap smear reports and convert the results into structured data that can be used to generate computerized alerts and reminders for clinicians. Future studies should focus on integrating accurate NLP models into CDSSs that will automatically alert providers of abnormal results in free-text reports. Wagholikar and colleagues 17 conducted a pilot study of a CDSS that used NLP to identify and extract results from free-text electronic Pap smear reports and make follow-up recommendations. The authors found that the CDSS recommendations were correct in 73 out of 74 test patients when compared with physician judgment. Also, two cases were correctly identified by the NLP-based CDSS for gynecological referral for abnormal results that were not initially identified by the physician. The one incorrect CDSS recommendation in the Wagholikar¹⁷ study and the excellent, but imperfect, NLP model performance in our study highlight a major issue going forward with NLP-enabled CDSSs; what level of accuracy is required before these systems can be safely used in patient care? For example, does the NLP model need to perform perfectly to be used for patients care or alternatively, is 100% sensitivity with less than perfect positive predictive value sufficient? Also, what type of governance structures should institutions develop to monitor and ensure the satisfactory performance of NLP-based CDSSs. To illustrate further, the performance characteristics of a previously high-performing NLP model may become unacceptable because the format of a free-text report in the EMR has changed. If this goes undetected, then patients' can be harmed when alerts do not fire appropriately. Therefore, as NLP-based CDSSs are implemented, institutions need to establish governance structures that are responsible for monitoring performance and upgrading NLP-based CDSSs as test report formats and recommendation guidelines change. There is growing literature showing that poorly designed or poorly implemented health information technology can result in patients harm²⁹; this is certainly the case with NLP-based CDSSs. For example, electronic health

records that generate excessive numbers of alerts for relatively minor result abnormalities may increase the frequency of missed test results due to information overload and alert fatigue. ^{30,31} To prevent alert fatigue, systems should permit a degree of customization by physicians that prioritize alerts based on urgency level. ^{32,33}

Our study found that significant numbers of abnormal mammogram and Pap smear results identified using NLP were not followed up in a timely manner. One woman with a BIRADS 4 result on her mammogram with a recommendation that several suspicious axillary masses be followed up via fine needle aspiration did not have the recommended procedure for another 8 months (251 days). Results of fine needle aspiration of the axillary masses were consistent with metastatic adenocarcinoma (breast primary). A second woman in our study with a BIRADS 5 result did not have recommended follow-up (biopsy) for approximately 8 months (239 days) and the results showed invasive ductal carcinoma. We also found instances of abnormal Pap smears with inadequate follow-up. Guidelines indicate that patients having Pap smear results with LSIL should be referred for colposcopy (if HPV testing positive or not performed) or for repeat Pap smear testing in one year (if HPV negative). However, 33% of patients with LSIL results in our study had no evidence of follow-up at 18 months from the incident abnormal Pap smear result. Our results are similar to that reported in the literature showing follow-up of abnormal Pap smear results is frequently inadequate. 11,34,35 For example, up to 28% of women don't receive timely follow-up of abnormal results: especially minority women. 11,12 Although multiple factors affect timeliness of follow-up for abnormal cancer screening results 13, the recognition of abnormal results by clinicians is the first step in the process - a process found to be suboptimal and susceptible to error^{7,8,13}. Poon et al⁸ found that 83% of physicians reported at least one delay in reviewing patients' test results during the previous 2-months; automated alerts and reminders may significantly reduce these delays. Presumably a well-designed NLP-based CDSS can help avoid or ameliorate some of the diagnostic delays we found in our study. For example, using structured and unstructured data from an EMR, a rules-based CDSS can generate physician alerts if patients with BIRADS 4-6 results have no evidence of subsequent follow-up (breast surgery clinic, biopsy, fine needle aspiration, etc.) within a prescribed time interval. Physicians and/or care managers can turn off alerts if they know patients have had follow-ups at outside institutions or if patients had subsequent follow-ups not included in the CDSS algorithm (example, palliative care or hospice).

There are several limitations in our study. The first limitation is that the NLP models we developed may not be readily portable to other institutions that have different formats and word pattern in their mammography and Pap smear reports. However, institution-to-institution variation in the report formats is likely minimized because of the ubiquity of result reporting standards for these tests. A second limitation is that we chose use cases (mammography and Pap smears) in which result reporting is standardized (BIRADS¹⁹ and Bethesda²⁰) and the number of possible result categories fairly small, with seven for mammography (table 1) and nine for Pap smears (table 2). Other types of reports, such as radiographs (other than mammograms), present much more of a challenge for NLP in that result reporting is less standardized and the number of possible results numerous. Consequently, NLP performance is highly dependent on the diagnostic radiology procedure being used, and recall (sensitivity) can range from 98% for detecting pulmonary embolus on

chest CT reports to 68% for intracranial hemorrhage on head CT reports.³⁶ The last limitation is that we relied on documentation in the electronic medical record to determine if abnormal mammogram and Pap smear results were followed up. It is possible that patients had appropriate and timely follow-up at outside institutions and this was not documented in our institution's EMR.

Finally, patients play an important role in the result follow-up process. For example, direct notification of test results to patients via letter, email, telephone, or patient portal might serve as a safety net to help facilitate timely follow-up of abnormal results and help empower patients to be active participants in their care.^{3,37–39} In fact, proposed Stage-3 meaningful use objectives for electronic health records stipulate that patients must be able to view their test results on line within 4 business days of the information becoming available to their physicians.²⁸ However, as the trend towards direct patient notification continues, systems need to be designed that facilitate patients' understanding of their test results.^{38,40}

CONCLUSION

In summary, this study outlines the development of NLP models that convert free-text results of mammography and Pap smear reports into a structured format that can be used to generate alerts and reminders for clinicians. Future directions include linking the NLP-derived Pap smear and mammography results to explicit evidence-based guidelines that can be provided to clinicians via an EMR-based CDSS and studying the impact of the CDSS on timely follow-up of abnormal cancer screening results.

Acknowledgments

FUNDING SOURCES: The project was funded by a National Institutes of Health grant 1UL1TR001111 through the North Carolina Translational and Clinical Sciences Institute.

References

- Hickner J, Graham DG, Elder NC, et al. Testing process errors and their harms and consequences reported from family medicine practices: a study of the American Academy of Family Physicians National Research Network. Qual Saf Health Care. Jun; 2008 17(3):194–200. [PubMed: 18519626]
- Plews-Ogan ML, Nadkarni MM, Forren S, et al. Patient safety in the ambulatory setting. J Gen Intern Med. Jul; 2004 19(7):719–725. [PubMed: 15209584]
- 3. Callen JL, Westbrook JI, Georgiou A, Li J. Failure to follow-up test results for ambulatory patients: a systematic review. J Gen Intern Med. Oct; 2012 27(10):1334–1348. [PubMed: 22183961]
- 4. Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. Ann Intern Med. Oct 3; 2006 145(7):488–496. [PubMed: 17015866]
- 5. NCQA. Patient-Centered Medical Home. 2011. http://www.ncqa.org/tabid/631/default.aspx. Accessed May 23, 2012.
- Lin JJ, Dunn A, Moore C. Follow-up of outpatient test results: a survey of house-staff practices and perceptions. Am J Med Qual. May-Jun;2006 21(3):178–184. [PubMed: 16679437]
- 7. Moore C, Saigh O, Trikha A, Lin J. Timely Follow-Up of Abnormal Outpatient Test Results: Perceived Barriers and Impact on Patient Safety. Journal of Patient Safety. 2008; 4(4):241–244.
- 8. Poon EG, Gandhi TK, Sequist TD, Murff HJ, Karson AS, Bates DW. "I wish I had seen this test result earlier!": Dissatisfaction with test result management systems in primary care. Arch Intern Med. Nov 8; 2004 164(20):2223–2228. [PubMed: 15534158]

9. Wahls TL, Cram PM. The frequency of missed test results and associated treatment delays in a highly computerized health system. BMC Fam Pract. 2007; 8:32. [PubMed: 17519017]

- 10. Zapka J, Taplin SH, Price RA, Cranos C, Yabroff R. Factors in quality care—the case of follow-up to abnormal cancer screening tests—problems in the steps and interfaces of care. Journal of the National Cancer Institute. Monographs. 2010; 2010(40):58–71. [PubMed: 20386054]
- 11. Peterson NB, Han J, Freund KM. Inadequate follow-up for abnormal Pap smears in an urban population. J Natl Med Assoc. Sep; 2003 95(9):825–832. [PubMed: 14527050]
- 12. Yabroff KR, Washington KS, Leader A, Neilson E, Mandelblatt J. Is the promise of cancer-screening programs being compromised? Quality of follow-up care after abnormal screening results. Med Care Res Rev. Sep; 2003 60(3):294–331. [PubMed: 12971231]
- 13. Zapka J, Taplin SH, Price RA, Cranos C, Yabroff R. Factors in quality care—the case of follow-up to abnormal cancer screening tests—problems in the steps and interfaces of care. J Natl Cancer Inst Monogr. 2010; 2010(40):58–71. [PubMed: 20386054]
- 14. Laxmisan A, Sittig DF, Pietz K, Espadas D, Krishnan B, Singh H. Effectiveness of an electronic health record-based intervention to improve follow-up of abnormal pathology results: a retrospective record analysis. Medical care. Oct; 2012 50(10):898–904. [PubMed: 22929995]
- 15. Hysong SJ, Sawhney MK, Wilson L, et al. Provider management strategies of abnormal test result alerts: a cognitive task analysis. Journal of the American Medical Informatics Association: JAMIA. Jan-Feb;2010 17(1):71–77. [PubMed: 20064805]
- 16. Souza NM, Sebaldt RJ, Mackay JA, et al. Computerized clinical decision support systems for primary preventive care: a decision-maker-researcher partnership systematic review of effects on process of care and patient outcomes. Implement Sci. 2011; 6:87. [PubMed: 21824381]
- Wagholikar KB, MacLaughlin KL, Henry MR, et al. Clinical decision support with automated text processing for cervical cancer screening. Journal of the American Medical Informatics Association: JAMIA. Sep-Oct;2012 19(5):833–839. [PubMed: 22542812]
- Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? Journal of biomedical informatics. Oct; 2009 42(5):760–772. [PubMed: 19683066]
- Radiology ACo. American College of Radiology Breast Imaging Reporting and Data System BI-RADS. Reston, VA:
- 20. Solomon D, Davey D, Kurman R, et al. The 2001 Bethesda System: terminology for reporting results of cervical cytology. JAMA: the journal of the American Medical Association. Apr 24; 2002 287(16):2114–2119. [PubMed: 11966386]
- 21. Solomon D, Davey D, Kurman R, et al. The 2001 Bethesda System: terminology for reporting results of cervical cytology. Jama. Apr 24; 2002 287(16):2114–2119. [PubMed: 11966386]
- 22. Johnson AJ. All structured reporting systems are not created equal. Radiology. Feb.2012 262(2): 726. author reply 726–727. [PubMed: 22282188]
- 23. Liu D, Zucherman M, Tulloss WB Jr. Six characteristics of effective structured reporting and the inevitable integration with speech recognition. Journal of digital imaging. Mar; 2006 19(1):98–104. [PubMed: 16270236]
- 24. Langlotz CP. Structured radiology reporting: are we there yet? Radiology. Oct; 2009 253(1):23–25. [PubMed: 19789252]
- 25. Johnson AJ, Chen MY, Swan JS, Applegate KE, Littenberg B. Cohort study of structured reporting compared with conventional dictation. Radiology. Oct; 2009 253(1):74–80. [PubMed: 19709993]
- 26. Schwartz LH, Panicek DM, Berk AR, Li Y, Hricak H. Improving communication of diagnostic radiology findings through structured reporting. Radiology. Jul; 2011 260(1):174–181. [PubMed: 21518775]
- Weiss DL, Langlotz CP. Structured reporting: patient care enhancement or productivity nightmare?
 Radiology. Dec; 2008 249(3):739–747. [PubMed: 19011178]
- 28. HHS(ONC). Request for Comment Regarding the Stage 3 Definition of Meaningful Use of Electronic Health Records (EHRs) 2013. http://www.healthit.gov/sites/default/files/hitpc_stage3_rfc_final.pdf. Accessed December 10, 2013.
- 29. Sittig DF, Singh H. Defining health information technology-related errors: new developments since to err is human. Arch Intern Med. Jul 25; 2011 171(14):1281–1284. [PubMed: 21788544]

 Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, Sittig DF. Information overload and missed test results in electronic health record-based settings. JAMA internal medicine. Apr 22; 2013 173(8):702–704. [PubMed: 23460235]

- Hysong SJ, Sawhney MK, Wilson L, et al. Understanding the management of electronic test result notifications in the outpatient setting. BMC medical informatics and decision making. 2011; 11:22. [PubMed: 21486478]
- 32. Singh H, Wilson L, Reis B, Sawhney MK, Espadas D, Sittig DF. Ten strategies to improve management of abnormal test result alerts in the electronic health record. Journal of patient safety. Jun; 2010 6(2):121–123. [PubMed: 20563228]
- 33. Singh H, Spitzmueller C, Petersen NJ, et al. Primary care practitioners' views on test result management in EHR-enabled health systems: a national survey. Journal of the American Medical Informatics Association: JAMIA. Jul-Aug;2013 20(4):727–735. [PubMed: 23268489]
- 34. Jones BA, Dailey A, Calvocoressi L, et al. Inadequate follow-up of abnormal screening mammograms: findings from the race differences in screening mammography process study (United States). Cancer Causes Control. Sep; 2005 16(7):809–821. [PubMed: 16132791]
- 35. McCarthy BD, Yood MU, Boohaker EA, Ward RE, Rebner M, Johnson CC. Inadequate follow-up of abnormal mammograms. Am J Prev Med. Jul-Aug;1996 12(4):282–288. [PubMed: 8874693]
- 36. Lakhani P, Kim W, Langlotz CP. Automated detection of critical results in radiology reports. Journal of digital imaging. Feb; 2012 25(1):30–36. [PubMed: 22038514]
- 37. Unruh KT, Pratt W. Patients as actors: the patient's role in detecting, preventing, and recovering from medical errors. International journal of medical informatics. Jun; 2007 76(Suppl 1):S236–244. [PubMed: 16829180]
- 38. Wald JS, Burk K, Gardner K, et al. Sharing electronic laboratory results in a patient portal—a feasibility pilot. Studies in health technology and informatics. 2007; 129(Pt 1):18–22. [PubMed: 17911670]
- 39. Hannan A. Providing patients online access to their primary care computerised medical records: a case study of sharing and caring. Informatics in primary care. 2010; 18(1):41–49.
- 40. Chen ET, Eder M, Elder NC, Hickner J. Crossing the finish line: follow-up of abnormal test results in a multisite community health center. Journal of the National Medical Association. Aug; 2010 102(8):720–725. [PubMed: 20806684]

IMPRESSION: BIRADS 5. HIGHLY SUGGESTIVE OF MALIGNANCY.
APPROPRIATE ACTION SHOULD BE TAKEN.

Spiculated 2.2-cm mass with associated microcalcifications and architectural distortion at 9 o'clock in the right breast, highly concerning for breast cancer. Ultrasound guided core needle biopsy is recommended.

FIGURE 1.

Sample Mammography Report.

RESULTS/INTERPRETATION

Atypical squamous cells, cannot exclude high-grade squamous intraepithelial lesion.

SPECIMEN ADEQUACY

Satisfactory for evaluation, endocervical/transformation zone component present.

FIGURE 2.

Sample Pap Smear Report.

TABLE 1Breast Imaging Reporting and Data System (BIRADS)

BIRADS Category	Follow-up Recommendation	% of Sample (N=421)
0: Incomplete	Need for further evaluation	3.6% (15)
1: Normal	Normal interval follow-up	13.5% (57)
2: Benign	Normal interval follow-up	57.9% (244)
3: Probably benign	A short-interval follow-up	12.1% (51)
4: Suspicious abnormality	A biopsy should be considered	8.6% (36)
5: Highly suggestive for malignancy	Biopsy or surgery should be performed	1.2% (5)
6: Biopsy-proven carcinoma	Appropriate action should be taken	3.1% (13)

Moore et al. Page 14

Table 2

Pap Report Results (501 results in 500 reports)[†]

Pap Results				
Negative				
Unsatisfactory for evaluation				
Abnormal Results (40)				
Atypical Squamous Cells of Undetermined Significance (ASCUS)	3.0% (15)			
Atypical Squamous Cells, Cannot Exclude High-Grade Squamous Cell Intraepithelial Lesion (ASC-H)				
Low-grade Squamous Intraepithelial Lesion (LSIL)				
High-grade Squamous Intraepithelial Lesion (HSIL)				
Atypical Glandular Cells (AGC)				
Squamous Cell Carcinoma (SCC)				
Endocervical Adenocarcinoma in Situ (AIS)				

 $\label{eq:TABLE 3} \textbf{Follow-up of Abnormal Mammograms (N=54)}$

BIRADS Category	Follow-up within 1 Month	Follow-up within 2 months	Description of Results not Followed up within 2 Months	
4	86.5% (31 of 36)	94.6% (34 of 36)	 Mammography report recommended fine needle aspiration of left axillary mass. Fine needle aspiration was performed 251 days later and the pathology results showed metastatic adenocarcinoma consistent with breast primary. Mammography report recommended biopsies of bilateral hypoechoic masses in each breast. The biopsies were performed 79 days later and the pathology results were benign. 	
5	80.0% (4 of 5)	80.0% (4 of 5)	Mammography report recommended biopsy of two breast masses suspicious for malignancy. Biopsy was performed 239 days later and pathology showed invasive ductal carcinoma.	
6	100.0% (13 of 13)	100.0% (13 of 13)	All BIRADS 6 results were followed up within 1 month	
TOTAL	89.1% (48 of 54)	94.6% (51 of 54)		

Moore et al. Page 16

 $\label{table 4} \textbf{TABLE 4}$ Follow-up Testing for Abnormal Pap Smear Results (N=40)

Abnormal Pap Smear Category	No Documented Follow-up Testing	If Follow-up, Average Months to Follow-up
Atypical Squamous Cells of Undetermined Significance (ASCUS)	47% (7 of 15)	8.1 months
Low-grade Squamous Intraepithelial Lesion (LSIL)	37% (7 of 19)	8.5 months
High-grade Squamous Intraepithelial Lesion (HSIL)	0% (0 of 2)	12.5 months
Atypical Glandular Cells (AGC)	0% (0 of 2)	0.6 months
Squamous Cell Carcinoma (SCC)	0% (0 of 1)	0.1 months
Endocervical Adenocarcinoma in Situ (AIS)	0% (0 of 1)	1.6 months