

Automated Detection of Radiology Reports that Document Non-routine Communication of Critical or Significant Results

Paras Lakhani^{1,2} and Curtis P. Langlotz¹

The purpose of this investigation is to develop an automated method to accurately detect radiology reports that indicate non-routine communication of critical or significant results. Such a classification system would be valuable for performance monitoring and accreditation. Using a database of 2.3 million free-text radiology reports, a rule-based query algorithm was developed after analyzing hundreds of radiology reports that indicated communication of critical or significant results to a healthcare provider. This algorithm consisted of words and phrases used by radiologists to indicate such communications combined with specific hand-crafted rules. This algorithm was iteratively refined and retested on hundreds of reports until the precision and recall did not significantly change between iterations. The algorithm was then validated on the entire database of 2.3 million reports, excluding those reports used during the testing and refinement process. Human review was used as the reference standard. The accuracy of this algorithm was determined using precision, recall, and F measure. Confidence intervals were calculated using the adjusted Wald method. The developed algorithm for detecting critical result communication has a precision of 97.0% (95% CI, 93.5–98.8%), recall 98.2% (95% CI, 93.4–100%), and F measure of 97.6% ($\beta = 1$). Our query algorithm is accurate for identifying radiology reports that contain non-routine communication of critical or significant results. This algorithm can be applied to a radiology reports database for quality control purposes and help satisfy accreditation requirements.

KEY WORDS: Critical results reporting, data mining, Joint Commission on Accreditation of Healthcare Organizations (JCAHO), natural language processing, online analytical processing (OLAP), quality assurance, quality control, radiology reporting

INTRODUCTION

When a radiologist interprets a study with a critical result, good medical practice requires that those findings are documented in the

radiology report and that the provider caring for the patient is notified in a timely manner. In these situations, the goal is to expedite delivery of this report using non-routine communication channels such as the telephone or hospital paging system to ensure receipt of these findings.

This practice has become more prevalent after the American College of Radiology (ACR) released a standard stating that radiologists should document non-routine communications in the body of the radiology report.¹ This practice is recommended for quality control and medical–legal reasons.^{2,3}

The Joint Commission, a governing body that regulates the accreditation process of hospitals, also requires that hospitals provide evidence of timely reporting of critical results.⁴ Thus, as part of hospital accreditation, the facility must provide documentation that radiologists are indeed using non-routine communication channels to transmit critical test results to referring physicians. Since there is not yet standard language that radiologists must use to report such findings, providing such evidence can be difficult,

¹From the Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA.

²From the Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA, USA.

For correspondence and copies of the algorithm: Paras Lakhani, Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA, USA; e-mail: paras.lakhani2@uphs.upenn.edu

Copyright © 2009 by Society for Imaging Informatics in Medicine

Online publication 14 October 2009

doi: 10.1007/s10278-009-9237-1

often requiring an imprecise and laborious manual review of a large body of reports that impedes the monitoring of such communications for quality control purposes.^{5,6}

The goal of this research project was to devise an automated method to search the text of radiology reports and identify those reports that contain documentation of communications with the referring provider.

Since the body of the radiology report consists of relatively unstructured narrative free text, these reports are generally less amenable to data abstraction via a computerized method, which is best suited for interpreting "coded" or well-defined data with standard relationships. One solution is to use natural language processing (NLP) systems that convert natural human language into discrete representations including first order logic, which are easier for computers to process.⁷⁻⁹ Some authors have used NLP techniques to manipulate radiology reports and then use that output for data-mining efforts.^{10,11}

However, we were previously successful in mining data directly from free-text radiology reports using simple rule-based queries based on standard languages.¹² Other authors have similarly used regular expressions for improving information retrieval.¹³ Thus, the purpose of this study was to determine if a simple rule-based algorithm using standard languages could accurately classify radiology reports containing documentation of communications. An accurate algorithm could then be used for information retrieval efforts and accreditation purposes.

MATERIALS AND METHODS

This study was Health Insurance Portability and Accountability Act compliant. We utilized a pre-existing de-identified database of radiology reports approved by the Institutional Review Board via exempted review.

Hospital Setting and Research Database Creation

The Hospital of the University of Pennsylvania employs over 100 radiologists including faculty

and trainees and performs approximately 300,000 diagnostic imaging examinations per year. We have used speech recognition technologies since 2000 to dictate radiology reports.

From 1997 through November of 2005, approximately 2.3 million digital diagnostic imaging procedures were performed at our institution archived on its PACS (GE Centricity 2.0, GE Medical Systems, Waukesha, WI) and radiology information system (RIS) (IDXrad v9.6, IDX, Burlington, VT).

Due to the load that indexing and searching would place on critical day-to-day clinical operations, a stand-alone research relational database management system (RDBMS) was created for data mining and research purposes. The radiology reports were duplicated and transferred onto a secondary stand-alone research RDBMS using Oracle 10g enterprise edition as the database server (Oracle Corporation, Redwood Shores, CA).

This radiology research database was accessible via structured query language (SQL). We opted for Oracle 10g enterprise edition, a more sophisticated information retrieval engine that included data-mining options, which extended SQL to enable more flexible and powerful text searching.

The database was loaded using a custom Perl script that connected to the clinical PACS database, established a mapping between the two database schemata, and performed the duplication of reports. The research database was created in 2004 and was updated on a weekly basis until November 2005.

Database Characteristics

This research database contains all of the radiology reports from our institution from November 1997 to 2005, for all modalities and body parts performed. It also contains only specific attributes that were deemed important for this and similar investigations. These attributes included the text of the radiology report, clinical history, comments, accession number, study date, study code, modality, and reading radiologist(s). Direct patient identifiers such as patient name, birth date, and medical record number were excluded from the database.

Report Indexing

Most databases are indexed after removing stop words (i.e., “if,” “and,” “or,” “but,” “the,” “no,” and “not”), which usually consist of common prepositions, conjunctions, and pronouns. The removal of stop words allows for more efficient indexing and querying of databases. In this study, however, we planned to develop SQL queries containing many common stop words. Therefore, the research database was re-indexed with all stop words included, which ensured that these words were searchable.

Algorithm Development

Queries were developed using SQL and designed to search for common phrases in

radiology reports that could indicate notifications. For example, the queries contained verbs such as “informed” or “notified” to select reports that contained these words in the radiology report text. Actual examples of some phrases from radiology reports that indicate documentation of communications are provided in Figure 1. Because many different words and phrases are used to indicate documentation of communications, we devised a query method that could select for many terms at the same time. To do this, we utilized full-text searching features of our information retrieval engine, which enabled us to search for multiple terms and phrases within the text of the radiology reports. An example of such a query would be: Select all of the radiology reports that *contain* the phrase

Examples of Actual Documentations of Communications in Radiology Reports:

1. At the conclusion of the examination, the *physician* caring for the patient was *notified* of these results.
2. The results were *communicated* with the *referring physician* at approximately 1:30AM.
3. *Doctor* [name] was *called* with these results following completion of the examination.
4. Case *discussed* with *Dr.* [name] of the *Trauma* Bay on 6/20/05 at 2010 hours.
5. I immediately *contacted* the *emergency* room following my review of the study.
6. These findings were *conveyed* by *telephone* to *Dr.* [name] at the time of dictation, 6/18/04.
7. *Dr.* [name] in the ICN, extension [telephone number] was *informed* of this.
8. Findings were *reported* to the *neurosurgical housestaff* at the time of study completion on 8/12/04.
9. The *emergency* room staff was made *aware* of the findings after the interpretation.
10. The case was *reviewed* with *Dr.* [name] of the Orthopedic Service at the conclusion of the examination.

Fig 1. Ten examples of non-routine communication of results to a healthcare provider extracted from radiology reports in our institution. Notification verbs and notification recipients are *italicized* and shown in *bold font*.

“informed” or “notified” using the appropriate SQL syntax. A query such as this could be expanded to include or exclude many more terms.

Given the complexity of human language, there exist a myriad ways in which a radiologist could document or dictate such communications in a radiology report. A complete list aiming to include all possible phrases and terms would contain many thousands of combinations. Rather than try to enumerate every possible combination, which would be an exhaustive effort, we initially analyzed a small sample of actual radiology reports (50 reports) containing documentation of non-routine communications and examined the specific phraseology used in those cases.

Based on those reports, we determined that documentations of communications contain three key elements (Fig. 2): (1) verb used to describe the communication (i.e., *discussed*, *notified*, *informed*, and *communicated*), (2) noun used to describe a recipient of the communication, such as a person (i.e., *physician* or *nurse*), or place (i.e., *medical intensive care unit* or *emergency department*), and (3) noun or pronoun referring to the diagnostic study itself (i.e., *MRI*, *the results*, *these findings*). Actual examples of sentences or phrases used by radiologists at our institution to document such communications are listed in Figure 1. Of these elements, all proper documentations of communications must contain the verb describing the act of communication itself (e.g., *discussed*, *notified*, and *informed*). Without such a verb, the phrase would have no rational meaning (e.g., “The results of the MRI referring physician” rather than “The results of the MRI were *discussed* with the referring physician”).

Therefore, we first created a list of verbs that radiologists used to document such communications

based on the initial sample of 50 reports. To ensure that the list would encompass nearly all potential relevant verbs, we added more verbs to the list after discovering commonly encountered synonyms using a standard thesaurus (for example, “convey” is a synonym of “communicated”). All synonyms were queried against the entire database and ranked in their order of frequency. The final list of verbs used in the query is provided in Table 1, ranked by their frequency in the entire database.

For every verb in Table 1, we selected 50 consecutive radiology reports containing those verbs and analyzed how those verbs were used in a phrase or sentence. Because there were 15 different verbs in our final query algorithm (Table 1), this resulted in an additional 750 radiology reports (15×50). Thus, we created a relatively large corpus of reports, most of which contained statements indicating documentation of communications. This proved to be a valuable resource in initially designing our query algorithm. Some of these verbs were very specific for indicating documentation of communications, such as “notified.” For example, 48 of 50 (96%) reports containing the word “notified” indicated documentation of communications, as in “the physician was *notified* of these results.” Other verbs were less specific, such as “informed,” which indicated documentation of communications in 43 of 50 reports (86%), mainly because of confounding statements such as “informed consent was obtained.”

Using a similar methodology, a list of potential notification recipients (i.e., *Dr.*, *physician*, *nurse*, *surgeon*, *MICU*, and *ER*) was created (Table 2). We created this list based on notification statements derived from both the initial sample of 50 reports and from the corpus of 750 reports

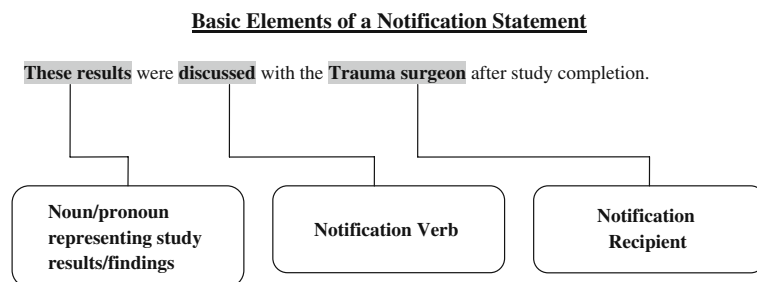


Fig 2. The three essential components of a notification statement are highlighted in *bold font*, which includes noun/pronoun representing study findings, notification verbs, and notification recipients.

Table 1. Notification Verbs

Verb	Frequency
Reviewed	118,905 (5.208%)
Discussed	111,584 (4.888%)
Reported	79,265 (3.472%)
Informed	63,805 (2.795%)
Called	22,882 (1.002%)
Notified	14,499 (0.635%)
Conveyed	11,611 (0.509%)
Communicated	9,229 (0.404%)
Given to	4,063 (0.178%)
Telephone (verb or noun)	2,643 (0.116%)
Was/Is Aware (adjective)	2,397 (0.105%)
Contacted	972 (0.043%)
Relayed	966 (0.042%)
Transmitted	496 (0.022%)
Alerted	61 (0.003%)

The notification verbs in the query algorithm are listed above, ranked by frequency in entire research database of 2,282,934 reports with corresponding percentages. These verbs are used to indicate documentation of communications. However, these verbs may be used in other contexts. Note that "telephone" may be used as a verb or noun. Also, "aware" is an adjective that is combined with the verb "was" or "is"

Table 2. Notification Recipients

Person	Team/Unit/Thing	Place
MD	Service	Nursery
Dr.	Team	Room
Drs.	Trauma	Clinic
Doctor	Critical (i.e. critical care unit)	Floor
Housestaff	Operating (i.e. operating room)	Unit
Houseofficer		Emergency
House (i.e. house staff)		ED
Referring		ER
Physician		MICU
Clinician		CICU
Resident		SICU
Intern		NICU
Extern		PACU
Fellow		APU
GYN		CCU
OB		PICU
CRNP		ICU
RN		
Nurse		
Surg%		
Neurosurg%		
Urolog%		
Pulmonolog%		
Cardiolo%		
Gastroenterolo%		
Neonatolo%		
Nephrolo%		
Oncolo%		
Gynecolo%		

The notification recipients in the query algorithm are listed above, which may represent a person, service/unit, or place. The percent (%) sign denotes a "wildcard" character. For example, a query containing the term "surg%" will select for words that begin with "surg" such as "surgeon," "surgery," or "surgical"

derived from the verbs in Table 1. Additional notification recipients were added to the list over time, as they were discovered during the query refinement process (see below). Over time, the list grew and eventually included over 50 terms (Table 2).

Algorithm Design

A simple query consisting of all of the verbs listed in Table 1 would yield imprecise results because those verbs could be used in many different contexts and are not specific for selecting documentation of communications. For example, the verb “reported” could be used in the phrase “These MRI results were *reported* to Dr. Smith,” indicating an actual documentation of communication, or used in a different context such as “There is a *reported* history of a lung nodule.” Thus, a query containing the verb “reported” would select both reports. Therefore, our query needed to distinguish between these possibilities and accurately select only reports indicating documentation of communications.

To overcome this problem, we utilized proximity searching in our algorithm. In Oracle, this is denoted by the “near” operator. This function selects reports containing different words that are within a certain distance of each other. For our purposes, we utilized this feature to select reports where notification verbs (Table 1) are within a certain distance from notification recipients (Table 2), such as “select all reports where ‘informed’ and ‘doctor’ are within 14 words of each other” utilizing the appropriate SQL syntax. By utilizing this feature, we were able to significantly improve the precision of our query algorithm.

We chose 14 words as the default maximum separation distance between our notification verbs and notification recipients. We chose this number after analyzing a large number of reports containing documentation of communications (the initial set of 50 reports, corpus of 750 reports described above and from the query refinement process as described below). The two exceptions include the verbs “called” and “reported,” which utilized a maximum separation distance of six words. This latter number was also determined after reviewing

radiology reports during query development and refinement.

We used more specific rules for certain verbs (“reported,” “reviewed,” “called,” and “transmitted”), which included proximity matching and/or certain two- to five-word phrases, as we found that to be reliable during the development and refinement process.

We also utilized additional full-text searching features, which included wildcards, equivalence, and standard Boolean operators. Wildcard characters were used as a rudimentary stemming method. An example is “surg%”. A SQL statement containing such would select for all words that begin with “surg,” such as “surgeon,” “surgeons,” “surgery,” and “surgical.” Thus, the use of wildcards reduced the length of a SQL statement and potentially increase the number of words selected.

The general schema for our query design is depicted in Figure 3.

Algorithm Refinement

After the initial query algorithm was developed, it underwent a rigorous process of iterative refinement to verify and improve accuracy. During this iterative refinement process, we applied the algorithm to select 100 consecutive reports from a specific week (for example, from Jan 1, 1998 to Jan 7, 1998). This provided an estimated confidence interval of about 10% per every sample of 100 reports.¹⁴ Each sample was subject to human review by a radiology resident in our department (PL), who analyzed and scored the results in a spreadsheet. The reports were then categorized in a binary fashion as either containing or not containing documentation of communications.

After each iteration, the algorithm was modified to include or remove terms and phrases to improve precision and recall. For example, the verb “given” was added to the list of notification verbs after discovering that it was used to document communications in a few reports, such as “the results of this study were given to the referring physician.”

After the algorithm was revised, it was applied to select a new set of consecutive reports from a different week. These reports were again subjected

General Schema of Query Algorithm

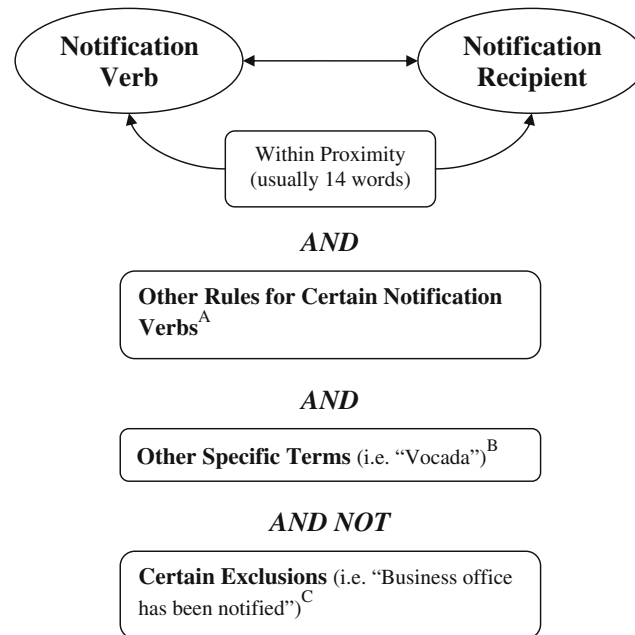


Fig 3. The general schema for the query algorithm includes notification verbs and recipients within a defined proximity, other rules and certain exclusions. A) The verbs “reported,” “reviewed,” “called,” and “transmitted” had additional or different rules, which consisted of proximity matching and/or specific two- to five-word phrases. B) “Vocada” is a specific third party message delivery system (Veriphy™, Nuance Communications, Burlington, MA) used by our institution to deliver some non-critical results. For example, “These results were communicated to the referring physician via Vocada at the conclusion of the study.” C) Certain phrases repeatedly occurred in radiology reports, which contained the notification verbs listed in Table 1 but were used in a context to indicate another meaning, as the example given above, which refers to a billing issue.

to human review, scored, and subsequently modified if necessary based on the new results. This process was repeated until the precision and recall of the query algorithm did not change by more than 0.5% between iterations.

Algorithm Validation

Once a final query algorithm was established, we determined precision by applying the final algorithm to 200 random reports from the entire database of approximately 2.3 million reports (excluding the initial test set and those specific dates used during the refinement process), using a randomization feature within Oracle. We chose 200 as the number of random reports to analyze because that provided an estimated 95% confidence interval of less than 7%.¹⁸ The 200 reports were manually scored by a radiology

resident in our department (PL) using the same process outlined above. Precision was calculated by dividing the number of reports that indicate documentation of communications by the 200 random reports sampled (see Fig. 4 and formula below).¹⁵

To determine recall, we selected 2,000 random reports that were not used in the development or refinement of the algorithm and were not identified by the query. In other words, these are radiology reports that do not contain any of the terms and phrases selected by the query. For example, if our query algorithm selected 5% of the radiology reports in the entire database, we would obtain a random sample of 2,000 reports from the remaining 95% of the database (Fig. 4). We chose a higher sample size for our recall calculation, since we estimated that this would result in an estimated 95% confidence interval of less than 7% based on

Accuracy of Query Algorithm

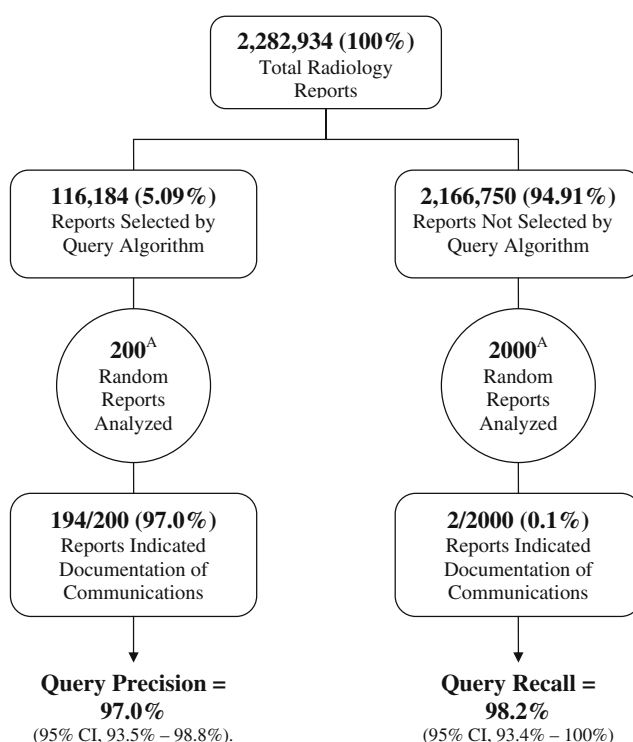


Fig 4. The query algorithm has a precision of 97.0% and recall of 98.2%. The query algorithm selected 116,184 (5.09%) radiology reports from the entire database of 2,282,934 reports. A) These random reports were pulled from nearly the entire database (99% of total database), as they exclude the small sample of reports analyzed during query development and refinement (which represent <1% of the total database) to prevent possibility of recurrence.

data obtained during the refinement process.¹⁴ These 2,000 random reports were also manually scored by a radiology resident in our department (PL) using the same process outlined above. The formulas used to calculate query precision and recall are shown below¹⁵:

Precision:

$$\frac{a}{a + b}$$

- a*: number of sampled reports containing documentation of communications.
b: number of sampled reports not containing documentation of communications.

Recall:

$$\frac{A}{C}$$

- A*: total number of radiology reports selected by query algorithm.

- C*: total number of radiology reports that indicate documentations of communications in entire database. This was estimated using the following formula: $C = (y \times B) + A$.¹⁶
y: percent of the 2,000 random sampled reports that indicate documentation of communications.
B: total number of radiology reports not selected by query algorithm.

The final actual 95% confidence intervals for query precision and recall were determined using the adjusted Wald (or Agresti-Coull) method.¹⁷ An *F* measure or weighted harmonic mean of precision and recall was calculated using equal weighting ($\beta=1$) between precision and recall using the formula¹⁸:

$$F = \frac{(1 + \beta^2) \times recall \times precision}{(\beta^2 \times precision) + recall}$$

RESULTS

The final list of notification verbs and recipients that constitute the terms in the query algorithm are provided in Tables 1 and 2, respectively.

A total of 116,184 radiology reports (5.09%) were selected by the algorithm as containing documentation of communications. The remaining 2,166,750 (94.91%) radiology reports were excluded by the query algorithm (Fig. 4).

In determining query precision, 194 of 200 (97.0%) random reports selected by the query algorithm had indicated documentation of communications. This resulted in a query precision of 97.0% (95% CI, 93.5–98.8%).

In determining query recall, 2/2,000 (0.1%) random reports not selected by the query algorithm had indicated documentation of communications. This resulted in a query recall of 98.2% (95% CI, 93.4–100%) using the formula described above (Fig. 4). The *F* measure of the algorithm is 97.6%, using an equal weighting between precision and recall.

DISCUSSION

In this study, we describe the development of a rule-based algorithm for identifying radiology reports that document communication of critical or significant results. This research was driven by a need to create an automated, precise query methodology that could determine the frequency of such communications for quality control purposes and help satisfy Joint Commission requirements.⁴

The precision (97.0%) and recall of (98.2%) of our algorithm compare favorably to that published elsewhere in the literature.^{13,19–21} Hand-crafted rule-based approaches such as ours have been shown to be capable of achieving relatively high accuracy for classifying text.²² The key drawbacks of rule-based approaches are that they usually require expert knowledge and can take a long time to generate, test, and validate. In this study, for example, we needed to manually review thousands of reports for development and validation and to achieve a narrow confidence interval for our precision and recall statistics. We also needed a very accurate method (greater than 90%) for classifying such reports, since our intention was

to use this to help us meet regulatory and accreditation requirements.

Regarding precision, six of the 200 sampled reports (3%) had not indicated documentation of communications. These reports contained terms defined in our query algorithm, but the context of such reports did not indicate documentation of communications. For example, a few reports had statements such as “This MRI was reviewed with Dr. [radiologist name],” where in fact the radiologist in question was the attending MRI radiologist, rather than the referring physician. Another example included a radiology report that stated “Dr. [name] of the service *will be notified* of this finding.” Since the future tense was used and no actual notification was made at the time of the report, it does not satisfy criteria for documentation of communications, since we cannot be exactly certain when and if a communication will be made. The ACR states that proper documentation of communications should indicate in writing that a notification has already been made.¹ In addition, proper communications should be made to a physician or healthcare provider/service caring for the patient. Therefore, reports that indicated notification directly to a patient, such as “findings and recommendations were discussed with the patient,” were not counted. These criteria are based on language on ACR and Joint Commission guidelines.^{1,4} The remaining three reports contained similar statements that were variations of the above examples.

Regarding query recall, 0.1% (2/2,000) of sampled reports not selected by the query algorithm had actually indicated documentation of communications. The first report had contained the phrase, “findings were reported by phone to Dr. Smith.” This was not selected by our query algorithm because “telephone” rather than “phone” was included in the query algorithm and because we employed more specific rules for the verb “reported” rather than simple proximity. For possible improvement, we would consider adding proximity searching with notification recipients for the verb “reported” and add the word “phone” to our algorithm. The second of the two reports included the statement “this information is indicated to Dr. [name] of the ED at 4:30 P.M. on 12/16/04.” This notification statement uses the verb “indicated,” which was not included in our query

algorithm because it represents a very uncommon usage of that verb.

Some terms/phrases were excluded from the query algorithm. These phrases contained notification verbs used in the algorithm but in a context that did not indicate documentation of communications to a healthcare provider or physician. One example is “business office has been notified” (Fig. 3). Reports containing such language referred to imaging studies performed by our department without available radiology interpretations in our RIS or PACS. In these cases, a standard statement was used to indicate that the patient not be billed for the study and to notify the “business office” for follow-up investigation. This macrostatement occurred very infrequently in the database, approximately 0.01% of all reports.

We did not need to employ negation detection in the algorithm, since it does not play a role in documentation of communications. While it may be possible for a radiologist to state “the referring physician was *not* notified of these results,” we did not encounter that in our extensive testing of the database.

Future Plans

We plan to apply this finalized query algorithm to our entire radiology reports database to track many quantitative metrics (i.e., frequency of reports containing documentation of communications by year, radiologist, modality, study code, indication, inpatient, or outpatient services, etc). These metrics can then be used to help satisfy Joint Commission requirements by monitoring our current reporting practices until we move to a common structured template for critical results communication.

Since the vast majority of these reports contain findings with significant or critical results (based on our analysis of 200 reports), we can use this algorithm to help build a large corpus of reports containing significant or critical findings. The few reports that had normal findings were clinically significant (i.e., no findings of acute injury in a trauma patient or live intrauterine pregnancy in a pregnant patient with vaginal bleeding), providing a reason why the radiologist contacted the referring physician or provider. Based on analyzing the corpus of such reports, similar algorithms could

then be created that automatically detect reports containing critical findings.

This query was developed using a standard language (SQL). Therefore, the syntax can be adapted to any full-text enabled SQL database, such as SQL server (Microsoft Corporation, Redmond, WA) or MySQL (Sun Microsystems, Santa Clara, CA), and the algorithm can be easily ported to other institutions.

Limitations

The algorithm was developed based on language used in radiology reports from our institution and therefore may be less accurate when used by other institutions. However, the timeframe of the analysis spanned 9 years (1997–2005), with over 100 different attending, fellow, and resident level radiologists producing reports during that time. Many of those radiologists were previously trained at other institutions, and thus the reports should encompass a relatively large breadth of reporting styles. For these reasons, we feel that the algorithm likely will be generalizable to other institutions. Finally, the algorithm does not handle all misspellings, although some are accounted for by using wildcards. Since this type of query algorithm can place a large load on the “radiology information system (RIS), it should be run during off-hours or on a stand-alone database separated from clinical operations. For query validation, manual review of the reports was performed by a radiologist and thus human errors due to fatigue and other factors are possible. To minimize this effect, the review process was done over a long period of time in small increments.

CONCLUSIONS

Our algorithm is accurate for identifying radiology reports that contain documentation of non-routine communications, with a precision of 97.0% and recall of 98.2%.

This algorithm can be applied to a radiology report database for quality-control purposes and to help satisfy accreditation requirements. Since this algorithm was developed using standard text-based queries, it can be ported to other institutions for their own monitoring and accreditation purposes.

REFERENCES

1. American College of Radiology: ACR practice guideline for communication of diagnostic imaging findings. In: Practice Guidelines & Technical Standards 2005. Reston, VA: American College of Radiology, 2005
2. Berlin L: Communicating radiology results. *Lancet* 367:373–375, 2006
3. Berlin L: Using an automated coding and review process to communicate critical radiologic findings: one way to skin a cat. *AJR Am J Roentgenol* 185:840–843, 2005
4. Patient safety requirement 2C (standard NPSG.2a). In: The 2007 Comprehensive Accreditation Manual for Hospitals: The Official Handbook. Oakbrook Terrace, Ill: Joint Commission Resources; 2007:NPSG-3–NPSG-4
5. Bates DW, Leape LL: Doing better with critical test results. *Jt Comm J Qual Patient Saf* 31:66–67, 2005
6. Singh H, Arora HS, Vij MS, Rao R, Khan MM, Petersen LA: Communication outcomes of critical imaging results in a computerized notification system. *J Am Med Inform Assoc* 14:459–466, 2007
7. Taira RK, Bashyam V, Kangaroo H: A field theoretical approach to medical natural language processing. *IEEE Trans Inf Technol Biomed* 11:364–375, 2007
8. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD: Unlocking clinical data from narrative reports: a study of natural language processing. *Ann Intern Med* 122:681–688, 1995
9. Langlotz CP: Automatic structuring of radiology reports: harbinger of a second information revolution in radiology. *Radiology* 224:5–7, 2002
10. Hripcsak G, Austin JH, Alderson PO, Friedman C: Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. *Radiology* 224:157–163, 2002
11. Dreyer KJ, Kalra MK, Maher MM, Hurier AM, Asfaw BA, Schultz T, Halpern EF, Thrall JH: Application of recently developed computer algorithm for automatic classification of unstructured radiology reports: validation study. *Radiology* 234:323–329, 2005
12. Lakhani P, Menschik ED, Goldszal AF, Murray JP, Weiner MG, Langlotz CP: Development and validation of queries using structured query language (SQL) to determine the utilization of comparison imaging in radiology reports stored on PACS. *J Digit Imaging* 19:52–68, 2006
13. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG: A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform* 34:301–310, 2001
14. Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH: *Biostatistics in Clinical Medicine*, New York, NY: Macmillan, 1987
15. Hersh WR, Detmer WM, Frisse ME: Information-Retrieval systems. In: H SE and E PL (Eds). *Medical Informatics*. New York, NY: Springer, 2001, pp 539–572
16. Lancaster FW: *Information Retrieval Systems: Characteristics, Testing and Evaluation*, 2nd edition. New York, NY: John Wiley & Sons Inc., 1979
17. Agresti A, Coull BA: Approximate is better than 'exact' for interval estimation of binomial proportions. *Am Stat* 52:119–126, 1998
18. Hripcsak G, Rothschild AS: Agreement, the F-Measure, and reliability in information retrieval. *J Am Med Inform Assoc*. 12:296–298, 2005
19. Goldstein I, Arzumtsyan A, Uzuner O: Three approaches to automatic assignment of ICD-9-CM codes to radiology reports. *AMIA Annu Symp Proc* 11:279–283, 2007
20. Imai T, Aramaki E, Kajino M, Miyo K, Onogi Y, Ohe K: Finding malignant findings from radiological reports using medical attributes and syntactic information. *Stud Health Technol Inform* 129:540–544, 2007
21. Mamlin BW, Heinze DT, McDonald CJ: Automated extraction and normalization of findings from cancer-related free-text radiology reports. *AMIA Annu Symp Proc* 420–424, 2003
22. Farkas R, Szarvas G: Automatic construction of rule-based ICD-9-CM coding systems. *BMC Bioinformatics* 11(9 Suppl 3):S10, 2008