Original Article

Application of Automated Pathology Reporting Concepts to Radiology Reports

Suzanne March, MBA, CMC[a]; George Cernile, BSc, CKE, PMP[b]; Kim West, BS[a]; Diane Borhani, MBA, CMC[a]; April Fritz, BA, RHIT, CTR[c]; Peter Brueckner, MD, FRCP, MBA[b]

Abstract: Objective: This study was designed to extend the concept of automated pathology reporting to radiology reports to find central nervous system (CNS) neoplasms that may currently go undetected. Methods: Existing E-Path software was modified to account for the structure and language of radiology reports. Logic was added to allow registries to configure whether they want only new reports or if they also want history, metastatic, and/or previously known reports. Five hospital registries and 3 central registries participated. Three quality-control (QC) studies were conducted with fine-tuning taking place between the studies. The first QC study included random samples of 1,500 reports from 3 data sources. The second and third QC studies each included 1 random sample from 2 different data sources. Results: The software was able to extract reportable CNS neoplasms with a high degree of specificity and sensitivity at 99% and 100% respectively, using the original set of coding rules. This rule set was favored by our hospital registries. Participating population-based registries preferred to receive only positive-new cases. The specificity and sensitivity for this category was 96% and 94% respectively. One hospital registry compared the cases found by the software to their registry database and found 13 additional CNS neoplasm cases in a 10-month period which represented an increase of 18%. Conclusion: Automated radiology reporting is a promising method of mining a previously untapped data source to find cases of CNS neoplasms that may be missed by conventional techniques.

Key words: *CNS neoplasms, brain tumors, automated cancer reporting, case ascertainment, radiology reports*

Introduction

The key to improving the cancer surveillance system in the United States is the timely collection and integration of patient information from multiple sources within a medical care environment and across multiple institutions.[1] Currently, about 95% of cancers are diagnosed through pathology reports from either hospital-based or independent laboratories.[2] A number of initiatives underway since 2000 have applied natural language processing (NLP) techniques to automate the process of identifying and submitting reportable pathology cases to central and hospital registries. Today, automated pathology reporting is the primary tool for extraction of case information from pathology reports for this purpose.[3,4] Currently, however, there is no similar software broadly available to extract potentially reportable cases from other data sources. The goal of our study was to extend automated pathology reporting to magnetic resonance imaging (MRI) of the brain and computed tomography (CT) scans of the head and neck to provide a cost-effective method of finding reports of CNS neoplasms that may currently go undetected.

We selected CNS neoplasms because the Benign Brain Tumor Cancer Registries Amendment Act (Public Law 107-260) passed October 29, 2002 imposed an unfunded mandate on all US population-based cancer registries to register these tumors.[5] Further, according to the Central Brain Tumor Registry of the United States (CBTRUS), only about 67% of all brain tumors had a histologically confirmed diagnosis, with substantial regional variation (range for participating cancer registries: 55%-97%). Of the non-malignant brain tumors, 56% were histologically confirmed, while 41% were confirmed radiologically.[6] The figures suggest that the historical heavy emphasis on pathology reports for casefinding could result in a substantial number of CNS neoplasms being missed.

Methods

For the study we sought the participation of 3 well-respected population-based (central) registries that would provide geographical diversity across the United States and whose data providers offered a rich source of MRIs of the brain and CT scans of the head and neck. Once a central registry was committed to participating in the project, that registry was commissioned to recruit a data provider within their catchment area. One central registry recruited 2 data providers. In addition, a not-for-profit health care provider participated but there was no participation from the corresponding central registry.

Gold Standard Reference Set

A.Fritz and Associates (AFA), highly respected experts in cancer coding, built a reference set of 10,000 de-identified radiology reports as the "gold standard" against which the software accuracy was measured. Two measurements of accuracy were used. *Sensitivity* quantifies the ability of the software to detect reportable CNS neoplasms, and *specificity*

[a]QuantumMark LLC, Reno, NV. [b]Artificial Intelligence In Medicine Inc., Toronto, Ontario, Canada. [c]A.Fritz and Associates LLC, Reno, NV.

Address correspondence to Suzanne March, MBA, CMC, QuantumMark LLC. Email: suzanne@quantummark.com

Figure 1. An Automated System for Reporting CNS Neoplasms

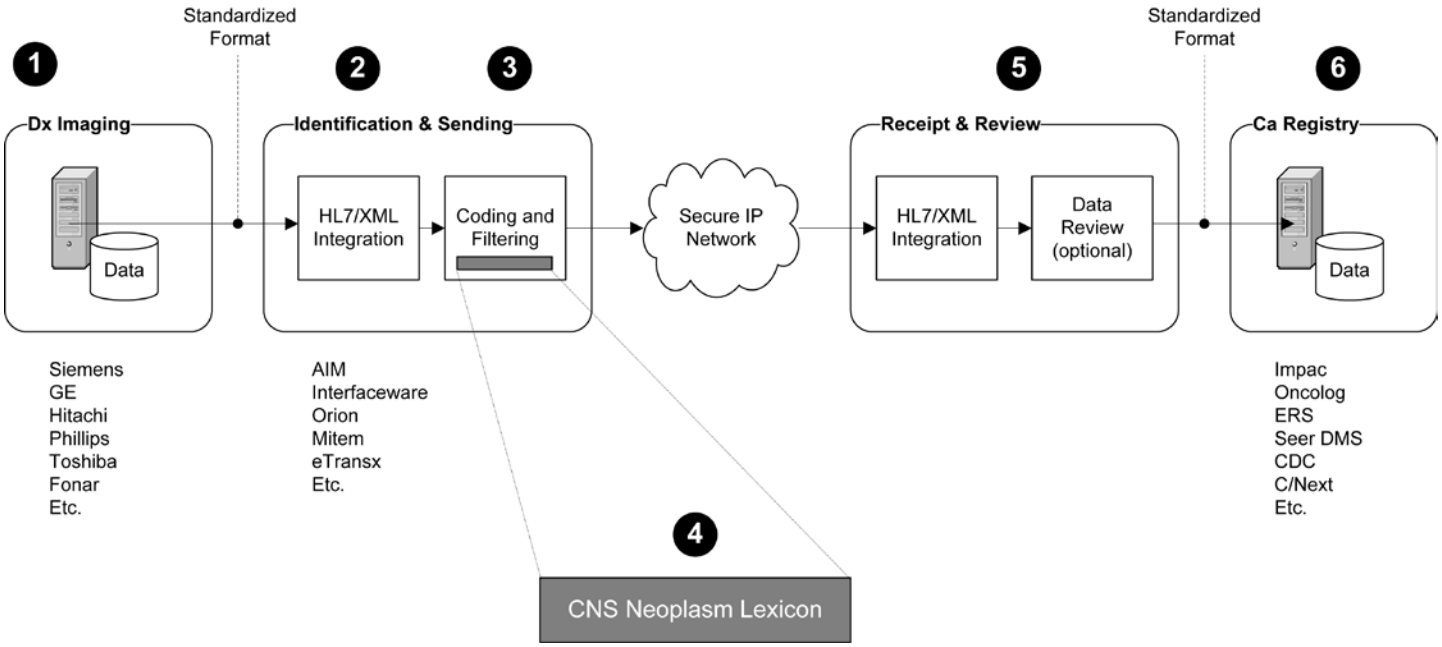

Table 1. Categorization of Selectable Radiology Reports

Code	*Description*	*Definition*	*Example*
0	Negative	No mention of cancer or CNS neoplasm	No significant intracranial abnormality
1	History of cancer	Mention of history of tumor or cancer	History of prostate cancer
			No other neoplasm
2	Metastatic disease	Reference to the presence of metastases	History of renal cancer with lesions on frontal lobe
3	Positive, but previously known	A report of any primary CNS neoplasm with wording to indicate that it was known to be present when the examination was ordered	Resected medulloblastoma
			Post-surgical change
4	Positive, new case	A report of any primary CNS neoplasm or other primary malignant tumor that did not contain reference to history of tumor or metastases	Pituitary gland microadenoma

Table 2. Sensitivity and Specificity Results for 3 Rounds of QC

	First QC study results by site						*Second QC study*		*Third QC study*	
	Central 1		Central 2		Hospital 1		Hospital 2		Central 3	
True positives	140		247		112		120		1083	
True negatives	1304		1109		1256		1286		735	
False positives	44		132		117		11		4	
False negatives	12		12		15		5		0	
Sample total	1500		1500		1500		1422		1822	
	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity
	97%	92%	89%	95%	91%	88%	99%	96%	99%	100%

measures its ability to reject reports that do not describe CNS neoplasms. Ideally, the sensitivity of the system should approach 100% to be sure that no reportable cases are missed. Alternatively, to the extent that the specificity of the system is below 100% reports will not be rejected by the system and will be false positives. Since we expect that all selected reports will be reviewed by a certified tumor registrar (CTR), the false-positive rate represents additional work for that person.

In preparation for developing the reference set, AFA and the development team collaborated to establish the set of coding rules that would be used for the manual review and that the developers would incorporate into the software.

Software Design

The system used in the study was based on E-Path software components from Artificial Intelligence In Medicine, Inc. The processing of diagnostic imaging reports required the development of a lexicon and knowledge base specific to that discipline and extensive heuristic testing followed by adjustments to enable the system to perform usefully. Radiology subject matter experts assisted by elucidating the meaning of terms that are used in reporting diagnostic imaging studies. Modifications were also made to the utilities used for reviewing, annotating and scoring reports.

Figure 1 is a schematic depiction of the automated system for reporting CNS neoplasms. The paragraph numbers below correspond to the labels in the diagram.

1. Diagnostic imaging systems such as MRI and CT scanners were used to image the head and neck to discover lesions that may be neoplasms of the central nervous system. Radiologists interpreted these images and prepared narrative reports describing the findings. The imaging reports were stored in a reporting system or database.
2. The modified ePath software analyzed the text content of the imaging reports to determine if a reportable finding is in the text. This required standardization of the format of the imaging report and an interface to the source data. A standard format allowed the report data to be passed from the integration engine to the coding and filtering component.
3. The coding engine included 2 components: a natural language processing (NLP) engine and a lexicon describing the concepts and vocabulary of a particular domain of interest, in this case radiology. The NLP engine used information in the lexicon to appropriately parse and codify concepts from their natural language expression. Once the narrative text of an imaging report had been converted to machine-readable form, the integration engine could easily discern which imaging reports should be forwarded to a cancer registry. These were transmitted over a secure network to a receiving system.
4. Development of a CNS neoplasm lexicon was the objective of this project. A lexicon is a collection of data that describes the language, concepts, and vocabulary of a particular domain of interest. It also contains coding system(s) to convert concepts to machine-readable values.
5. Imaging reports were received at the registry using an integration/messaging engine, much like the engine used to extract the reports from their original source. However, before passing the data onto the cancer registry system, the software allowed a human at the registry to make the final determination as to whether or not it should be included in the registry database.
6. The cancer registry software/database is the ultimate destination of the imaging reports. Here, the reports would be stored in a "suspense" file to be further incorporated into cancer case abstracts. A standard input format utilizing HL7 and/or XML technologies would be used to achieve compatibility with the different varieties of cancer registry software.

Software Development

During the iterative testing period, the developers compared the results from the software to the reference set built by AFA. Using groups of approximately 2,000 reports at one time, they reviewed every discrepancy between the manual and computer coding. Facilitated meetings were held with the CTRs from AFA to discuss and resolve the discrepancies. In some cases, the CTRs decided to change their coding to better comply with a coding rule, but most of the time the software was modified to improve its accuracy. Once the accuracy was high for a set of 2,000, a new set was selected and a new baseline of sensitivity and specificity was established. After 3 datasets and a total of 6,000 test records processed with 17 test iterations executed, the sensitivity and specificity were 99% and 96% respectively. This was deemed an acceptable level to begin testing at the participating registry sites.

Registry implementations took place in September 2010. Once the system was installed, the registries began receiving files containing the reportable radiology reports selected by the system in accordance with the established selection rules. Registries were asked to review the reports to determine if they agreed with the software and inform the developers of disagreements. They were also asked to record the amount of time they spent reviewing the reports and to compare the positive cases found by the software to their registry database to determine if the tool picked up cases not found by existing registry procedures.

After the registries had used the software for a few weeks, it became apparent that there were significant differences in how registries would eventually use the tool in a production environment, and that 1 universal set of coding rules would not be sufficient. In some cases, the registry wanted only cases that represent new cases of CNS neoplasms. Other registries wanted to receive reports of all neoplasms, whether or not they had been previously diagnosed and regardless of the tissue of origin. To accommodate these differing requirements, an automated classification system was introduced to assign reportable reports into 4 categories as described in Table 1. The negative (not reportable) reports constitute a fifth category.

New logic was added to the software to allow registries to select which categories of reports they wished to receive. To maintain comparability during the study period,

Table 3. Initial QC Study Results from Categorization of Selectable Reports												
	Central Registry 1				*Central Registry 2*				*Hospital Registry 1*			
Round One	**Manual**	**Auto**	**Match**	**% Match**	**Manual**	**Auto**	**Match**	**% Match**	**Manual**	**Auto**	**Match**	**% Match**
History	81	11	9	11%	65	36	17	26%	31	60	1	3%
Metastatic	3	23	3	100%	6	32	6	100%	28	56	22	79%
Positive known	40	121	32	80%	135	281	116	86%	30	84	17	57%
Positive new	28	29	4	14%	53	30	13	25%	38	29	13	34%
Total	152	184			259	379			127	229		

however, all participating registries continued to receive and process all categories. The remaining 4,000 test reports in the reference set were used to perform iterative testing of the new functionality. After the registries had used the revised software for 2 months, we began the quality control phase of the project.

QC Studies

Three sites participated in the first QC study. Each site provided a random set of 1,500 reports from MRIs of the brain and CT scans of the head and neck. A CTR at the site reviewed the reports and categorized them, assigning codes 0-4 in accordance with the classification described in Table 1. Then, after the same reports were processed by the software, a Microsoft Excel spreadsheet was created to enable comparison of the 2 sets of results. Using the CTR coding as the benchmark for correctness, the reports were defined as true positives, true negatives, false positives and false negatives and the sensitivity and specificity were calculated. For this calculation, categories 1-4 were considered positive. Separately, it was determined how successful the software was at distinguishing between the 4 categories of positive.

The developers reviewed every report where the software's result differed from that of the CTR to determine the causes of the discrepancy and also attempted to identify recurring patterns. Facilitated sessions were held to discuss the findings and agree on the solutions, which included a combination of changing the manually applied code to better conform to the selection rules or modifying the software by changing the lexicon, casefinding logic, or which sections of the report were processed by the system. Software modifications were introduced in 6 iterative stages to make it easier to determine if accuracy improved. To determine if the improvements in sensitivity and specificity achieved by modifying the software were general (rather than specific to the data set used to make the adjustments), a second QC study was conducted using new data. More modifications were made to the software and then a third QC study was conducted, again with a new dataset.

Results

Sensitivity and Specificity Results

The number of true and false positives and negatives and the calculated specificity and sensitivity results for the 3 QC studies are documented in Table 2.

Table 3 shows the software's ability to categorize selectable reports into the 4 categories of positive. The column label *Manual* is the CTR's categorization. *Auto* is the software's categorization and *Match* is the number of times the software classification agreed with the manual review.

At each site, the total number of cases categorized using the software exceeded the total number categorized manually. The difference is due to overselection (false positives) of reports by the software. Significant development work was expended to improve the initial match results. After 6 rounds of iterative modification of the software, the average matching results for history of cancer, metastatic disease, positive but previously known and positive new were 77%, 98%, 65%, and 52%, respectively.

To further quantify the software's ability to categorize reports, during the second and third QC studies, each of the 4 positive categories was isolated and the sensitivity and specificity for each category was computed. Results are documented in Table 4.

Case-Ascertainment Results

Hospital 1 provided the most extensive comparison between cases found by the software and cases found by their traditional methods. Table 5 documents the performance over a 10-month period from November 2010 through August 2011. The figure in the rightmost column of the total line is the average new cases found per month.

Discussion

Conclusions regarding the tool's ability to select reportable CNS neoplasms with a minimum of false-positives depend on the way in which the tool will be used in a production environment. It is clear from the study that it is realistic, using the original set of coding rules, to extract reportable CNS cases with a high degree of specificity and sensitivity at 99% and close to 100%, respectively. The coding rules include selecting histories, metastases and positive previously known as well as new cases, a configuration that appealed to our hospital registry participants.

The central registries in the study, on the other hand, were interested only in new cases. By the third QC study, significant improvement was achieved for the positive-new category with specificity at 96% and sensitivity at 94%. For the other categories the modifications continued to produce mixed results. Sensitivity improved for positive, but previously known and metastatic disease specificity decreased. In the third QC study, history of cancer resulted in lower scores for both. Further work is required to accomplish

Table 4. Second and Third QC Results Isolated by Classification (Hospital 2 and Central 3, Respectively)								
	Positive new		*Positive known*		*History of cancer*		*Metastatic disease*	
	QC2	**QC3**	**QC2**	**QC3**	**QC2**	**QC3**	**QC2**	**QC3**
True Positive	22	291	8	273	46	184	16	101
True Negative	1382	1452	1390	1353	1352	1460	1400	1678
False Positive	4	59	10	67	21	120	4	37
False Negative	14	20	14	129	3	58	2	6
Sample Total	1422	1822	1422	1822	1422	1822	1422	1822
Specificity	100%	96%	99%	95%	98%	92%	100%	98%
Sensitivity	61%	94%	36%	68%	94%	76%	89%	94%

Table 5. Hospital 1 Comparison of Cases Found By Method						
Month	*Found by software*	*Found traditionally*	*Missed by software*	*Additional cases found by software*	*Net new cases*	*Percent increase in cases found*
November 2010	7	5	0	2	2	40
December 2012	7	6	0	1	1	17
January 2012	6	4	0	2	2	50
February 2012	9	7	0	2	2	29
March 2012	9	8	0	1	1	13
April 2012	8	6	0	2	2	33
May 2012	9	9	0	0	0	0
June 2012	17	14	0	3	3	21
July 2012	3	4	1	0	-1	-25
August 2012	5	5	0	0	0	0
Totals	80	68	1	13	12	18

consistently high sensitivity and specificity for individual categories in isolation. However, we believe the most typical configurations will be either selection of positive new only or selection of all categories. The high degree of accuracy for these configurations suggests that the software can be a useful tool for registries to find CNS neoplasms.

With regard to the tool's ability to discover CNS cases that might go undetected by current methods, the results were very promising. The 10-month study conducted by Hospital 1 found 13 new CNS cases not found by their traditional methods of using E-Path and monthly hospital discharge lists. One case that was found traditionally was not found by the software, for a net increase of 12 additional cases. Overall, this represented an 18% improvement in CNS case ascertainment. While it is important to be cautious about generalizing the Hospital 1 experience, preliminary data is becoming available from other study participants. For example, after reviewing 91 reports selected by the software during the test period, Registry 1, a central registry, identified 20 reports (22%) they expected to see in their registry database but did not. The timing of the review was approximately 5 months after the end of 2011 and reports were still being received. A second review will be necessary to determine if they would have ultimately been found.

Conclusion and Future Possibilities

Automated reporting of CNS neoplasms from radiology reports is a reliable method of mining a previously untapped data source. As was the case for the automated pathology reporting software, increased use of the tool will provide the experience necessary to further refine the lexicon and continuously improve the software's performance.

Further research is required to determine if the use of radiology reports for casefinding can be extended to other cancers. During the study, we explored extending the software to pancreas and biliary tract radiology reports because, like CNS, we suspected that advanced cases could be missed if no follow up pathology report was ordered. However, because the radiology reports included many other organs, from colon, bowels, bladder, spleen, heart, lungs, liver, etc, the linkage between the diagnosis and the affected organ was difficult to determine programmatically.

The Radiological Society of North America (RSNA) set a goal in 2005 to collaborate with radiology professional organizations to develop a comprehensive set of anatomic and pathologic terms called RadLex.[7] The first version of the *RadLex Playbook* was recently released on November 1, 2011.[8] Since the 2007 American College of Radiology Intersociety

Conference, work has been underway to use structured radiology reports to improve communication of radiology procedures employing both consistently ordered sections and standardized language.[9] For a variety of reasons, adoption of standardized radiology reports has been slow.[10,11] But as the obstacles to standardization are overcome, we believe automated radiology reporting can be successfully applied to cancers beyond CNS.

Acknowledgements

This project was funded in part with federal funds from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services under Contract No. HHSN261200900040C. We would especially like to acknowledge our hospital and central registries for their willingness to support our project with domain expertise, technical support, data review and analysis. We know you had many competing demands on your time and your commitment to our project was greatly appreciated.

References

1. U.S. Department of Health and Human Services. Solicitation for SBIR Contract Proposals, Topic 233 Development of Software Systems to Facilitate the Use of Electronic Data Records in the Collection of Population-Based Cancer Surveillance Data, PHS. 2007;1:35.
2. Dale D, Golabek JK, Chong N. The Impact of E-path Technology on Ontario Cancer Registry Operations. J Registry Manage. 2002;29(2):52-56.
3. Center for Disease Control and Prevention and the National Program of Cancer Registries. NPCR-AERRO ePath Reporting Activities. Available at: http://www.cdc.gov/cancer/npcr/informatics/aerro/activities/epath.htm. Accessed May 15, 2012.
4. Artificial Intelligence In Medicine Inc. Cancer Surveillance Solution. Available at: http://www.aim.on.ca/cancer-surveillance.php. Accessed May 15, 2012.
5. Center for Disease Control and Prevention and the National Program of Cancer Registries. Appendix B: Benign Brain Tumor Cancer Registries Amendment Act. Available at: http://www.cdc.gov/cancer/npcr/pdf/btr/Amendment_Act.pdf. Accessed May 15, 2012.
6. Central Brain Tumor Registry of the United States. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004-2008. 2012:11.
7. Radiological Society of North America. About RadLex. Available at: http://www.rsna.org/RadLex.aspx. Accessed May 15, 2012.
8. Brooks L, Morley M. Initial Release of RSNA's RadLex Playbook Finds Quick Adoption. RSNA Press Release. November 28, 2011.
9. Kahn C Jr, Langlotz C, Burnside E, et al. Toward Best Practices in Radiology Reporting. Radiology. 2009;252(3):852-856.
10. Weiss D, Langlotz C. Structured Reporting: Patient Care Enhancement or Productivity Nightmare? Radiology. 2008;249(3):739-747.
11. Langlotz C. Structured Radiology Reporting: Are We There Yet? Radiology. 2009;253(1):23-25.