Identifying Fractures in BioSense Radiology Reports

Achintya N. Dey, M.A., Haobo Ma, M.D., M.S., Armen Asatryan, M.D., M.P.H., Roseanne English, B.S., and Jerome Tokars, M.D., M.P.H.

Centers for Disease Control and Prevention, National Center for Health Informatics, Division of Emergency Preparedness

OBJECTIVE

The purposes of this study are to validate a keyword-based text parsing algorithm for identifying fractures and compare radiology results with chief complaint and ICD-9 final diagnoses.

BACKGROUND

BioSense is a bio-surveillance system designed and operated by the Centers for Disease Control and Prevention (CDC) to provide information helpful in dealing with public health emergencies. The BioSense system receives near real-time chief complaint and/or diagnosis data from >360 hospitals, and some hospitals send additional data such as text radiology reports, microbiological laboratory reports, and pharmeutical receipt.

METHODS

We studied 17,647 visits with an extremity film and final ICD-9 coded diagnosis from 13 hospitals during March 1st, 2006 through December 31st, 2006. To identify fractures from text radiology reports, we designed a keyword-based text parsing program using SAS software. We used negation words like "absence," "fails to reveal," "negative," "healed," "lack of" etc. to eliminate readings indicating no fracture; and used words such as "fracture" and "fx" to find fractures. To validate this algorithm, we used it to identify 100 randomly selected radiology reports with fracture and 100 without fracture. To validate the text parsing method, each of these 200 radiology reports were reviewed by a human reader. Next, the 200 radiology reports were compared with chief complaints and discharge ICD-9 diagnoses and agreement assessed using the Kappa statistic. ICD-9 diagnosis codes in the range 800-829 indicated fracture.

RESULTS

Among 100 records with fracture identified by text parsing, human review found evidence of fracture in 91, no fracture in 2, and uncertain results in 7. Among 100 records with no fracture identified by text parsing, human review showed fracture in 2 and no fracture in 98. Thus the text parsing algorithm had 0.98 sensitivity and 0.98 specificity compared with the gold standard human review. Next we assessed chief complaint data. Among 100 records with fracture found by text parsing, only 12 had a chief complain report of fracture; among 100 with no fracture found by text parsing, none had chief complain report of fracture (Kappa=0.12). Next, we assessed ICD-9 coded diagnosis. Among 100 records with fracture found by text parsing, 85 had a diagnosis of fracture; among 100 with no fracture found by text parsing, 8 had a diagnosis of fracture (Kappa=0.77). Finally we applied the text parsing algorithm to all 17,647 visits with a text radiography reading. Of these, 73% were emergency department patients, 19% were inpatients, and 8% were outpatients; and 3,891 (22%) had radiographs with fractures. The five most common sites of fracture were hand (12%), wrist (11%), ankle (10%), foot (10%), and finger (9%).

CONCLUSIONS

We developed a simple text parsing algorithm that can identify fractures from X-ray reports with acceptable accuracy. The method has potential usefulness for rapid identification of patients with fractures during public health emergencies. Radiology reports showed poor agreement with chief complaints but good agreement with discharge ICD-9 codes; disagreements may have occurred because of inaccuracies in the text parsing algorithm, missing radiology or diagnosis data, or problems with accurate linkage of radiology reports with final diagnoses in our electronic system. Also, earlier studies have shown disagreements between radiographic interpretation by clinicians (whose diagnoses are reflected in ICD-9 codes) vs. radiologists (whose readings are reviewed by our system). One reason for this may be the subtle findings of fracture were missed by clinicians but reported by the radiologist. We are currently assessing errors made by the text parsing method in an attempt to improve it. We also plan to assess natural language processing systems that may improve accuracy and extract additional information such as findings other than fracture and degree of certainty.