Biosurveillance Evaluation of SNOMED CT’s Terminology (BEST Trial): Coverage of Chief Complaints

Peter L. ELKIN, MD1, Steven H. BROWN, MD2, Andrew BALAS, MD1, PhD, Zelalem TEMESGEN, MD2, Dietlind WAHNER-ROEDLER, MD1, David FROEHLING, MD1, Mark LIEBOW, MD1, Brett TRUSKO, PhD1, S. Trent ROSENBLOOM, MD2, Greg POLAND, MD1

1Mayo Clinic College of Medicine
2Vanderbilt University
3Old Dominion University

Abstract. The current United States Health Information Technology Standards Panel’s interoperability specification for biosurveillance relies heavily on chief complaint data for tracking rates of cases compatible with a case definition for diseases of interest (e.g. Avian Flu). We looked at SNOMED CT to determine how well this large general medical ontology could represent data held in chief complaints. In this experiment we took 50,000 records (Comprehensive Examinations or Limited Examinations from primary care areas at the Mayo Clinic) from December 2003 through February 2005 (Influenza Season). Of these records, 36,097 had non-null Chief Complaints. We randomly selected 1,035 non-null Chief Complaints and two Board-certified internists (one Infectious Diseases specialist and one general internist) reviewed the mappings of the 1,035 chief complaints. Where the reviewers disagreed, a third internist adjudicated. SNOMED CT had a sensitivity of 98.7% for matching clinical terms found in the chief complaint section of the clinical record. The positive predictive value was 97.4%, the negative predictive value was 89.5%, the specificity was 81.0%, the positive likelihood ratio was 5.181 and the negative likelihood ratio was 0.016. We conclude that SNOMED CT and natural language parsing engines can well represent the clinical content of chief complaint fields. Future research should focus on how well the information contained in the chief complaints can be relied upon to provide the basis of a national strategy for biosurveillance. The authors recommend that efforts be made to examine the entire clinical record to determine the level of improvement in the accuracy of biosurveillance that can be achieved if we were to incorporate the entire clinical record into our biosurveillance strategy.

Keywords. Concept representation, Epidemiological Research, Standards

Acknowledgements. This work has been funded in part by grants from the National Library of Medicine LM06918 and the Centers for Disease Control and Prevention PH000022 and HK00014. The authors would like to thank Luka Bajzer for his technical support of this research.
Introduction

As bioterrorism has become a real and increasing threat to our society, early detection holds the promise of limiting the damage resulting from such an attack. Bioterrorism represents a significant threat to the public health of nations. In warfare it is said that if you kill one adversary, you take one person from the battlefield; but if you make one adversary ill, you take many people from the battlefield, as it takes additional human resources to protect, relocate, and care for the infected individual, even without counting the effects on others that might be secondarily infected. Given the potential of this strategy to further the goals of our enemies, coupled with the minimal infrastructure needed to deliver biologic weapons, the potential for this type of attack represents a serious threat to a country’s national security.

One important part of any national defense strategy is to have early detection for outbreaks of illness consistent with exposure to a biologic weapon. An alternate approach, termed electronic biosurveillance, creates a secondary use for information in computer databases that is collected primarily for clinical care and other health-related purposes. In a common approach referred to as syndromic surveillance, data are assigned (“binned”) to syndrome categories (e.g., respiratory or gastrointestinal). The data are then analyzed to provide estimates of disease burden by geographic location over time that can be used for early event detection and health situational awareness during an emergency.

BioSense, the U.S. CDC’s system for electronic biosurveillance, was initiated in 2003. Data are received from national healthcare organizations including the Department of Defense (DoD), Veterans’ Affairs (VA), and others. The ICD-9 diagnosis codes and CPT procedure codes from these sources are binned into 11 syndromes. Data analyses are displayed in the BioSense Application, which can be viewed through the CDC secure data network by CDC, state and local health departments, and DoD and VA personnel. The BioIntelligence Center (BIC) at CDC consists of personnel who monitor the BioSense application daily for data anomalies, assist state and local health department personnel in data interpretation, and contribute to improving the BioSense application.

In mid-2005, CDC began to develop the capacity to receive real-time data from hospital facilities across the United States. This initiative strengthens BioSense by emphasizing access to real-time clinically rich data. Individual patient records are stripped of obvious identifiers such as names, addresses and telephone numbers at the data source before being sent to CDC. Patient populations include those treated in emergency department, hospital inpatient, and hospital-affiliated outpatient settings. Hospital-level data includes a daily hospital census. Patient-level data of interest includes patient chief complaint (text or Coded Element with no coding system specified), physician diagnosis[1] (ICD9-CM codes), selected ED-specific data (e.g., vital signs) (Numbers or as an Equation), microbiology laboratory orders (coded test names, using local coding system) and results (LOINC coded tests, SNOMED for Organisms and unspecified coding scheme for other results), radiology orders (local codes only) and results (Coding system to be determined), and pharmacy orders (Drug names are free text; note that in the data model they site RxNorm as one of the standards but it is not actually implemented at present). As of March 2006, chief
complaint, diagnosis, and selected demographic data were being received from hospitals distributed across the United States.[2]

The Health Information Technology Standards Panel (HITSP)[3] formed a technical committee on biosurveillance in 2006 whose name was later changed to the Public Health Technical Committee. In 2006, the work program produced an interoperability specification focused on the needs of biosurveillance. The specification followed was aimed at satisfying a use case for biosurveillance provided by the American Health Information Community (AHIC) led by Secretary Leavitt.[4] Our work was further informed by a set of data elements thought important to the use case by AHIC and its subgroup charged to develop recommendations regarding biosurveillance data elements.

Given this advice in September of 2006 a first version of the specification was published. Under the current timetable the specification will see some minor revisions and is scheduled to be recognized by Secretary Leavitt by November 2007. Once recognized by the Secretary, companies who wish to do business in this area with the federal government must conform to the specification.

The Certification Commission for Health Information Technology (CCHIT) creates certification criteria for electronic health records. The Interoperability Specifications developed within HITSP are considered input to the certification process and cycle. These standards based specifications are intended to drive interoperability among and between electronic medical record and other Health IT systems.

The Biosurveillance Interoperability Specification relies heavily on chief complaint data for carrying out biosurveillance.[5] As much of the chief complaint data nationally is not recorded in a structured and codified fashion, we chose to investigate how well the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) covered the concepts in chief complaints found in Mayo Clinic health records.

Method

SNOMED-CT is a large scale clinical nomenclature, which is description logic, based and is produced by the College of American Pathologists. The terminology has greater than 370,000 concepts and 1,000,000 terms. In our lab we add another 790,000 terms as synonyms to make the terminology more clinically friendly (no new concepts are added).

In this experiment we selected 50,000 history and physical records or limited exams from Adult Medicine or Pediatric outpatient visits. All visits occurred in December through February (our typical influenza season). Records were collected in an most recent to oldest fashion starting in 2005 and extending back to 2003 to complete the cohort of 50,000 records. From these source clinical records 13,903 had null or blank chief complaint sections of the record, leaving 36,097 records with data in the chief complaint field. From these a random sample of 1035 records was selected for review.

All records were processed using the Multi-threaded Clinical Vocabulary Server (MCVS) developed at the Mayo Clinic. The server uses natural language processing to assign free text elements to a controlled representation. Previous research has shown the mappings to be of high accuracy with a sensitivity (recall) of 99.7%, a positive
predictive value (precision) of 99.8% and a specificity of 97.9%. Two independent reviewers (a general internist and an Infectious Diseases sub-specialist) reviewed each item. Where the reviewers disagreed a third individual (general internist) adjudicated.

The method for the reviewer judgments (see Figure 1) is outlined as:

1. Is the Chief Complaint term a reasonable term? (Yes--TP or FN; No--TN or FP)
   A. Yes. Does SNOMED CT match the term exactly?
      i. Yes: Code a TP for SNOMED CT’s ability to code the term
         ii. No: Look up the SNOMED CT term in a SNOMED CT Browser (MCVS Browser).
            a. A Match is identified: Code a TP for the SNOMED CT.
            b. Does not match: Code a FN for the SNOMED CT term.
   B. No, it is not a reasonable term: Did the MCVS find something in SNOMED CT which maps to this nonsensical term?
      i. Yes: Code a FP for SNOMED CT.
      ii. No: Code a TN for SNOMED CT.

| CE: Leg ulcers in both the lower extremities. | - Ulcer of lower extremity (disorder) [95344007] [K]
| - [has Finding Site]
| . Both lower extremities (body structure) [4180000] [M] |

Figure 1: The review tool demonstrating a complete match using a compositional expression.

**Final Model of just CC Section To Predict Case Status of Influenza**

<table>
<thead>
<tr>
<th>Symptom' (In order of best predictor)</th>
<th>Multivariate Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) GOLD STANDARD 2=Cough</td>
<td>7.81 (4.12, 14.8) [&lt;=.001]</td>
</tr>
<tr>
<td>2) GOLD STANDARD 3=Fever</td>
<td>3.90 (2.22, 6.84) [&lt;=.001]</td>
</tr>
<tr>
<td>3) GOLD STANDARD 9=Infiltrates</td>
<td>0.12 (0.00, 0.74) [0.018]</td>
</tr>
<tr>
<td>4) GOLD STANDARD 8=Sore Throat</td>
<td>2.51 (0.77, 8.16) [0.126]</td>
</tr>
<tr>
<td>C-statistic</td>
<td>0.714</td>
</tr>
</tbody>
</table>

'Contribution to regression is mutually exclusive to that of the “best” symptom (and next best symptom, etc.)

Table 1: Final Optimum Case Definition for surveilling Influenza from Chief Complaints
Results:

![ROC Curve Final Model for Just CC Section](image)

Figure 2: ROC curve representing the various models of Influenza identification from within Chief Complaints.

Overall, SNOMED CT was found to have excellent coverage of the knowledge found in clinical chief complaints. The sensitivity (recall) was 98.7% and the positive predictive value (precision) was 97.4% (see Table 1).

<table>
<thead>
<tr>
<th></th>
<th>CE</th>
<th>CID</th>
<th>Partial</th>
<th>Nothing</th>
<th>Total</th>
<th>Total: FP Fixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>tp</td>
<td>452</td>
<td>120</td>
<td>288</td>
<td>0</td>
<td>860</td>
<td>897</td>
</tr>
<tr>
<td>tn</td>
<td>16</td>
<td>1</td>
<td>71</td>
<td>6</td>
<td>94</td>
<td>102</td>
</tr>
<tr>
<td>fp</td>
<td>11</td>
<td>0</td>
<td>56</td>
<td>2</td>
<td>69</td>
<td>24</td>
</tr>
<tr>
<td>fn</td>
<td>3</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Sum</td>
<td>482</td>
<td>121</td>
<td>424</td>
<td>8</td>
<td>1035</td>
<td>1035</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.993407</td>
<td>1</td>
<td>0.969697</td>
<td>Undefined</td>
<td>0.986239</td>
<td>0.98679868</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.592593</td>
<td>1</td>
<td>0.559055</td>
<td>0.75</td>
<td>0.576687</td>
<td>0.80952381</td>
</tr>
<tr>
<td>PPV</td>
<td>0.976242</td>
<td>1</td>
<td>0.837209</td>
<td>0</td>
<td>0.925727</td>
<td>0.973941368</td>
</tr>
<tr>
<td>NPV</td>
<td>0.842105</td>
<td>1</td>
<td>0.8875</td>
<td>1</td>
<td>0.886792</td>
<td>0.894736842</td>
</tr>
<tr>
<td>Pos LR</td>
<td>2.438362</td>
<td>Undefined</td>
<td>2.199134</td>
<td>Undefined</td>
<td>2.32981</td>
<td>5.180693069</td>
</tr>
<tr>
<td>Neg LR</td>
<td>0.011126</td>
<td>0</td>
<td>0.054204</td>
<td>Undefined</td>
<td>0.023863</td>
<td>0.016307513</td>
</tr>
</tbody>
</table>

Table 2: Results of SNOMED CT mappings of the clinical content coverage of Chief Complaints. Judgments as to the degree of match with the input phrase are generated by MCVS (CID = Concept ID, CE = Compositional Expression). tp = true positive, tn = true negative, fp = false positive, fn = false negative, ppv = positive predictive value and POS LR = positive likelihood ratio.)
Some common systematic errors, which were corrected after the first set of reviews, included eliminating proper names from the abstraction and one point of erroneous synonymy (“Pt” mapped to “Pregnancy Detection” instead of “Patient”). The numbers in bold reflect the data after these two sources of error were corrected and the data under total reflect the numbers prior to making these corrections.

Discussion:

Bioterrorism and epidemic infectious diseases such as H5N1 Avian Influenza are a serious and ever present threat to our nation’s public health. In this study, SNOMED CT has demonstrated good coverage for the clinical thoughts / concepts used in chief complaints. It remains unproven as to whether or not the information contained in chief complaints is adequate for the purposes of biosurveillance. A Receiver Operator Characteristics Curve was generated from all the potential models of Influenza surveillance from the chief complaint sections of the record (see Figure 2). A c-Statistic of 0.714 which was obtained from the ROC curve is better than current biosurveillance models for Influenza but not consistent with other common tests used in medical practice such as Well’s Criteria for Deep Pulmonary Emboli or an exercise stress test which each have c-Statistics around 0.800. This gap raises the question as to whether whole record surveillance is superior to surveillance using chief complaints alone. More research is needed to establish an evidence-based direction in terms of biosurveillance data for our nation. The authors specifically recommend the support of research looking at how data from the entire clinical record can best be utilized to maximize our ability to perform effective biosurveillance.

References: