ICD9-CM Claims Data are Insufficient for Influenza Surveillance

Peter L. Elkin[1], **Steven H Brown**[2]

[1]Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, SUNY.

[2]Department of Biomedical Informatics, Vanderbilt Universit. Department of Veteran's Afairs.

Abstract

Background: Influenza and Influenza like illness are representative of a class of epidemic infectious diseases that have important public health implications. Early detection via biosurveillance can speed lifesaving public heath responses. In the United States, biosurveillance is typically conducted using ICD9 coded visit diagnoses and uncoded chief complaint data.

Objective: To determine the accuracy of ICD9 diagnoses using laboratory confirmed cases as the gold standard.

Design: A six-year retrospective cohort study.

Setting: A tertiary referral center.

Patients: All 3,825 patients with an ICD9-CM diagnosis of Influenza and all 1455 patients with laboratory confirmed Influenza.

Results: Of the 3,828 patients assigned ICD9-CM visit codes indicating a diagnosis of Influenza, 2,825 were not confirmed by laboratory testing and 1,003 patients under went laboratory testing. Only 664 (66.2%) tested positive for Influenza. Of the 1,455 patients who tested positive for Influenza 45.6% were identified by ICD9-CM code.

Conclusion: ICD9-CM had a low 66.2% Positive Predictive Value (precision) for Influenza and a low 45.6% Sensitivity (recall) for Influenza in patients tested for Influenza. ICD9 coded visit diagnoses/claims data are insufficient alone to serve as the basis for Influenza Surveillance.

Primary Funding Source: CDC grants PH00022 and HK00014.

Keywords

Biosurveillance; Influenza; Public Health; Health Informatics; ICD9-CM

Contact information: Peter L. Elkin, Steven H Brown. elkinp@buffalo.edu.

Introduction

Influenza is a health threat which presents in yearly outbreaks, resulting in between 250,000 and 500,000 dealths a year out of approximately three to five million cases. Pandemics (larger outbreaks) are less frequent. [1] Some forms of Influenza are more virulent such as H1N1. Many agents of Bioterrorism and natural epidemics such as H1N1 and Avian Influenza present as Influenza-like illnesses. The CDC defines an influenza-like illness as fever plus either cough and/or sore throat in the absence of a known cause other than influenza. [2] In a systematic review of the literature on the clinical diagnosis of human influenza, no sign or symptom was useful in differentiating between influenza and a similar influenza-like illness, for example an Anthrax attack, when all age groups were included. [3] This can be an important consideration for the planning of a public health response. For patients over 60 years of age the combination of fever, cough, and the acute onset of symptoms had a likelihood ratio of 5:4 for the diagnosis of influenza versus a similar influenza-like illness.

In the United States the Biosense system uses ICD9-CM codes and the data from chief complaint fields to perform daily biosurveillance. Information received can be reported to the CDC's emergency operations center and can then trigger a public health response. In Europe, researchers have also advocated ontology based biosurveillance for use in surveilling antibiotic resistance in the Artemus project led by Lovis et al. [4]

In a prior study, we showed that whole record biosurveillance using SNOMED CT and natural language processing (NLP) was superior to chief complaint surveillance alone for Influenza surveillance. [5] In that study NLP using SNOMED CT had a sensitivity of 92.9%. It was also previously reported by Elkin et al that only 8% of patients diagnosed with Influenza undergo laboratory testing. We also showed that SNOMED CT had good content coverage of the information found in chief complaint fields. [6]

The safety of the public health is a global and ubiquitous responsibility. Detecting an influenza epidemic early can trigger a public health response and save lives. For example, timely vaccination can prevent further cases and antiviral treatment can reduce mortality by as much as 36%. [67] Influenza biosurveillance is an important part of global health strategy. Biosurveillance techniques should be optimized and continuously improved by using the most appropriate and accurate methods available. In this study we determine the accuracy of ICD9 encoded visit diagnoses, one of the pillars of current U.S. influenza biosurveillance strategy, using laboratory confirmed cases as the gold standard

Methods

This six year retrospective cohort study was approved by the Institutional Review Board (IRB# 1836–05). Cases came primarily from Minnesota, Wisconsin, North Dakota, South Dakota, and Iowa. No hospital or outpatient locations were excluded from the study (e.g. ED, Inpatient, Outpatient, etc.).

3,828 eligible patients were all patients who were diagnosed with Influenza by ICD9-CM coding (ICD9-CM codes 487 and its sub-codes 487.XX) between October 2000 and March

2006. Patients are consented once for use of their clinical record data and are approached to ensure continued consent as they present for each annual visit. Only patients who have not denied consent for the use of their clinical record were included in this study.

These patients were compared by medical record number with a database of patients laboratory tested for Influenza. All patients were either tested by viral culture or by polymerase chain reaction (PCR). Testing was for both Influenza A and for Influenza B. There were 1,455 patients who were tested and tested positive for Influenza in the total population during the same time period.

The data was analyzed to determine the positive predictive value and sensitivity of ICD9-CM codes for the Identification of true Influenza cases

Results

20,068 patients were enrolled in the study and of these 17,243 patients were tested for Influenza. In our population of patients seen from October 2000 and March 2006, 1,455 patients tested positive for Influenza. Of these cases, 664 were diagnosed by ICD9-CM code.

Of the population of 3,828 patients that had ICD9-CM diagnoses of Influenza, 1,003 were laboratory tested for Influenza. Of the 1,003 patients laboratory tested for influenza, 664 tested positive and 339 tested negative (Figure 1).

The analysis of the data showed that ICD9-CM had a sensitivity (recall) of 45.6% and a positive predictive value (precision) of 66.2% for the identification and retrospective surveillance of Influenza (Table 1). The Negative predictive value and the Specificity for ICD9 were good at 95.1% and 97.9% respectively.

Limitations

This data was from one large academic medical institution (Mayo Clinic) and in that way may not generalize. Respiratory Syncytial Virus testing was done very uncommonly and may account for some portion of the iCD9-CM codes for Influenza that tested negative.

Discussion

ICD9-CM encoded visit diagnoses yielded suboptimal results in their ability to identify true cases of Influenza. From our data, ICD9-CM encoded visit diagnoses alone should not be used for the biosurveillance of Influenza or Influenza like illness. Likewise, biosurveillance relying on laboratory results alone would also have low sensitivity (though high specificity) - we have previously found that only 8% of patients who are diagnosed with Influenza actually have confirmatory laboratory testing.

Biosurveillance is an activity that needs to favor Sensitivity and ICD9 has been shown in this experiment to be insensitive. Clearly, more complex models are needed.

In Georgia, the current U.S. influenza biosurveillance using chief complaint data and ICD9 coded visit data has a sensitivity of 19.1% and a positive predictive value of 73.7%. [8] Methods such as whole record surveillance using SNOMED CT have yielded better results with sensitivities of 92.9% in prior publications. In a previous study we demonstrated that SNOMED CT had good coverage, 98.7% sensitivity for the clinical content of chief complaints. [6]

Influenza remains an important global public health problem and accurate surveillance coupled with a prompt public health response can decrease both morbidity and mortality from this disease.

Influenza like illness may represent exposure to agents of bioterrorism such as inhalational anthrax and can be an important indicator of a population's exposure to such agents, especially if they occur in the non-Influenza season.

From our data, ICD9-CM alone should not be used for the biosurveillance of Influenza or Influenza like illness. As only 8% of patients are laboratory tested for Influenza who are diagnosed with Influenza, laboratory surveillance alone although specific would be quite insensitive.

The formal Ontological understanding of the clinical record stands as our best single tool in the fight to contain emerging infectious diseases. Biosurveillance leading to detection and an early public health response to Influenza can decrease mortality and morbidity of the Influenza season. [5] NLP based Ontological surveillance using SNOMED CT has also been successfully used for fully automated electronic quality monitoring (eQuality). [9–11]

Further research is needed to identify the optimal combination of Ontological and direct coding system evidence to predict Influenza and other emerging infectious diseases.

Conclusion

ICD9-CM had a low sensitivity (recall) and positive predictive value (precision) for the identification of true cases of Influenza in a population of patients tested for Influenza. However the negative predictive value and specificity of ICD9-CM was strong. ICD9-CM should not be used alone for the surveillance of Influenza.
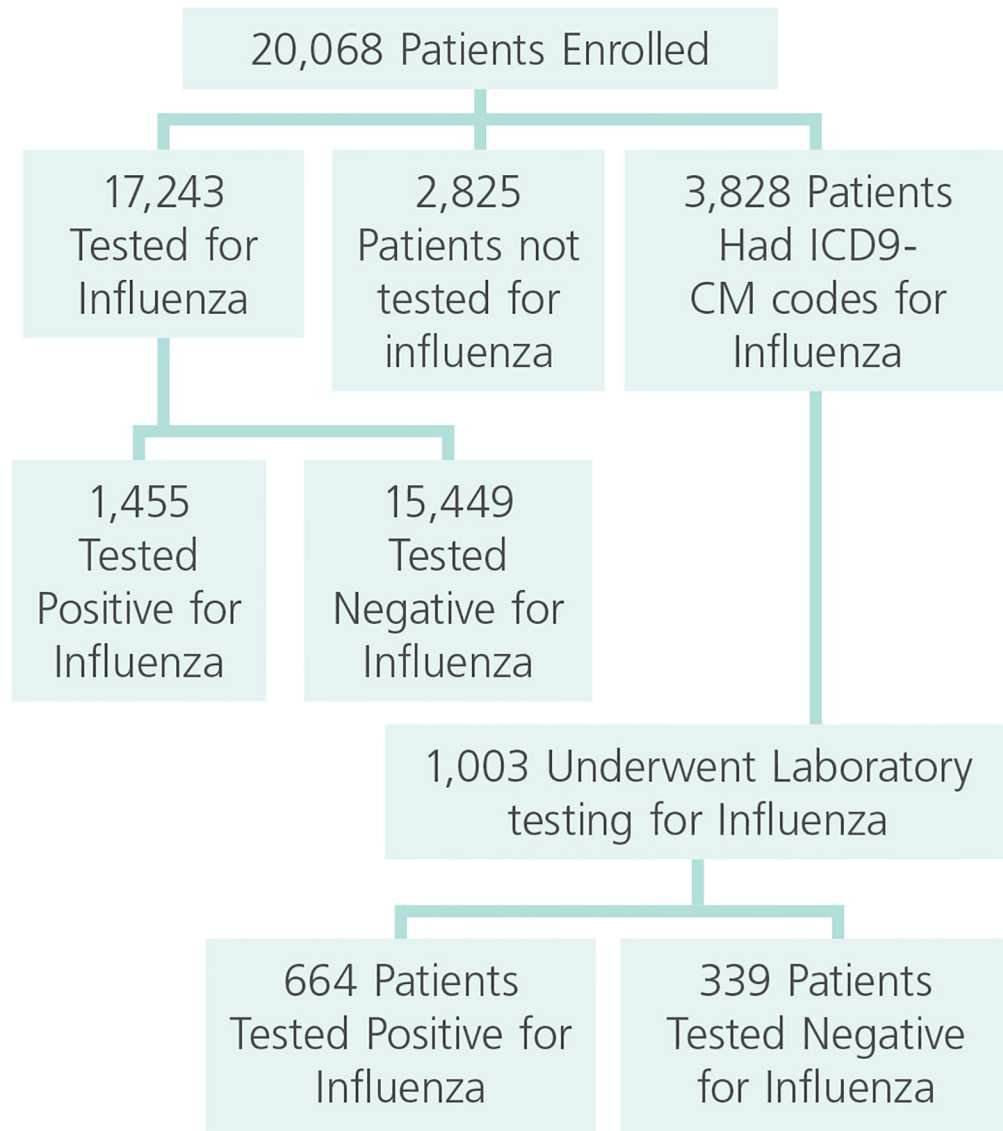
Acknowledgements

References

1. "Influenza (Seasonal) Fact sheet N°211". who. int. 3 2014 Retrieved 25 November 2014.

2. US Centers for Disease Control and Prevention. CDC Seasonal Influenza (Flu) - Flu Activity and Surveillance. US Centers for Disease Control; 2011.

3. Call SA, Vollenweider MA, Hornung CA, Simel DL, McKinney WP. Does this patient have influenza? JAMA. 2005;293(8):987–97. [PubMed: 15728170]

4. Teodoro D, Pasche E, Gobeill J, Emonet S, Ruch P, Lovis C. Building a transnational biosurveillance network using semantic web technologies: requirements, design, and preliminary evaluation. J Med Internet Res. 2012 5 29;14(3): e73. [PubMed: 22642960]

5. Elkin PL, Froehling D, Wahner-Roedler D, Brown SH, Bailey K. "Comparison of NLP Biosurveillance Methods for Identifying Influenza from Encounter Notes"; Ann Intern Med. 2012 1 3;156(1 Pt 1):11–8. [PubMed: 22213490]

6. Elkin PL, Brown SH, Balas A, Temesgen Z, Wahner-Roedler DL, Froehling D, Liebow M, Trusko B, Rosenbloom ST, Poland G. Biosurveillance evaluation of SNOMED CT's terminology (BEST Trial): coverage of chief complaints. Stud Health Technol Inform. 2008; 136:797–802. [PubMed: 18487829]

7. Adisasmito W, Chan PK, Lee N, Oner AF, Gasimov V, Aghayev F, Zaman M, Bamgboye E, Dogan N, Coker R, Starzyk K, Dreyer NA, Toovey S. Effectiveness of antiviral treatment in human influenza A(H5N1) infections: analysis of a Global Patient Registry. J Infect Dis. 2010;202(8):1154–60. [PubMed: 20831384]

8. http://www.thci.org/syndromic/conference2009/abstractsubmissions/ Whitney_Evaluation_of_Syndromic_Surveillance_Influenza-Like_Illness.doc

9. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, Dittus RS, Rosen AK, Elkin PL, Brown SH, Speroff T. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA. 2011 8 24;306(8):848–55. [PubMed: 21862746]

10. Brown SH, Elkin PL, Fielstein E, Speroff T. eQuality for all – extending automated quality measurement from free text. AMIA Annu Symp Proc. 2008 11 6:71–5.

11. Brown SH, Speroff T, Fielstein EM, Bauer BA, Wahner-Roedler DL, Greevy R, Elkin PL. eQuality: Automatic assessment from narrative clinical reports. Mayo Clin Proc 2006 81(11):1472–1481 [PubMed: 17120403]

```
                        ┌──────────────────────────┐
                        │  20,068 Patients Enrolled │
                        └──────────────────────────┘
```

20,068 Patients Enrolled

17,243 Tested for Influenza

2,825 Patients not tested for influenza

3,828 Patients Had ICD9-CM codes for Influenza

1,455 Tested Positive for Influenza

15,449 Tested Negative for Influenza

1,003 Underwent Laboratory testing for Influenza

664 Patients Tested Positive for Influenza

339 Patients Tested Negative for Influenza

Figure 1:
Enrollment and results flowchart.

Table 1.

Two by Two table of ICD9s performance as a surveillance tool for Influenza.

	PCR Positive Influenza	**PCR Negative Influenza**	
ICD9 Positive for Influenza	664	339	PPV 0.662
ICD9 Negative for Influenza	791	15,449	NPV 0.951
Total	Sensitivity 0.456	Specificity 0.979	