

# Automated influenza case detection for public health surveillance and clinical diagnosis using dynamic influenza prevalence method

Fuchiang Tsui<sup>1,2</sup>, Ye Ye<sup>1,2</sup>, Victor Ruiz<sup>1</sup>, Gregory F. Cooper<sup>1,2</sup>, Michael M. Wagner<sup>1,2</sup>

<sup>1</sup>Real-time Outbreak and Disease Surveillance Laboratory (RODS), Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA

<sup>2</sup>Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA, USA

Address correspondence to Fuchiang Tsui, E-mail: tsui2@pitt.edu

## ABSTRACT

**Objectives** To assess the performance of a Bayesian case detector (BCD) for influenza surveillance and clinical diagnosis.

**Methods** BCD uses a Bayesian network classifier to compute the posterior probability of a patient having influenza based on 31 findings from narrative clinical notes. To assess the potential for disease surveillance, we calculated area under the receiver operating characteristic curve (AUC) to indicate BCD's ability to differentiate between influenza and non-influenza encounters in emergency department settings. To assess the potential for clinical diagnosis, we measured AUC for diagnosing influenza cases among encounters having influenza-like illnesses. We also evaluated the performance of BCD using dynamically estimated influenza prevalence, and measured sensitivity, specificity and positive predictive value.

**Results** For influenza surveillance, BCD differentiated between influenza and non-influenza encounters well with an AUC of 0.90 and 0.97 with dynamic influenza prevalence ( $P < 0.0001$ ). For clinical diagnosis, the addition of dynamic influenza prevalence to BCD significantly improved AUC from 0.63 to 0.85 to distinguish influenza from other causes of influenza-like illness.

**Conclusions and policy implications** BCD can serve as an influenza surveillance and a differential diagnosis tool via our dynamic prevalence approach. It enhances the communication between public health and clinical practice.

**Keywords** Bayesian case detector, dynamic prior, emergency department notes, influenza, influenza-like illness

## Introduction

Automated case detection applies criteria defined in a case definition or a disease model to identify the existence of a single individual with a disease, which finds its principle applications in disease surveillance and clinical decision support. In disease surveillance, the application detects an individual with a disease. In clinical decision support, it may aid diagnosis.

Availability of data is a prerequisite for automated case detection in both disease surveillance and clinical decision support; therefore, the range of applications of automated case detection is shaped by data availability. Research on automated case detection occurs in three main data-based areas.

The first area exploits chief complaints, i.e. brief textual descriptions characterizing the reason why a person has sought medical attention, recorded by clinicians or

registration personnel.<sup>1–4</sup> Sensitivity and specificity of case detection based on chief complaints are poor, except when the nosological definitions are broad enough to match the precision with which patients tend to describe their reason for encounter, e.g. 'respiratory illness'. When analyzed temporally and spatially, the counts of cases in these broad categories are helpful in tracking the relative level of disease activities during epidemics and may provide early detection of selected types of outbreaks, such as an outbreak of cryptosporidiosis.<sup>5</sup>

**Fuchiang Tsui**, Associate Professor

**Ye Ye**, Doctoral Student

**Victor Ruiz**, Doctoral Student

**Gregory F. Cooper**, Professor

**Michael M. Wagner**, Professor

The second area exploits data from electronic laboratory systems.<sup>6–12</sup> For example, Public Health England's (PHE's) Data Mart system monitors the weekly count of isolates of respiratory pathogens from hospitals or specialist laboratories when additional typing is required.<sup>13</sup> There are two issues with laboratory tests that can impact surveillance accuracy. First, laboratory tests may have different sensitivities and specificities.<sup>14</sup> Those tests with low sensitivity and high specificity could lead to an underestimation of disease burden. Second issue is the sampling incompleteness. Due to limited resources, definitive laboratory testing is not always performed for individual encounters.

The first two areas are the basis for many deployed public health disease surveillance systems; but typically fall into two extreme ends of the accuracy spectrum for case detection. Systems using chief complaints usually have large population coverage, but can only identify broad categories (e.g. nosological symptoms) with poor diagnostic precision.<sup>15</sup> Conversely, electronic laboratory systems are able to detect specific diseases (e.g. salmonella) with greater diagnostic precision, but the detection is restricted to the tested population. BioSense<sup>16</sup> developed by the Centers for Disease Control and Prevention and the Electronic Clinical Laboratory Reporting System (ECLRS) developed by the New York State Department of Health,<sup>9</sup> are examples of the two ends, respectively.

The third area, which is the focus of this paper, exploits unstructured data in clinical notes, which are free-text notes authored by treating clinicians in order to document patient encounters. A key advantage of the clinical notes is that it is rich in diagnostic information. Its main limitation is that ~80% of clinical information is locked within the free-text, which must be analyzed by natural language processing (NLP).<sup>17</sup> An NLP parser can extract whether a patient had or did not have a set of targeted clinical findings based on what a clinician has written in a clinical note.

In contrast to automated case detection, there are several manual case reporting systems. PHE uses sentinel physician reporting of influenza-like illness (ILI) cases to monitor influenza activity.<sup>13</sup> The World Health Organization's Global Influenza Program uses case reporting to monitor virological and epidemiological influenza trends around the world.<sup>18</sup>

Greaves<sup>19</sup> summarized that an ideal surveillance system for management of emerging infection would be accurate, timely, electronic, use agreed on case definition(s), run continuously for long periods, be able to detect local outbreaks or clusters of cases, and easily used to follow up on cases and contacts.

Previous research on automated case detection of influenza from clinical notes date back to studies by Elkin *et al.*<sup>20</sup> and Tsui *et al.*<sup>21–24</sup> Elkin applied logistic regression modeling for influenza detection from emergency department (ED)

encounter notes processed by the Multithreaded Clinical Vocabulary Server system. Tsui developed an automated Bayesian case detector (BCD) that leverages free-text ED notes to provide an estimated probability that a patient has influenza given a set of NLP-extracted clinical findings. BCD uses a Bayesian network (BN) model to represent a diagnostic medical knowledgebase. An inference engine in BCD assigns finding values extracted from an NLP parser to a BN model and applies Eq. (1) to compute the posterior probability of a disease given the disease's prior probability.

$$P(D|f_1, f_2, \dots, f_n) = \frac{P(f_1, f_2, \dots, f_n|D)P(D)}{P(f_1, f_2, \dots, f_n|D)P(D) + P(f_1, f_2, \dots, f_n|\bar{D})P(\bar{D})} \quad (1)$$

where, the set  $\{f_1, f_2, \dots, f_n\}$  represents symptoms and signs extracted from notes of each ED encounter,  $P(D)$  is a disease prior probability (or prevalence), and  $(P(D) + P(\bar{D}) = 1)$ .

We previously built two Bayesian classifiers for influenza case detection.<sup>23</sup> A BN-expert classifier, defined by two board-certified physicians, comprised 31 clinical findings. This classifier had a near-Naive BN structure, assuming conditional independence among findings given the disease, except for the lab-confirmed influenza finding, which depended on both disease and a nasal swab test order finding. The conditional probability tables (CPTs) were elicited to represent physicians' subjective probabilities that express the strength of dependence between findings and disease.

The second classifier, a BN-EM classifier, had the same finding set and structure as the BN-expert classifier. To incorporate both the experts' subjective assessments and collected data, we estimated the CPT of BN-EM classifier by updating the CPT of the BN-expert classifier through the Expectation-Maximization maximum a posteriori algorithm applied to a large training dataset (468 influenza and 29 004 non-influenza encounters). Both two classifiers used a constant influenza prior ( $P(D) = 0.1$ ).

An initial study of BCD was conducted for influenza detection using Topaz<sup>22,25,26</sup>, an NLP tool developed at the University of Pittsburgh, and a high differentiation between influenza cases and non-influenza controls drawn from a low-influenza summer period was found (area under the ROC curve (AUC) for BN-expert classifier: 0.96; AUC for BN-EM classifier: 0.97).<sup>22</sup> This study did not provide a good indication of the performance of BCD for influenza detection that we would expect in an operational ED setting, since it did not sample non-influenza controls over the course of the entire year in the test dataset. Moreover, it did not evaluate whether BCD's differentiation ability varies

among different influenza priors, age groups and among patients with ILI (e.g. patients having fever and a symptom of cough or sore throat).<sup>27</sup>

The current study serves as an extension to our previous study by expanding the analyses of BCD performance in more realistic ED settings for public health surveillance among different age groups, and clinical differential diagnosis between influenza and non-influenza ILI encounters. We further compared case detection using ED notes with case detection using chief complaints, and explored the use of dynamic influenza priors (influenza priors changed over time). We believe BCD is the first decision support system that has the potential to support both public health surveillance and clinical differential diagnosis. One use case for BCD's public health surveillance application would be to provide influenza surveillance through reporting individual posterior influenza probabilities,  $P(\text{influenza}|\text{data of patient}_i)$  and likelihoods,  $P(\text{data of patient}_i|\text{influenza})$ , to a regional disease surveillance agency or an outbreak detection system. One use case for BCD's clinical diagnosis application would be to function like a rapid influenza test to support the decision of test-versus-treat recommendations for individual patients.

## Methods

We obtained ED encounter records from four hospitals at the University of Pittsburgh Medical Center (UPMC) from 2008 to 2011. All the data were received through real-time Health Level 7 (HL-7)<sup>28</sup> communication.

### Training dataset

The training dataset for the BN-EM classifier contained 468 PCR-confirmed cases of influenza (*PCR+*), which were recorded between January 2008 and August 2010. The training dataset also contained 29 004 non-influenza controls, which were recorded between July 2010 and August 2010, excluding PCR-test positives. A training case consisted of Topaz-extracted influenza findings from an ED note and the binary diagnosis of *PCR+* or not *PCR+*.

### Test dataset

We created a test dataset comprised of 176 *PCR+* influenza cases and 1 620 non-influenza controls between September 2010 and December 2011. To better reflect the general ED population, controls were uniformly, randomly sampled from all ED encounters with no positive PCR results, and the sampling period included influenza seasons.

## Study design

In this study, we evaluated the BCD with two core components—Topaz and BN-EM classifier—for both disease surveillance and clinical decision support. Appendix A describes details of the classifier model.

We measured BCD's ability to detect *PCR+* influenza cases in the general ED population, as well as its detection ability in each of the three age groups: children (ages 0–17), younger adults (ages 18–64) and older adults (65 years and older). To benchmark BCD's performance against a predominant approach to automated case detection in public health surveillance, we compared detection abilities and time latencies of using ED notes (from clinicians captured in a dictation system) with using chief complaints (from triage nurses captured in a registration system). The time latency was indicated by the time delay between the two data types (ED notes and chief complaints); the time delay was defined as the time difference between the receipt time of the chief complaint of an ED encounter ( $t_0$ ) and the receipt time of the earliest ED note associated with the encounter ( $t_1$ ), i.e. time delay ( $\Delta t$ ) =  $t_1 - t_0$ .

To assess BCD performance on a clinically relevant differentiation task for clinical decision support, we evaluated BCD's differentiation abilities to identify influenza cases among patients with ILI symptoms. We used clinical findings extracted by Topaz to determine whether a patient has ILI or not. Our previous study showed that Topaz extracts fever (accuracy 0.91) and cough (1.0) with high accuracy, but exhibits poor accuracy for sore throat (0.52).<sup>23</sup> Among the 176 *PCR+* influenza cases between September 2010 and December 2011, there were 110 cases having ILI. Using these influenza cases and randomly selected non-influenza controls, we formed four test datasets with different influenza prevalence in ILI population: 10% (dataset A: 110 influenza cases with ILI and 990 non-influenza ILI controls), 20% (dataset B: 110 cases and 440 controls), 30% (dataset C: 110 cases and 256 controls) and 50% (dataset D: 110 cases and 110 controls).

Previous BCDs used a fixed influenza prior of 0.1 in BNs (i.e.  $P(D) = 0.1$  and  $P(\bar{D}) = 0.9$  in Eq. (1)). To reflect the changes of influenza prevalence over the course of a year (i.e.  $P(D)$  and  $P(\bar{D})$  changed over time), we developed a simple dynamic (influenza) prior,  $P_d(\text{influenza})$ , and measured its differentiation abilities. We estimated the influenza prevalence on the encounter registration date by using the proportion of PCR-confirmed influenza cases among all ED encounters in the previous week (last 7 days), as shown by Eq. (2). If there were no confirmed influenza cases in the previous week,  $P_d(\text{influenza})$  would be set to 0.

$$P_d(\text{influenza}) = \frac{\text{number of PCR-confirmed influenza cases in the previous week}}{\text{number of all ED encounters in the previous week}} \quad (2)$$

It is common for clinicians to record a positive influenza test result in an ED note. To compare BCD's differentiation ability with laboratory tests' diagnostic ability, we removed encounters of which ED notes mentioned a lab order or a confirmed lab result. We then calculated five sets of sensitivity, specificity and positive predictive values (PPV) by changing the probability threshold for classifying an ED encounter with ILI symptoms into an influenza or a non-influenza encounter.

We employed two standard metrics for estimating BCD's differentiation ability: the AUC and the Brier Skill Score (BSS).<sup>29</sup> AUC measures differentiation ability of a test for correct classification of a disease (0.5 corresponding to a random guess and 1 corresponding to a perfect differentiation). BSS measures the difference between the score for the (BCD) forecast and the score for the unskilled standard forecast, e.g. persistently guessing a patient has or does not have a disease, normalized by the disease prior of the dataset. BSS ranges from negative infinity to 1. A BSS of 1 represents a perfect forecast whereas a BSS of 0 or less represents an unskilled forecast.

## Results

This section presents the evaluation results of BCD applied to influenza surveillance and clinical decision support.

### Potential of BCD for disease surveillance

We evaluated BCD performance in influenza surveillance for all ages and for individual age groups (Table 1).

#### BCD's differentiation ability

Using 176 PCR-confirmed influenza cases and 1 620 randomly selected non-influenza controls as the test dataset, the AUC of BCD with BN-expert classifier and BN-EM classifier were 0.82 (95% CI: 0.79–0.86) and 0.90 (95% CI: 0.87–0.93), respectively. These results indicate the ability of BCD to detect influenza in general ED patients.

With dynamic priors, AUCs of the BN-expert and BN-EM classifiers statistically significantly ( $P < 0.0001$ ) increased to 0.95 and 0.97, respectively. Figure S1 (available online) shows the ROCs. Figure S2 (available online) shows the adjusted ROCs by removing those weeks without

laboratory confirmed influenza cases, i.e.  $P_d(\text{influenza}) = 0$ , demonstrating that the BCD using BN-EM classifier with the input of ED notes and dynamic priors remained the best performance in influenza season (AUC: 0.93, 95% CI: 0.90–0.95), compared with BCDs with other configurations.

#### BCD performance for three age groups

Table 1 summarizes the two Bayesian classifiers' performance in overall population and individual age-group populations. Both of the two classifiers had significantly lower performance in the older adults group compared with the performance in the other groups.

Table S1 and Figures S3–S5 (available online) show the frequency of influenza-related findings (extracted from ED notes by Topaz) in different age groups in the test datasets. Compared with adults (Figures S4 and S5), infant and children were more likely to have signs and symptoms, such as fever, chill, ILI and anorexia, and have nasal swab tests (Figure S3). Compared with younger adults (Figure S4), older adults (65+) were less symptomatic: they were less likely to have fever, cough, headache and sore throat (Figure S5). Moreover, this group had an almost equal likelihood of not having fever, headache, fatigue, sore throat and wheezing in both influenza cases and non-influenza controls. Therefore, it would be very difficult to use the absence of clinical findings to rule out influenza. This helps explain why the BCD did not have a sound performance for these older patients (AUC = 0.63).

#### BCD performance when using chief complaints from triage nurses

With the 1 796 encounters as a test dataset, we found that classifiers using chief complaints from triage nurses had poor differentiation ability (AUC = 0.62, 95% CI: 0.58–0.66), which was significantly lower than classifiers using ED notes (AUC = 0.90) ( $P < 0.0001$ ). Among these 1 796 encounters, 1 340 encounters (101 PCR+ influenza cases and 1 239 non-influenza controls) did not have any influenza-related findings extracted from their chief complaints.

Figures S1 and S2 show the ROCs of BN-EM and BN-expert classifiers using chief complaints. With a dynamic prior, the differentiation ability of BCD increased from 0.62 to 0.95 (95% CI: 0.93–0.96) (Figure S1). After we removed those weeks without laboratory confirmed influenza cases, the differentiation ability of the BCD with the input of chief complaints and dynamic priors dropped from 0.95 to 0.83 (95% CI: 0.79–0.86) (Figure S2), which was still significantly higher than the performance of BCD relying on a constant prior and chief complaints (AUC = 0.62).

Based on the data received from UPMC through real-time HL-7 communication, the median time for receipt of a chief



**Table 1** Performance of BCD using two classifiers and two types of priors for disease surveillance in all and difference age groups.

Classifier	Overall AUC (95% CI), BSS [P(influenza) = 0.1]	Overall AUC (95% CI), BSS [dynamic influenza prior]	Age-based AUC (95% CI), BSS [P(influenza) = 0.1]		
			Infants and children [0–17 years old] n = 87 encounters	Adults [18–64 years old] n = 1376 encounters	Older adults [65 years and older] n = 333 encounters
BN-expert	0.82 (0.79–0.86), –0.422	0.95 (0.94–0.97), –0.422	0.87 (0.78–0.96), –2.660	0.86 (0.82–0.89), –0.276	0.59 (0.46–0.72), –0.664
BN-EM	0.90 (0.87–0.93), –0.043	0.97 (0.96–0.98), –0.043	0.94 (0.88–0.99), –1.495	0.94 (0.92–0.96), 0.087	0.63 (0.49–0.78), –0.415

complaint was 36 seconds while the median time for receipt of an initial ED note was 6.44 h. The median time delay between the receipt of a chief complaint and receipt of an earliest ED note was 6.41 h. There was no statistically significant difference between time delays for influenza encounters (median of time delays: 6.01 h) and non-influenza encounters (median of time delays: 6.46 h). The median length of ED stays was 2.85 h in the test dataset. An ED note was normally available after the patient had been discharged.

### Potential of BCD for clinical decision support

We evaluated BCD performance using the BN-EM classifier for clinical decision support using a constant influenza prior and dynamic influenza priors. Table 2 lists the AUCs of BCD's ability to distinguish influenza from other causes of ILI.

When we used a constant influenza prior that assumed influenza prevalence to be 0.1 for the entire test period, BCD had poor differentiation ability to detect influenza among ILI encounters (AUC ranged from 0.56 to 0.60).

With dynamic priors, BCD's differentiation ability was greatly increased (AUC ranged from 0.86 to 0.88) (Table 2 and Figure S6). For BCD using dynamic priors, we further calculated five sets of sensitivity, specificity and PPV for each of the four datasets (Table 2) after removing encounters of which ED notes mentioned lab order or confirmed lab results. Sensitivities and specificities varied slightly in the test datasets with different prevalence, but PPV increased as the influenza prevalence increased. When more than 30% ILI visits were influenza cases, PPV reached above 0.6.

## Discussion

### Main finding of this study

Leveraging routinely collected electronic health records (EHRs), BCD demonstrated potential for enhancing the communication between public health and clinical practice. For public health surveillance, BCD automatically captured influenza cases from general ED encounters to facilitate potential outbreak detection. The dynamic prior approach

can significantly improve conventional chief-complaint based influenza surveillance.

For clinical decision support, BCD further enhanced disease differential diagnosis using dynamic priors (population disease preference). It has potential to serve as a rapid test like a decision support tool for test-versus-treat recommendations for individual patients.

For influenza surveillance, BCD performed better using narrative ED notes than using chief complaints.

### What is already known on this topic

Current public health surveillance systems rely on manual or automated case detection. Sentinel physician reporting of ILI is one example of manual surveillance. There are three areas of automated case detection: syndromic surveillance, electronic lab reporting and state-of-the-art free-text based disease surveillance.

### What this study adds

This study introduced and evaluated BCD for influenza surveillance and clinical decision support. BCD has potential to improve disease surveillance. Compared with sentinel physician reporting, BCD significantly reduces the manual reporting burden and improves timeliness from weekly to (near) real-time. Compared with syndromic surveillance using ED chief complaints, BCD captures cases with a higher accuracy by using NLP to extract more clinical findings from ED notes. BCD reached a good differentiation ability (AUC between 0.9 and 0.95) to identify influenza from general ED encounters. With the randomly selected controls across 18 months, we expect to see similar performance in an operational ED setting. Compared with electronic laboratory reporting that has a small selective population, BCD has a much larger population coverage (all hospital or ED encounters), and the coverage are unlikely to be impacted by laboratory testing policies that may change over the course of a year.

BCD is an automated case detection system. It employs a probabilistic case definition (a BN) to represent patterns of

**Table 2** Performance of BCD using BN-EM classifier to detect influenza among ILI encounters

Test datasets (prevalence, number of influenza cases, number of non-influenza ILI controls)	Constant prior <sup>a</sup>	Dynamic prior <sup>b</sup>			
	AUC (95% CI)	AUC (95% CI)	(Sensitivity, specificity, positive predictive value) <sup>c</sup>		
A: 10%, 110, 990	0.58 (0.53–0.63)	0.86 (0.84–0.89)	(0.30,0.95,0.37), (0.60,0.85,0.28), (0.70,0.81,0.27), (0.80,0.79,0.27), (0.90,0.76,0.27)		
B: 20%, 110, 440	0.60 (0.55–0.65)	0.88 (0.85–0.91)	(0.34,0.95,0.60), (0.60,0.87,0.51), (0.70,0.83,0.48), (0.80,0.82,0.50), (0.90,0.79,0.49)		
C: 30%, 110, 256	0.56 (0.50–0.62)	0.88 (0.84–0.91)	(0.32,0.95,0.71), (0.60,0.85,0.61), (0.70,0.83,0.61), (0.80,0.82,0.62), (0.90,0.78,0.60)		
D: 50%, 110, 110	0.56 (0.48–0.63)	0.86 (0.80–0.91)	(0.24,0.95,0.81), (0.60,0.85,0.77), (0.70,0.81,0.76), (0.80,0.79,0.77), (0.90,0.75,0.76)		

<sup>a</sup>When using a constant prior in Bayesian network classifier, we assumed that influenza prevalence was 0.1 over the course of a year.

<sup>b</sup>When using a dynamic prior in Bayesian network classifier, we used the proportion of PCR-confirmed influenza cases among all ED encounters in previous week as the estimation of daily influenza prevalence.

<sup>c</sup>To compare performance of BCD with clinical diagnostic tests, we removed encounters of which ED notes mentioned lab order or confirmed lab results. We then calculated five sets of sensitivity, specificity and positive predictive values.

a disease's symptoms and/or signs, which could be quickly captured by either using machine learning algorithms or experts' clinical knowledge. BCD can further update a probabilistic case definition when more cases become available and when disease prevalence changes over time. Moreover, BCD can serve as an automated, bi-directional communication channel between public health agencies (or regional outbreak detection systems) and clinical practices, i.e. automated case reporting to public health agencies and automated delivery of disease prevalence to clinical practices. Similar to the conventional case definition, BCD's probabilistic case definition is portable across regions without the concern of patients' privacy. Additionally, the probabilistic case definition takes disease prevalence into account, which may vary across different regions.

Elkin *et al.*<sup>20</sup> compared prediction performance between narrative ED notes and chief complaints (available from a section in ED notes as a surrogate of triage chief complaints) and found an AUC increase from 0.65 to 0.76. We did a similar comparison and found that the AUC significantly increased from 0.62 to 0.95, and we used chief complaints directly from triage nurses (instead of ED notes), which are commonly collected by current syndromic surveillance systems. Moreover, we measured the timeliness of ED notes (6.44 h from an ED registration). The results indicated that ED notes could be more timely for outbreak detection, compared with lab confirmed reports in daily and weekly ILI sentinel clinician reporting.

We demonstrated the performance of influenza case detection increased after the use of dynamic priors, taking seasonal changes into account. The daily prior was estimated with the influenza prevalence in the previous week among ED encounters. These priors can be readily calculated from

routinely collected EHRs and electronic laboratory reporting required by EHR Meaningful Use, which is part of the Affordable Care Act.<sup>30</sup>

With dynamic priors, the AUC of BCD using chief complaints was significantly increased from 0.62 to 0.83. Nationwide chief-complaint based influenza surveillance systems may benefit from our dynamic prior approach. Case detection performance could be largely boosted by incorporating region-specific daily influenza prevalence, which may be estimated based on the proportion of laboratory-confirmed encounters in healthcare facilities that cover the vast majority of the population residing in the region.

BCD has potential to function as a rapid test when integrated with a clinical notes system. The sensitivity and specificity pairs (sensitivities: 60–90% and specificities: 75–87%) using dynamic influenza priors indicate that BCD has comparable differentiation ability to a rapid test. Furthermore, this study found that when influenza prevalence among ILI encounters increased, the PPV increased, which indicates that BCD can be used for clinical diagnosis during influenza season to reduce false positives. Lee *et al.*<sup>31</sup> showed that point-of-care tests (costing U.S. \$22) with sensitivity of 0.25 and specificity of 0.75 were economically dominant (cost less and were more effective than doing nothing) for patients aged 65–85 years with ILI for seasonal influenza when prevalence of influenza is 0.2 or 0.3 among ILI patients. Since BCD already achieved a sensitivity of 0.32 or more and a specificity of 0.95 when prevalence of influenza is 0.2 or 0.3 among ILI patients (in all age groups), BCD can potentially be more economically dominant than doing nothing for senior patients even if we assume BCD costs \$22 per estimation, which would be far less in practice. In the future, we can maximize the usability of the tool by embedding BCD

into a dictation or report writing system such as PowerNote in Cerner<sup>®</sup> to provide real-time decision support for clinicians regarding test-or-treat recommendations prior to discharge.

### Limitations of this study

This study did not show a viable change to current public health practices. We did not study the workflow for public health users to use BCD. In the future, a comprehensive evaluation should be conducted by first presenting BCD to users (e.g. public health officials) and conduct surveys and usability testing in laboratory setting. After integrating BCD into the workflow of public health agencies, we may administer post-deployments interviews and surveys to assess real-world usage of the system.

We did not systematically compute economic costs, which requires additional simulations.<sup>31</sup> Potential future work includes employing the Monte Carlo decision analytic computer simulation models<sup>31</sup> to conduct a comprehensive economic comparison with point of care tests in both seasonal and pandemic influenza scenarios for different patient ages and different risks of hospitalization and mortality.

Another future direction is to explore integrating BCD with hospital EHR systems and public health surveillance systems. Close integration with an EHR system will allow BCD to collect additional data types such as radiology notes, medications, and laboratory results. Doing so would enable BCD to add additional (e.g. respiratory) diseases such as Respiratory Syncytial Virus to a predictive model for better differential diagnosis based on additional clinical findings. Enhanced communication with a public health surveillance system will likely increase BCD performance by using updated population disease prevalence, in addition to BCD's routine disease reporting to the surveillance system.

### Conclusion

BCD has potential to serve as a tool for both public health surveillance and clinical differential diagnosis. It has several advantages—diagnostic ability, timeliness, employing a BN model as a case definition, (near) real-time processing of EHRs and facilitating outbreak detection. With dynamic priors, BCD can further improve its performance in case detection and differential diagnosis. BCD has potential to serve as a communication channel between public health agencies and clinical practice. We recommend the Meaningful Use<sup>32</sup> to include ED notes to improve case detection accuracy.

### Acknowledgments

We would like to thank our honest broker, Hoah-Der Su. We thank reviewers of the Journal of Public Health for their insightful suggestions and comments.

### Funding

This research was funded by Grant R01LM011370 from the U.S. National Library of Medicine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. National Library of Medicine or the U.S. National Institute of Health.

### Conflicts of interest

None.

### Contributors

FT and YY drafted the article. FT, YY and VR analyzed data. GFC and MMW provided insights and revised the article.

### Supplementary data

Supplementary data are available at the *Journal of Public Health* online.

### References

- Ivanov O, Gesteland PH, Hogan W. Detection of pediatric respiratory and gastrointestinal outbreaks from free-text chief complaints. *AMIA Annu Symp Proc*. 2003; 318–322. Date accessed: 4/12/2017. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1480317/>.
- Wagner MM, Espino J, Tsui F *et al*. Syndrome and outbreak detection using chief-complaint data—experience of the Real-Time Outbreak and Disease Surveillance project. *Morb Mortal Wkly Rep* 2004;28–31.
- Chapman WW, Dowling JN, Wagner MM. Classification of emergency department chief complaints into 7 syndromes: a retrospective analysis of 527,228 patients. *Ann Emerg Med* 2005;46(5):445–55.
- Tsui F-C, Espino JU, Dato VM *et al*. Technical description of RODS: a real-time public health surveillance system. *J Am Med Inform Assoc* 2003;10(5):399–408.
- Berger M, Shiau R, Weintraub JM. Review of syndromic surveillance: implications for waterborne disease detection. *J Epidemiol Community Health* 2006;60(6):543–50.
- Panackal AA, M'ikanatha NM, Tsui F-C *et al*. Automatic electronic laboratory-based reporting of notifiable infectious diseases at a large health system. *Emerg Infect Dis* 2002;8(7):685–91.
- Overhage JM, Grannis S, McDonald CJ. A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. *Am J Public Health* 2008;98(2):344–50.
- Effler P, Ching-Lee M, Bogard A *et al*. Statewide system of electronic notifiable disease reporting from clinical laboratories: comparing automated reporting with conventional methods. *J Am Med Assoc* 1999;282(19):1845–50.

- 9 Chang H-G, Noonan-Toly C, Chen J-H, Backenson B. Early detection of possible outbreaks from electronic laboratory reports. *Online J Public Health Inform* 2014;**6**(1):e131.
- 10 Vogel J, Brown JS, Land T *et al*. MDPHnet: secure, distributed sharing of electronic health record data for public health surveillance, evaluation, and planning. *Am J Public Health* 2014;**104**(12):2265–70.
- 11 Klompas M, McVetta J, Lazarus R *et al*. Integrating clinical practice and public health surveillance using electronic medical record systems. *Am J Prev Med* 2012;**42**(6):S154–62.
- 12 CDC. *Electronic Laboratory Reporting*. Date accessed: 6/5/2017. <http://www.cdc.gov/chrmeaningfuluse/elr.html> (5 June 2017, date last accessed).
- 13 Public Health England. *Sources of UK Flu Data: Influenza Surveillance in the UK (Part of: Seasonal Influenza: Guidance, Data and Analysis and Children's Health)*. First published: 1/7/2014). Date accessed: 4/30/2017. <https://www.gov.uk/guidance/sources-of-uk-flu-data-influenza-surveillance-in-the-uk>.
- 14 WHO Global Influenza Programme. *A Manual for Estimating Disease Burden Associated With Seasonal Influenza*. Date accessed: 4/30/2017. [http://www.who.int/influenza/resources/publications/manual\\_burden\\_of\\_disease/en/](http://www.who.int/influenza/resources/publications/manual_burden_of_disease/en/).
- 15 Wagner MM, Moore AW, Aryel RM. *Handbook of Biosurveillance*. Academic Press, Burlington, MA, USA, 2011.
- 16 Bradley CA, Rolka H, Walker D *et al*. BioSense: implementation of a national early event detection and situational awareness system. *MMWR Morb Mortal Wkly Rep* 2005;**54**(Suppl):11–9.
- 17 Kho AN, Rasmussen LV, Connolly JJ *et al*. Practical challenges in integrating genomic data into the electronic health record. *Genet Med* 2013;**15**(10):772–8.
- 18 WHO. *Influenza Surveillance and Monitoring*. Date accessed: 4/30/2017. [http://www.who.int/influenza/surveillance\\_monitoring/en/](http://www.who.int/influenza/surveillance_monitoring/en/).
- 19 Greaves F. What are the most appropriate methods of surveillance for monitoring an emerging respiratory infection such as SARS? *J Public Health (Bangkok)* 2004;**26**(3):288–92.
- 20 Elkin PL, Froehling DA, Wahner-Roedler DL *et al*. Comparison of natural language processing biosurveillance methods for identifying influenza from encounter notes. *Ann Intern Med* 2012;**156**(1\_Part\_1):11–8.
- 21 Tsui F, Espino J, Sriburadej T *et al*. Building an automated Bayesian case detection system. *Paper Presented at: Conference of the International Society for Disease Surveillance*. ISDS, Park city, Utah, 2010.
- 22 Tsui F, Wagner M, Cooper G *et al*. Probabilistic case detection for disease surveillance using data in electronic medical records. *Online J Public Health Inform* 2011;**3**(3):1–17.
- 23 Ye Y, Tsui F, Wagner M *et al*. Influenza detection from emergency department reports using natural language processing and Bayesian network classifiers. *J Am Med Inform Assoc* 2014;**21**(5):815–23.
- 24 Pineda AL, Ye Y, Visweswaran S *et al*. Comparison of machine learning classifiers for influenza detection from emergency department free-text reports. *J Biomed Inform* 2015;**58**:60–9.
- 25 Harkema H, Dowling JN, Thornblade T *et al*. ConText: an algorithm for determining negation, experienter, and temporal status from clinical reports. *J Biomed Inform* 2009;**42**(5):839–51.
- 26 Chu D. Clinical feature extraction from emergency department reports for biosurveillance. *Master's Thesis*. Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh 2007.
- 27 CDC. *Overview of Influenza Surveillance in the United States*. 2016. Date accessed: 6/5/2017. <https://www.cdc.gov/flu/weekly/overview.htm>.
- 28 HL7. *Health Level Seven International*. Date accessed: 4/12/2017. <http://www.hl7.org/>.
- 29 Wilks DS. *Statistical Methods in the Atmospheric Sciences*, Vol. 100. Academic Press, Kidlington, OX, UK, 2011.
- 30 Hinrichs SH, Zarcone P. The Affordable Care Act, meaningful use, and their impact on public health laboratories. *Public Health Rep* 2013;**128**(2\_suppl):7–9.
- 31 Lee BY, McGlone SM, Bailey RR *et al*. To test or to treat? An analysis of influenza testing and antiviral treatment strategies using economic computer modeling. *PLoS one* 2010;**5**(6):e11284.
- 32 CDC. *Meaningful Use*. Date accessed: 6/5/2017. <https://www.cdc.gov/chrmeaningfuluse/introduction.html>.

## Appendix A

### Description of the Bayesian Classifier Model

Unlike conventional case definitions with sets of diagnostic rules, our Bayesian model processed probable cases based on underlying statistics. It comprised 31 NLP-extracted findings from ED notes, including abdominal pain, anorexia, arthralgias, cervical lymphadenopathy, chest pain, chill, conjunctivitis, cough, cyanosis, diarrhea, dyspnea, fatigue, fever, headache, hemoptysis, hoarseness, influenza-like illness, lab confirmed influenza, nasal swab order, malaise, myalgias, nausea, pain on eye movement, photophobia, pneumonia, rhinorrhea, sore throat, suspected influenza, viral infection, viral syndrome and wheezing.<sup>23</sup> Each of these findings contributed to the influenza probability with a different weight in the format of conditional probabilities. For example, the probability of an influenza case having nasal swab order,  $P(\text{nasal swab order}|\text{influenza}) = 0.94$ ,  $P(\text{nasal swab order}|\text{non-influenza}) = 0.54$ .