Available online at www.sciencedirect.com
ScienceDirect
Journal of Biomedical Informatics 41 (2008) 224–231
Journal of Biomedical Informatics
www.elsevier.com/locate/yjbin

Estimating the joint disease outbreak-detection time when an automated biosurveillance system is augmenting traditional clinical case finding

Yanna Shen [a,b,*], Christina Adamou [b], John N. Dowling [b], Gregory F. Cooper [a,b]

[a] *Intelligent Systems Program, University of Pittsburgh, 5113 Sennott Square, 210 S. Bouquet St., Pittsburgh, PA 15260, USA*
[b] *Department of Biomedical Informatics, University of Pittsburgh, M-183 VALE, 200 Meyran Ave., Pittsburgh, PA 15260, USA*

Received 21 December 2006
Available online 21 November 2007

Abstract

The goals of automated biosurveillance systems are to detect disease outbreaks early, while exhibiting few false positives. Evaluation measures currently exist to estimate the expected detection time of biosurveillance systems. Researchers also have developed models that estimate clinician detection of cases of outbreak diseases, which is a process known as *clinical case finding*. However, little research has been done on estimating how well biosurveillance systems augment traditional outbreak detection that is carried out by clinicians. In this paper, we introduce a general approach for doing so for non-endemic disease outbreaks, which are characteristic of bioterrorist induced diseases, such as respiratory anthrax. We first layout the basic framework, which makes minimal assumptions, and then we specialize it in several ways. We illustrate the method using a Bayesian outbreak detection algorithm called PANDA, a model of clinician outbreak detection, and simulated cases of a windborne anthrax release. This analysis derives a bound on how well we would expect PANDA to augment clinician detection of an anthrax outbreak. The results support that such analyses are useful in assessing the extent to which computer-based outbreak detection systems are expected to augment traditional clinician outbreak detection.

Keywords: Disease outbreak detection; Automated biosurveillance systems; Clinical case finding; Evaluation

1. Introduction

Electronic biosurveillance refers to the systematic collection and automated analysis of electronically available data with the intent of detecting outbreaks of disease rapidly [1]. These electronic data can come from emergency department visits, over-the-counter medication sales, ambulatory care visit records, and other sources.

The traditional public health system relies heavily on clinicians to report suspected cases of disease outbreaks. The main reason for implementing a biosurveillance system is to improve the timeliness of outbreak detection relative to depending solely on clinicians to detect outbreaks. Due to the lack of actual outbreaks of diseases under surveillance, however, evaluations of improvement are scarce. Therefore, there exists considerable uncertainty as to whether and how much a biosurveillance system would improve the timeliness of detecting disease outbreaks, relative to traditional case detection by clinicians, which is known as *clinical case finding*.

Automated biosurveillance systems and clinical case finding are likely to exhibit different behaviors in detecting disease outbreaks. Consider, for example, an outdoor release of anthrax spores by a bioterrorist. As a consequence, suppose that early in the outbreak a modest number of respiratory anthrax cases appear in emergency departments (EDs) that serve the area of the population downwind of the point of anthrax release. These anthrax cases might not be identified via case finding upon presentation to the ED because they might be easily confused with other respiratory diseases that are much more common. On the other hand, a biosurveillance system that monitors the chief complaints of patients that visit all these EDs may be able to detect an increase in respiratory cases

* Corresponding author.
E-mail address: shenyn@cbmi.pitt.edu (Y. Shen).

1532-0464/$ - see front matter © 2008 Published by Elsevier Inc.
doi:10.1016/j.jbi.2007.11.002

that follows a telltale downwind pattern. Such a pattern may have a high positive predictive value for a windborne, bioterrorist-induced outbreak disease. Such a situation would favor early detection by the biosurveillance system, relative to clinical case finding.

Conversely, consider a single patient who presents to an ED with a rash characteristic of smallpox. Suppose the chief complaint is coded by the triage nurse as "rash". Such a case might well be detected by an ED clinician, but would be unlikely to be detected by a biosurveillance system that is monitoring the chief complaints of patients that present to EDs in the region. Such a situation clearly favors early detection by clinical case finding, relative to computer-based detection.

Generalizing the examples just provided, automated biosurveillance systems are expected to detect an outbreak relatively early when (1) the outbreak cases present with clinical features (typically symptoms and signs) that are only weakly diagnostic in individual patients (i.e., have a low positive predictive value for each such patient case), but are strongly diagnostic when viewed as a spatio-temporal pattern in the population, and (2) such clinical features are available in a timely manner in electronic form. Conversely, traditional outbreak detection by clinicians is likely to do relatively well if early in the outbreak only a few patient cases present with clinical features that have a high positive predictive value for the outbreak disease. Since the outbreak detection performance of biosurveillance systems is likely to differ from that of clinical case finding—at least in some situations—there exists the potential for such computer-based detection to complement clinical case finding.

In this paper we model how well automated biosurveillance systems are expected to improve traditional clinical case finding. We believe this is an informative perspective from which to evaluate biosurveillance systems. In particular, the paper focuses on evaluating the detection of non-endemic, bioterrorist-induced diseases, such as inhalational anthrax.

2. Background

In evaluating biosurveillance systems, investigators have used several measures of detection performance. Some common performance measures include the direct reporting of sensitivity and false positive rates. Sensitivity (true positive rate) estimates the chance that a future outbreak will ever be detected. The false positive rate is the expected number of false alerts of an outbreak per unit time; it is sometimes called the *false alert rate*. Receiver operating characteristic (ROC) curves, such as those presented by Burkom [2], are commonly used summaries for assessing the tradeoff between sensitivity and false positive rates in detecting disease outbreaks.

The ability to detect outbreaks in a timely manner is an issue of central importance. Fawcett and Provost introduced a framework for evaluating the performance of activity monitoring algorithms in financial transaction systems [3]. The Activity Monitoring Operating Characteristic (AMOC) curve that they defined has become a useful and popular method for assessing the performance of systems that detect outbreaks of disease [4–6]. As it is typically applied in evaluating biosurveillance systems, the AMOC curve plots the expected time to detection (since the outbreak began) versus the false alert rate. As the threshold used for detecting an outbreak decreases, the time to detection generally decreases, but the false alert rate generally increases. Kleinman and Abrams describe additional measures for assessing biosurveillance performance [7].

The closest prior work to that reported here is by Buckeridge et al. [8]. They compared the outbreak detection performance between automated syndromic surveillance and clinical case finding. In particular, they did so using a cumulative sum computer algorithm and simulations that model (1) inhalational anthrax outbreaks and (2) clinical case finding performance of anthrax cases. Their results include a report of the mean detection benefit of using the computer algorithm to augment clinical case finding at a given level of specificity. We discuss their results in more detail in Section 5.

The current paper differs from Buckeridge et al. in that it provides a general formulation in the form of an equation for estimating disease outbreak detection time when an automated biosurveillance system is augmenting traditional clinical case finding. In contrast, Buckeridge et al. use stochastic simulation to investigate the extent to which a biosurveillance system is expected to augment clinical case detection in a particular outbreak scenario involving inhalational anthrax. They do not provide a general formulation of joint computer-clinician outbreak detection.

3. Methodology

In this section we layout a general equation for measuring joint detection performance when a biosurveillance system is augmenting traditional clinical case finding. We then specialize it in several ways under assumptions.

We assume that if the expected false alert rate of computer-based outbreak detection and clinical case finding is zero, then a computer generated alert and a clinical case diagnosis will be treated identically in terms of the resulting public health response. This assumption can be relaxed; however, we believe it is useful to first introduce a formulation that incorporates this symmetry, which then can be extended in future, to more complex models. The example introduced in Section 4 assumes that clinical case finding has a zero false positive rate, although we discuss there how this assumption can be readily relaxed.

We now introduce a general model of clinician outbreak detection. We believe that versions of this model can serve as useful approximations for estimating the expected time that clinicians[1] will take to detect a particular type of out-

[1] More generally, the model could be applied to estimate the detection time of diagnosticians of any kind, including public health officials.

break of some disease D. In this paper we assume that D is a non-endemic disease in which the number of cases under a non-outbreak condition is zero or nearly so. Many potential bioterrorist-induced diseases, such as inhalational anthrax and smallpox, satisfy this assumption. For such diseases, the identification of a single case is sufficient to indicate an outbreak of the disease.

Let time(i) be a function that maps patient case i to the time at which that patient presented with outbreak disease D to clinicians. Assume that patient cases with D are consecutively numbered from 1 to M, where case 1 denotes the first patient and M denotes the last patient to present to clinicians with D during the outbreak. Let K denote background knowledge and information, which could include clinician skills, clinician networks of interaction, and the clinical findings of each patient.

Let $P(\tau_i|K)$ designate the probability that case i would be detected by a clinician at time τ_i and no other cases would be detected by clinicians on or before τ_i. Thus, each of the remaining $M-1$ cases either was not ever diagnosed as having D or was diagnosed by clinicians as having D at a time later than τ_i.

Let U denote the greatest amount of time beyond which it is not useful to detect the outbreak. Let $P(\sim U|K)$ represent the probability that clinicians never detect the outbreak at all by time U, which means that none of the M cases are diagnosed as having D.

Now consider outbreak detection by the computer outbreak detection algorithm. Let r be a threshold such that if the computer algorithm's probability of an outbreak is above r, then an alert is raised. For that threshold, there will be some false alerting rate of the algorithm and some time t_r at which the algorithm's outbreak probability is expected to initially exceed r during an outbreak (or a simulated outbreak). If the algorithm's probability never exceeds r, then let $t_r = U$. Many values of threshold r are considered.

Eq. (1) expresses the expected joint detection time (EJDT) for a given alerting threshold r. The first term on the right hand side is computer algorithm's detection time, assuming that clinicians never detect the outbreak on their own (i.e., $\min(t_r, U)$), multiplied by the probability of that situation occurring (i.e., $P(\sim U|K)$). The sum in Eq. (1) considers the situation in which clinicians do detect outbreak disease D. Each possible case is considered as the possible first case in which clinicians detect D. The integral derives the expected time to detect a given case i. The min function within the integral captures the notion that we wish to represent the earliest detection of the case, regardless of whether it was first detected by the computer algorithm (at time t_r) or by clinicians (at time τ_i).

$$\mathrm{EJDT}(r) = \min(t_r, U)P(\sim U|K) + \sum_{i=1}^{M} \int_{\tau_i=\mathrm{time}(i)}^{U} \min(t_r, \tau_i)P(\tau_i|K)\mathrm{d}\tau_i \quad (1)$$

Eq. (1) is quite general. It makes no assumptions about how the computer algorithm or the clinicians detect an outbreak, or the type of data that is used in doing so. For example, Eq. (1) does not assume that the cases are independent of each other. It also does not assume that clinicians work independently of each other in detecting an outbreak disease.

It may be challenging to apply Eq. (1) directly in practice, however. Therefore, we next introduce specialized versions. If we assume that people with D are diagnosed independently of each other, then we can represent the probabilities $P(\tau_i|K)$ and $P(\sim U|K)$ as a product of probabilities. Let $P(\tau|B(i))$ denote the probability that patient outbreak case i will be diagnosed as having D at time τ, given background knowledge and information $B(i)$. For example, $B(i)$ would typically include the information that case i was first seen at time(i). It might also contain additional information, such as the type and severity of the symptoms of patient case i, as well as information about the particular clinician who cared for that patient.

With these assumptions, Eq. (1) can be specialized to Eq. (2). In this equation, the first term expresses the situation in which clinicians never detect the outbreak in a useful amount of time. Each value of j in the sum expresses the situation in which the jth case is detected and no cases are detected before it.

$$\mathrm{EJDT}(r) = \min(t_r, U) \cdot \prod_{i=1}^{M} \left(1 - \int_{\tau=\mathrm{time}(i)}^{U} P(\tau|B(i))\mathrm{d}\tau\right) + \sum_{j=1}^{M} \int_{\tau'=\mathrm{time}(j)}^{U} \min(t_r, \tau') \cdot P(\tau'|B(j)) \cdot \prod_{\substack{z=1 \\ z\neq j}}^{M} \left(1 - \int_{\tau''=\mathrm{time}(z)}^{\tau'} P(\tau''|B(z))\mathrm{d}\tau''\right)\mathrm{d}\tau' \quad (2)$$

To further simplify the modeling and analysis, in Eq. (3) we assume that the background information $B(i)$ only includes the information that case i was first seen at time(i).

$$\mathrm{EJDT}(r) = \min(t_r, U) \cdot \prod_{i=1}^{M} \left(1 - \int_{\tau=\mathrm{time}(i)}^{U} P(\tau|\mathrm{time}(i))\mathrm{d}\tau\right) + \sum_{j=1}^{M} \int_{\tau'=\mathrm{time}(j)}^{U} \min(t_r, \tau') \cdot P(\tau'|\mathrm{time}(j)) \cdot \prod_{\substack{z=1 \\ z\neq j}}^{M} \left(1 - \int_{\tau''=\mathrm{time}(z)}^{\tau'} P(\tau''|\mathrm{time}(z))\mathrm{d}\tau''\right)\mathrm{d}\tau' \quad (3)$$

If we model time as being discrete, we can convert the integrals in Eq. (3) into sums.

4. An example application of the methodology

In this section, we describe an example application that involves deriving an upper bound on the time required to detect an outbreak of windborne inhalational anthrax. We first describe the model of clinician outbreak detection

that we used. Next we summarize the outbreak detection system we applied. We then give an overview of how we simulated cases of patients with inhalational anthrax and overlaid those cases onto background data of real cases of patients who visited emergency departments during a period when there were no known outbreaks of disease occurring. Finally, we describe the results from applying Eq. (3) to derive an upper bound on the expected time required for clinicians and the outbreak detection system (working in parallel) to detect a simulated outbreak of windborne anthrax.

We derive an upper bound because it is more straightforward and conservative than modeling a point estimate. Also, as a practical matter, if the upper bound turns out to be low, it provides support that outbreak detection is likely to be effective. We emphasize, however, that the purpose of this example is to illustrate a basic application of Eq. (3), rather than focus on the specific details of the example.

4.1. The model of clinician outbreak detection

Our model of clinician outbreak detection of anthrax uses the model developed and described by Adamou, et al. [9]. We assume that there are two ways in which a patient infected with inhalational anthrax can be detected as having the disease by a clinician. One is based on the observation of characteristic signs on a chest X-ray, such as mediastinal widening. The other is by the appearance of a positive blood culture for *Bacillus anthracis*. If a diagnosis were made based on the chest X-ray, we assume it would occur within 4 hours of the patient presenting to the clinician. If it were made based on a blood culture, we assume it would occur within 48 hours of presentation.

We modeled the probability of a chest X-ray (and alternatively a blood culture) leading to an outbreak disease diagnosis as being sensitive to the time at which a patient case presented to a clinician. Later cases are more likely to be advanced, and thus more likely to be diagnosed. Table 1 in the Appendix shows these probabilities, which were derived from a model that was constructed by author CA and author JND. Author JND is an infectious disease specialist, who estimated the model's parameters based on the literature and his clinical beliefs [9].

We set the parameter U to be $\text{time}(M) + 48$, where $\text{time}(M)$ is the time (in hours) that the last patient case with anthrax (from the simulation) presented to the ED. Beyond this time, we assume detection would not occur, and even if it did, it would be of little or no help to any patient who is diagnosed so late. Before this time, we assume it is possible that detection could occur and that it would benefit the treatment of a patient who has the outbreak disease.

Our model of clinician outbreak detection provides an upper bound for two reasons. First, our infectious disease specialist was more confident in estimating an upper bound on detection time from chest X-rays and blood cultures than in estimating a mean detection time or a lower bound on it. Second, Eq. (3) assumes that clinicians detect outbreaks independently of each other. In reality, when outbreak cases would start to appear on a given clinic or emergency department, clinicians would likely begin noticing each other's cases, and thus, their disease detection behavior would no longer be independent. The assumption of diagnostic independence yields an upper bound on clinician detection performance. Since in this example we model an upper bound on clinician performance, the EJDT will also be an upper bound on joint computer and clinician detection performance.

4.2. PANDA

We used PANDA (Population-wide ANomaly Detection and Assessment) as the outbreak detection algorithm [4]. PANDA uses a spatio-temporal, multivariate Bayesian approach to biosurveillance. In particular, a spatio-temporal, causal Bayesian network is used to model an outbreak due to the windborne spread of anthrax. Each person in the population being monitored for an outbreak is modeled using a subnetwork. The primary clinical information about each patient is whether he or she presented to the ED with a respiratory chief complaint (e.g., a cough). These subnetworks are connected through a common set of nodes that represent the disease outbreak conditions, such as the hypothesized location and time of release of anthrax spores. Since the resulting Bayesian network requires millions of nodes to model a medium-sized U.S. metropolitan population, PANDA uses several optimization methods to keep the model size manageable and the inference time tractable.

4.3. Creating simulated datasets

We evaluated the joint detection performance of PANDA and the clinician model on datasets produced by overlaying simulated anthrax cases onto a background of actual ED cases obtained from several hospitals in Allegheny County, Pennsylvania. All personal identifying information was removed from these actual ED cases. The simulated cases of anthrax were produced by the 2004 version of the BARD simulator [10].

BARD uses a Gaussian plume model and weather conditions to estimate the distribution of spore concentrations over a geographic region. Based on the spore concentrations in a given zip code, BARD uses a clinical model to simulate the number of patients who will contract inhalational anthrax over time. In particular, BARD produces a list of anthrax cases, where each case consists of a date-time field and a zip code.

For our experiments, we applied the datasets generated in [4] in order to compare the joint detection performance (clinicians + PANDA) with the detection performance of PANDA alone and the clinicians alone. These datasets were produced by providing several parameters to the BARD simulator, such as the weather conditions and

parameters for the location and height of a simulated anthrax release.

The generated datasets correspond to simulated releases of anthrax at dosages 1.0 kg. We generated 96 datasets (anthrax cases overlaid onto actual ED cases), each with an unique combination of release date, wind direction, wind speed, release location, release height, and quantity of spores released. For each month in 2002, eight random release times were selected for use by the simulator. Thus, a total of $8 \times 12 = 96$ different datasets were generated by overlaying the anthrax cases produced by the simulator onto a background of actual ED cases from 2001 to 2002 [4].

4.4. Deriving the Expected Joint Detection Time (EJDT)

We applied Eq. (3) using each of the 96 datasets described above. For the remainder of this paragraph, we consider one of the 96 datasets. We combined cases in the dataset (outbreak cases overlaid onto real cases) as input to PANDA, and PANDA's posterior probability of an outbreak for the dataset was recorded in order to determine its detection time and false positive rate for various probability detection thresholds. In particular, the detection time was the time from the first anthrax case presenting to the ED until the threshold was crossed by the posterior probability of anthrax that was output by PANDA. The false positive rate was derived as FP/M, where FP is the number of false positives that occurred when PANDA monitored the real data during an M month period before the simulated release. The value of M varied in each of the 96 datasets.

We used the simulated anthrax cases in the dataset in applying an instance of Eq. (3) to determine clinician detection. The clinician detection time was derived as the sum of the presentation time of a patient with anthrax (relative to the presentation time of the first anthrax case) plus the time required by the clinician to diagnose that case (if ever), which we modeled as requiring either 4 hours (for a chest X-ray based diagnosis) or 48 hours (for a blood-culture based diagnosis) after presentation to the ED.

The above process of applying Eq. (3) was performed using each of the 96 datasets. These datasets together produced a curve that plots the EJDT as a function of PANDA's false alert rate.[2]

4.5. Results

Fig. 1 shows the AMOC curves of an upper bound on the expected time to detection (relative to the presentation time of the first anthrax case) as a function of false alert rate. The plot of the upper bound on clinician expected detection time is a horizontal line at approximately 62 hours. The expected detection time of PANDA alone is also shown. An upper bound on the joint detection time, EJDT, is slightly lower than the expected detection time by PANDA. In particular, at zero false alert per month, an upper bound on EJDT is 21.8 h and the expected detection time by PANDA is 22.7 h, which suggests that the joint detection time is expected to be at least 54 minutes faster than PANDA's detection time. A point estimate of the mean clinician detection time would be lower, and thus, the EJDT would be lower as well.

We performed a sensitivity analysis in order to gain additional insight regarding clinician and EJDT detection time. In particular, we assume that a patient with inhalational anthrax, who presents to a clinician, would with probability p be diagnosed upon presentation as having anthrax. Under this model, Eq. (3) reduces to Eq. (4).

$$\mathrm{EJDT}(r) = \min(t_r, U) \cdot (1-p)^M + \sum_{j=1}^{M} \min(t_r, \mathrm{time}(j)) \cdot p \cdot (1-p)^{j-1} \quad (4)$$

In Fig. 2, we plot the expected joint detection time at different levels of p, under the assumption that p is constant. As p increases, EJDT decreases as expected. Specifically, if $p > 0.01$ and the false alarm rate is zero per month, then EJDT is less than 14.9 h. At this level of clinician detection performance, the EJDT is more than 7 h sooner than using PANDA alone. Also, interestingly, when $p = 0.002$ the joint detection performance is very close to that shown in Fig. 1.

As a second type of sensitivity analysis, we developed a variation on the analysis plotted in Fig. 2. Let d denote the number of cases required for clinicians to first detect the outbreak disease. We plot the expected joint detection time as a function of d in Fig. 3. This plot makes no assumptions about individual clinician performance in detecting a disease outbreak. For d less than 200 cases, the EJDT decreases substantially as d decreases, and clinician detection contributes significantly to the joint detection performance. When d is more than 300 cases, the EJDT tends to remain at approximately 23.7 h when there is zero false alert rate per month. These results suggests that PANDA contributes significantly to the joint detection performance when d is larger than 200.

Furthermore, the number of cases d that yields the joint performance shown in Fig. 1 is approximately 148 cases, which shows that clinician detection is on the cusp of having a significant impact on overall detection performance of the joint computer-clinician detection system.

[2] Since we assumed in this example that clinicians did not falsely diagnose anyone as having anthrax, the clinicians did not contribute to the false alert rate. If we wished to model clinicians as having a constant false alert rate of q per month for diagnosing anthrax, we could simply add q to the false alert rate (*far*) for PANDA, which would shift the EJDT versus *far* plot to the right by q units. Since q seems negligible, we assumed it to be 0 in the example here.

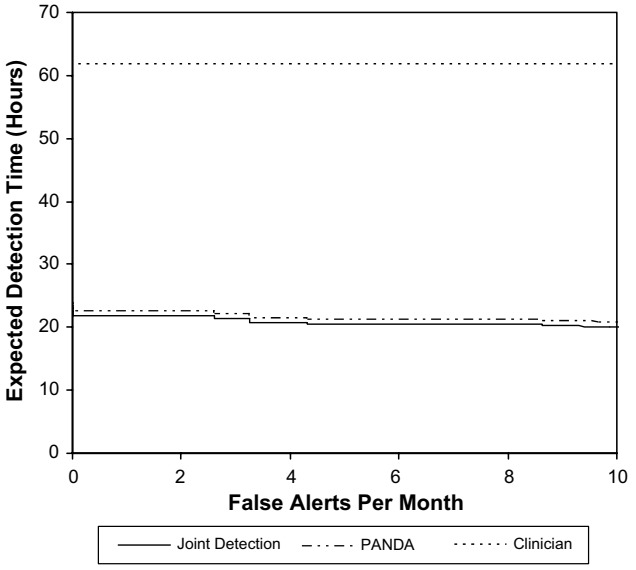

Fig. 1. Plots showing the detection performance of PANDA, an upper bound on the expected detection performance of clinicians, and an upper bound on the expected detection performance when PANDA is augmenting clinician detection (joint detection). Detection time is relative to the first simulated anthrax case who presented to the ED (= time zero).

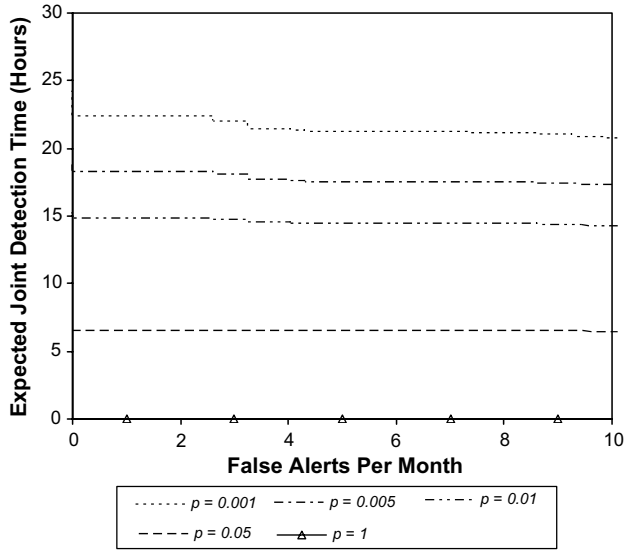

Fig. 2. Plots showing the expected joint detection time at different levels of clinician detection proficiency (p).

5. Discussion of the results relative to previous work

In this section we compare the results in the previous section to selected results obtained by Buckeridge et al. [8], who estimated detection performance of inhalational anthrax based on a computer algorithm and clinical case finding. In particular, they estimated how much less time would be required for outbreak detection if a computer algorithm were being applied to monitor emergency department syndromic data. Their evaluation used models of clinical case finding and of inhalational anthrax that

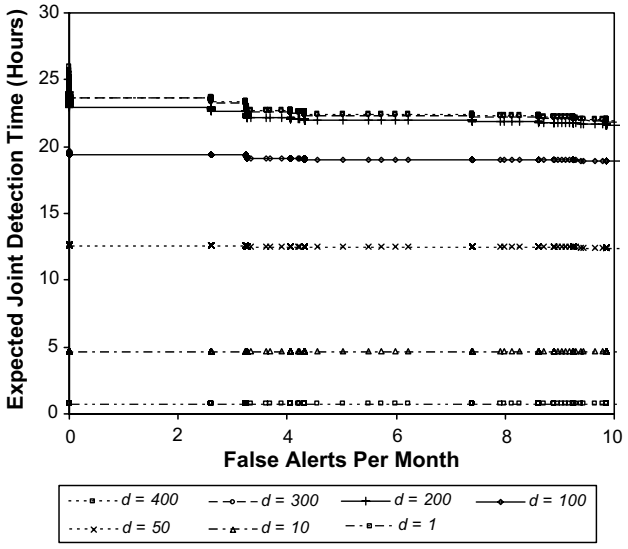

Fig. 3. Plots showing the expected joint detection time as a function of the number of cases (d) that are needed for clinicians to detect an outbreak.

were derived independently from those we derived and described in Section 4. They also used a different computer detection algorithm and a different set of actual ED cases on which to inject simulated anthrax cases.

In particular, Buckeridge et al. reported that their model of clinical case finding had an average detection time of 3.7 days after release. In Section 4.5, we reported an upper bound on the expected clinician detection time as 62 h, which is approximately 1.1 days faster than their model.

Buckeridge et al. define the detection time by clinical case finding to be the duration between the release of anthrax spores and the first positive blood culture. In contrast, we ignore the incubation period and calculate the clinician detection time as the time between the first anthrax case presenting to clinicians and the first characteristic signs on a chest X-ray or the first positive blood culture. Therefore, we do not consider the 1.1-day difference to be surprising.

The computer augmentation time in detection, which is called the *detection benefit* in Buckeridge's paper, is computed as the difference in the timeliness between the computer algorithm and clinical case finding. The augmentation time is influenced by the false alert rates. In particular, at a false alert rate of approximately 1 every 10 days, their mean computer augmentation time is 1 day while our mean augmentation time is 1.7 days; at a false alert rate of approximately 1 every 40 days, their mean computer augmentation time is 0.32 days while our mean augmentation time is 1.6 days.

The computer algorithm used by Buckeridge et al. differed from the one used here. In particular, they applied a temporal detection algorithm based on cumulative sums of residuals, whereas we applied a Bayesian spatio-temporal algorithm. Also, they used ICD9 diagnoses as the primary ED patient feature. We used chief complaints. Buckeridge and colleagues developed a point estimate of

clinical case finding. We used an upper bound, in part to be more conservative in our analysis and in part because the clinical case finding model is not the main focus of this paper. Given these differences, the differing results that are summarized in this section are not surprising.

6. Summary and future work

This paper introduced a general framework in the form of an equation for estimating how well biosurveillance systems are expected to augment traditional clinical case finding in detecting an outbreak of a non-endemic infectious disease. The method yields the expected joint detection time (EJDT) of a computer-based biosurveillance system that is working in parallel with traditional clinician outbreak detection. We presented the basic method at several different levels of generality. The general mathematical framework that we introduced for evaluating joint clinician-machine joint detection can be applied in evaluating other types of joint clinician-machine detection, such as the joint detection of nosocomial infectious disease outbreaks.

We applied specialized versions of the method to an example in order to derive an upper bound on joint outbreak detection of inhalational anthrax, and we performed a sensitivity analysis. The example provides support that the EJDT measure can be useful for assessing the extent to which computer-based outbreak detection systems are expected to augment traditional clinician outbreak detection.

We also can use EJDT as a performance measure that guides the development of outbreak detection algorithms that complement the expected detection performance of clinicians. It would be best to develop algorithms that are augmentative of, rather than redundant with, clinician detection. For example, for a given false alert level it might be that computer algorithm A is expected to detect outbreak disease D earlier than algorithm B, but the EJDT of B is less than A, because B better complements clinician performance. If so, it would be best to use algorithm B. This perspective suggests a new approach toward developing outbreak detection algorithms.

A primary direction for future research is to incorporate additional background information into the term K in Eq. (1). Such information could include additional clinical context about patient features, where those patients are seen, and which clinicians see them. The specific formulation of Eq. (1) would then follow from the particular form of K. It will be useful to perform extensive evaluations of such extensions, using a variety of computer algorithms, outbreak simulations, and simulations of clinical case finding.

Acknowledgments

This research was funded by a grant from the National Science Foundation (NSF IIS-0325581). We thank John Levander with programming support for portions of this research. We thank Bill Hogan for making the BARD simulations available. We also thank our colleagues in the Bayesian Biosurveillance project for their comments and suggestions.

Appendix .

Table 1 shows the probabilities that were used in the example described in Section 4. These probabilities were derived from a model developed by authors CA and JND; the model is described in detail in [9], including the literature used to estimate the model parameters. According to the model, a patient with inhalational anthrax, who presents with respiratory symptoms, would be diagnosed either at 4 hours or less (by chest X-ray), at 48 h or less (by blood culture), or never. The variable *dt* (diagnosis time) in Table 1 represents the diagnosis time relative to the day that the patient presented to the ED. The variable *day* in Table 1 denotes the number of days since a patient with inhalational anthrax became exposed to (and presumptively infected by) an outdoor point release of anthrax spores. As an example, if an anthrax patient case presents to a clinician on the 5th day after being infected, that case would have 0.1% chance to be diagnosed with anthrax at up to 4 h after presentation to the ED (by way of a chest X-ray), a 1.4% chance of being diagnosed at up to 48 h (by way of a blood culture), and a 98.5% chance of not being diagnosed at all as having anthrax.

Table 1
Probability parameters used in the example in Section 4

day	$P(dt\|day)$		
	$dt = 4$ h	$dt = 48$ h	$dt =$ never
1	0.000	0.011	0.989
2	0.000	0.011	0.989
3	0.000	0.012	0.988
4	0.000	0.013	0.987
5	0.001	0.014	0.985
6	0.003	0.034	0.963
7	0.010	0.054	0.936
8	0.032	0.123	0.845
9	0.060	0.172	0.768
10	0.094	0.206	0.700
11	0.150	0.260	0.589
12	0.212	0.314	0.474
13	0.267	0.339	0.394
14	0.338	0.359	0.303
15	0.406	0.361	0.233
16	0.495	0.350	0.155
17	0.592	0.318	0.090
18	0.643	0.292	0.065
19	0.670	0.278	0.053
20	0.680	0.272	0.048

References

[1] Mandl KDJMO, Wagner MM, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc 2004;11(2):141–50.
[2] Burkom HS. Biosurveillance applying scan statistics with multiple, disparate data sources. J Urban Health 2003;80:i57–65. 2, suppl 1.
[3] Fawcett T, Provost F. Activity monitoring: noticing interesting changes in behavior. In: Proceedings of the fifth international conference on knowledge discovery and data mining 1999:53–62.
[4] Cooper GF, Dash DH, Levander JD, et al. Bayesian biosurveillance of disease outbreaks. In: Proceedings of the twentieth conference on uncertainty in artificial intelligence. 2004:94–103.
[5] Neill DB, Moore AW, Cooper GF. A Bayesian spatial scan statistic. Adv Neural Inform Process Syst 2006;18:1003–10.
[6] Wong W-K, Moore AW, Cooper GF, Wagner MM. What's strange about recent events (WSARE): an algorithm for the early detection of disease outbreaks. J Mach Learn Res 2005;6:1961–98.
[7] Kleinman KP, Abrams AM. Assessing surveillance using sensitivity, specificity and timeliness. Stat Methods Med Res 2006;15(5):446–64.
[8] Buckeridge DL, Owens DK, Switzer P, Frank J, Musen MA. Evaluating detection of an inhalational anthrax outbreak. Emerg Infect Dis 2006;12(12):1942–9.
[9] Adamou C, Cooper GF, Wong W-K, Dowling JN, Hogan WR. Modeling clinician detection time of a disease outbreak due to inhalational anthrax. Pittsburgh: Department of Biomedical Informatics Report Series, University of Pittsburgh; 2007. Report No.: DBMI-07-352.
[10] Hogan WR, Cooper GF, Wagner MM. A Bayesian anthrax aerosol release detectors. RODS Technical Report. 2004.