Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Biomedical Informatics 38 (2005) 99–113
Journal of Biomedical Informatics
www.elsevier.com/locate/yjbin

Algorithms for rapid outbreak detection: a research synthesis

David L. Buckeridge[a,b,*], Howard Burkom[c], Murray Campbell[d], William R. Hogan[e], Andrew W. Moore[f], for the BioALIRT Project[1]

[a] Palo Alto VA Health Care System, Palo Alto, CA, USA
[b] Stanford Medical Informatics, Stanford University, Stanford, CA, USA
[c] Applied Physics Laboratory, Johns Hopkins University, Laurel, MD, USA
[d] IBM T.J. Watson Research Center, Yorktown Heights, NY, USA
[e] The RODS Laboratory, University of Pittsburgh, Pittsburgh, PA, USA
[f] Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

Received 4 June 2004
Available online 30 November 2004

Abstract

The threat of bioterrorism has stimulated interest in enhancing public health surveillance to detect disease outbreaks more rapidly than is currently possible. To advance research on improving the timeliness of outbreak detection, the Defense Advanced Research Project Agency sponsored the Bio-event Advanced Leading Indicator Recognition Technology (BioALIRT) project beginning in 2001. The purpose of this paper is to provide a synthesis of research on outbreak detection algorithms conducted by academic and industrial partners in the BioALIRT project. We first suggest a practical classification for outbreak detection algorithms that considers the types of information encountered in surveillance analysis. We then present a synthesis of our research according to this classification. The research conducted for this project has examined how to use spatial and other covariate information from disparate sources to improve the timeliness of outbreak detection. Our results suggest that use of spatial and other covariate information can improve outbreak detection performance. We also identified, however, methodological challenges that limited our ability to determine the benefit of using outbreak detection algorithms that operate on large volumes of data. Future research must address challenges such as forecasting expected values in high-dimensional data and generating spatial and multivariate test data sets.
Published by Elsevier Inc.

Keywords: Surveillance; Outbreak detection; Public health informatics; Evaluation; Timeliness; Spatial analysis; Data mining; Bioterrorism

1. Introduction

The threat of bioterrorism has stimulated interest in public health surveillance. Government agencies and researchers are particularly interested in enhancing the ability of surveillance systems to detect outbreaks rapidly [1]. This interest stems, in part, from the large projected savings resulting from detecting outbreaks rapidly in some scenarios, particularly scenarios associated with bioterrorism [2]. The potential improvement in timeliness of outbreak detection over more traditional public health methods, such as telephone calls from concerned clinicians, defines the usefulness of a surveillance system for rapid outbreak detection.

Early outbreak detection has always been of interest to public health. Most outbreaks are recognized from accumulated case reports or by alert clinicians [3], and surveillance systems have traditionally enhanced detection by identifying outbreaks spread across multiple reporting sites. In contrast, many bioterrorism surveillance systems are designed to detect large outbreaks in single communities before a clinician can diagnose the

* Corresponding author. Fax: +1 650 725 7944.
E-mail address: david.buckeridge@stanford.edu (D.L. Buckeridge).
[1] See the acknowledgments for a full list of contributors.

1532-0464/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.jbi.2004.11.007

index case and then report to public health. Researchers and public health practitioners use terms such as "syndromic" and "pre-diagnostic" to identify surveillance focused on very early detection of outbreaks [4]. The shift in focus to very early detection places new demands on systems. Early detection systems must use data collected from individuals prior to the assignment of a definitive diagnosis. These data sources, such as emergency department chief complaints and over-the-counter pharmaceutical sales, are by nature less specific than diagnostic data. Surveillance systems that use these types of data sources require outbreak detection algorithms that can analyze non-specific signals and use as much available information as possible to avoid false positives while improving timeliness.

To advance research on improving timeliness of outbreak detection, the Defense Advanced Research Project Agency (DARPA) sponsored the Bio-event Advanced Leading Indicator Recognition Technology (BioALIRT) project beginning in October 2001 [5]. This project brought together researchers from four universities, two companies, the US military, and an independent evaluation contractor, providing a forum for the interchange and critical consideration of early research results. Table 1 lists the teams and web addresses where readers may find more information including publications and software. Project teams focused on two approaches to improving the timeliness of outbreak detection: (i) identifying new non-traditional data sources with early alerting potential; and, (ii) developing and evaluating outbreak detection algorithms for surveillance data. Another important research focus not mentioned in this paper was the development of methods for ensuring privacy in surveillance. The purpose of this paper is to synthesize some major themes in research on detection algorithms conducted through the BioALIRT project. Papers to describe the entire project in greater detail, and to summarize research on surveillance data sources conducted within the project, are under development. In this paper, we do not present detection algorithms in detail, but rather we convey our research findings subdivided by the types of information available for surveillance. We begin by presenting a practical classification for detection algorithms that guided much of our work. Then we synthesize our research on detection algorithms according to our practical classification and discuss approaches to evaluating detection algorithms. Finally, we close by considering our progress towards addressing the challenges in detection algorithm research and by identifying promising areas for future research.

Table 1
BioALIRT project teams and web sites that provide details on the research groups and access to available software

Team	Web site
Carnegie Mellon University, Computer Science, Auton Lab	www.autonlab.org
General Dynamics, Advanced Information Systems	www.gd-ais.com
IBM Watson Research	www.watson.ibm.com
Johns Hopkins University, Applied Physics Lab	www.jhuapl.edu
Johns Hopkins University, School of Public Health	www.jhsph.edu
Potomac Institute for Policy Studies	www.potomacinstitute.org
Stanford University, Stanford Medical Informatics	www.smi.stanford.edu
University of Pittsburgh, RODS Laboratory	www.health.pitt.edu/rods
Walter Reed Army Institute of Research	www.wrair.army.mil

2. Framing the problem

Surveillance is the process of systematic data collection, ongoing analysis and interpretation, and dissemination of results to those who need to know [6]. Surveillance analysis should provide information to guide public health decision-making, and the improvement in decision-making and public health outcomes determines the ultimate worth of improved detection performance. To allow careful consideration of analysis methods though, we limit our focus in this paper to detection algorithms, which were one aspect of the BioALIRT program. We do not examine the broader question of how these detection algorithms influence public health decision-making. In addition, although problems encountered in syndromic surveillance motivate our work, we do not address the larger question of the utility of syndromic surveillance in general. Indeed, the majority of findings presented are applicable to outbreak detection in other types of surveillance.

Discussions and reviews of outbreak detection algorithms tend to consider detection algorithms in terms of their historical foundations such as statistical process control [7], change-point detection [8] or spatial statistics [9]. This approach facilitates a review as algorithms fit neatly into these categories, but this perspective does not address the practical application of these algorithms. We suggest an approach to considering outbreak detection algorithms that instead considers the types of information encountered in surveillance. The intent is to provide a classification of algorithms based on their functional characteristics that is meaningful to surveillance practitioners and researchers.

Surveillance analysts and researchers conduct many analyses under the broad heading of outbreak detection. In practice, we have found that a few information contexts encompass most surveillance environments (Table 2). The number of data sources, the availability of covariate data, and the availability of spatial information define these contexts. Traditionally, most surveillance analysis has used only a single data source with

Table 2
Common surveillance contexts and sample scenarios

	Spatial regions		
	None	Few	Many
Single data source			
No individual covariates (single time-series)	Daily counts of ED visits for a single hospital	Daily counts of ED visits by hospital location for all hospitals in a city	Daily counts of ED visits by patient location for all ZIP codes in a city
Individual covariates	Daily ED visit records (e.g., age, gender, chief complaint) for a single hospital	Daily ED visit records (e.g., age, gender, chief complaint) by hospital location for all hospitals in a city	Daily ED visit records (e.g., age, gender, chief complaint, patient ZIP) for all hospitals in a city
Multiple data sources			
No individual covariates (multiple time-series)	As with single data source plus daily counts of pharmaceutical prescriptions from the same ED	As with single data source plus daily counts of pharmaceutical prescriptions from each ED	As with single data source plus daily counts of pharmaceutical prescriptions by patient location
Individual covariates	As with single data source plus daily pharmaceutical prescription records from the same ED	As with single data source plus daily pharmaceutical prescription records from each ED	As with single data source plus daily pharmaceutical prescription records by patient location

no covariate data (i.e., a single time-series, the upper-left cell in Table 2). One might encounter data of this type when analyzing counts of emergency department (ED) visits to a single hospital, or more typically, case reports of a reportable condition [10]. When spatial information is available by case, we make a distinction between the situation where there are only a few geographic locations and the one where there are many geographic locations. One might encounter data from a few locations when analyzing counts of ED visits from a small number of hospitals, and data from many spatial locations when analyzing ED visits by patient home ZIP code. Analysis of multiple data sources generally requires additional considerations, most notably an approach to combining the data sources. In the simplest case, one might encounter multiple data sources when analyzing aggregate counts from different data streams for the same institution or geographic region (for example, ED visit counts and pharmaceutical prescription counts for the same hospital). The addition of spatial information and other covariates results in more complex analytic contexts.

Outbreak detection in public health surveillance has traditionally used single time-series due to limitations of data and algorithms. Contexts represented by the cells in Table 2 to the right and down from this traditional context contain progressively more information. Enhanced public health information systems are beginning to provide the data necessary for surveillance in these more information-intensive contexts [11–13]. A major aim of the algorithm research in the BioALIRT project is to expand the knowledge of outbreak detection algorithms that can use large volumes of information effectively. An underlying hypothesis of this research is that algorithms that use the additional information found in spatial location, other covariates, and multiple data sources can detect outbreaks more rapidly than algorithms that do not use this additional information.

3. Research synthesis

3.1. Detection algorithms by surveillance analysis context

Surveillance to detect outbreaks entails analyzing observed surveillance data prospectively to detect a meaningful change from the expected range of data values. Due to its prospective nature, with new data arriving continuously or at regular intervals, analysis for outbreak detection differs from the retrospective analyses performed in many public health and epidemiological settings, where all the data are available at the outset of the analysis. Surveillance analysis also faces challenges not encountered in other public health and epidemiological settings due to the need to model and forecast expected data values over time and space.

Another important component of analysis for outbreaks is the approach used to determine what constitutes a meaningful change from expected values. Assessing the performance of detection algorithms is another important consideration and there are many metrics for measuring the detection performance of an algorithm, some of which we discuss below. Regardless of the metric used, the goal of outbreak detection is to identify outbreaks rapidly with few false alarms.

The algorithm research conducted through the project included evaluation of existing algorithms applied in realistic surveillance settings, the extension of existing algorithms, and the development of new algorithms. Existing algorithms for prospective outbreak detection were the foundation for much of our work. We do not attempt to review the literature on outbreak detection algorithms exhaustively. Instead, we refer interested readers to recent reviews of this topic [7–9] and endeavor here to make our description of algorithms accessible to readers without extensive experience in this area. We divide our synthesis of algorithm research into three sections that contain the contexts discussed earlier. Table 3 shows the algorithms examined by research teams in each surveillance context.

3.2. Single data source without spatial information

Many researchers have examined the use of statistical process control (SPC) methods for surveillance analysis of a single time-series [14,15]. These methods were designed to monitor the number of defective products in a manufacturing process, and to identify an increase that might imply the onset of a problem in the manufacturing process. Given the vast SPC literature, research on SPC methods within the project tended to focus on the practical issues of applying and extending existing algorithms for outbreak detection rather than on the development of new algorithms.

Project researchers found the exponentially weighted moving average (EWMA) to be a simple and robust SPC method for surveillance of sparse data [16]. We noted that theoretically defined alerting thresholds for the EWMA and other SPC methods [17] tend to produce false alarm rates in ranges that are not useful for public health practice, possibly because public health surveillance data tend to violate assumptions of SPC methods as we discuss below. Instead, we found alert thresholds empirically derived from historical algorithm output to be more robust. In other words, thresholds derived assuming Gaussian residuals give frequently suboptimal and occasionally poor performance for reasons that include statistical problems such as non-stationarity and correlation, as well as data quality issues such as late and missing reports. Adaptive, empirically derived thresholds do not solve all these problems, but they do help.

Table 3
Common surveillance contexts and analytic approaches evaluated or developed by project teams

	Spatial regions		
	None	Few	Many
Single data source			
No individual covariates (single time-series)	SPC,[a] Regression + SPC, KF[b] + diagnostics, wavelet analysis	Region or neighborhood prediction + single region methods, unsupervised clustering + single region methods	Global spatial clustering + cumulative sum, space–time scan
Individual covariates	BCD,[c] WSARE[d]	BCD, WSARE, BARD[e]	Space–time scan, BCD, WSARE, BARD
Multiple data sources			
No individual covariates (multiple time-series)	Unsupervised clustering, combine *p-values* from single series, MSPC[f]		
Individual covariates			Multivariate space–time scan

[a] SPC, statistical process control.
[b] KF, Kalman filter.
[c] BCD, biosurveillance using a change-point detector.
[d] WSARE, what's strange about recent events.
[e] BARD, Bayesian aerosol release detector.
[f] MSPC, multivariate statistical process control.

It is possible to enhance the sensitivity of an EWMA and other SPC methods to a gradually increasing signal by incorporating a short guard band (e.g., 2 days) to avoid contamination of the baseline with an outbreak signal [18]. Most anomaly-detection methods operate by comparing observed data values to expected ones derived from a baseline interval. If there is no gap between the baseline and test intervals, then the start of a signal can upwardly bias the expectation calculated from the baseline and cause the algorithm to fail to flag the remainder of the signal. Introducing a lag, or guard band, between the baseline interval and the recent interval helps to prevent the loss of sensitivity that may result from a gradually increasing outbreak contaminating the baseline in this way. While intuitively obvious, it is also worth noting that the depth of memory parameter in the EWMA method (and analogous parameters in other methods) affects performance for different types of outbreak signals. For example, the traditional choice of a value of around 0.3 for the memory parameter [16] results in good performance for detection of a lognormal epidemic curve with a short median incubation period (e.g., 2 days). However, a memory parameter closer to 1.0 (i.e., replacement of EWMA with a Shewhart chart) is preferable for the detection of single-day spikes or more scattered signals since little data smoothing then occurs.

Single time-series that contain a greater frequency of events tend to exhibit substantial temporal structure such as a day-of-week cycle. We found that, in this situation, pre-processing before application of a SPC method can improve detection performance. An approach used by many research teams was to predict the daily count using a linear or Poisson regression model, and then to apply a SPC method to the regression residuals. For example, to predict sales of over-the-counter (OTC) pharmaceuticals, one research team used a Poisson model with variables for trend, day of the week, seasonality, and temperature, plus an autoregressive term. They then used the EWMA method to detect changes in the regression residuals. The Poisson model does not require constant variance and may be more robust than other linear models for forecasting in this setting [19]. Another research team used an adaptive Kalman filter [20] to forecast expected values for each day followed by the application of standard diagnostic methods to the forecast residuals. Initial results suggest that this approach may work well for non-stationary data, but further research is required. Another pre-processing approach used by one research team was to perform a wavelet analysis of the time series [21]. Researchers used the wavelet model to forecast values for each day and to produce residuals which they analyzed for deviations from expectation.

Methods for monitoring single time-series cannot easily exploit additional information available from covariates in individual records. Researchers in this project developed two approaches for exploiting covariate information. The Biosurveillance using a change-point detector (BCD) method and the what's strange about recent events? (WSARE) method [22,23] were both developed for surveillance of a single data source with covariate information. The BCD method is an extension of a standard change-point surveillance method, the likelihood test [24], that accommodates covariate information by representing the distribution of covariate values as a Bayesian network. The algorithm then uses the network to infer the probabilities of the recently observed data under assumptions that an outbreak has and has not occurred. A likelihood ratio is then calculated from these two probabilities, and a p-value is empirically derived by using the empirical Monte Carlo strategy which is commonly adopted to determine statistical significance when the sampling distribution is intractable [25]. The BCD method performed well in a recent blinded algorithm evaluation using real data [26].

The WSARE method searches for irregularities in covariate values using rules [22,23]. An example of a rule is "Gender = Male AND Home Location = NW." This rule indicates that WSARE is determining whether the number of recent cases involving male patients in the northwest geographic area is unusually high. Each day, WSARE considers all possible rules involving single or pairs of covariates and selects the most statistically significant rule for the current time period using a randomization test to guard against multiple-hypothesis testing errors. To account for trends in the baseline data such as day-of-week and seasonal effects, WSARE uses a Bayesian network to produce the baseline distribution by taking the joint distributions of the data and conditioning on attributes that are responsible for the trends. In simulations designed by its authors, WSARE produced better detection times and slightly higher false-positive rates than algorithms that do not examine covariate information [22].

In general, we found that algorithms used for surveillance analysis in other domains, such as SPC algorithms intended for use in a manufacturing, require adaptation for use in public health surveillance, due mainly to the characteristics of surveillance data. Healthcare administration data in particular contain regular (e.g., day-of-week) and irregular (e.g., seasons and holidays) temporal patterns in addition to unexplained irregular variation (Fig. 1). In some situations it is possible to model many of these features using regression methods, but regression modeling generally requires considerable baseline data and we found that models developed for one data set or geographic location did not easily transfer to other data sets or locations. SPC methods are attractive due to their simplicity, and the true benefit of more complex regression-based methods over simpler SPC methods is unknown in real settings. Another

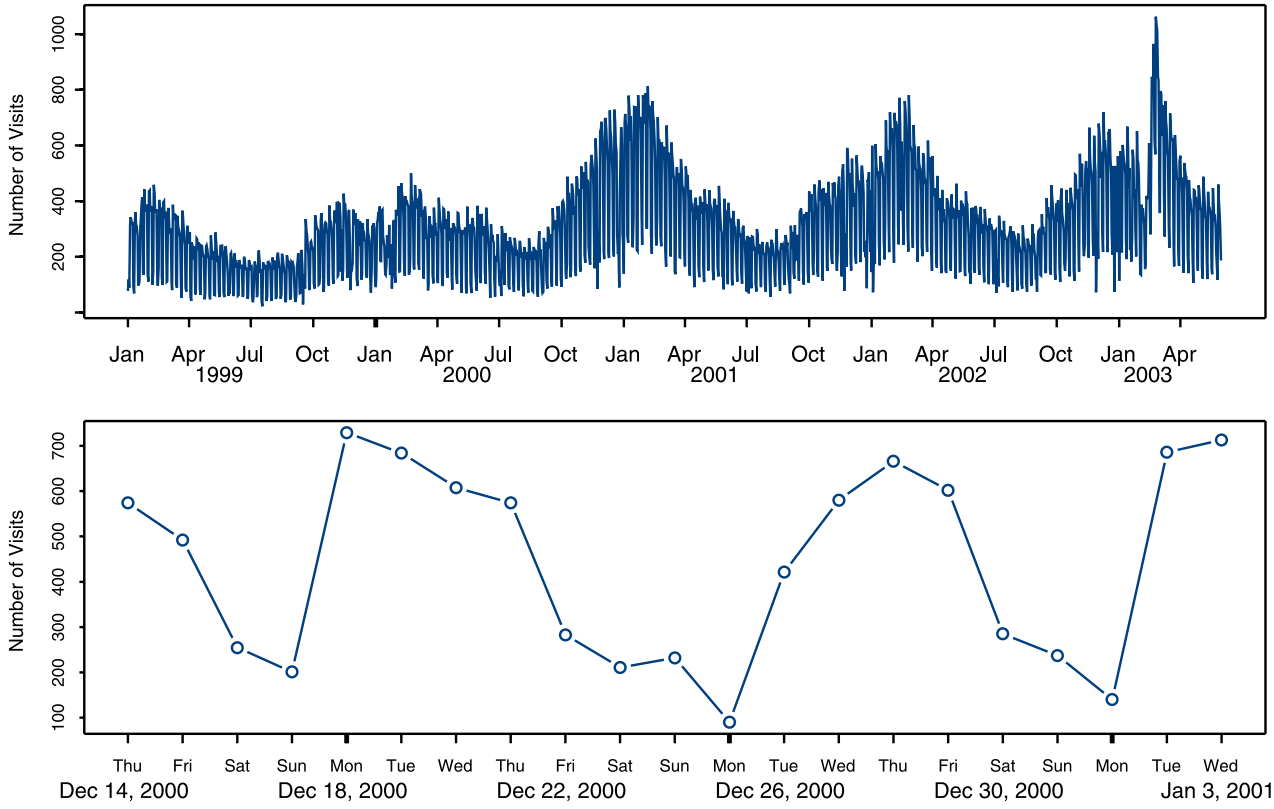

Fig. 1. Example of univariate time-series data encountered in public health surveillance. These data are aggregate daily counts of ambulatory clinic visits for respiratory conditions in a metropolitan area. The top plot illustrates the seasonal and long-term variation in the data. The bottom plot shows the same data at a higher temporal resolution to illustrate the day-of-week and holiday effects.

important observation from our work with single time-series surveillance is that different algorithms are suited to detecting different types of outbreak signals. This observation likely extends to other surveillance contexts and may seem self-evident, but the issue of matching algorithms to types of outbreak signals has received little explicit consideration in the public health surveillance literature. For example, a signal resulting from a bioterrorism attack may differ considerably from a signal due to an influenza epidemic because of the spatial distribution of victims, the temporal distribution of primary cases, and other scenario-dependent features. A final observation is that data quality influences the performance of an outbreak detection algorithm. Inconsistent reporting of cases may strongly impact algorithm performance, and we found that using rates instead of counts can help to address this problem. When reliable denominator data are not available, which is often the case in syndromic surveillance, we found it helpful to use ratios. These included the ratio of syndromic counts to all counts and the ratio of syndromic counts to daily counts of reporting providers. The ratio of syndrome counts to reporting providers was particularly effective in reducing the late reporting problem, when the necessary data were available.

3.3. Single data source with spatial information

In contrast to surveillance analysis of a single time-series, there is not an extensive body of literature describing algorithms suitable for prospective analysis of a single data source with spatial information [9]. Research teams therefore devoted a considerable amount of effort to examining the application of existing methods in this context and to developing new methods. Different approaches exist for using spatial information in surveillance analysis. Some approaches use spatial location as a categorical variable similar to any other covariate such as age. More detailed approaches account for unique characteristics of spatial information such as topology and geometry [27]. In general, methods for cluster detection tend to consider topology explicitly through adjacency or distance between regions. Geometry, or the expected shape of an outbreak, is less frequently a prime consideration and many methods make simplifying assumptions about geometry to improve analytic tractability.

A straightforward approach to using spatial information is to apply a separate SPC method within each spatial region. For example, if daily counts of emergency department visits are available for 20 ZIP codes, this

approach would entail following the results of 20 SPC methods, one for each ZIP code. If temporal structure is present within the time-series in spatial regions then pre-processing of counts within each region before application of an SPC method may improve performance. In this approach, critical thresholds for declaring an outbreak must account for the multiple tests performed to avoid excessive false alerts. When analytical corrections for the multiple comparisons are not easily determined, decision thresholds are usually set with Monte Carlo simulation using historical data. This approach does not consider the unique characteristics of spatial information, and we examined variations on this approach that incorporate information on the topology of regions. For example, we calculated smoothed region forecasts by modeling each region forecast as a function of values in neighboring regions. We also aggregated connected spatial regions and then applied SPC methods to the smaller number of aggregate regions. Raubertas [28] was one of the first researchers to use neighboring regions in a similar approach to forecast values in prospective surveillance analysis. Some research teams noted that this type of smoothing can decrease sensitivity for localized signals and that there is no data-sensitive, generalizable way to form neighborhood groupings. In an attempt to avoid the arbitrary aspect of neighborhood definition, one research team examined the use of unsupervised clustering methods including *k*-means and hierarchical clustering to aggregate connected regions [29]. Initial results suggest that such clustering can improve detection performance for outbreaks spread across a small number of spatial regions.

When many spatial regions are under surveillance, the application of a separate SPC method for each region can become unwieldy. Also, aggregating regions prior to analysis lose information. We examined two general approaches to detecting outbreaks in data from large numbers of spatial regions [30]. One approach is to detect increased global clustering or the general propensity for connected regions to exhibit similarly increased rates of health-related events. The other approach is to detect focused clusters of regions with increased rates of events, or the propensity for increased rates of events to occur around a single location. To detect increased global clustering we found it useful to apply a method along the lines of those described by Rogerson [31,32] and Jarpe [33]. This approach entails applying a global spatial clustering algorithm to the surveillance data at each time point and following changes in the spatial clustering statistic over time using a SPC or change-point detection method. In essence, this reduces a spatial surveillance problem to a univariate time-series surveillance problem. We examined Tango's method for global clustering followed by a cumulative sum [31] in a number of settings. Important modifications to this approach for practical use included the calculation of expected values from historical observations as opposed to population estimates, and modification to the statistic to enable consideration of multiple counts in a temporal interval as opposed to continuous updating following the arrival of each case. One drawback of global cluster detection approaches is that they identify the existence of clustering but do not identify the location of clustering [34]. In addition, many global clustering algorithms require a priori specification of a parameter corresponding to the scale of the clustering, although Tango [35] describes a method for searching over all possible clustering scales.

To detect focused clusters in data from a large number of spatial regions, we examined the space–time scan statistic [36,37]. This method identifies the location of the most anomalous cluster relative to the expected spatial case distribution. It does not require a priori specification of likely cluster locations or cluster size. As with the global cluster detection methods, an important issue encountered when using the space–time scan statistic for prospective surveillance is the need for stable, updated daily estimates of the expected spatial distribution of counts in order to avoid finding spurious clusters, although an approach based purely on permutation of observed values may avoid this [38]. We found test performance was acceptable when using recent data as a baseline with a temporal guard band to avoid bias towards the null hypothesis. As discussed earlier, the guard band separates the baseline data used to compute expected values from the current data under examination. This prevents an outbreak signal from affecting both the baseline and recent data, and thus reducing algorithm sensitivity. An attractive feature of the space–time scan statistic is that it can incorporate covariate information contained in individual records. This enhancement requires the calculation of expected values within each region for each combination of strata (e.g., age group, gender). The BCD and WSARE methods described in the previous section also incorporate covariate and spatial information. These algorithms, however, treat spatial location as a categorical variable, and this does not recognize the full value of spatial information. Another method developed through the BioALIRT project is the Bayesian Aerosol Release Detector (BARD), which aims to detect patterns in the data consistent with the aerosol release of a disease agent [39]. The BARD method uses a Gaussian model of agent dispersion, recent meteorological data, and Bayesian models of disease symptoms and the data under normal conditions to estimate the probability that recently observed surveillance data indicate an aerosol release of a disease agent into the community.

In general, available detection algorithms use spatial information differently. The simplest approach is to treat spatial location as categorical variable. This approach does not harness the full potential of spatial

information, but it does make minimal assumptions and allows any algorithm capable of considering categorical covariates to incorporate spatial information. Exploiting topology or connectedness requires making some assumptions about the influence of inter-region distance on disease clustering under normal and abnormal conditions. There is no well-defined set of methods for modeling space–time variation in surveillance data under normal conditions. We relied upon a historical baseline in many cases, but there is a need for research on this topic. Use of a space–time scan statistic minimizes the assumptions about the topology of clustering under outbreak conditions, but imposes assumptions about the geometry of disease patterns. For example, the space–time scan statistic assumes a contiguous cluster in the form of a circle or cylinder. While there is some work to generalize scan statistics to other shapes [40,41], it is important to note that in most situations it is computationally intractable to search over all possible space–time shapes to detect outbreaks. Ideally, an analyst would use epidemiological knowledge to limit the search to space–time shapes of practical interest, and further research is required to define these shapes. The BARD method is an example of how an algorithm can use knowledge about disease-agent exposure to limit the search to a specific geometric shape, namely an ellipsoid corresponding to a Gaussian plume. As with the need to understand the relationship between temporal signals and types of outbreaks, there is also a need to understand the relationship between space–time signals and types of outbreaks.

Space–time analytic methods are attractive for outbreak detection because they may help to direct attention to specific regions of interest, but their effective use faces some challenges. For example, application of space–time analytic methods requires calculating a stable estimate of the spatial distribution, and updating the distribution as it changes over time. Also, the spatial error inherent in relying on a single location for each record, usually home address, may limit the relationship between measured spatial location and true spatial disease patterns. For example, spatial analysis by home address may offer little benefit over purely temporal analysis if exposure occurs in the middle of a workday when few people are at home. Misspecification of exposure location is one of many potential sources of error in an analysis, but understanding the contribution of spatial information to outbreak detection in realistic surveillance settings is an important topic for future research.

3.4. Multiple data sources with and without spatial information

Public health agencies have traditionally not had access to multiple data sources suitable for surveillance of a single outcome. By "Multiple data sources" we mean separately collected data sources (e.g., clinical visits and pharmaceutical prescriptions) relevant to a single outcome (e.g., influenza cases), and this is distinct from a single data source with covariate information (Table 2). Analysis of multiple surveillance data sources is becoming feasible as surveillance data are increasingly available electronically, but the research literature on the analysis of multiple surveillance data sources is limited. We examined two approaches to the analysis of multiple data sources. One approach is to compute algorithm outputs for each data source separately with univariate methods, and then to combine the separate results using a second method. The other approach is to apply a multivariate algorithm to data from all sources simultaneously.

A straightforward approach to combining results from separate univariate analyses of different data sources is to pool the p-values from the separate analyses using a "consensus" method such as the one defined by Edgington [42]. This approach assumes that the data sources are independent, and in practice, false-positive detections tend to increase when more correlated series are pooled in this manner. One approach to pooling p-values that accounts for correlation among data sources is a Bayesian network [43]. Initial results with this approach are encouraging and this is a promising area for future research. An approach to conducting simultaneous analysis of multiple data sources is to use multivariate SPC (MSPC) methods that explicitly account for the covariance among data sources. We examined some MSPC methods for this purpose but found them overly sensitive to changes in the covariance structure, such as a fall-off in counts from a single data source, which are not of interest in a surveillance context. More specifically, in a blinded comparison of detection algorithms, we found that MSPC methods tended to have greater sensitivity than univariate methods, but the MSPC methods also produced more false alarms, which weakened their overall performance [26]. A modified version of the Hotelling's T^2 statistic following Ye et al. [44] avoided this oversensitivity to some extent, and future research on directional adaptations of the newer methods seems warranted. In situations where many related data sources were available (e.g., sales data for dozens of classes of over-the-counter pharmaceuticals) we explored approaches to reducing the number of data sources. Reduction was accomplished through unsupervised clustering algorithms, expert opinion, or a combination of the two. In general, data reduction improved detection performance when highly correlated data sources could be combined. We did not rigorously evaluate the impact on detection performance of different approaches to data reduction, pre-processing, or the potential effects on real-time decision-making

secondary to delays in data acquisition. These are important areas for future investigation. When covariates and spatial information were present, we examined the use of a multivariate space–time scan statistic [45]. This method searches for significant clusters with a stratified scan statistic defined by the addition of log likelihoods computed for the separate data types. Monte Carlo simulation studies demonstrate that this statistic detects anomalous counts across data sources while retaining sensitivity to spatial anomalies in individual data sources [45]. Another option in this context is to apply an algorithm such as WSARE, although we did not explicitly examine this over the course of the project.

Heuristically, combing all available information in a multivariate algorithm should give better results than univariate methods applied to data from each source followed by a hypothesis test based on separate outputs. In practice though, univariate surveillance analysis is easier to implement and there are a number of issues that make simultaneous multivariate analysis difficult in practice. The central problem is similar to that faced when using spatial information for surveillance. Definition of meaningful deviations in a multivariate sense requires not only careful modeling of data under the null hypothesis, but also explicit specification of data characteristics under alternative hypotheses. In univariate and spatial analyses, most analysts implicitly rely on the "omnibus" alternative hypothesis, that something is not normal. With spatial and multivariate data though, more specific alternative hypotheses are required to guard against detection of statistically significant changes from baseline that are not of epidemiological importance. Specification of alternative hypotheses requires knowledge of outbreak patterns, though, and this knowledge is not generally available. For example, one research team has experience using the Kalman filter formalism to analyze multiple data sources such as radar returns in an attempt to detect patterns attributable to distant aircraft or ballistic missiles [20]. This analytic approach is analogous to following multiple health-related data source in order to detect patterns attributable to epidemics. We were not able to apply the Kalman filter formalism to multiple data source surveillance though, because there is insufficient knowledge of the dynamic patterns in surveillance data attributable to disease outbreaks. Understanding outbreak patterns in high-dimension data and translating these patterns into specific outbreak detection hypotheses is an important area of future research.

3.5. The evaluation of outbreak detection algorithms

The research teams in the BioALIRT project invested considerable effort in the evaluation of outbreak detection algorithms. These efforts included development of test data, work with evaluation metrics, and participation in blinded algorithm evaluation studies. Siegrist describes in detail the methods for the most recent blinded evaluation [26]. Briefly, the blinded evaluation involved presenting each algorithm development team with test data sets containing unlabelled outbreaks. Each team prospectively applied their algorithms to the test data and declared the location of suspected outbreaks before the true locations of outbreaks were released to allow evaluation of detection performance. In this section, we summarize our experience with test data and evaluation metrics.

3.6. Test data

We defined test data as data containing known outbreak signals used to define the performance characteristics of algorithms. The role played by test data in algorithm research is a crucial one. The validity of conclusions regarding algorithm performance rests upon the validity of the test data used to define performance. Ideal test data contain a sufficient number of outbreak and non-outbreak periods for precise calculation of sensitivity, specificity, and timeliness, and a variation in outbreak signal that covers some range of plausible outbreak type and size. The signal for each outbreak should be clearly defined in terms of onset, completion, and possibly the nominal and desired alert times or points of successful detection. Test data can be described as: (i) wholly authentic; (ii) wholly simulated; or, (iii) simulated outbreaks superimposed onto authentic data (Table 4).

Wholly authentic data are appealing due to their face validity. They allow evaluation of the true effectiveness of detection algorithms with real data. Algorithms must address the challenges of real surveillance data and the outbreak signals reflect true disease activity in the population. The project used wholly authentic test data for a blinded algorithm evaluation in 2003 [26]. Despite the intuitive appeal of this approach there are practical problems. A central problem is the definition of what constitutes an outbreak. In general, it is possible to derive outbreaks directly from the data alone or to use external information to define outbreaks. One approach to identifying outbreaks from the data alone is for a committee of experts to identify changes in the data that would warrant epidemiological investigation [26]. This is a time-consuming procedure whose reliability is unknown. An algorithm to nominate potential outbreaks for review by an expert committee can semi-automate the process, but the selection of the nominating algorithm may bias the evaluation. In principle, it is also possible to automate this process fully by using an algorithm that retrospectively examines data for outbreaks. This approach might be useful if there are large volumes of data to examine and there is a sufficiently large training set of known outbreaks to guide the retrospective

Table 4
Advantages and disadvantages to different methods for generating test sets for evaluation of outbreak detection algorithms

Type of test set	Advantages	Disadvantages
Wholly authentic	Face validity; authentic background and outbreak signal	Resources required to define outbreaks; validity and reliability of outbreak identification may be poor and difficult to assess; limited number and variety of outbreaks
Wholly simulated	Exact specification of outbreak signal; large number of test sets possible; can be simple to develop; enables sensitivity analyses	Complexity of simulating background and outbreak signal; validity may be poor and difficult to assess; can require many parameter values
Simulated outbreaks superimposed onto authentic data	Greater face validity than wholly simulated test sets; exact specification of outbreak signal; large number of test sets possible; enables sensitivity analyses	Complexity of simulating outbreak signal; validity may be poor and difficult to assess; can require many parameter values

detection algorithm. Approaches to using external information to define outbreaks include consulting other surveillance systems in the same region or contacting public health agencies. Regardless of the approach taken to identifying outbreaks, the number and variety of existing outbreaks limit authentic test data. In some cases, such as for outbreaks resulting from bioterrorism, insufficient data exist to calculate precise estimates of detection performance. In cases where sufficient data do exist, such as for influenza outbreaks, the resources required to describe outbreaks adequately for use in an evaluation study will limit the number and variety of outbreaks available for a study. A final and important limitation of using wholly authentic data is that privacy concerns often limit the ability of multiple research groups to share test data for comparative evaluation of algorithms developed by different groups.

Wholly simulated test data are appealing for algorithm evaluation because they allow exact specification of the outbreak signal, perfect knowledge of the outbreak onset, and evaluators can create large amounts of test data. The BioALIRT project used wholly simulated test data for a blinded algorithm evaluation in 2002 and a number of research teams used this approach during algorithm development. Although wholly simulated data allow statistical precision in algorithm research, validation of both the background and the outbreak signal makes this approach problematic. This problem is non-trivial even in the single time-series context, and increasing the complexity of a simulation model to generate spatial and other covariate information magnifies the importance of this issue. To allow for meaningful evaluation of diverse algorithms, both normal and outbreak data must be simulated in a manner that ensures sufficient complexity and validity in terms of factors such as spatial patterns, temporal patterns, and joint distributions of variables. As a simulation model grows to meet these requirements, the number of parameters increases, the ability to verify the model becomes difficult, and ultimately it becomes more difficult to ensure the validity of the simulated data.

Superimposition of simulated outbreaks onto authentic data improves on purely simulated data by eliminating background modeling questions. The project did not use this approach in a blinded algorithm evaluation, but it was explored and research continues on this topic [46]. The main benefit of the superimposed or 'injected' outbreak approach is that the normal or baseline data need not be simulated. However, in addition to the need for a valid outbreak signal, algorithms must not be able to identify the simulated outbreak signal too easily in the background data. As a trivial example, a simulated signal superimposed on physician visit data may be easily detected if the signal does not account for the day-of-week effect in health care utilization. Sufficiently accurate modeling of the outbreak effects on the data is an important concern for this approach as it is for wholly simulated data. However, since early alerting is the primary purpose of these algorithms, one may limit the modeling complexity by bounding the focus of the simulation to the initial part of an outbreak and making plausible assumptions about individual and health care provider behavior before public and media reaction further skew the data [46].

In general, the approach to generating test sets should be tailored to the evaluation, and all the approaches outlined above play an important role in research on outbreak detection algorithms (Table 4). Wholly simulated test data are likely to be most useful for evaluation of algorithms under development and for comparative evaluation of algorithms with relatively simple data requirements (e.g., algorithms that operate on a single time-series). Wholly authentic test data are likely to be most useful when the data contain a sufficient quantity and variety of outbreaks of the type the algorithms are designed to detect, and when resources exist to identify outbreaks in the data. Simulated outbreaks superimposed onto authentic data are likely to be most useful when there is a need to evaluate algorithms and surveillance systems rigorously under a range of assumptions in a realistic setting. Regardless of the approach used to generate test data,

algorithm evaluations do not necessarily reproduce the context in which surveillance systems operate. In other words, algorithm evaluations help one to understand the performance of algorithms for finding anomalies, but cannot characterize the influence of algorithm performance on public health decision-making. Finally, it is worth noting that outbreak signals of different fidelity and complexity are needed to evaluate different algorithms. In a comparative algorithm evaluation, this implies that the most complex algorithm dictates the required fidelity and complexity of the outbreak signal. For example, evaluation of a time-series algorithm requires a test data set with a signal that is valid only in the time domain, whereas evaluation of an algorithm, such as WSARE, that examines the joint distribution of many variables requires a more complex outbreak signal that is valid in terms of time, space, and other covariates.

3.7. Evaluation metrics

Sensitivity and specificity are two commonly used evaluation metrics for outbreak detection algorithms (Table 5). When calculated over a range of parameter settings for an algorithm, the set of sensitivity/specificity values can be plotted to determine the receiver operating characteristic (ROC) curve and the area under the ROC curve [47]. While these metrics summarize the overall ability of an algorithm to detect outbreaks, they do not evaluate the timeliness of detection. Given that a main motivation of the project is to develop algorithms for rapid detection, we adopted the activity monitoring operating characteristic (AMOC) [48]. Fig. 2 shows an example AMOC curve. An AMOC curve associates a timeliness score to each false alert rate of interest. For our score, we used the median time to alarm following the onset of an outbreak, and we calculated this score for a few false-positive values chosen to be practical from the public health perspective. This measure gives a concise timeliness summary, but it does combine short- and long-duration outbreaks without discrimination. We found the use of *median* timeliness in the setting more robust than *mean* timeliness, as it is difficult to account for missed detections when calculating mean timeliness. One possible approach is to use the duration of the outbreak, or the interval between the beginning of the outbreak and likely detection by another means, such as reporting by concerned clinicians. In contrast, when using median time it is possible to assign missed detections an infinite value without unduly influencing calculation of timeliness as long as an algorithm detects more than half of the outbreaks.

A limitation of both the ROC and AMOC curves is that they do not evaluate the ability of an algorithm to identify the geographic location of an outbreak. One approach to evaluating the spatial accuracy of an algorithm that identifies cluster locations is to use the free response operating characteristic (FROC) curve [49]. We explored the use of this evaluation metric [45] but did not use it in any blinded algorithm evaluations. There are a number of other evaluation metrics described in the literature. Evaluation studies of change-point detection algorithms tend to employ other metrics [8] including average run length (ARL), predictive value (PV) of an alarm, and probability of successful detection (PSD). We did not routinely employ these evaluation metrics in our research, but others have found them to be useful.

Table 5
Metrics used for evaluation of outbreak detection algorithms

Evaluation metric	Description
Sensitivity	The probability of alarm given that an outbreak occurs
Specificity	The probability of no alarm given that an outbreak does not occur
Predictive value	The probability that an alarm is truly an outbreak
ROC[a] curve	The curve defined by plotting sensitivity (or true positive rate) against 1– specificity (or false-positive rate) for a range of algorithm parameter settings
Area under ROC curve	Summarizes the detection performance of an algorithm. Values over 0.5 indicate that the algorithm is better than a random detection scheme
AMOC[b] curve	The curve defined by plotting a summary measure of time to alarm given an outbreak occurs against false-positive rate for a range of algorithm thresholds
FROC[c] curve	The curve defined by plotting the fraction of outbreak locations detected against false-positive detection rate for a range of algorithm thresholds
ARL[d]	The expected time until the first detection (ARL_0); the expected time until an alarm when there is an outbreak at the initiation of surveillance (ARL_1)
PSD[e]	Probability of an alarm before some critical point in the outbreak given that the outbreak is detected

[a] ROC, receiver operating characteristics.
[b] AMOC, activity monitoring operating characteristics.
[c] FROC, free response operating characteristics.
[d] ARL, average run length.
[e] PSD, probability of successful detection.

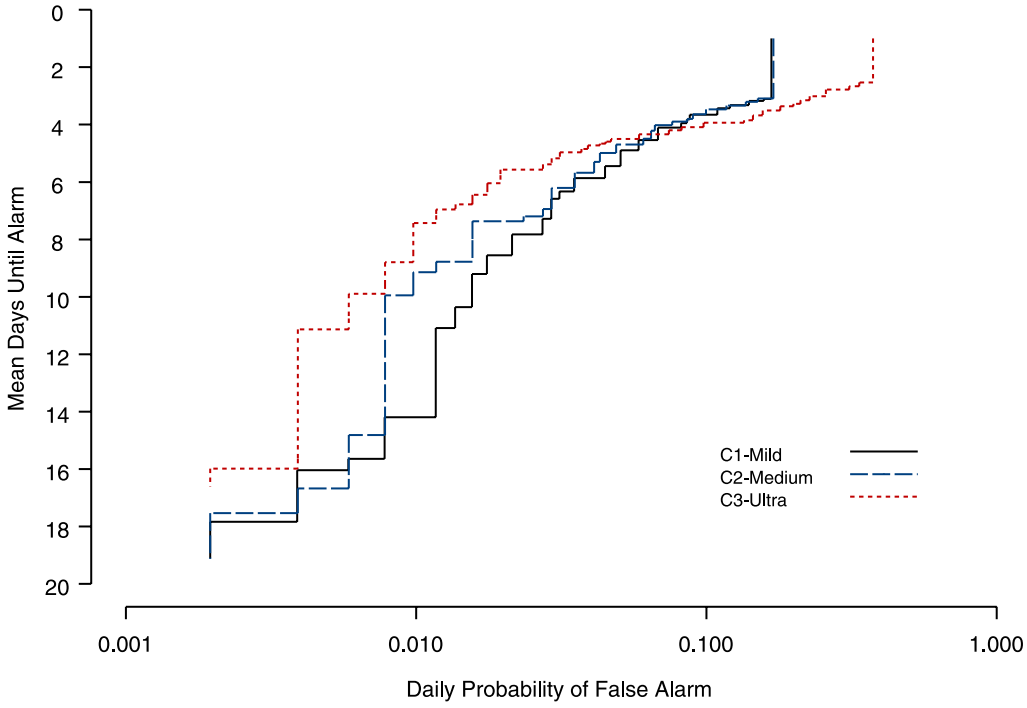

Fig. 2. Activity monitoring operating characteristics (AMOC) curves for three algorithms.

4. Discussion

Algorithms for outbreak detection can play an important role in surveillance. We have described a practical classification for outbreak detection algorithms that is based on the information requirements and functional characteristics of algorithms. We found this framework useful in our algorithm research and we intend for this classification to be accessible to surveillance practitioners and to demonstrate the practical application of algorithms for researchers. Our framework also provides a foundation for the development of a more formal classification, or ontology, of outbreak detection algorithms that might facilitate selecting algorithms to accomplish specific surveillance tasks [50]. Assessment and comparison of algorithms require objective evaluation methods. Rigorous evaluations of outbreak detection performance are rarely reported in the literature, and the methodology for such evaluation is not well developed. We noted particular deficiencies in methods for generating realistic test data and in evaluation metrics for assessing the timeliness and spatial precision of outbreak detection algorithms.

In this paper, we have synthesized research on outbreak detection algorithms conducted by the academic and industry groups in the DARPA BioALIRT project. Algorithms used for outbreak detection in public health traditionally operate on low-dimensional data, typically univariate time-series. We identified many algorithms suitable for use in this context. The main shortcomings of these algorithms were the modifications required to apply them to real surveillance data, and the lack of published evaluation of these algorithms in realistic settings. There are few existing algorithms well suited to outbreak detection in more information-intensive contexts. Some approaches exist to exploit spatial information (such as using patient residence to find significant case clusters), but very few methods prospectively seek anomalies in the distributions of other covariates. Researchers in this project developed methods capable of using covariate information for prospective surveillance and there remains a pressing need for development of outbreak detection algorithms that can operate in information-intensive contexts. There is also a need for rigorous comparative evaluation to determine the practical benefit of using covariate and spatial information from multiple data sources in surveillance. Such evaluation will require development of test data for this purpose, and detailed studies in a prospective setting such as others have conducted for retrospective algorithms [34].

In synthesizing our research on outbreak detection algorithms we noted a number of specific findings that relate to the process or practice of analysis for outbreaks. We found that when analyzing data sources without spatial information, there are many potentially applicable methods and the main issues are adapting these methods to prospective analysis in realistic settings, and selecting the method most appropriate for each setting. For example, reliable denominator data are often not available for surveillance data sources. Algorithms that require denominator data to account

for changes in risk may require considerable adaptation to operate in this setting. We also found that one must consider the vagaries of available data sources in selecting outbreak detection algorithms. For example, the amount of historical data available and the frequency of counts should influence the selection of a method. Regression methods require more baseline data than statistical process control (SPC) methods, and simpler SPC methods are likely to outperform more complex methods in settings with a low frequency of data counts.

We also noted some specific findings about the process of analyzing data sources with spatial information, and the process of analyzing multiple data sources. A fundamental consideration in the analysis of a data source with spatial information is whether, and how, to use topological information on adjacency or distance between sub-regions. Any outbreak detection method suitable for use with covariate information can use spatial information if topology is not considered. Methods that use topology may be more powerful, but they require an understanding of the nature of the spatial distribution, including how it changes over time. In settings where it is difficult to understand the spatial distribution, a simpler method that does not use topology may be preferable. A similar situation exists in the analysis of multiple data sources; namely whether to use the full covariance structure between data sources in an analysis. Simple strategies for combining results from multiple univariate analyses do not make use of the full covariance structure, but they may outperform more complex strategies that simultaneously analyze multiple data sources, especially if the full covariance structure is not well understood.

At a more general level, we also identified challenges that future research on outbreak detection algorithms should address. A common problem is to understand high-dimensional data under normal conditions sufficiently to forecast expected values with precision. In some situations, data exhibit complex patterns under normal conditions and this complexity may preclude direct application of methods developed in other domains where normative data behavior is better understood. Further work is required to define the principal factors influencing normal data behavior and to evaluate forecast models that incorporate these factors, especially in spatial and multivariate analysis. Another problem is to understand the data effects of outbreaks sufficiently to recognize epidemiologically meaningful deviations from normal. As the complexity of the data under analysis grows, for example by incorporating spatial and other covariate information, there are more epidemiologically irrelevant ways to observe anomalous patterns, and we need to understand how to explain away irrelevant patterns in an automated fashion as much as possible. There is a need to move beyond non-specific alternative hypotheses to improve specificity and provide informative detection results that are meaningful to public health practitioners. In addition to facilitating outbreak detection, improved knowledge of the data effects of outbreaks should also lead to an improved understanding of the etiology and management of outbreaks.

Our research project considered outbreak detection using a range of data types, but we were not able to consider all possible surveillance settings. For example, we tended to limit our research on spatial and space–time methods to approaches suitable for analyzing count data collected from spatial regions. As a result we did not consider many of the spatial analysis methods suited to spatial point patterns, although some of the methods we used are applicable to point patterns as well as regions. This limitation was a function of the available surveillance data, which for privacy reasons are usually aggregated to spatial regions such as ZIP codes or census regions.

The goal of this paper is to synthesize our research on outbreak detection algorithms. To accomplish this, we have had to limit detail and scope. For example, we have omitted many details about individual methods in order to allow consideration of the broader themes that allow a synthesis of research across different project teams. We have also limited our scope to the detection of deviations from expected in surveillance data, and we have not discussed issues such as adjustment of data to account for reporting delay. In a larger sense, we have also limited our scope by not considering the interplay between detection of deviations in surveillance data and public health decision-making. While the important areas of alert follow-up and consequence management are beyond this scope, outcomes such as intervention decisions and the costs, morbidity and mortality attributable to an outbreak, are how the systems that employ these algorithms must ultimately be measured.

5. Conclusions

The research conducted for this project has examined how algorithms can use large volumes of electronically available data to improve the timeliness of outbreak detection. An underlying hypothesis of this research is that algorithms that use the additional information found in covariates, spatial location, and multiple data sources can detect outbreaks more rapidly than algorithms that do not use this additional information. Our results tend to support this hypothesis, and we can identify at least two reasons why we are not able to offer stronger support. First, incorporating spatial and covariate information into a prospective analysis for a disease outbreak raises a number of methodological challenges. Forecasting expected values and identifying epidemiologically significant changes from

expectation in high-dimensional data are two challenges we identified. Second, it is difficult to demonstrate convincingly the performance of spatial and multivariate algorithms in realistic settings. There are few authentic data sets with known outbreaks available to evaluate these algorithms, and generation of spatial and multivariate test data sets through simulation is a difficult problem. Future research on detection algorithms and methods for evaluating outbreak detection must address these challenges.

We have tried to lay the foundation for research that will describe whether, and how, surveillance for rapid outbreak detection should move from analyzing a single time-series to analyzing covariate and spatial information collected from multiple data sources. Our research results provide initial support for such a transition. The results reported in the paper address the focused problem of anomaly detection in surveillance data. The larger problem of how these results can improve public health decision-making is an area that requires future research.

Acknowledgments

Many individuals contributed to the research described in this paper: Charu Aggarwal (IBM), Ronald Bloom (General Dynamics [GD]), Karen Cheng (GD), Greg Cooper (University of Pittsburgh [Pitt]), David Crary (GD), Denver Dash (Pitt), Jeremy Espino (Pitt), Vijay Iyengar (IBM), John Levander (Pitt), Chung-Sheng Li (IBM), Gene McClellan (GD), Mark Musen (Stanford), Milind Naphade (IBM), Daniel Neill (Carnegie Mellon University [CMU]), Robin Sabhnani (CMU), David Siegrist (Potomac Institute of Policy Research), Jeff Schneider (CMU), Paul Switzer (Stanford), Rich Tsui (Pitt), Walter Reed Army Institute of Research, Weng-Keen Wong (CMU), and Kun-Lung Wu (IBM). The Defense Advanced Research Projects Agency (DARPA) supported this research, and Ted Senator was instrumental as project manager in providing guidance and an environment for conducting this research.

References

[1] Wagner MM, Tsui FC, Espino JU, Dato VM, Sittig DF, Caruana RA, et al. The emerging science of very early detection of disease outbreaks. J Public Health Manag Pract 2001;7(6):51–9.

[2] Kaufmann AF, Meltzer MI, Schmid GP. The economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable?. Emerg Infect Dis 1997;3(2):83–94.

[3] Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V, CDC working group. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. Morb Mortal Wkly Rep 2004; 53(RR-5):1–11.

[4] Mandl KD, Overhage JM, Wagner MM, Lober WB, Sebastiani P, Mostashari F, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc 2003;11(2):141–50.

[5] The Defense Advanced Research Agency. Bio-surveillance system SOL BAA 01-17. Arlington, VA; 2000.

[6] World Health Organization. Report on technical discussions at the twenty-first world health assembly on national and global surveillance of communicable diseases. Geneva: World Health Organization; 1968. Report No.: A21/Technical Discussion/5.

[7] Brookmeyer R, Stroup DF. Monitoring the health of populations: statistical principles and methods for public health surveillance. New York: Oxford; 2004.

[8] Sonesson C, Bock D. A review and discussion of prospective statistical surveillance in public health. J R Stat Soc Ser A 2003;166(1):5–21.

[9] Lawson A. Large scale: surveillance. In: Lawson A, editor. Statistical methods in spatial epidemiology. New York: Wiley; 2001. p. 197–206.

[10] Koo D, Wetterhall SF. History and current status of the National Notifiable Diseases Surveillance System. J Public Health Manag Pract 1996;2(4):4–10.

[11] Lombardo J, Burkom H, Elbert E, Magruder S, Lewis SH, Loschen W, et al. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics ESSENCE II. J Urban Health 2003;80(2 Suppl. 1): i32–42.

[12] Lewis MD, Pavlin JA, Mansfield JL, O'Brien S, Boomsma LG, Elbert Y, et al. Disease outbreak detection system using syndromic data in the greater Washington, DC area. Am J Prev Med 2002;23(3):180–6.

[13] Tsui FC, Espino JU, Dato VM, Gesteland PH, Hutman J, Wagner MM. Technical description of RODS: a real-time public health surveillance system. J Am Med Inform Assoc 2003;10(5):399–408.

[14] Tillett HE, Spencer IL. Influenza surveillance in England and Wales using routine statistics. Development of 'cusum' graphs to compare 12 previous winters and to monitor the 1980/81 winter. J Hyg (Lond) 1982;88(1):83–94.

[15] Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting *Salmonella* outbreaks. Emerg Infect Dis 1997;3(3):395–400.

[16] Hunter JS. The exponentially weighted moving average. J Qual Technol 1986;18(4):203–7.

[17] Lucas JM, Saccucci MS. Exponentially weighted moving average control schemes: properties and enhancements. Technometrics 1990;32:1–29.

[18] Burkom H, Elbert Y, Feldman A, Lin J. Role of data aggregation in biosurveillance detection strategies with applications from ESSENCE. In: Syndromic surveillance: reports from a national conference, 2003, New York, NY. MMWR 2004;53(Suppl.):67–73.

[19] Williamson GD, Weatherby HG. A monitoring system for detecting aberrations in public health surveillance reports. Stat Med 1999;18(23):3283–98.

[20] Harvey A. The Kalman filter and its applications in econometrics and time series analysis. Methods Oper Res 1981;44:3–18.

[21] Zhang J, Tsui FC, Wagner MM, Hogan WR. Detection of outbreaks from time series data using wavelet transform. Proc AMIA Symp 2003:748–52.

[22] Wong W-K, Moore A, Cooper G, Wagner M. Rule-based anomaly pattern detection for detecting disease outbreaks. In: AAAI-02. Edmonton, Alberta; 2002. p. 217–23.

[23] Wong WK, Moore A, Cooper G, Wagner M. WSARE: What's strange about recent events? J Urban Health 2003;80(2 Suppl. 1): i66–75.

[24] Lai TL. Sequential changepoint detection in quality control and dynamical systems. J R Stat Soc Ser B 1995;57(4):613–58.
[25] Barnard GA. Discussion of Bartlet, "The spectral analysis of point processes". J R Stat Soc Ser B 1963;25:264–96.
[26] Siegrist D, Pavlin JA. BioALIRT biosurveillance testbed evaluation. In: Syndromic surveillance: reports from a national conference, New York, NY. MMWR 2004;53(Suppl.):152–8.
[27] Renz J. Qualitative spatial reasoning with topological information. Heidelberg: Springer-Verlag; 2002.
[28] Raubertas RF. An analysis of disease surveillance data that uses the geographic locations of the reporting units. Stat Med 1989;8(3):267–71. discussion 279–81.
[29] Duda RO, Hart PE, Stork DG. Pattern classification. 2nd ed. New York: Wiley; 2000.
[30] Besag J, Newell J. The detection of clusters in rare diseases. J R Stat Soc Ser A 1991;154:143–55.
[31] Rogerson PA. Surveillance systems for monitoring the development of spatial patterns. Stat Med 1997;16:2081–93.
[32] Rogerson PA. Monitoring point patterns for the development of space–time clusters. J R Stat Soc Ser A 2001;164:87–96.
[33] Jarpe E. Surveillance of the interaction parameter of the Ising model. Commun Stat Theory Methods 1999;28(12):3009–27.
[34] Kulldorff M, Tango T, Park PJ. Power comparisons for disease clustering tests. Comput Stat Data Anal 2003;42:665–84.
[35] Tango T. A test for spatial disease clustering adjusted for multiple testing. Stat Med 2000;19(2):191–204.
[36] Kulldorff M. A spatial scan statistic. Commun Stat Theory Methods 1997;26(6):1481–96.
[37] Kulldorff M, Athas WF, Feurer EJ, Miller BA, Key CR. Evaluating cluster alarms: a space–time scan statistic and brain cancer in Los Alamos, New Mexico. Am J Public Health 1998;88(9):1377–80.
[38] Heffernan R, Mostashari F, Das D, Besculides M, Rodriguez C, Greenko J, et al. New York city syndromic surveillance system. In: Syndromic surveillance: reports from a national conference, 2003, New York, NY. MMWR 2004;53(Suppl.):25–7.
[39] Hogan WR, Cooper G, Wagner MM. A Bayesian anthrax aerosol release detector. Pittsburgh, PA: RODS Laboratory, University of Pittsburgh; 2004.
[40] Patil GP, Taillie C. Upper level set scan statistic for detecting arbitrarily shaped hotspots: Center for Statistical Ecology and Environmental Statistics. University of Pittsburgh; 2002. Report No.: 2002-0601.
[41] Iyengar VS. On detecting space–time clusters. In: Tenth ACM SIGKDD international conference on knolwedge discovery and data mining. Seattle, WA; August 22–25 2004.
[42] Edgington ES. A normal curve method for combining probability values from independent experiments. J Psychol 1972;82:85–9.
[43] Russell S, Norvig P. Artificial intelligence: a modern approach. Upper Saddle River, NJ: Prentice-Hall; 1995.
[44] Ye N, Cheng Q, Emran S, Vilbert S. Hotelling's T^2 multivariate profiling for anomaly detection. In: IEEE workshop on information assurance and security. West Point, NY; 2000.
[45] Burkom HS. Biosurveillance applying scan statistics with multiple, disparate data sources. J Urban Health 2003;80(2 Suppl. 1): i57–65.
[46] Buckeridge DL, Burkom H, Moore AW, Pavlin JA, Cutchis PN, Hogan WR. Evaluation of syndromic surveillance systems: development of an epidemic simulation model. In: Syndromic surveillance: reports from a national conference, 2003, New York, NY. MMWR 2004;53(Suppl.):137–43.
[47] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:839–43.
[48] Fawcett T, Provost F. Activity monitoring: noticing interesting changes in behavior. In: Fifth ACM SIGKDD international conference on knowledge discovery and data mining; 1999.
[49] Chakraborty D. Statistical power in observer-performance studies. Comparison of the Receiver-Operating characteristic and free-response methods in tasks involving localization. Acad Radiol 2002;9:147–56.
[50] Buckeridge DL, Musen MA, Switzer P, Crubezy M. An analytic framework for space–time aberrancy detection in public health surveillance data. Proc AMIA Symp 2003:120–4.