

Research article

Open Access

## Modeling and detection of respiratory-related outbreak signatures

Peter F Craigmile\*<sup>1</sup>, Namhee Kim<sup>1</sup>, Soledad A Fernandez<sup>2</sup> and Bema K Bonsu<sup>3</sup>

Address: <sup>1</sup>Department of Statistics, 404 Cockins Hall, 1958 Neil Avenue, Columbus, OH 43210. USA, <sup>2</sup>Center for Biostatistics, M200 Starling Loving Hall, 320 West 10th Avenue, Columbus, OH 43210. USA and <sup>3</sup>Department of Pediatrics, Division of Emergency Medicine, Columbus Children's Hospital, 700 Children's Drive, Columbus, OH 43205. USA

Email: Peter F Craigmile\* - pfc@stat.osu.edu; Namhee Kim - kim.1862@osu.edu; Soledad A Fernandez - soledad.fernandez@osumc.edu; Bema K Bonsu - bonsub@columbus.rr.com

\* Corresponding author

Published: 5 October 2007

Received: 12 January 2007

Accepted: 5 October 2007

*BMC Medical Informatics and Decision Making* 2007, **7**:28 doi:10.1186/1472-6947-7-28

This article is available from: <http://www.biomedcentral.com/1472-6947/7/28>

© 2007 Craigmile et al.; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Time series methods are commonly used to detect disease outbreak signatures (e.g., signals due to influenza outbreaks and anthrax attacks) from varying respiratory-related diagnostic or syndromic data sources. Typically this involves two components: (i) Using time series methods to model the baseline background distribution (the time series process that is assumed to contain no outbreak signatures), (ii) Detecting outbreak signatures using filter-based time series methods.

**Methods:** We consider time series models for chest radiograph data obtained from Midwest children's emergency departments. These models incorporate available covariate information such as patient visit counts and smoothed ambient temperature series, as well as time series dependencies on daily and weekly seasonal scales. Respiratory-related outbreak signature detection is based on filtering the one-step-ahead prediction errors obtained from the time series models for the respiratory-complaint background.

**Results:** Using simulation experiments based on a stochastic model for an anthrax attack, we illustrate the effect of the choice of filter and the statistical models upon radiograph-attributed outbreak signature detection.

**Conclusion:** We demonstrate the importance of using seasonal autoregressive integrated average time series models (SARIMA) with covariates in the modeling of respiratory-related time series data. We find some homogeneity in the time series models for the respiratory-complaint backgrounds across the Midwest emergency departments studied. Our simulations show that the balance between specificity, sensitivity, and timeliness to detect an outbreak signature differs by the emergency department and the choice of filter. The linear and exponential filters provide a good balance.

### Background

Well-known, as well as previously uncharacterized infections continue to (re)emerge around the globe. To avoid

casualties from outbreaks of these infections and from the potential criminal uses of bioagents, surveillance systems are needed that have the capacity to identify such

outbreaks accurately and rapidly. The accuracy and timeliness of biosurveillance systems rests on the ability to model the uncertainty, severity, and aberrancy of clinical symptoms that are likely to portend disease outbreaks as expressed through the data monitoring system. Shmueli [1] summarizes the problems that biosurveillance systems, in general, pose to traditional statistical monitoring: (a) biosurveillance data may not be independent or stationary; (b) non-traditional data are assumed to contain earlier signature of an outbreak but this signal is weaker compared to actual diagnosis data; (c) since there are no data that contain bioterrorist outbreaks, outbreak patterns particularly as they would manifest in non-traditional data streams are unknown; (d) biosurveillance data are assumed to have no bioterrorist outbreaks, but the natural outbreaks add up to the background noise. The key issue at hand is to design statistical modeling and detection methods that can address these problems.

Among children, respiratory symptoms are an attractive target for surveillance. These are a prominent feature of many childhood epidemics and an early presentation of diseases like avian influenza, severe acute respiratory syndrome (SARS), and inhalational anthrax that have recently come to the public's attention. Unfortunately, respiratory complaints are also a feature of many common childhood illnesses, reducing the ability of biosurveillance systems to detect epidemics of greater public health concern. What is needed, therefore, is clinical information that is readily accessible and pre-processed in a manner that reflects the severity and aberrancy of respiratory symptoms. Using such data, discrimination between common childhood diseases and more serious respiratory epidemics would be possible.

Chest radiographs (X-rays), because they are readily available and are generally ordered by clinicians to evaluate respiratory complaints that are atypical or severe, have the potential to act as such a bio-monitoring and validation tool. In addition, detection based on models of radiograph ordering can indicate when in-depth follow-up is needed, as may occur when ordering of radiographs by clinicians is excessive for a given time of the year. Such in-depth review of radiographs may confirm clinical suspicions of an emerging epidemic or signal the need to perform a targeted review of medical charts to identify anomalous findings or groupings of aberrant findings that might herald the early stages of a respiratory-related outbreak.

In this article we consider time series methods for the modeling and detection of respiratory-related outbreak signatures based on chest radiograph ordering patterns from a number of pediatric emergency departments (EDs) located in the Midwestern region of the United States.

These models include ambient temperature records collected in each city, as a covariate. We use the temperature series as a surrogate measure of the annual influenza season. Also, a patient visit count series is included in the models to account for variations between-EDs (like ED sizes) and within-EDs (day-of-week, for example). Addressing the fact that the underlying process is neither "independent or stationary", our interest is to model the underlying "respiratory-complaint background", using the available covariates and significant temporal dependencies present in the data. Without modeling the spatial dependence directly, we investigate whether or not there is evidence of spatial homogeneity in the statistical models across cities. We use filter-based prediction methods to indicate evidence of respiratory-related outbreaks (e.g., due to anthrax attack) using chest radiograph data. We describe the form and function of various filters that are commonly used to detect outbreak signatures. Using a stochastic model for an anthrax attack, we assess the performance of these methods. Since there are no data that contain outbreak patterns, the use of a model is key to providing realistic outbreak patterns that can accurately be used to evaluate these statistical detection methods.

Reis, Mandl, and others use time series methods to detect evidence of disease outbreaks at Boston Children's Hospital [2-4], modeling specific clinical complaints as deterministic trend and seasonalities plus a stationary autoregressive moving average (ARMA) process. They repeat this process for the visit counts instead of considering a joint modeling procedure. The detection algorithm filters one-step-ahead prediction errors [5], and looks for values in this residual process that exceed a predetermined threshold. Their simulations are based on less realistic deterministic outbreak models. Ivanov, et al. [6] use the Exponentially Weighted Moving Average smoother to measure timeliness, sensitivity, and specificity of free-text chief complaints (information describing patient's status on the ED visit). They indicate that the methods are good for detecting relatively large seasonal outbreaks, but not for small outbreaks. Burkom, et al. [7] extend this approach using Bayes Belief Networks to improve detection sensitivity and timeliness. There are also a number of wavelet-based and smoothing-based methods that can be used to monitor and detect abnormalities of unknown form, occurring over different time scales [1,8,9]. The use of scan statistics [10-12] has also gained popularity in recent years. Many of the methods based on scan statistics use fixed-effect models for biosurveillance (also see, e.g., [13] and [14] for other methods based on fixed-effect models).

Our manuscript is organized as follows: we start by summarizing the data of interest, and propose a statistical model for the ED data in each city. Using the growing lit-

erature on the subject, we outline a stochastic model for an anthrax outbreak along with a healthcare utilization model for simulating people entering the ED. We then describe the methodology and theory for detecting unexpected outbreak signatures in time series sources using filters. We examine the form and function of the filters used for detection. In the results section, we explore the time series models obtained for each Midwest ED included in our study. We assess these detection methods using a simulated anthrax attack, and we end with a discussion.

## Methods

### A chest radiograph model for respiratory-complaints

The data of interest consist of daily counts of ED visits and chest radiographs taken between January 1st, 2003 and September 9th, 2004, in five metropolitan children's hospitals in the Midwest of the USA (Minneapolis/St. Paul, Milwaukee, Chicago, Akron, and Columbus), supplemented with time series of daily average temperature, obtained from the Average Daily Temperature Archive at The University of Dayton [15]. These series are shown in Figure 1. There seems to be strong seasonal component in the chest radiograph and visit series, for all the cities, which is negatively correlated with temperature. Although it could be argued that we do not have enough years of data to prove this empirically, it is expected that a higher number of respiratory complaints is associated with colder temperatures.

We now discuss two aspects of the statistical model for these daily chest radiograph counts: distribution and scale. Although the outcome variable of interest is counts, like other researchers in the field [1-4,6-9], we model the data using a Gaussian rather than Poisson process. The reasons are: (i) A normal approximation to the Poisson random variable is reasonable when the mean number of radiograph counts is large, as in this study; (ii) Gaussian time series models are easier to fit, diagnose, and interpret; (iii) The theory for filter-based methods commonly used for detecting the outbreak signature in such data is well developed for Gaussian processes. In some applications, count data are transformed when Gaussian models are used. Transformations, such as log or square root, are used to parameterize multiplicative models or to stabilize the variance. We use the original scale instead of transforming the data, like other researchers in the area, for ease of interpretation. This is especially important when we analyze filter-based detection methods in this article. For the same reason, we do not model the number of chest radiographs per daily ED visits (i.e., the proportion of radiographs). The filter-based theory, allows us to *directly* assess the effect of additive stochastic outbreak signatures upon the radiograph series.

In our study, the first ten months of data (Jan to Oct, 2003) were used as the training set, while the remaining ten months (Nov, 2003 to Sep, 2004) were used, as a test dataset, to evaluate the model and detection methods.

Since the goal is to predict the number of chest radiographs, for each city we fit a linear time series regression model, using the number of chest radiographs as the response, and the number of visits and temperature as predictor variables. We smooth the temperature series, because we believe that long-range temporal trends are more predictive of chest radiograph counts. Let  $\{R_{k,t}\}$  denote the number of chest radiographs for the ED in city  $k$  ( $k = 1, 2, \dots, 5$ ) on day  $t$  and let

$$R_{k,t} = \beta_{0,k} + \beta_{V,k} V_{k,t} + \beta_{T,k} T_{k,t} + X_{k,t} \quad (1)$$

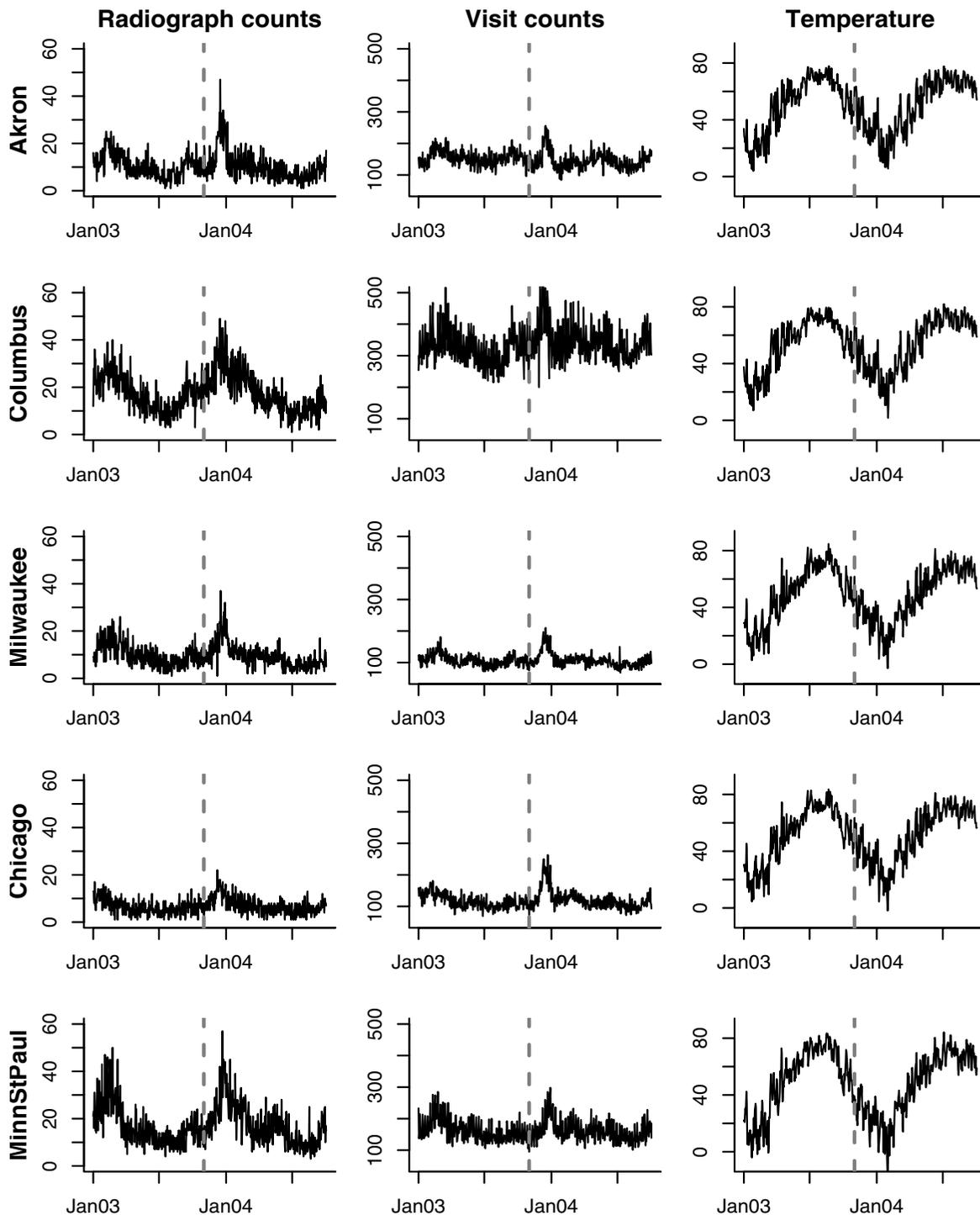
Here  $\{V_{k,t}\}$  are the visit counts for city  $k$  and  $\{T_{k,t}\}$  is the smoothed time series of temperatures for city  $k$ , filtered by taking a thirty day moving average to remove intra-month variation. To complete the model we assume that  $\{X_{k,t}\}$  is a zero mean stationary time series (we discuss the consequences of this assumption in the Discussion), that we shall represent using a seasonal autoregressive integrated moving average process (SARIMA). The SARIMA model (defined in the Appendix) allows for simultaneous modeling of dependencies on both the day as well as the weekly seasonal scales. To aid in the comparison of the dependencies across cities, the order of the time series model, as determined by choosing the autoregressive ( $p_k$  and  $P_k$ ) and moving average orders ( $q_k$  and  $Q_k$ ), will be the same for each city  $k$ .

### A stochastic model for an anthrax outbreak

We now propose a simple stochastic model for an inhalational anthrax outbreak, based on the work of Buckeridge, et al. and Brookmeyer et al. [16,17]. As proposed by these authors, our model incorporates two elements:

1. A stochastic model of infection and progression of the disease.
2. A model of health-care-utilization that, on a day-by-day-basis, tracks the behavior of each infected individual.

Any inhalational anthrax outbreak starts with the *dispersion* of the anthrax spores. Once spores are inhaled by a subject, in the *incubation* stage spores either *germinate* or are *cleared* out the lung. For the spores that germinate during incubation, the later stages of the disease are the *prodromal* and *fulminant* stages, followed by *death*. Buckeridge, et al. [16] model the spread of spores over a grid covering the Norfolk, Virginia region using a Gaussian plume model. Their model considers the source and



**Figure 1**  
**Time series plots.** For each of the five metropolitan children's hospitals, plots of the daily chest radiograph counts (left), daily ED visit counts (middle) and daily temperature (right). The vertical dotted line indicates the separation between the training and test data.

strength of the anthrax attack, along with prevalent wind directions.

Instead of using the region-based approach of Buckeridge, et al., we use the individual-based infection scheme of Brookmeyer, et al. [17]. Although it is known that anthrax spores can survive for long periods of time in the environment, we consider a small scale scenario. Since the population of interest is the individuals attending children's emergency departments, we assume an outbreak that affects a fixed number,  $N$  say, of children. Brookmeyer, et al. [17] define the infection probability of an individual exposed to inhalational anthrax using a competing risk model, modeling the dynamics of spore clearance and germination. Let  $\theta$  represent the hazard rate per unit time (days, say) that a spore is cleared from the lung and  $\lambda$  be the rate of germination. Suppose that each individual inhales a dose of  $D$  spores. Then, the probability that at least one spore germinates is called the attack rate (AR) and is calculated using a Poisson approximation, as

$$AR = 1 - e^{-D\lambda/(\lambda+\theta)}$$

For a given attack rate, the probability that at least one of the  $D$  spores germinates within  $t$  days is given by

$$F(t) = 1 - (1 - AR)^{(1 - \exp(-(\lambda+\theta)t))}$$

Note that the limit,  $\lim_{t \rightarrow \infty} F(t) = AR$ . Based on a statistical analysis of an anthrax outbreak that occurred in Sverdlovsk, Russia, Brookmeyer, et al. [17] estimate hazard rates of between 0.05 and 0.11 for  $\theta$  (which is compatible with  $\theta = 0.07$  obtained from animal studies for an AR of 0.5). The value of  $\lambda$  is not estimated in their data analysis – based on animal studies they found that the rate  $\lambda$  lies between  $5 \times 10^{-7}$  and  $10^{-5}$ . Buckeridge, et al. [16] propose log normal models for the duration of the prodromal and fulminant stages, based on the 2001 anthrax attack in the United States. The median duration of the prodromal stage was 12.18 days, with a dispersion of 1.41, and the median duration of the fulminant stage was 1.5 days with a dispersion again of 1.41.

Next, we describe a simplified health-utilization model for people entering the ED, based on ideas discussed in [16]. During the incubation stage, we assume that no infected people enter the ED. Non-infected people that enter the ED for other chest problems are part of the background data. During the prodromal and fulminant stages we assume a simple Markov model of utilization: each infected subject is a Bernoulli event, independent across days. At the prodromal stage people enter the ED with probability  $P_d$  on a weekday and  $P_w$  on weekends. At the fulminant stage the probability of entering the ED,  $P_f$  say, is larger than the probabilities in the prodromal stage. The

reason is that at the fulminant stage, the anthrax symptoms are similar to those of a heart attack and therefore people enter the ED with higher probability. The differentiation between weekday and weekend is irrelevant at this stage. We suppose that a small percentage of people in the prodromal or fulminant stages are misdiagnosed and thus need to re-enter the system. Also, we assume that people can potentially be misdiagnosed a maximum of two times during the same attack (10% in the first visit and 5% in the second visit). The probabilities of entering the ED after being misdiagnosed are increased by an additive factor,  $C$ , for every additional entry. Our model allows for a small probability of drop-out, to account for other ways of leaving the system (e.g., pharmacy visit). The health-utilization model could easily be extended to include, e.g., varying probabilities of entering the ED by stage/time, and/or more advanced ways to exit the system.

### Filter-based outbreak signature detection

The main idea of filter-based methods is to create a detection process  $\{D_{k,t}\}$ , for each time point  $t$ , which is a weighted average of the diagnostic or syndromic data to be used for the detection of an outbreak signature. The weights that appear in  $\{D_{k,t}\}$  are defined using the form of the time series model and a filter,  $\{a_i\}$  say. Extreme positive values of the detection process at a time point indicate a possible outbreak signature. The definition of the detection processes has its origin in process control [5], and is popularly used in the context of biosurveillance [1-4,6,8]. Naturally, the form of the filter has an important effect in the detection process.

A common parametric approach that we follow in this study is to first obtain the residual process by subtracting off the non-stationary part of the model (i.e., the effect of the covariates). We then define the filter  $\{c_{k,j}\}$ , that first decorrelates (whitens) this residual process using the one-step-ahead prediction errors and second filters this whitened series using  $\{a_i\}$  to yield the detection process  $\{D_{k,t}\}$ . Commonly used examples of the filter  $\{a_i\}$  include the differencing, moving average, or exponential filter. For a fixed value of  $\alpha$ , let  $1 - \alpha$  denote the specificity (the probability of no detection, when there is no actual outbreak signature). We declare evidence of an outbreak signature at time  $t$  if the detection process at that time point,  $D_{k,t}$ , exceeds a threshold  $\tau_{k,\alpha}$  calculated using the data or the process. Further details are given in the Appendix.

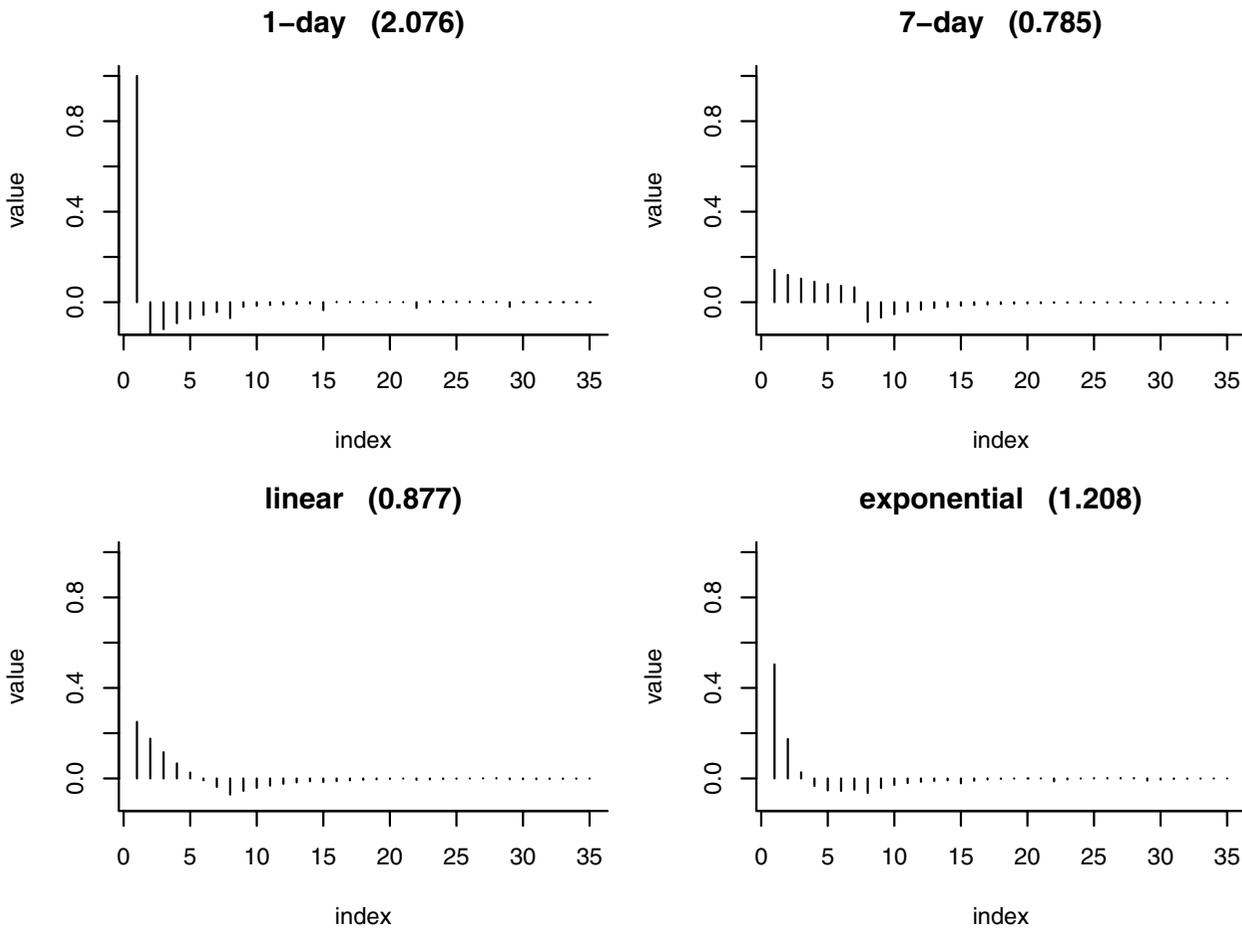
To understand the detection process  $\{D_{k,t}\}$  for different choice of filters,  $\{a_i\}$ , one should consider the filter involved in the calculation of  $\{D_{k,t}\}$  using the residual series,  $\{X_{k,t}\}$ . Using the results from the Appendix, at time point  $t$ ,

$$D_{k,t} = \sum_{h=0}^{L-1} a_h \sum_{j=0}^m c_{k,j} X_{k,t-h-j} = \sum_{l=0}^{m+L-1} f_{k,l} X_{k,t-l}. \quad (2)$$

In this expression,  $\{f_{k,l}\}$  is the filter defined by the convolution of the filters  $\{a_l\}$  and  $\{c_{k,l}\}$ . Hence, we conclude that the detection of outbreaks not only depends on the choice of filter, but on the statistical properties of the time series model which defines  $\{c_{k,l}\}$ . We will now focus on the effect of  $\{a_l\}$  (we will investigate the effect of changing the time series model in the Simulations section). As Reis, et al did [2], we examine four different filters,  $\{a_l\}$ , each of which is a form of difference filter. Each filter is an average of a number of days close to the time point minus a weighted average of the remaining values in the past:

1. 1-day filter:  $\{a_l\} = (1, 0, 0, 0, 0, 0, 0)$ ;
2. 7-day filter:  $\{a_l\} = (1, 1, 1, 1, 1, 1, 1)/7$ ;
3. Linear filter:  $\{a_l\} = (7, 6, 5, 4, 3, 2, 1)/28$ ;
4. Exponential filter:  $\{a_l\} = (64, 32, 16, 8, 4, 2, 1)/127$ .

The filters  $\{f_{k,l}\}$  for the detection processes vary according to the amount of autocorrelation within each time series. As an illustration, Figure 2 displays the  $\{f_{k,l}\}$  filter obtained when  $m = 28$  for each of the four  $\{a_l\}$  filters defined above, using a SARIMA model based on the Akron data (details of the SARIMA model are shown in the Results section).



**Figure 2**  
**Filter plots.** Plots of the four filters  $\{f_{k,l}\}$  that can be used to calculate the detection process  $D_{k,t}$  for the Columbus series. The number in parentheses after the filter name is  $\tau_{k,\omega}$  calculated using a normal approximation.

1. When  $\{a_t\}$  is a 1-day filter, the  $\{f_{k,t}\}$  filter consists of the current time point of the process minus a smaller weighted decaying average of the past values. The weekly seasonal terms in the SARIMA model are reflected in this filter.
2. When  $\{a_t\}$  is the 7-day filter,  $\{f_{k,t}\}$  is a weighted average of the last 7 days (with most weight on the current day), minus a weighted decaying average of the remaining past days. The weekly seasonal terms are not as strong, compared to the filter that is the convolution of the 1-day filter.
3. When  $\{a_t\}$  is the linear filter,  $\{f_{k,t}\}$  is an average of 6 days in the past. Far more weight is put on the current day relative to the previous 6 days. From this we subtract an average of values previous to the 6 days. Most of the weight from the second average comes from around days 8–10 in the past.
4. When  $\{a_t\}$  is the exponential filter,  $\{f_{k,t}\}$  is a combination of the 4 most recent days (mostly the current day, very little by the fourth day). We subtract an average of the past values (mostly days 6 – 11 in the past).

In Figure 2, the numbers within the parentheses at the top of each panel show the value of the threshold  $\tau_{k,0.03}$ , calculated using the normal approximation used in the Appendix (equation 10), assuming an innovation variance of  $\sigma_k^2 = 1$  for the process. We examine the effects upon the sensitivity, specificity, and timeliness to detect anthrax outbreaks in the simulations that we present later.

## Results

### Time series modeling

We fit the chest radiograph model given by (1) to the first ten months of data for each ED in city  $k$ . After specifying a model for  $\{R_{k,t}\}$ , we have a regression model with time series errors, that can be fit using standard maximum likelihood methods. We selected the order of the SARIMA model for the time series errors,  $\{X_{k,t}\}$ , as defined by (4) in the Appendix using standard identification techniques based on the sample autocorrelation and partial autocorrelation (e.g., [18]). To facilitate the comparison of the time series models across cities, we restricted to same order of model for each series. The orders  $p_k = 1$  and  $q_k = 1$ , correspond to a single autoregressive and moving average term on the daily scale, equivalent to observing an autoregressive process of first order with measurement error ([19], Exercise 2.9). With a seasonal period of seven days ( $s = 7$ ), we set  $P_k = 1$  and  $Q_k = 1$ , so that the random seasonal component is a combination of an autoregressive and moving average term, each of first order over a period of seven days. The seasonal component of the time

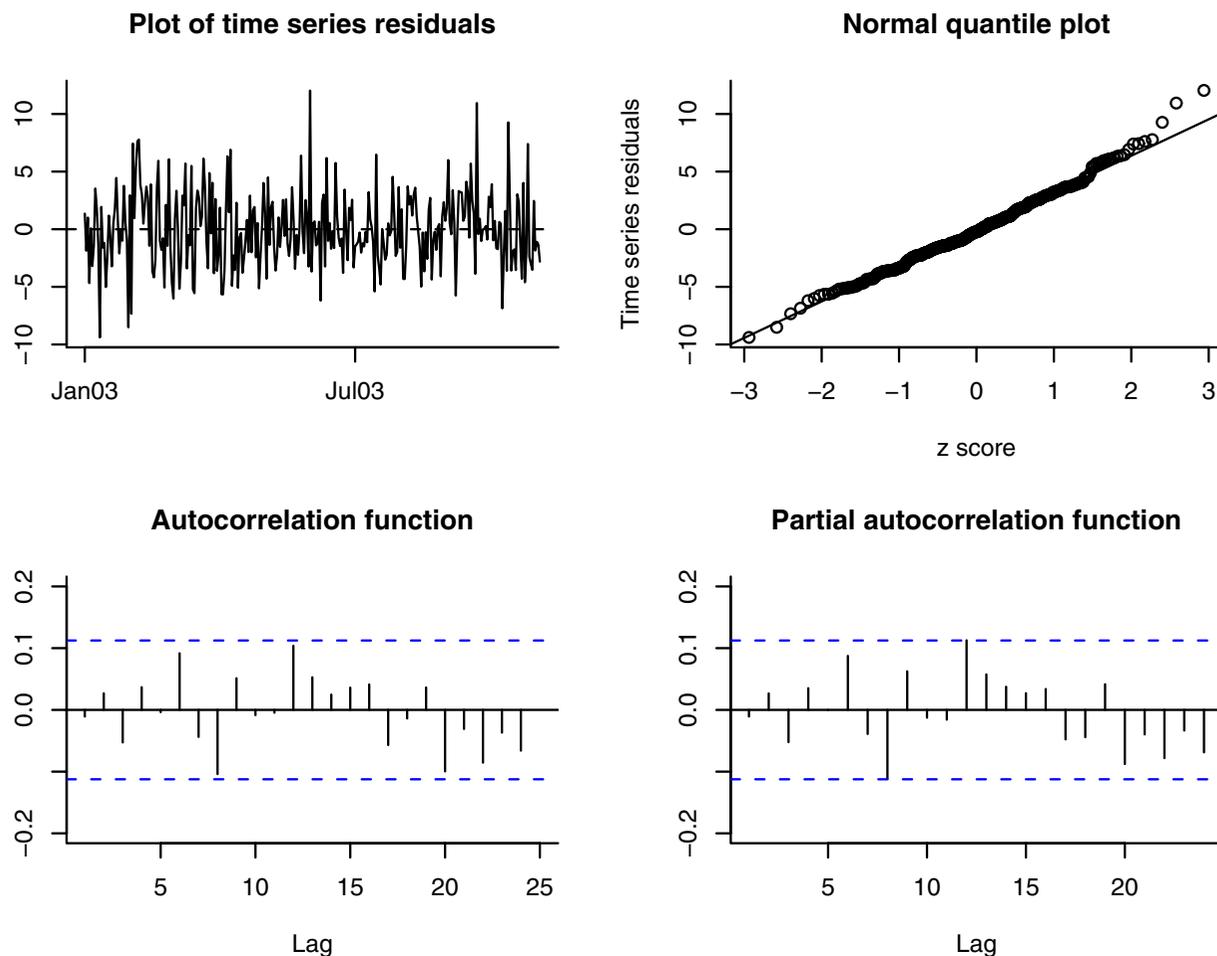
series model corresponds to an evolution of a first order autoregressive process with a measurement error over the weeks.

The model fit for the data in each city was assessed using diagnostic plots (time series plots, normal quantile plots, autocorrelation and partial autocorrelation plots, and the spectral estimates) of the estimated innovations of the time series component. Figure 3 shows some of these diagnostic plots for the Akron data (the residuals plots for the other cities were similar). Except for a few extremely positive values, the estimated time series looks stationary. The normal approximation is good, as evidenced by the straight line on the normal quantile plot. The plots of the sample autocorrelation and partial autocorrelations functions of the time series innovations lie inside the dotted confidence bounds for a white noise process (i.e., are samples of uncorrelated time series errors).

The parameter estimates for each city are summarized in Table 1. The standard errors for each parameter are shown in parentheses after each estimate. We note some homogeneities in the results across the EDs for each city. The intercept of the model,  $\hat{\beta}_{0,k}$ , is positive for all cities, being largest in magnitude for Columbus, and smallest in magnitude for Chicago. The parameter estimate,  $\hat{\beta}_{V,k}$ , relating visit counts and the radiograph counts is significantly different from zero, and positive, indicating, as one would expect, that a large number of ED visits are associated with a larger number of radiographs. The parameter estimate,  $\hat{\beta}_{T,k}$ , relating the smooth temperature and radiograph counts is negative, indicating a significant negative association between these two quantities across all cities. The values of the autoregressive and moving average term on the daily scale ( $\hat{\phi}_{k,1}$  and  $\hat{\theta}_{k,1}$ , respectively) are similar in value across all cities. Comparing these parameters, we can see that the strength of daily correlations is weakest for Milwaukee. Except for Chicago, the weekly seasonal autoregressive and moving average terms ( $\hat{\Phi}_{k,1}$  and  $\hat{\Theta}_{k,1}$ , respectively) are similar in value. By looking at the pointwise 95% confidence intervals for these two parameters, we conclude that there is no significant weekly variation in the Chicago series. As expected the values of  $\hat{\sigma}_k^2$ , the estimated variance of the time series innovations, differ by city.

### Simulations

In the simulations described in this section we used the following experimental design. For each city, we fit the



**Figure 3**  
**Time series residual plots.** Summary diagnostic plots of the time series residuals based on the first ten months of training data from the Akron ED.

regression time series model using the first ten months of data (training data). We used the second half of the data from November 1st, 2003 onwards and added outbreak-related counts to test for the detection of an outbreak signature.

We simulated an outbreak, as previously described, 500 times on 500 individuals using an attack rate of 0.5. The first day of the outbreak was randomly picked from a uniform distribution in the period December 5th, 2003 to June 22nd, 2004. (Making the earliest date later than November 1st, 2003, guarantees enough data for the filter-based methods to work with a prediction window of  $m = 28$  days). In the absence of other information (e.g., Buckeridge, et al. [16] do not report their probabilities) we

set the ED utilization probabilities to be:  $P_d = 0.25$  (week-day entry probability at the prodromal stage);  $P_w = 0.4$  (weekend entry probability at the prodromal stage) and  $P_f = 0.80$  (entry probability at the fulminant stage). The daily drop-out probability was set at 0.05. We assumed that with probability 0.9, the infected persons that enter the ED for the first time during a given outbreak receive a chest radiograph, i.e., 10% of the infected persons are misdiagnosed at the first visit. In a subsequent visit, 5% of the infected persons that re-enter the ED are misdiagnosed. The misdiagnosis additive factor,  $C$ , was assumed to be 0.05. We added the counts from the ED visits and subsequent number of chest radiographs generated from the healthcare utilization model on each day during the outbreak period. Let  $\{O_{k,t}\}$  denote the number of chest radi-

**Table 1: Parameter estimates obtained from the chest radiograph model for each of the different cities studied. The standard errors are shown in parentheses**

	Akron	Columbus	Milwaukee	Chicago	MinnStPaul
$\hat{\beta}_{0,k}$	10.98 (2.51)	12.73 (1.71)	10.84 (2.14)	3.18 (1.68)	9.15 (2.86)
$\hat{\beta}_{V,k}$	0.05 (0.01)	0.06 (0.00)	0.06 (0.01)	0.05 (0.01)	0.11 (0.01)
$\hat{\beta}_{T,k}$	-0.15 (0.03)	-0.24 (0.02)	-0.14 (0.02)	-0.05 (0.02)	-0.22 (0.03)
$\hat{\phi}_{k,1}$	0.91 (0.05)	0.92 (0.07)	0.79 (0.13)	0.96 (0.05)	0.95 (0.03)
$\hat{\theta}_{k,1}$	-0.74 (0.07)	-0.83 (0.11)	-0.69 (0.14)	-0.90 (0.07)	-0.89 (0.05)
$\hat{\Phi}_{k,1}$	0.88 (0.16)	0.88 (0.04)	0.83 (0.25)	-0.49 (0.34)	0.84 (0.08)
$\hat{\Theta}_{k,1}$	-0.84 (0.17)	-0.99 (0.06)	-0.77 (0.28)	0.42 (0.36)	-0.73 (0.10)
$\hat{\sigma}_k^2$	10.71	18.09	12.45	6.46	18.33

ographs attributable to the anthrax attack on day  $t$ . Since each simulated outbreak will be different, we start by summarizing the distribution of  $\{O_{k,t}\}$ . Figure 4 shows a plot of the 0.025, 0.5 (i.e., median), and 0.975 quantiles calculated over the 500 simulated realizations for each day  $t$ . Examining the progression of the quantiles over time allows us to explore the center and tails of the extra radiograph count distribution. The counts increase rapidly in the first week, then stay fairly constant (except for the spikes), for a week and then slowly drop to zero after the second week. The small spikes are due to the different probabilities of entry in the prodromal stage. Although the shapes at each quantile are similar, the magnitude and duration of the extra radiographs counts differ. The counts drop to zero at different time points for the different quantiles. They drop to zero after 3 weeks for the 0.025 quantile, after 4 weeks for the 0.5 quantile, and after 7 weeks for the 0.975 quantile. These patterns in the observed time-varying distribution of the outbreak signature are a strong motivation to not use deterministic outbreak patterns, as used by some authors, since deterministic outbreak patterns do not give a realistic assessment of detection methods.

We declare that there is evidence of an outbreak signature in the radiograph data at time  $t$  if the outbreak process  $D_{k,t}$  defined by (2), exceeds a given threshold,  $\tau_{k,\alpha}$ . Suppose that  $\{Y_{k,t}\}$  is the process defined as the sum of  $\{O_{k,t}\}$ , the extra radiograph counts attributable to the outbreak, and the radiograph process  $\{R_{k,t}\}$  that follows model (1). Removing the non-stationary part due to the estimated effect of the covariates yields

$$W_{k,t} \equiv Y_{k,t} - (\beta_{0,k} + \beta_{V,k}V_{k,t} + \beta_{T,k}T_{k,t}) \approx O_{k,t} + X_{k,t}.$$

By applying the filter to this series we obtain the detection process that we would observe in the presence of extra counts attributable to an outbreak:

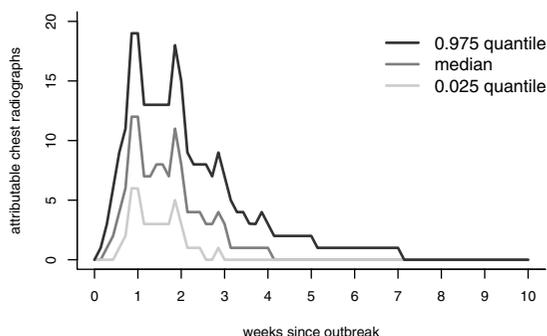
$$\sum_{l=0}^{m+L-1} f_{k,l}W_{k,t-l} \approx \sum_{l=0}^{m+L-1} f_{k,l}O_{k,t-l} + \sum_{l=0}^{m+L-1} f_{k,l}X_{k,t-l}.$$

To examine the effect of the filter upon the outbreak signature we examine the standardized quantity

$$g_{k,t} = \sum_{l=0}^{m+L-1} f_{k,l}O_{k,t-l} / \tau_{k,\alpha}$$

for the filters that we defined previously. We set  $\alpha = 0.03$ , which corresponds to a false alarm rate of one per month [2]. Then, as shown in the Appendix, the threshold,  $\tau_{k,\alpha}$  is chosen by solving  $P(D_{k,t} > \tau_{k,\alpha}) = \alpha$  for  $\tau_{k,\alpha}$ . Either we can estimate this value from the data using the  $1 - \alpha$  quantile of the  $\{D_{k,t}\}$  process of non-outbreak-based training data, or via a normal approximation, given by (10). There were some differences between the values of  $\tau_{k,\alpha}$  for the two methods in the simulations we studied. We chose the normal approximation method as it tended to preserve the specificity across the filters and EDs that we considered. Scaling  $g_{k,t}$  by  $\tau_{k,0.03}$  and using the same simulated radiograph realizations due to the outbreak, allow us to compare the filtered signals consistently across filters.

The filtered series will be different for each simulated outbreak pattern realization. Just as in Figure 4 with the radiograph series, we calculate the 0.025, 0.5 (median), and 0.975 quantiles of the filtered radiograph series on each day  $t$ , based on the model fit to the Akron ED series. Figure 5 shows the time-varying quantiles for each of the four filters. Relative to the no-outbreak situation, positive values



**Figure 4**  
**Quantiles of the chest radiograph distribution attributable to an anthrax outbreak.** Plots of the day-by-day quantiles of the radiograph counts attributable to an anthrax attack, based on 500 simulated anthrax outbreaks.

denote those time points for which there would be a greater chance of setting off a detection. Negative values denote the time periods that would actually decrease the chance of a detection. Except for the 0.975 quantile of the 1-day filter, each quantile of the filtered signal drops below zero for a period, and stabilizes at zero after that.

There are some similarities in the shapes of the three quantiles presented in Figures 4 and 5. Namely, the filtered signals increase over the first week, and then slowly decrease. The spikes in the distributions of the original signal are preserved for the 1-day filter, smoothed out for the 7-day filter, and partial smoothed over for the linear and exponential filters. Smoothing out the spikes tends to lead more sustained peaks (i.e., wider periods of higher sensitivity). The power of detection is smaller for the 1-day and 7-day than the linear and the exponential (as witnessed by smaller maximum heights in each of these two curves). For all filters, a drop below zero occurs immediately after the peaks; the counts go slowly back to zero. Periods for which the filtered signals are below zero inhibit detection and thus can mislead the prediction of the end of the outbreak signature.

We calculated the proportion of true non-detects in the absence of an outbreak signature, based on the test data (the specificity), as well as the proportion of true detects during the outbreak (the sensitivity) averaging over the 500 simulated anthrax outbreaks, for each filter and city. We calculated the actual specificity and sensitivity using the test data, for values of  $\alpha$ , in the range 0.01 to 0.10 in steps of 0.005. Figure 6 displays the difference of actual specificity (calculated using the training data) and the nominal specificity (predetermined  $1 - \alpha$ ) for different values of  $\alpha$ . Curves above or below the horizontal dotted line

at zero, indicate departures from the nominal  $\alpha$ . The Chicago series preserves the nominal value of the specificity (curves are clustered around the zero line). For the Minneapolis-St.Paul series the actual specificity is biased downwards, and for the other cities, the departure from the nominal value changes as  $\alpha$  increases. The choice of filter affected the calibration. To understand the tradeoff between the specificity and the sensitivity we show the receiver operator characteristic (ROC) curves, for each city, in Figure 7. For all cities, the 1-day filter has the poorest sensitivity. Except for Milwaukee, the 7-day filter tends to have higher sensitivities compared to the other filters, but also tends to have the poorest specificity. The exponential filter balances the specificity-sensitivity tradeoff.

Tables 2 and 3 display, respectively, the median and maximum detection times for each filter and city for our model. The 1-day filter performed poorly. The other filters are more comparable. We observed (results not shown) that the time to detection increases if the attack rate is reduced and also if the probabilities of entering the ED is delayed.

We now compare the performance of our model, as defined by equation (1), with three other models, in order to investigate the effect of different covariates and time series components upon outbreak signature detection. Remember  $\{R_{k,t}\}$  are the number of chest radiographs for the ED in city  $k$  on day  $t$ ,  $\{V_{k,t}\}$  are the ED visit counts, and  $\{T_{k,t}\}$  are the thirty day smoothed time series of temperature. For each day,  $t$ , and day of the week,  $d = 1, \dots, 7$ , let  $D_{d,t}$  be an indicator function that is one if that day is the  $d$ th day of the week, and zero otherwise. The four models we compare are:

1. Our covariates plus SARIMA errors model:  $R_{k,t} = \beta_{0,k} + \beta_{V,k}V_{k,t} + \beta_{T,k}T_{k,t} + X_{k,t}$  where  $\{X_{k,t}\}$  is the SARIMA model used in the Results section.

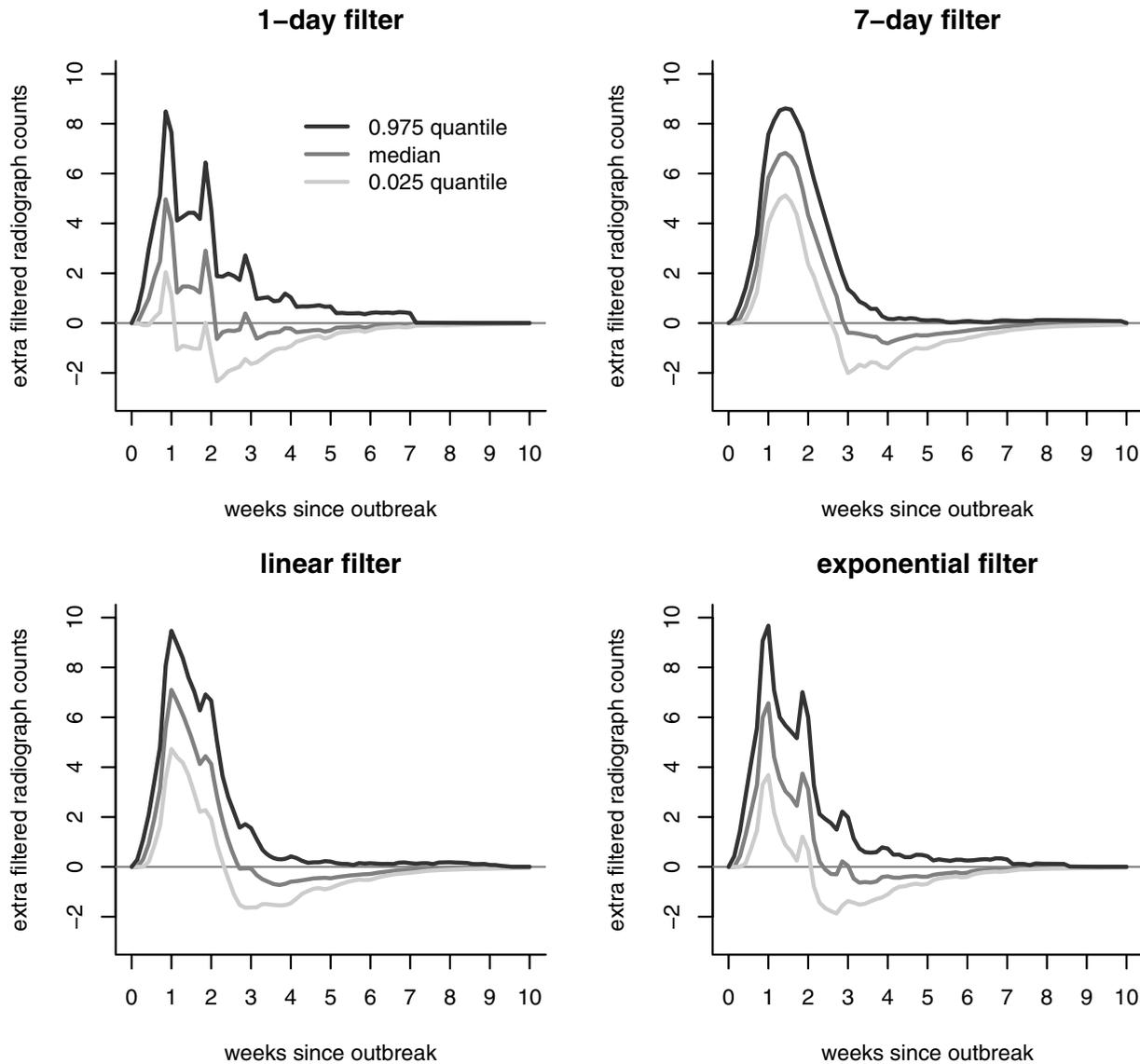
2. Covariates (with seasonality) with autoregressive moving average (ARMA) errors model:

$$R_{k,t} = \beta_{0,k} + \beta_{V,t}V_{k,t} + \beta_{T,k}T_{k,t} + \sum_{d=1}^6 \beta_{D,d,k}D_{d,t} + X_{k,t},$$

where  $\{X_{k,t}\}$  is an ARMA model (equation (4), without the  $\Phi_k$  and  $\Theta_k$  terms) with orders  $p_k = 1$  and  $q_k = 1$ . Instead of modeling the weekly effect using a random seasonal component we use day-of-the week as a fixed effect (covariate).

3. Covariates and no time series errors:

$$R_{k,t} = \beta_{0,k} + \beta_{V,t}V_{k,t} + \beta_{T,k}T_{k,t} + \sum_{d=1}^6 \beta_{D,d,k}D_{d,t} + X_{k,t},$$

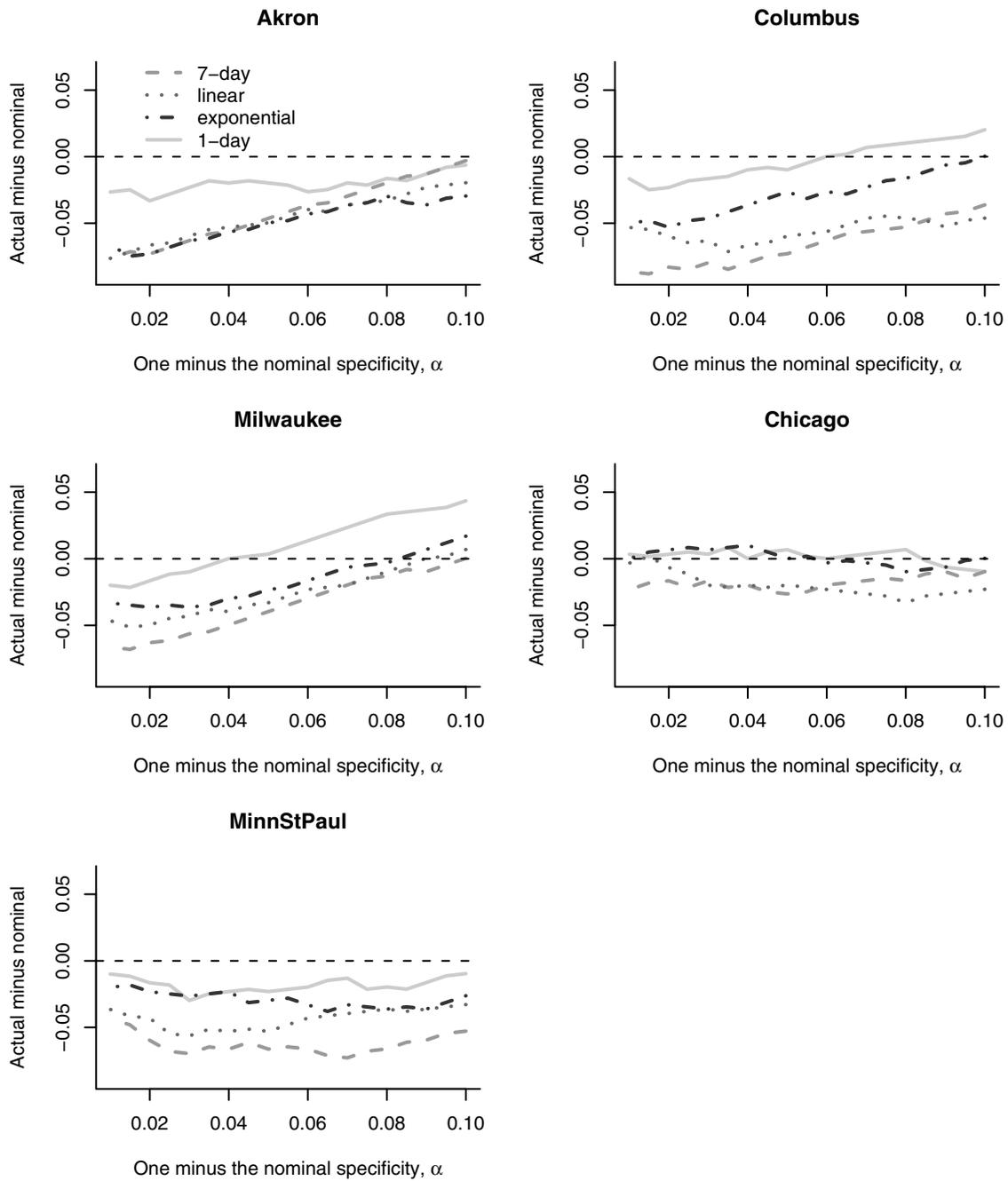


**Figure 5**  
**Quantiles of the filtered chest radiograph distribution attributable to an anthrax outbreak.** Plots of the day-by-day quantiles of the filtered radiograph counts attributable to an anthrax attack, based on 500 simulated anthrax outbreaks. The filtering operation, as defined by (3), is based on the first ten months of training data from Akron, in combination with one of four filters (1-day, 7-day, linear and exponential). Each filtered signal is scaled by  $\tau_{k, 0.03}$ , calculated using a normal approximation, for comparisons to be made across filters.

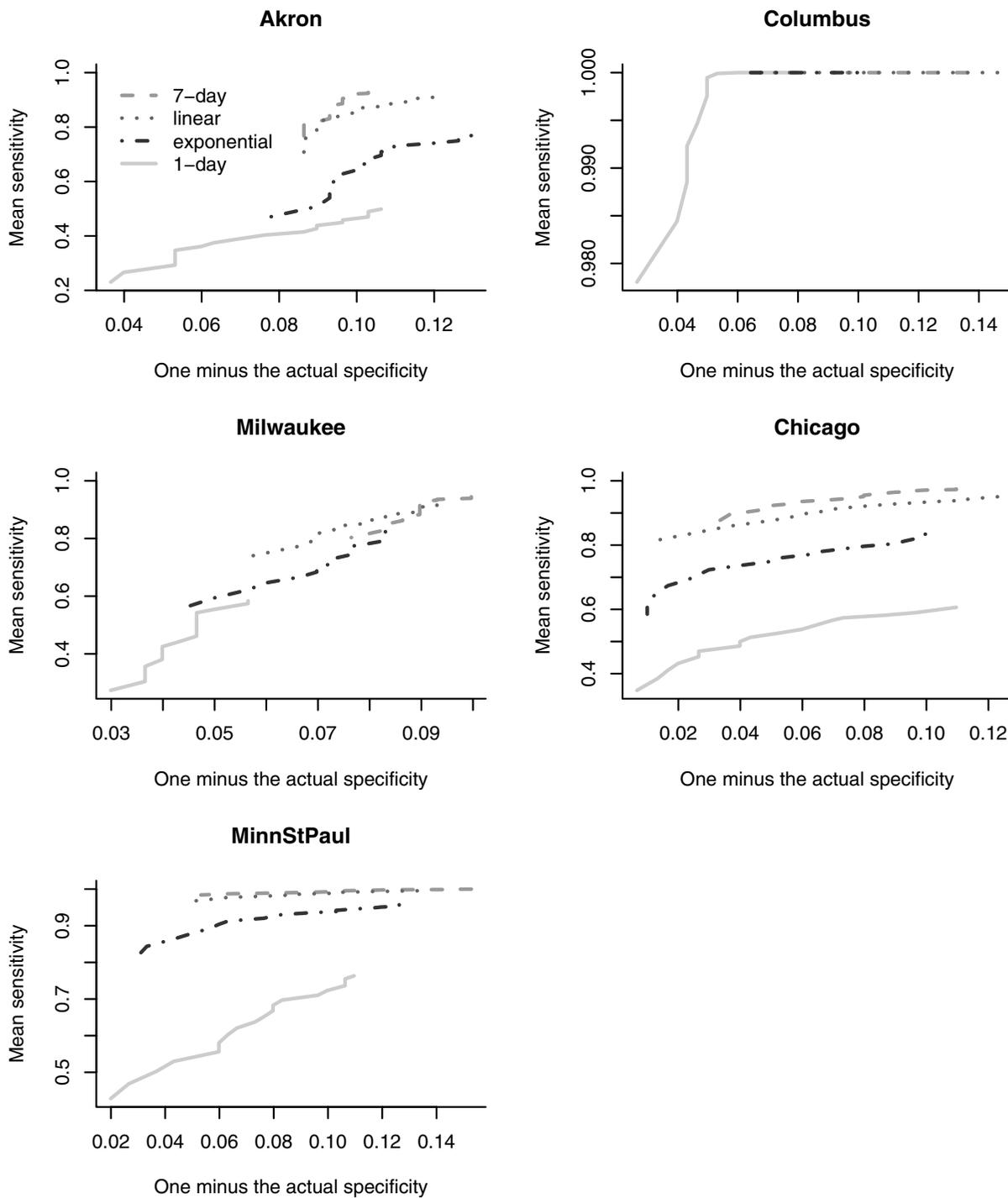
where  $\{X_{k,t}\}$  is a mean zero white noise process with innovation variance  $\sigma_k^2$ .

4. No covariates and ARMA errors:  $R_{k,t} = \beta_{0,k} + X_{k,t}$  where  $\{X_{k,t}\}$  is the same ARMA model as in Model 2.

Figure 8 displays the difference of actual specificity (calculated using the training data) and the nominal specificity (the predetermined  $1 - \alpha$ ) for different values of  $\alpha$  for these four models. For illustration we use the linear filter, which represents a compromise between the actual specificity and sensitivity (Across all four models, the 1-day filter always had specificities closest to nominal and the 7-



**Figure 6**  
**The biases in the specificity.** A plot of the actual specificity (the specificity calculated using the training data) minus the nominal specificity ( $1 - \alpha$ ) for different values of  $\alpha$ . The value of the detection threshold  $\tau_{k,\alpha}$  was calculated using a normal approximation.



**Figure 7**  
**Receiver operating characteristic curves.** Receiver operating characteristic curves (averaged over 500 simulations of an anthrax outbreak) for the detection of the anthrax outbreak signature in the chest radiographs at each of the five metropolitan children's hospitals.

day filter had specificities furthest away). Except for Columbus, Model 1 achieved a specificity closer to the nominal value across all EDs (Model 3 outperforms Model 1 in Columbus). Except in Chicago, Model 4 had the largest magnitude of bias.

Figure 9 compares the receiver operator characteristic (ROC) curves for the four models across the five EDs using the linear filter (Across all models, the sensitivity for the 1-day filter was the lowest while the 7-day and linear filters tended to have the highest sensitivity). To readily compare the models, we use one minus the nominal specificity on the x-axis. Except for Akron and Minneapolis/St. Paul, the random weekly and daily time series components in Model 1 yield higher sensitivities, compared to including a day-of-week covariate plus ARMA errors in Model 2, and day-of-week covariate plus no ARMA errors in Model 3. This is due to the insignificance of the day-of-week fixed effect in Models 2 and 3. In Akron, Model 3 has higher sensitivities than Model 1, suggesting that the time series errors are less relevant in outbreak signature detection for the test data. In Minneapolis/St. Paul Models 1, 2, and 3 have equivalent sensitivities. Across all cities either Model 2 or Model 4 had the lowest sensitivities, indicating that these models are less reliable for outbreak signature detection.

In terms of sensitivity, except for Chicago, Model 3 tended to outperform Model 2 for all filters in the cities we studied. This is counter-intuitive, as we would expect that a model that includes the significant ARMA time series component (Model 2) should outperform one that did not contain any time series components (Model 3). In practice there is a tradeoff between estimation and prediction. Table 4 displays the Akaike information criterions (AICs) for the four models, fit to the training data of the five EDs. For each ED, a smaller AIC denotes a model that better fits the data. Model 3 performed better than Model 2 in one-step-ahead predictions (Figure 9), but Model 2 better explains the training data than Model 3 (Table 4). Using the AIC we find Model 1 fits best to the training data, even though it does not always yield the highest sensitivity.

## Discussion

Our intention in this study was to find a flexible set of statistical models that could be applied across a number of emergency departments. We employ time series models that include covariates, such as patient visit counts and ambient temperature, as well as random seasonal terms. We use chest-radiograph ordering data from emergency departments of five regional Midwest children's hospitals to detect signatures of respiratory outbreaks. We include visit counts series as a covariate in the chest radiograph model to account for variations due to, for example, ED

sizes, changes of staff within the ED, and even some seasonalities across the time period of interest. We use the temperature series as a surrogate measure of the influenza season – the colder months in the western hemisphere. This is a more accurate measure of the influenza season than using a fixed covariate such as a sinusoid. To reflect uncertainty in the variation of the influenza background over seasons, these models allow for randomness in the seasonal components. The use of random seasonal components is an advantage over traditional fixed effect models, since temporal patterns are not assumed to repeat precisely the same way. Thus, signature detection capabilities are improved for the majority of EDs – sensitivity is higher. For increased accuracy and timeliness, the use of our model for the data analysis should represent one component of an integrated detection system. Once a signal is triggered by any of our models, we recommend the use of clinical follow-up to corroborate or refute the emergence of a bona fide epidemic. For example, radiographs and medical charts will need to be reviewed to identify highly anomalous findings or groupings of aberrant findings that would be expected to be present at early stages of outbreaks. We believe that the approach utilized in this work will aid in this process and is more appropriate than models using fixed periodicities that do not have the ability to capture the underlying variabilities across seasons.

Of note, our study shows that there are similarities in the chest radiographs series from different EDs that can, for the most part, be modeled by similar time series models. Similarities of the time series model across EDs have a number of ramifications for detection of outbreak signatures. First, by borrowing information across the different EDs we can build more complicated multivariate time series models, possibly involving the joint modeling of chest radiograph and visit counts across locations. Second, we could use these models to jointly detect outbreak signatures across large spatial regions. In this context, Diggle et al. [20] use spatio-temporal Cox models to identify anomalies (real and artificial outbreaks) in the space-time distribution of gastrointestinal infections. But, some caution is needed because one potential drawback of aggregating data spatially is that the chance of detection can be reduced (data from unaffected areas will mask the outbreak signature, and increase the detection time) [21]. For localized outbreaks, there is still some utility in building models that borrow strength across EDs, even if the joint detection of outbreak signatures is not meaningful. In terms of these localized outbreaks, a geographically close site could act as a benchmark to judge detections at other sites. For example, under certain circumstances an epidemic detection signal triggered in Columbus, but not in Akron, could imply that some unusual event had occurred in Columbus. It should be noted, however, that even though there are similarities in the time series models, our

**Table 2: Median time to detect an outbreak for each city and filter**

	1-day	7-day	linear	exponential
Akron	3	1	1	2
Columbus	1	1	1	1
Milwaukee	4	1	1	1
Chicago	2	1	1	1
Minn/St. Paul	2	1	1	1

simulations demonstrate that the sensitivity and timeliness to detect outbreak signatures using chest radiograph counts were different across EDs. These differences may limit the utility of our models in comparing signals across different EDs.

Our study has a number of limitations. We assume that the resulting series  $\{X_{k,t}\}$ , after accounting for these covariates, is stationary. We also ignore the fact that an individual may have multiple ED visits. Furthermore, these data do not contain any known anthrax outbreaks. Instead, outbreak patterns are simulated using a simple stochastic model. Although more care needs to be taken when summarizing simulation results, we agree with Buckeridge, et al. [16] and Brookmeyer, et al. [17] that stochastic outbreaks are more realistic than deterministic ones. Other results (not presented here) indicate that the results are different (and possibly misleading) with a smooth deterministic outbreak rather than a stochastic one. We utilize the four filters used by Reis, et al. [2]. Although filter design is an active area of research, especially in engineering, we decided to only study these four simple and fairly easy to understand filters. In the future we could extend our analysis to different choices of filter. We view wavelet-based biosurveillance methods [1,8] as extensions of this idea to different linear time-invariant filters. An important issue in filter-based-prediction methods is the choice of covariates and the correlation structure for the time series model. Clearly, the properties of the time series model could change with time. One way to improve the prediction of outbreak signatures is to use time-varying models, with parameters that slowly evolve over time. The use of time-varying parameter models is analogous to a periodic review of the parameter values. Window-based estimation methods could be used to re-

estimate the model parameters. In this case, the choice of window length is critical as it represents a compromise between the efficiency to estimate the model parameters, and the sensitivity and specificity of detection. Other issues to be taken into account include negative singularities (as defined by [8]) that can cause a drop in the sensitivity. Abnormal points (like holidays or severe weather) can cause "false alarms". Our solution to this problem is to jointly model chest radiograph and visit counts. So far, we have assumed that the visit counts are observed without error. Related to these issues, we are investigating methods that can be used to incorporate the uncertainty in parameter estimation upon the detection procedure.

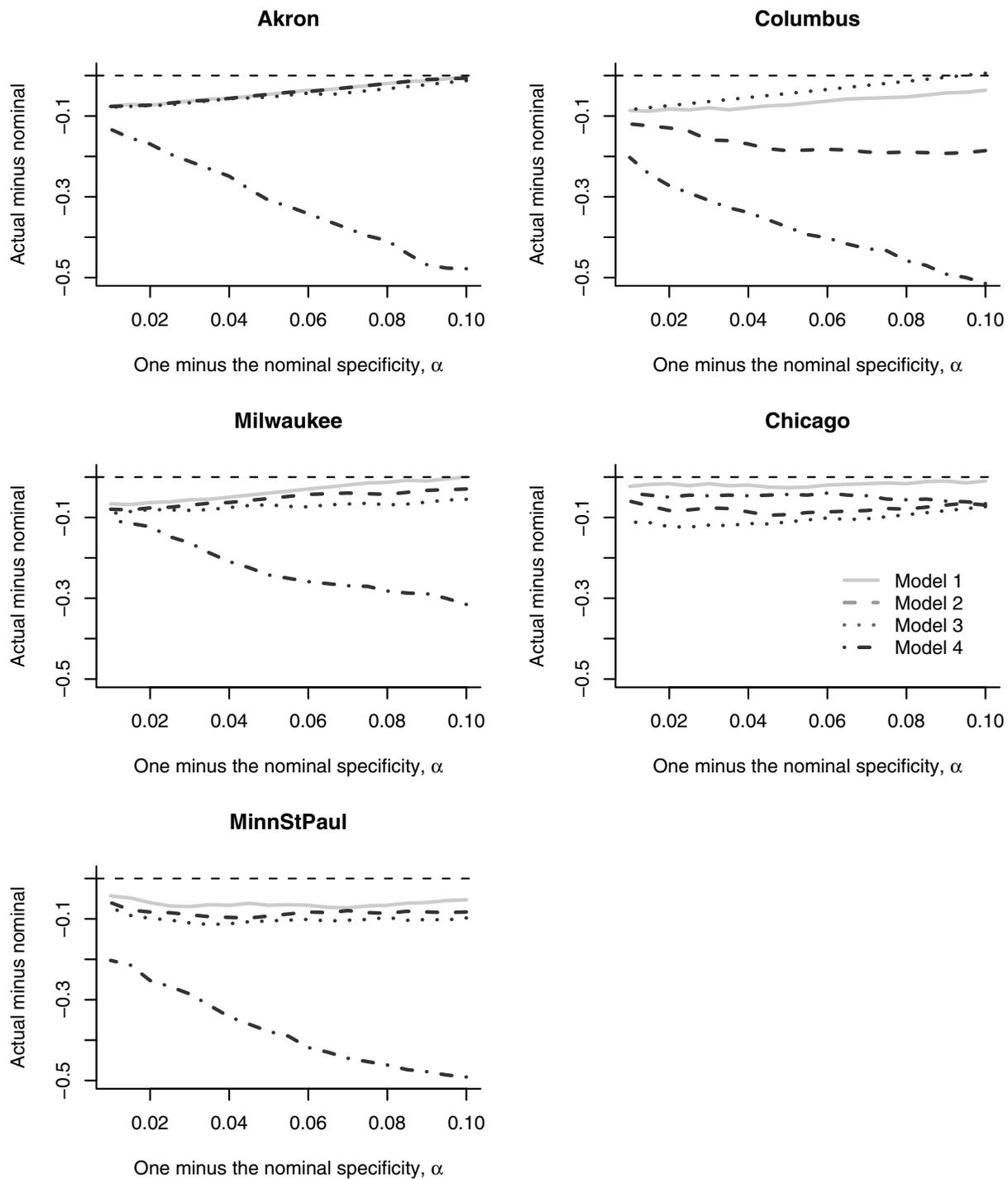
**Conclusion**

We present in this paper a stochastic model of chest radiograph ordering patterns and temperature as an adjunct to a biosurveillance system for detecting emerging respiratory-related epidemics, focusing on a potentially high-impact public health hazard such as inhalational anthrax. We show that in time series analysis of respiratory-related data it is important to capture important seasonal effects that are present in such data, as well as to consider the influence of important covariates that can be easily obtained and incorporated into the models. We also demonstrate the importance in assessing the sensitivity and specificity of these methods, of utilizing more realistic stochastic, rather than deterministic, models for outbreaks patterns. We demonstrate spatial homogeneity in chest radiograph data across EDs and suggest ways in which these observations may be used to improve regional biosurveillance for (re)emergent infections.

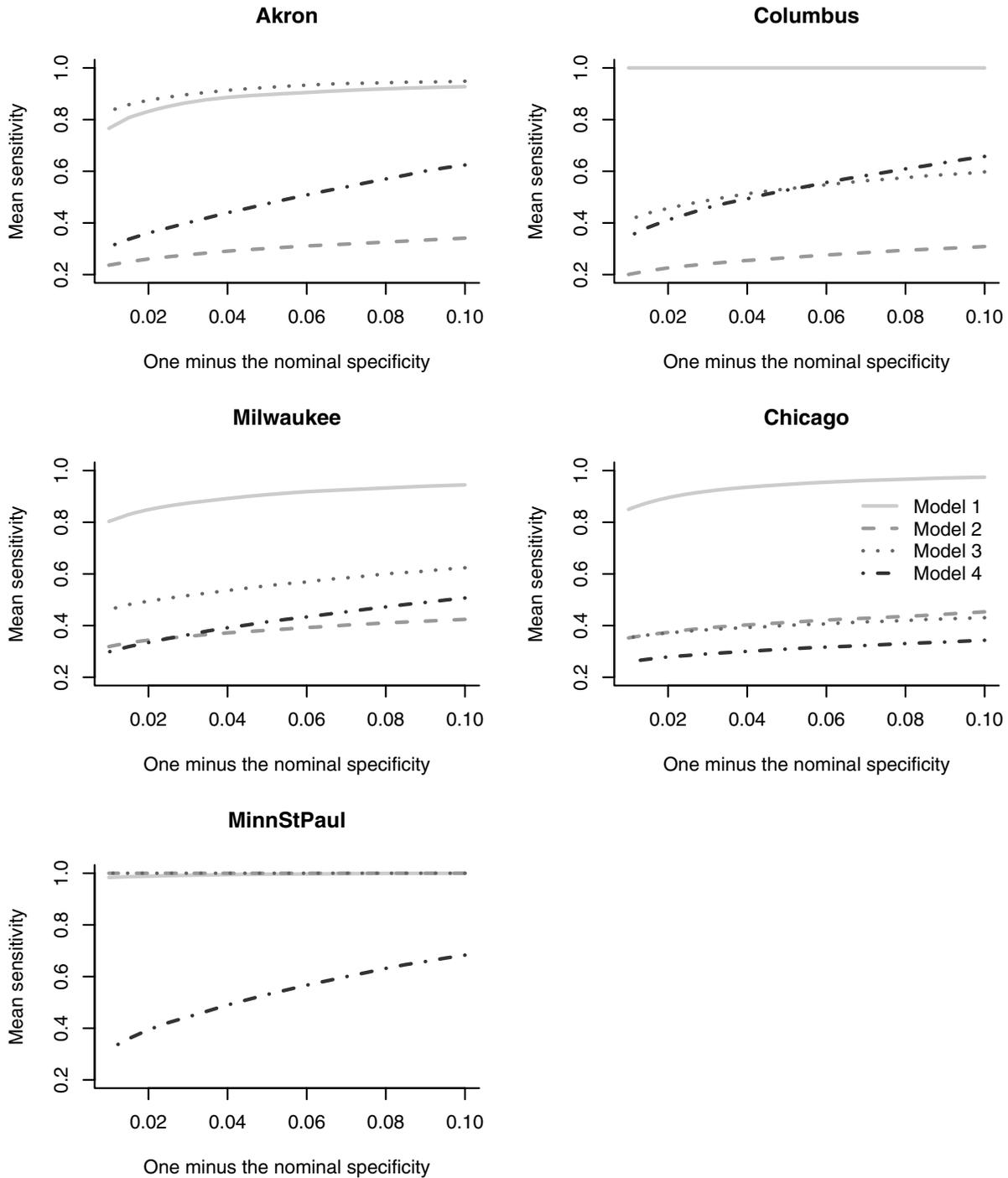
Regardless of the choice of filter, our simulations demonstrate that the specificity calculated using the training data

**Table 3: Maximum time to detect an outbreak for each city and filter**

	1-day	7-day	linear	exponential
Akron	9	9	7	7
Columbus	2	1	1	1
Milwaukee	9	7	7	7
Chicago	7	4	4	5
Minn/St. Paul	6	2	4	5



**Figure 8**  
**The biases in the specificity for the four models.** A plot of the actual specificity (the specificity calculated using the training data) minus the nominal specificity ( $1 - \alpha$ ) for different values of  $\alpha$  using a linear filter, for four different models. The value of the detection threshold  $\tau_{k,\alpha}$  was calculated using a normal approximation.



**Figure 9**  
**Receiver operating characteristic curves for the four models.** Receiver operating characteristic curves (averaged over 500 simulations of an anthrax outbreak) for the detection of the anthrax outbreak signature in the chest radiographs at each of the five metropolitan children's hospitals using the four different models. The linear filter is used in each case.

varied across the five EDs studied (Figure 6). The tradeoff between specificity and sensitivity also varied by the ED (Figures 6 and 7). The 1-day filter, while closely matching the nominal specificity, performed poorly in terms of maximizing the sensitivity to detect an outbreak signature. The 7-day filter, by smoothing over the outbreak signature (Figure 5), maximized the sensitivity, but at a compromise to the actual specificity obtained (Figures 6 and 7). The linear and exponential filters provided a balance between the specificity and sensitivity. Detection using our covariate-based seasonal time series model performed well across all EDs compared with fixed-effects regression models or time series models that omitted seasonal terms and/or covariates (Figures 8 and 9).

**Competing interests**

The author(s) declare that they have no competing interests.

**Authors' contributions**

Dr. Bonsu provided the motivation, scientific background, and data for this work. Under supervision, N. Kim developed the preliminary simulation model for an anthrax attack. Dr. Craigmile worked on the filtering theory used in the paper, and together with N. Kim and Dr. Fernandez analyzed the data, planned and carried out the simulations, and wrote the paper. Dr. Craigmile took the lead with writing. All authors jointly edited this work.

**Appendix**

**The SARIMA model**

In city  $k$ , we define a SARIMA  $(p_k, 0, q_k) \times (P_k, 0, Q_k)$  with period of seasonality  $s$  for the time series  $\{X_{k,t}\}$  using the notation of Section 6.5 of Brockwell and Davis [19]. Letting  $B$  denote the backshift operator defined by  $B^r Z_{k,t} = B^{r-1} Z_{k,t-1}$  for  $r > 0$ , with  $B^0 Z_{k,t} = Z_{k,t}$ , we define  $X_{k,t}$  by

$$\phi_k(B)\Phi_k(B^s)X_{k,t} = \theta_k(B)\Theta_k(B^s)Z_{k,t}.$$

In this model the parameter  $s$  denotes the period of the seasonality. The characteristic polynomials for the SARIMA model are

$$\begin{aligned} \phi_k(z) &= 1 - \phi_{k,1}z - \dots - \phi_{k,p_k}z^{p_k}; \\ \Phi_k(z) &= 1 - \Phi_{k,1}z - \dots - \Phi_{k,P_k}z^{P_k}; \\ \theta_k(z) &= 1 + \theta_{k,1}z + \dots + \theta_{k,q_k}z^{q_k}; \\ \Theta_k(z) &= 1 + \Theta_{k,1}z + \dots + \Theta_{k,Q_k}z^{Q_k}. \end{aligned}$$

The terms  $\phi_k(z)$  and  $\theta_k(z)$  define an autoregressive moving average process on the unit time scale, whereas the terms  $\Phi_k(z)$  and  $\Theta_k(z)$  define an autoregressive moving average process on a time scale of  $s$  units. Thus we can model dependencies simultaneously on two different time

scales. It is customary to restrict to the class of *causal* time series models (e.g., [[19], Chapter 2]). The SARIMA model is causal (i.e., can be represented in terms of a moving average (MA) process of only past events) if  $\phi_k(z) \neq 0$  and  $\Phi_k(z) \neq 0$ , for complex valued  $z$  such that  $|z| \leq 1$ .

To complete the model we assume that  $\{Z_{k,t}\}$  is a mean zero white noise process with innovation variance  $\sigma_k^2$ .

**Filtering**

A discrete-time filter is a set of coefficients,  $\{f_l: l = \dots, -2, -1, 0, 1, 2, \dots\}$ , where  $\sum_l |f_l| < \infty$ . The new time series  $\{Y_t\}$  obtained by filtering the time series  $\{X_t\}$  using  $\{f_l\}$  is defined by the convolution

$$Y_t = \sum_{l=-\infty}^{\infty} f_l X_{t-l},$$

for each  $t$ . The actual operation that a filter carries out upon a time series depends on the values of the filter coefficients. Simple examples of a filter include the differencing filter defined by

$$f_l = \begin{cases} 1 & l = 0; \\ -1 & l = 1; \\ 0 & \text{otherwise,} \end{cases}$$

which corresponds to  $Y_t = X_t - X_{t-1}$  for each  $t$ , and the simple moving average filter of length  $2M + 1$  whose filter is defined by

$$f_l = \begin{cases} 1/(2M+1) & l = -M, \dots, M; \\ 0 & \text{otherwise.} \end{cases}$$

Thus,  $Y_t = (2M + 1)^{-1} \sum_{l=-M}^M X_{t+l}$ , for each  $t$ .

Suppose that the radiograph series  $\{R_{k,t}\}$  follows a time series model given by (1). Let  $\gamma_{X,k}(\cdot)$  denote the autocovariance function of the stationary error component  $\{X_{k,t}\}$ . Consider only the last  $m$  values of the process  $R_{k,t}$ ,  $\{R_{k,t-1}, \dots, R_{k,t-m}\}$ . Given the model parameters, by linearity of the prediction operator, the best linear one-step predictor of  $R_{k,t}$  is given by

$$\hat{R}_{k,t} = \beta_{0,t} + \beta_{V,k}V_{k,t} + \beta_{T,k}T_{k,t} + \hat{X}_{k,t} \tag{5}$$

where

$$\hat{X}_{k,t} = \sum_{j=1}^m b_{k,j}X_{k,t-j}, \tag{6}$$

**Table 4: Akaike information criterions (AICs) for Models 1–4 fit to the training data for each ED**

City	Model 1	Model 2	Model 3	Model 4
Akron	1600	1618	1790	1618
Columbus	1768	1803	1845	1852
Milwaukee	1645	1658	1664	1670
Chicago	1446	1474	1481	1475
Minn/St. Paul	1763	1777	1791	1915

denotes the one-step ahead prediction for the SARIMA process and  $\{b_{k,1}, \dots, b_{k,m}\}$  are obtained from the solution of the set of  $m$  linear equations:

$$\gamma_{X,k}(h) = \sum_{j=1}^m b_{k,j} \gamma_{X,k}(h-j), \quad h = 1, \dots, m. \quad (7)$$

The one-step prediction error process,  $\{E_{k,t}\}$  is

$$E_{k,t} = R_{k,t} - \hat{R}_{k,t} = X_{k,t} - \hat{X}_{k,t} = X_{k,t} - \sum_{j=1}^m b_{k,j} X_{k,t-j} = \sum_{j=0}^m c_{k,j} X_{k,t-j},$$

where we define  $c_{k,0} = 1$  and  $c_{k,j} = -b_{k,j}$  for  $j = 1, \dots, m$ . The process  $\{E_{k,t}\}$  is a filtering of  $\{X_{k,t}\}$  and so if  $\{X_{k,t}\}$  is stationary, by the linear time invariant filtering result for stationary processes,  $\{E_{k,t}\}$  is also stationary ([18], Theorem 4.10.1). Moreover for large enough  $m$ , if  $\{X_{k,t}\}$  is an invertible time series process then we can approximate  $\{X_{k,t}\}$  by an autoregressive, AR( $m$ ) process ([18], Theorem 4.4.3). Using this approximation we can argue that  $\{E_{k,t}\}$  is approximately a mean zero Gaussian independent and identically distributed process with variance  $\gamma_{X,k}(0)$ .

Let  $\{a_0, \dots, a_{L-1}\}$  denote a pre-specified filter of width  $L$ . We define the detection process  $\{D_{k,t}\}$  by filtering the error process  $\{E_{k,t}\}$  using this filter. Thus,

$$D_{k,t} = \sum_{h=0}^{L-1} a_h E_{k,t-h}. \quad (8)$$

By the filtering result for stationary processes, stationarity of  $\{X_{k,t}\}$  implies that  $\{D_{k,t}\}$  is also a stationary process. For large enough  $m$ , since  $\{E_{k,t}\}$  is approximately a white noise process,  $\{D_{k,t}\}$  is approximately a moving average process of order  $L - 1$  with coefficients  $\theta_0 = 1, \theta_j = a_{j-1}/a_0$  for  $j = 1, \dots, L - 1$ , and innovation variance  $a_0^2 \gamma_{X,k}(0)$ .

We declare evidence of an outbreak signature (test positive), at time  $t$  if the observed detection value  $D_{k,t}$  is larger than a given threshold. For a fixed value of  $\alpha$ , let  $1 - \alpha$  denote the specificity (the probability of a true-negative). Similarly, the sensitivity is defined to be the probability of

true positive. The threshold,  $\tau_{k,\alpha}$  is chosen by solving  $P(D_{k,t} > \tau_{k,\alpha}) = \alpha$  for  $\tau_{k,\alpha}$ . Either we can estimate this value from the data using the  $1 - \alpha$  quantile of the  $\{D_{k,t}\}$  process of non-outbreak-based training data, or if  $\{X_{k,t}\}$  is Gaussian then for large  $m$ ,

$$P(D_{k,t} > \tau_{k,\alpha}) \approx P\left(Z > \frac{\tau_{k,\alpha}}{\sqrt{\gamma_{X,k}(0) \sum_{h=0}^{L-1} a_h^2}}\right) = \alpha, \quad (9)$$

where  $Z$  is a standard normal random variable. The solution for the threshold under this approximation is

$$\tau_{k,\alpha} = \left(\gamma_{X,k}(0) \sum_{h=0}^{L-1} a_h^2\right)^{1/2} \Phi^{-1}(1 - \alpha), \quad (10)$$

where  $\Phi^{-1}(\cdot)$  denotes the inverse cumulative distribution function for a standard normal random variable.

The filter-based detection method requires knowledge of the "true" values of the model parameters. In our work, we replace the parameters with their maximum likelihood estimates.

### Acknowledgements

We would like to thank Elaine Damo at Columbus Children's Hospital for collecting the data, and Dr. Steven MacEachern for helpful discussions about our work. We also thank the two referees for their extensive comments that improved this article.

### References

1. Shmueli G: **Wavelet-Based Monitoring in Modern Biosurveillance**. Tech rep 2005 [<http://www.rhsmith.umd.edu/faculty/gshmueli/biosurveillance2005.pdf>]. Smith School of Business, University of Maryland
2. Reis B, Pagano M, Mandl K: **Using temporal context to improve biosurveillance**. *Proceedings of the National Academy of Sciences* 2003, **100(4)**:1961-1965.
3. Reis B, Mandl K: **Time series modeling for syndromic surveillance**. *BMC Medical Informatics and Decision Making* 2003, **3(2)**: [<http://www.biomedcentral.com/1472-6947/3/2>].
4. Mandl K, Overhage J, Wagner M, Lober W, Sebastiani P, Mostashari F, Pavlin J, Gesteland P, Treadwell T, Koski E, Hutwagner L, Buckenridge D, Aller R, Grannis S: **Syndromic surveillance: a guide informed by the early experience**. *Journal of the American Medical Informatics Association* 2004, **11(2)**:141-150.

5. Box GEP, Luceno A: *Statistical control by monitoring and feedback adjustment* New Jersey: John Wiley and Sons; 1997.
6. Ivanov O, Gesteland P, Hogan W, Mundorff M, Wagner M: **Detection of Pediatric Respiratory and Gastrointestinal Outbreaks from Free-Text Chief Complaints.** In *Proceedings of the Annual Fall Symposium of the American Medical Informatics Association* Madison: Omni Press; 2003:318-322.
7. Burkom HS, Elbert Y, Feldman A, Lin J: **Role of data aggregation in biosurveillance detection strategies with applications from essence.** *Morbidity and Mortality Weekly Report (MMWR)* 2004, **53**:67-73.
8. Zhang J, Tsui FC, Wagner M, Hogan W: **Detection of Pediatric Respiratory and Gastrointestinal Outbreaks from Free-Text Chief Complaints.** In *Proceedings of the Annual Fall Symposium of the American Medical Informatics Association* Madison: Omni Press; 2003:748-752.
9. Naumova EN, O'Neil E, MacNeill I: **A System for Early Outbreak Detection and Signature Forecasting.** *Morbidity and Mortality Weekly Report* 2005, **54**:77-83.
10. Ismail N, Pettitt A, Webster R: **'Online' monitoring and retrospective analysis of hospital outcomes based on a scan statistic.** *Statistics in Medicine* 2003, **22**:2861-2876.
11. Wallenstein S, Naus J: **Scan Statistics for Temporal Surveillance for Biologic Terrorism.** *Morbidity and Mortality Weekly Report* 2004, **53**:74-78.
12. Naus J, Wallenstein S: **Temporal Surveillance Using Scan Statistics.** *Statistics in Medicine* 2006, **25**:311-324.
13. Jackson ML, Baer A, Painter I, Duchin J: **A simulation study comparing aberration detection algorithms for syndromic surveillance.** *BMC Medical Informatics and Decision Making* 2007, **7**(6): [<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=17331250>].
14. Brillman JC, Burr T, Forslund D, Joyce E, Picard R, Umland E: **Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance.** *BMC Medical Informatics and Decision Making* 2005, **5**(4): [<http://www.biomedcentral.com/1472-6947/5/4/>].
15. **Average Daily Temperature Archive at The University of Dayton** [<http://www.engr.udayton.edu/weather/citylistUS.htm>]
16. Buckeridge D, Burkom H, Moore A, Pavlin J, Cutchis P, Hogan W: **Evaluation of syndromic surveillance systems – design of an epidemic simulation model.** *Morbidity and Mortality Weekly Report* 2004, **53**:137-143 [<http://www.cdc.gov/mmwr/pdf/wk/mm53su01.pdf>].
17. Brookmeyer R, Johnson E, Barry S: **Modelling the incubation period of anthrax.** *Statistics in Medicine* 2005, **24**:531-542.
18. Brockwell PJ, Davis RA: **Time Series. Theory and Methods.** Second edition. New York: Springer-Verlag; 1991.
19. Brockwell PJ, Davis RA: **Introduction to time series and forecasting.** Second edition. New York: Springer-Verlag; 2002.
20. Diggle P, Rowlingson B, Su T: **Point process methodology for on-line spatio-temporal disease surveillance.** *Environmetrics* 2005, **16**(5):423-434.
21. Fienberg S, Shmueli G: **Statistical Issues and Challenges Associated with Rapid Detection of Bio-terrorist Attacks.** *Statistics in Medicine* 2005, **24**(4):513-529.

## Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1472-6947/7/28/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

