

# Algorithm for Statistical Detection of Peaks — Syndromic Surveillance System for the Athens 2004 Olympic Games

Urania G. Dafni,<sup>1,2</sup> S. Tsiodras,<sup>1,2</sup> D. Panagiotakos,<sup>1</sup> K. Gkolfinopoulou,<sup>1,2</sup> G. Kouvatseas,<sup>1</sup> Z. Tsourti,<sup>1</sup> G. Saroglou<sup>2</sup>  
<sup>1</sup>University of Athens, Athens, Greece; <sup>2</sup>Hellenic Center for Infectious Diseases Control, Ministry of Health, Athens, Greece

**Corresponding author:** Urania G. Dafni, Division of Public Health, Department of Nursing, University of Athens, 123 Papadiamantopoulou St., Athens 11527, Greece. Telephone: 30-210-746-1456; Fax: 30-210-746-1489; E-mail: dafni@cc.uoa.gr.

## Abstract

**Introduction:** No generally accepted procedure exists for detecting outbreaks in syndromic time series used in the surveillance of natural epidemics or biologic attacks.

**Objectives:** This report evaluates the usefulness for syndromic surveillance of the Pulsar approach, which is based on removing long-term trends from an observed series and identifying peaks in the residual series of surveillance data with cutoffs determined by using a combination of peak height and width.

**Methods:** Simulations were performed to evaluate the Pulsar method and compare it with other approaches. The daily syndromic counts in emergency departments of four major hospitals in the Athens area during August 2002–August 2003 were analyzed for two common syndromes. A standardized residual series was generated by omitting trends and noise in the original data series; this series was examined for the presence of peaks (i.e., points having magnitude higher than at least one of three probabilistically determined cutoffs). The whole process was iterated, and the baseline was recalculated by assigning reduced weight to the identified peaks.

**Results:** For the specific simulation schema used, the Pulsar method fared well when compared with other approaches in meeting the performance criteria of sensitivity, specificity, and timeliness.

**Conclusions:** Although the suggested algorithm needs further validation regarding the correspondence between detected peaks and true biologic alerts, the Pulsar technique appears effective for observing peaks in time series of syndromic events. The simplicity of the algorithm, its ability to detect peaks based not only on height but also on width, and its performance in the simulated data sets make it a promising candidate for further use in syndromic surveillance.

## Introduction

Syndromic time series are used in surveillance of natural epidemics or biologic attacks. CDC and the New York City Department of Health and Mental Hygiene used syndromic surveillance systems for detection of biologic terrorism after the September 2001 terrorist attacks (1–3). Almost simultaneously, other systems emerged (4), including those developed by the Boston Department of Health (5) and the University of Pittsburgh (4,6), CDC's drop-in surveillance systems (7), the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) (8,9), and others (10–12).

The Athens 2004 Olympic Games (August 13–29, 2004) have made critical the need for a real-time surveillance system that can alert public health officials to unexpected communicable-disease outbreaks and likely clinical presentations of a biologic terrorist attack, as has been used for other major athletic events (13–17). Therefore, in July 2002, a drop-in syndromic surveillance system was established in Greece similar

to that used during the Salt Lake 2002 Olympic Winter Games (13,18).

Different outbreak-detection algorithms are used in operating syndromic surveillance systems (2,19–23). Ideally, all alert mechanisms generate an alert whenever the number of observed events exceeds the expected number of events while minimizing the frequency of false alerts. However, no generally accepted procedure exists for outbreak detection in syndromic surveillance (24). This paper proposes an algorithm for statistical detection of peaks. The method is based on removing long-term trends from the series of observations and identifying peaks in the residual series of data. This approach was developed for studying episodic hormonal secretion and has been used for other applications (25–27). An important feature of the proposed algorithm is that it generates alerts, taking into consideration both height and breadth of signals. The proposed method was applied in the Athens Olympic syndromic surveillance system database (18) and was compared through simulations with other methods currently applied in syndromic data series (19–23,28).

## Methods

### Data Acquisition

Drop-in syndromic surveillance in emergency departments (EDs) of major hospitals was first established in Greece by the Hellenic Center for Infectious Diseases Control in July 2002. The project's primary aims were to assess system feasibility and data-collection timeliness, establish a 2-year background database, and enhance collaboration with and sensitization of ED personnel of major hospitals (18). During August 2002–August 2003, the syndromic surveillance system operated in eight hospitals and one major health-care center in the greater Athens area. Surveillance was conducted for the following 10 syndromes: 1) *respiratory infection with fever*; 2) *bloody diarrhea*; 3) *gastroenteritis (diarrhea, vomit) without blood*; 4) *febrile illness with rash*; 5) *meningitis, encephalitis, or unexplained acute encephalopathy/delirium*; 6) *suspected acute viral hepatitis*; 7) *botulism-like syndrome*; 8) *lymphadenitis with fever*; 9) *sepsis or unexplained shock*; and 10) *unexplained death with history of fever*. These syndrome categories were used by the Salt Lake City Department of Health for syndromic surveillance during the 2002 winter Olympics (13). Trained personnel visited EDs and identified syndromic cases from chief complaints as recorded in ED visit books. All syndromes identified daily in the ED were recorded, as were the total number of visits. Data were entered into a database, and data management and analysis were performed centrally. In the work presented here, the time series for the two most commonly encountered syndromes (*respiratory infection with fever* and *gastroenteritis [diarrhea, vomit] without blood*) are used.

### Algorithm Description

The Pulsar method is based on identifying peaks in the syndromic time series that exceed a specified threshold. Long-term changes are first screened out, and then peaks are identified in the screened series. This approach has been previously suggested for studying episodic hormonal secretion (25).

First, a baseline is defined for the original syndromic series by using the locally weighted smoothing scatterplots method (LOWESS) (29), in which a fixed proportion of observations (the smoothing parameter) is used, and a baseline value is calculated from the observations closest in time to the point. Weights are assigned to the observations, depending on their distance from the point. The fraction of observations in the window is selected so that the window's average width minimizes the bias-corrected Akaike's information criterion (AIC), which incorporates both the tightness of the fit and the model complexity. This criterion often selects better models than AIC in small samples (30). Then, a weighted nonparametric

regression of syndromic counts versus time within the window provides the initial baseline value estimate for that time point. After the initial estimation of baseline values, new weights giving less influence to observations far from the corresponding baseline values are assigned, and the weighted regression is repeated. This procedure produces baseline estimates that are not influenced by extreme outlier observations.

A residual series, containing short-term variations but not trends, is obtained by subtracting the smoothed data from the original counts and is standardized by dividing the residuals by an estimate of the noise level, to yield a scaled residual series, expressed in signal-to-noise units. The peaks in the standardized residual series are identified on the basis of a combination of height and width, with no assumption for the shape of the peak. To be classified as a peak, an elevation should either be substantially high, even if it is narrow, or span multiple points in width, even if it is moderately high. For a point in the signal-to-noise series to be considered part of a peak, it should exceed a certain cutoff value  $G(1)$ ; or it should exceed a lower cut-off value  $G(2)$  along with one adjacent point; or it should exceed an even lower cut-off value  $G(3)$  along with two adjacent points; and so forth. The specific choices of  $n$  and  $G(n)$ s depend on the time series used for calibration purposes, the relative choice between higher but narrow peaks as opposed to lower but broad ones, and the desired false-alert rate. After the initial identification of peaks, the baseline is recalculated. Reduced weight is assigned to observations previously identified as part of a peak. Iterations of the whole process are performed until the same assignment of points to peaks is achieved.

### Algorithm Customization

In the 13-month syndromic series, LOWESS smoothing was applied with optimal smoothing parameter equal to 15% for *respiratory infection with fever* and 52% for *gastroenteritis (diarrhea, vomit) without blood*. Alternative estimates were used for the standardization, including the standard deviation and the mean absolute deviation in the original series, as well as the 7-day moving standard deviation and the 7-day mean absolute deviation in the simulated series. The latter were based either on the seven most recent observations to the current time point or on the tenth to fourth most recent observations (i.e., not taking into account the three most recent ones). The procedure is performed iteratively to weigh down extreme values and detect outliers appearing in clusters. In this data set, extreme clustered observations do not appear to exist, and two iterations were sufficient to obtain a smoothed series (the resulting detected peaks of the two iterations differ by <2.5%).  $G(1)$ ,  $G(2)$ , and  $G(3)$  cutoffs were chosen under the assump-

tion of normality for the standardized residual series to derive 97% specificity in the whole series and take into account the effect of multiple testing on the significance level. The threshold is given by  $G(n) = \text{probit}(1 - \alpha^*[n/6]/d)$  where  $d$  = number of days that a false alert occurs with probability  $\alpha = 0.10$  and  $n = 1$  or 2 or 3, whereas the factor  $n/6$  provides the necessary adjustment for multiple testing (25).

## Alternative Methods

The Pulsar approach was evaluated by comparison through simulations with other commonly used syndromic surveillance methods (19–23,28). All parameters for each model used in the comparisons were set so that the specificity (true nonalerts/nonoutbreaks) in the original time series was fixed at 97%, assuming no outbreak condition (20,21). For each method, the day of an outbreak on which an alert was first generated was recorded. Sensitivity (true alerts/outbreaks) across all simulated series for each syndrome and the timeliness for each method (i.e., the percentage of the first alert per day of outbreak) were compared among the alternative approaches. The three performance criteria (sensitivity, specificity, and timeliness) were reported and compared through the Wilcoxon signed rank or Friedman nonparametric tests. Bonferroni-adjusted  $\alpha^*$  are reported. The methods mentioned here have been used in syndromic surveillance and were evaluated in this syndromic data series.

The temporal aberration detection (TAD) approach used by the Early Aberration Reporting System (EARS), a program provided by CDC to all interested health departments, uses cumulative sum (CUSUM) methods from the quality-control literature. CUSUM compares the proportion of syndrome counts to total visits on each of the most recent 3 days to the mean proportion plus 1 standard deviation, during a 7-day moving baseline. CUSUM of positive differences is calculated based on a 3-day interval, and an alert is considered to occur if it exceeds 2 standard deviations (2,22,23). Time-series methods (e.g., autoregressive integrated moving average [ARIMA] time-series models) were proposed for describing 10-year syndromic data from a major Boston-area hospital (20,21). Different filters were evaluated in data sets with simulated outbreaks, using a fixed specificity rate of 97%. The linear 7-day filter proved superior in simulations (20). Standard one-sided CUSUM methods have also been proposed for detecting outbreaks in surveillance data (19,28).

## Simulation Schema

For evaluation of the performance of the proposed methodology, 100 simulated series were created. The original time

series of counts is considered to include no outbreaks. A scenario involving a terrorist attack depends on the biologic agent, quality, and quantity released; the method of dispersion; and population characteristics. A 4-day outbreak was chosen to represent a probable period between symptom presentation and diagnosis (i.e., the window of opportunity for possible earlier detection because of syndromic surveillance) (31). However, different durations of that window are also possible.

Each simulated time series was produced by randomly injecting 4-day-long outbreaks to the original time series of daily counts for each syndrome of interest with probability of 15% per day (leading to 18.5 4-day outbreaks on average among simulated series). An outbreak led to duplication of the observed counts of the syndrome for that day (respiratory infection: median size = 27; 5<sup>th</sup> and 95<sup>th</sup> percentiles: 24, 29 and gastroenteritis: median size = 15; 5<sup>th</sup> and 95<sup>th</sup> percentiles: 14, 16). Two adjacent outbreaks were forced to be  $\geq 15$  days apart to ensure that a previous outbreak did not adversely affect the alert-detection mechanism of the next (20). The detection algorithms should detect an outbreak as if it is the first one that occurs in the original time series.

An outbreak was considered successfully detected if an alert was generated on  $\geq 1$  day of the outbreak. Alternative patterns of outbreaks were also examined, including 1) constant increase for all 4 days, equal to the median counts of the syndrome (23.5 for *respiratory infection with fever* and 15.5 for *gastroenteritis*) or 2) constant increase for all 4 days, equal to the 75<sup>th</sup> percentile of the counts of the syndrome (35 for *respiratory infection with fever* and 22 for *gastroenteritis*); 3) linear increase for the 4 days: (increase of one median/day); 4) exponential increase for the 4 days: increase of 1, 1.5, 2.5, and 4 medians for day 1–4, respectively; or 5) exponential increase for the first 3 days (1, 1.5, 2.5 medians) and subsequent decrease on day 4. All statistical computations were performed by using SAS<sup>®</sup> software, version 8.2 (32).

## Results

The original 13-month time series for four major hospitals in metropolitan Athens sharing the same catchment area for the *respiratory infection with fever* and *gastroenteritis (diarrhea, vomit) without blood* syndromes were used to illustrate and evaluate the proposed method. A total of 305,039 ED visits (mean: 770/day) were recorded during August 2002–August 2003 in these hospitals. The corresponding mean total syndrome counts were 26 and 15 per day for each syndrome, respectively.

The six different standardization estimates already described for the Pulsar algorithm, leading to different threshold speci-

fications, were compared. The best approach for both syndromes with respect to the achieved sensitivity was the one that used the standard deviation in the original series (Table 1; see Model 1). The corresponding parameter  $d$  to the G(1), G(2), and G(3) thresholds was 0.5 and 1 for *respiratory infection with fever* and *gastroenteritis (diarrhea, vomit) without blood* syndromic series, respectively, whereas  $\alpha$  was set equal to 0.10 for both syndromes. The standardized residuals from the original time series and from a sample simulated series (number 10) for each syndrome along with the thresholds are illustrated (Figure 1).

The TAD approach was used both on the count series and on the proportion of counts of syndromes to total ED visits (2,22,23). The results for the count series were superior to those for the proportion series. The fixed specificity of 97% in the original time series of counts, for both syndromes, was reached by using 3 standard deviations for the alert mechanism (Model 1) instead of the 2 used in EARS (Model 2) (2,22,23) (Table 2).

ARIMA models were used for describing the 13-month original series of syndromic data (20,21). For *respiratory infection with fever*, the autoregressive (AR) order, the moving average (MA) order, and the integration (I) order were all equal to 4 days. Weekend was also statistically significant and used as an explanatory variable in the model. For *gastroenteritis (diarrhea, vomit) without blood*, AR = 4 days, MA = 2 days, and I = 4 days. The filters evaluated were seventh-order MA (Model 1), seventh-order linear average (Model 2), and seventh-order exponential average (Model 3). The threshold was again set so that specificity of 97% was achieved in the original time series, and the best filter regarding sensitivity was the seventh-order MA filter (Model 1). The corresponding thresholds are equal to  $\text{probit}(1-\alpha/7)$  and  $\text{probit}(1-\alpha/8)$  for each syndrome, respectively, with  $\alpha$  equal to 0.10.

For the one-sided CUSUM method used here, a 7-day moving average and standard deviation used for standardization proved superior to the standard approach (19,28). The cumulative sum was calculated by  $S_t = \max\{0, (S_{(t-1)} + z_t - k)\}$ , where  $k = 0.5$ . The specified threshold  $h$  was set so that specificity 97% was achieved in the original time series, and the corresponding values for the two syndromes were set to 3.5 and 2.75, respectively (Model 1). A second approach employing values from the literature that actually minimize the average run length (ARL) of the process was also used ( $k = 0.5$  and  $h = 2.5$ ) (19) (Table 2).

The sensitivity and specificity of the alternative methods (TAD, the time-series approach, and CUSUM) were compared (Table 2). Performance criteria for the best models with respect to sensitivity for each approach, among the ones using a set specificity of 97% in the original time series, are directly compared (Figures 2 and 3). Box-plots of the model's sensitivity and specificity (Figure 2) and timeliness (Figure 3) are presented.

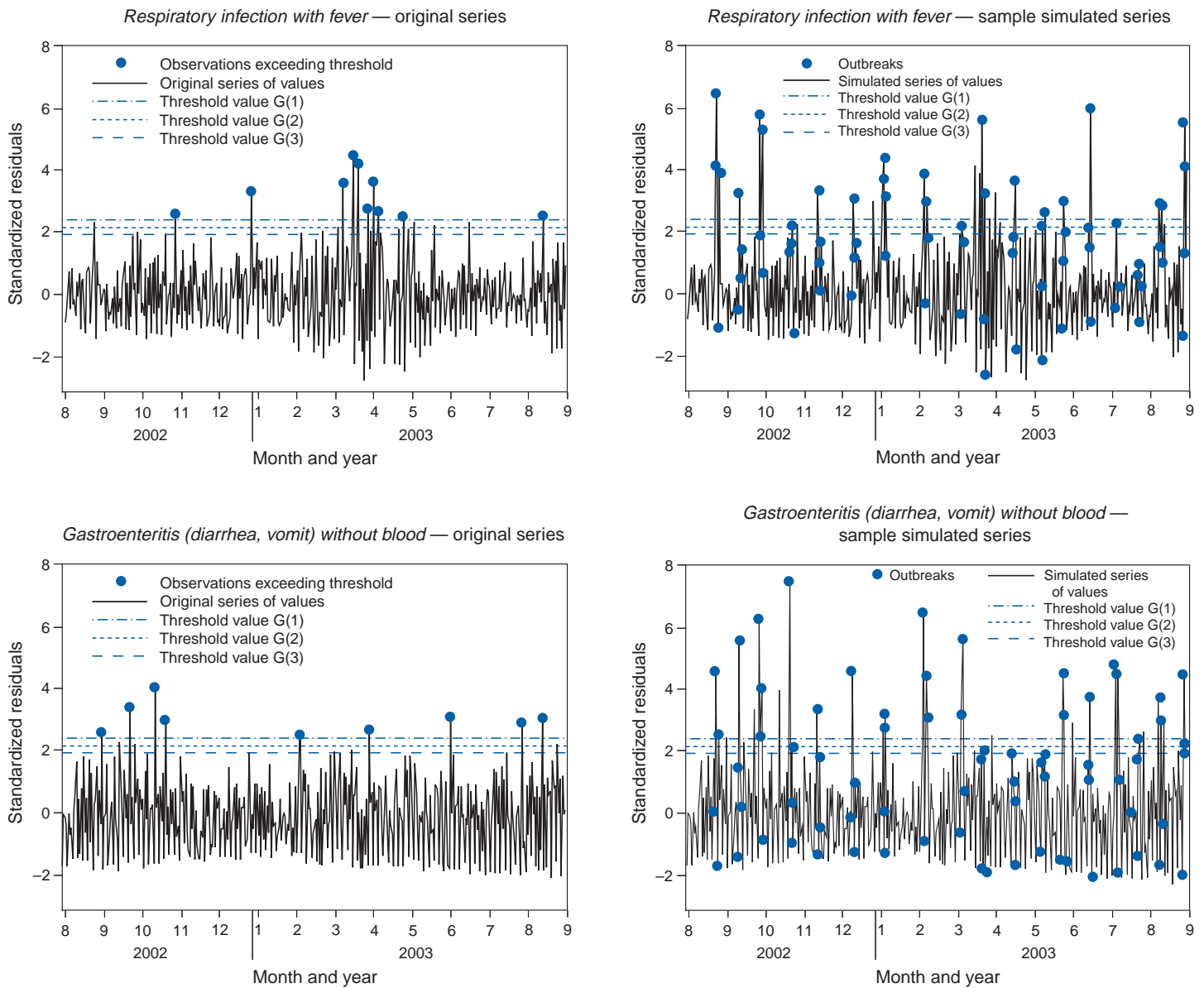
The Pulsar approach fared well in comparison with the other methods for each evaluation criterion. In particular, mean sensitivity was statistically significantly higher (Bonferroni  $\alpha^* = 0.0056$ ) for the Pulsar approach when compared with the other approaches for both syndromes (Wilcoxon signed rank,  $p < 0.001$  for all comparisons). Furthermore, mean specificity for the Pulsar method was significantly higher (Bonferroni  $\alpha^* = 0.0056$ ) than the specificity of the one-sided CUSUM method (Wilcoxon signed rank,  $p < 0.001$ ). This finding holds for both syndromes examined. In addition, in the case of *respiratory infection with fever*, the specificity of Pulsar was significantly higher than the specificity of TAD (Wilcoxon signed rank,  $p < 0.001$ ), whereas in the case of *gastroenteritis*, the specificity of Pulsar was higher than the specificity of the ARIMA approach (Wilcoxon signed rank,  $p < 0.001$ ). No other signifi-

**TABLE 1. Sensitivity and specificity of the Pulsar approach for *respiratory infection with fever* and *gastroenteritis (diarrhea, vomit) without blood* syndromes (simulated series)**

Method	Syndrome							
	<i>Respiratory infection with fever</i>				<i>Gastroenteritis (diarrhea, vomit) without blood</i>			
	Sensitivity		Specificity		Sensitivity		Specificity	
	Mean (SE*)	Median (Min–Max†)	Mean (SE)	Median (Min–Max)	Mean (SE)	Median (Min–Max)	Mean (SE)	Median (Min–Max)
<b>Pulsar analysis</b>								
Model 1§	0.854 (0.0071)	0.85 (0.667–1)	0.981 (0.0004)	0.981 (0.972–0.991)	0.812 (0.0074)	0.818 (0.625–1)	0.982 (0.0003)	0.981 (0.975–0.991)
Model 2¶	0.744 (0.0081)	0.737 (0.5–0.9)	0.982 (0.0004)	0.981 (0.972–0.991)	0.768 (0.0088)	0.771 (0.55–1)	0.978 (0.0004)	0.978 (0.972–0.988)
Model 3**	0.656 (0.009)	0.649 (0.45–0.895)	0.985 (0.0004)	0.985 (0.975–0.997)	0.536 (0.0112)	0.538 (0.211–0.789)	0.991 (0.0004)	0.991 (0.981–1)
Model 4††	0.523 (0.0104)	0.5 (0.263–0.842)	0.987 (0.0004)	0.988 (0.976–0.997)	0.496 (0.0109)	0.5 (0.2–0.789)	0.99 (0.0004)	0.991 (0.981–1)
Model 5§§	0.718 (0.0089)	0.722 (0.5–0.9)	0.993 (0.0004)	0.994 (0.979–1)	0.701 (0.0101)	0.706 (0.375–0.944)	0.991 (0.0004)	0.991 (0.981–1)
Model 6¶¶	0.719 (0.0093)	0.737 (0.5–0.9)	0.992 (0.0004)	0.994 (0.982–1)	0.7 (0.0104)	0.706 (0.375–0.944)	0.99 (0.0005)	0.991 (0.978–1)

\* Standard error.  
 † Minimum–Maximum.  
 § Standardization by the standard deviation of the original series.  
 ¶ Standardization by the mean absolute deviation of the original series.  
 \*\* Standardization by the 7-day moving standard deviation of the simulated series.  
 †† Standardization by the 7-day mean absolute deviation of the simulated series.  
 §§ Standardization by the 7-day moving standard deviation of the simulated series with a 3-day lag.  
 ¶¶ Standardization by the 7-day mean absolute deviation of the simulated series with a 3-day lag.

FIGURE 1. Time-series plots of standardized\* residuals for the Pulsar method — G(1), G(2), G(3) thresholds and outbreaks



\* Standardization by the standard deviation of the original series.

cant differences regarding specificity between the Pulsar method and the others were identified for either syndrome. Timeliness for the first day (proportion of alerts at the first day of an outbreak) differed significantly among the four approaches (Friedman test,  $p < 0.001$ ) for both syndromes. Timeliness of the Pulsar method was lower than the timeliness of the ARIMA model (Wilcoxon signed rank,  $p < 0.001$ ; Bonferroni  $\alpha^* = 0.0056$ ). However, for *respiratory infection with fever*, Pulsar's timeliness was higher than TAD and CUSUM ( $p < 0.001$ ), and for *gastroenteritis*, Pulsar's timeliness

was higher than TAD's ( $p < 0.001$ ). Results were similar when the mentioned alternative patterns of outbreaks were used.

## Discussion

This paper proposes an algorithm for outbreak detection in the context of syndromic surveillance time-series data, based on alert criteria for both height and breadth of signals (25). The performance of the Pulsar approach and other suggested methods for outbreak detection (19–23,28) were assessed

**TABLE 2. Sensitivity and specificity for alternative outbreak-detection approaches for *respiratory infection with fever and gastroenteritis (diarrhea, vomit) without blood syndromes (simulated series)***

Method	Syndrome									
	<i>Respiratory infection with fever</i>					<i>Gastroenteritis (diarrhea, vomit) without blood</i>				
	Sensitivity		Specificity			Sensitivity		Specificity		
	Mean (SE)*	Median (Min–Max) <sup>†</sup>	Mean (SE)	Median (Min–Max)	Mean (SE)	Median (Min–Max)	Mean (SE)	Median (Min–Max)	Mean (SE)	Median (Min–Max)
<b>Temporal aberration detection</b>										
Model 1 <sup>§</sup>	0.774 (0.0098)	0.778 (0.526–0.947)	0.977 (0.0008)	0.978 (0.956–0.99)	0.701 (0.0088)	0.696 (0.471–0.889)	0.981 (0.0007)	0.981 (0.956–0.997)		
Model 2 <sup>¶¶</sup>	0.912 (0.0067)	0.914 (0.737–1)	0.955 (0.001)	0.955 (0.931–0.975)	0.831 (0.0082)	0.833 (0.611–1)	0.957 (0.001)	0.956 (0.924–0.978)		
<b>Auto-regressive integrated moving average</b>										
Model 1 <sup>††</sup>	0.738 (0.0083)	0.737 (0.5–0.9)	0.981 (0.0005)	0.981 (0.969–0.997)	0.679 (0.01)	0.684 (0.35–0.867)	0.979 (0.0005)	0.979 (0.966–0.991)		
Model 2 <sup>§§</sup>	0.667 (0.0088)	0.667 (0.5–0.895)	0.978 (0.0005)	0.978 (0.969–0.994)	0.629 (0.0105)	0.637 (0.35–0.833)	0.976 (0.0005)	0.975 (0.963–0.988)		
Model 3 <sup>¶¶¶</sup>	0.55 (0.0096)	0.55 (0.3–0.737)	0.979 (0.0005)	0.979 (0.969–0.997)	0.57 (0.0106)	0.579 (0.294–0.8)	0.975 (0.0003)	0.975 (0.966–0.985)		
<b>Cumulative sum</b>										
Model 1 <sup>***</sup>	0.711 (0.011)	0.706 (0.444–0.895)	0.926 (0.0023)	0.923 (0.877–0.978)	0.728 (0.0095)	0.737 (0.474–0.941)	0.94 (0.0015)	0.941 (0.906–0.975)		
Model 2 <sup>**†††</sup>	0.84 (0.0088)	0.842 (0.6–1)	0.87 (0.0029)	0.869 (0.803–0.946)	0.781 (0.0091)	0.789 (0.556–0.947)	0.927 (0.0016)	0.927 (0.887–0.96)		

\* Standard error.

† Minimum–Maximum.

§ Specificity set at 97%.

¶ Threshold of 2 standard deviations.

\*\* Models presented only for purposes of illustration.

†† 7-day moving average filter.

§§ 7-day linear filter.

¶¶¶ 7-day exponential filter.

\*\*\* Specificity set at 97%.

††† Threshold set so that  $k = 0.5$  and  $h = 2.5$ .

through simulations on the basis of direct comparison of sensitivity, specificity, and timeliness. For these performance criteria, Pulsar appears to be at least as effective as the other methods.

The Pulsar approach, first suggested for studying of episodic hormonal secretion, was successfully used in the context of syndromic surveillance data. Syndromic data are expressed initially in signal-to-noise units; then, through an iterative process, peaks are identified. Point elevations that are substantially high or elevations only moderately high but spanning multiple points in width are identified as peaks. The thresholds for peak detection are determined probabilistically on the assumption of normally distributed residuals. The idea of stochastically determining the thresholds is extended to the other methods under comparison. The thresholds are chosen so that a specificity of 97% is achieved in the original syndromic time series (20,21).

In the simulated data sets, the 97% specificity was most closely reproduced when using the Pulsar method as compared with the other methods (Tables 1 and 2). Sensitivity for the chosen Pulsar model (Model 1) for *respiratory infection with fever* ranged from 67% to 100%, with a mean of 85%, whereas sensitivity for *gastroenteritis (diarrhea, vomit) without blood* ranged from 62.5% to 100%, with a mean of 81%. The mean sensitivity for the Pulsar approach was higher than the sensitivity for the other methods. This method compared well with the others as far as specificity. All methods held specificity close to the 97% benchmark, with the exception of the one-sided CUSUM. In all methods evaluated, the higher percentage of alerts was generated on the first day of the out-

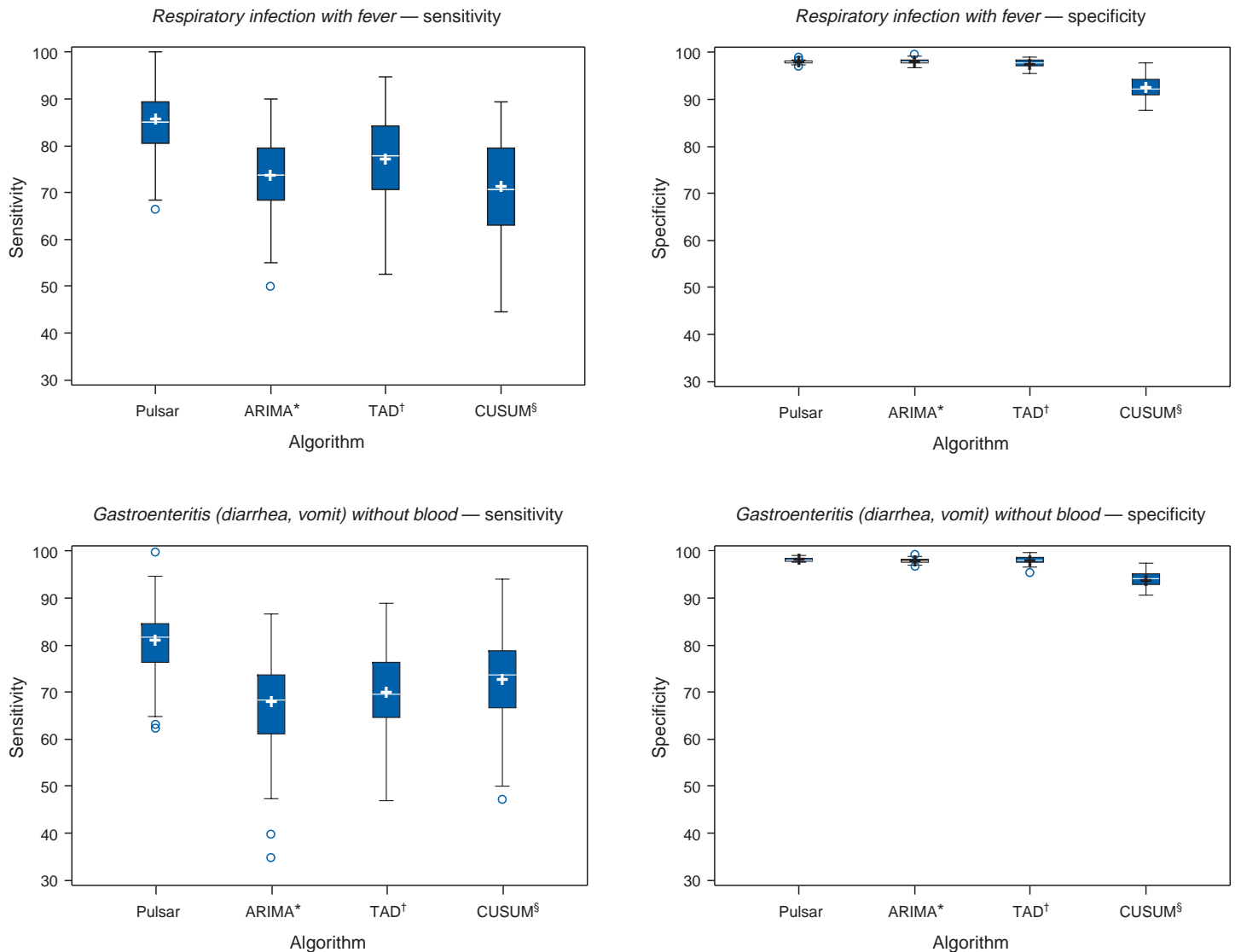
break with the exception of the TAD model, for which alerts occurred with similar frequency on the first 3 days of the outbreak. The ARIMA model exhibited the best timeliness results, followed by the Pulsar approach.

Of note, the performance evaluation criteria led to uniformly worse results for all methods when applied to the daily proportion of syndrome counts to total visits as opposed to syndrome counts. Methods adapted to proportion are under investigation and could be evaluated simultaneously. In addition, a specific simulation schema was used to compare methods, with varying outbreak sizes of fixed duration affecting the generalization of the comparison under other simulation settings. However, the critical comparison is always the one based on the detection performance of real outbreaks (33). Finally, this analysis did not consider other methods that have been proposed for analysis of syndromic data (34,35), including spatial statistical methods (e.g., spatial scan statistic, Bayesian approaches, and multivariate methods) (36–40).

## Conclusion

The performance results of the Pulsar method are overall comparable with the other methods examined for the specific simulation schema used. The simplicity of the algorithm, its ability to be modified regarding choice of standardization and distributional assumptions for the signal-to-noise ratio, and its ability to detect peaks based not only on height but also on width (which more closely addresses the epidemic shapes that one would expect to last for >1 day) make it a promising candidate for further use in syndromic surveillance. The abrupt

**FIGURE 2. Box plots of sensitivity and specificity of 100 simulated series for *respiratory infection with fever and gastroenteritis (diarrhea, vomit) without blood syndrome counts***



\* Autoregressive integrated moving average.

† Temporal aberration detection.

§ Cumulative sum.

increase in population anticipated for the Athens 2004 Olympic Games will provide an ideal prospective surveillance setting for comparing the behavior of all proposed methods regarding alert mechanisms.

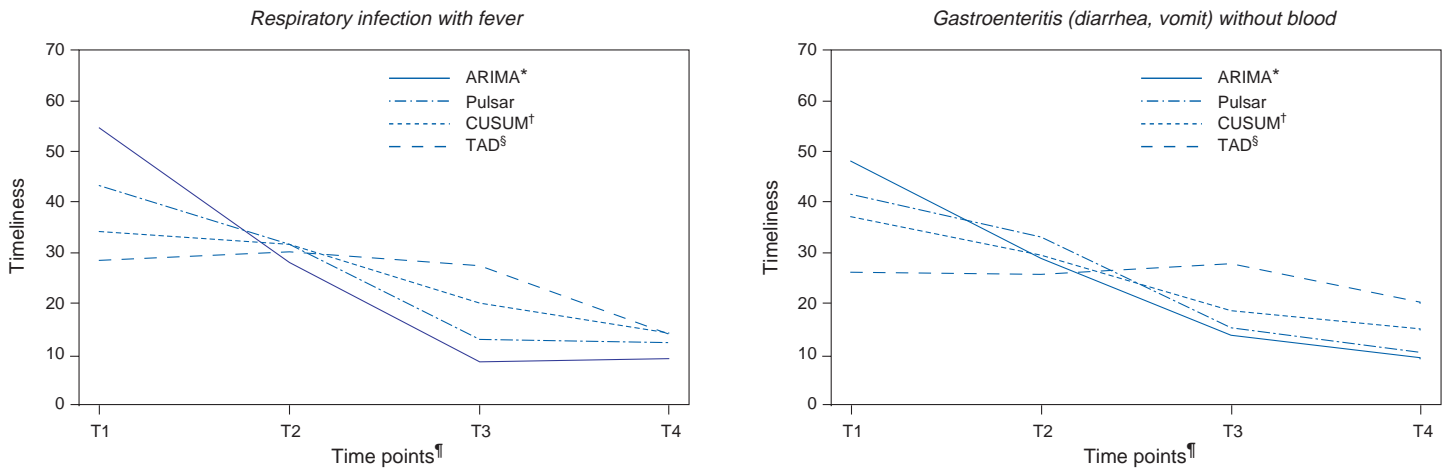
#### Acknowledgments

The authors thank the members of the syndromic surveillance team, Dimitris Papamihail, Aggeliki Lambrou, and Ioannis Karagiannis, as well as the >50 health professionals who made every effort to gather quality data from the EDs.

#### References

1. Mostashari F, Fine A, Das D, Adams J, Layton M. Use of ambulance dispatch data as an early warning system for communitywide influenza-like illness, New York City. *J Urban Health* 2003;80 (2 Suppl 1):i43–9.
2. Das D, Weiss D, Mostashari F, et al. Enhanced drop-in syndromic surveillance in New York City following September 11, 2001. *J Urban Health* 2003;80(2 Suppl 1):i76–88.
3. CDC. Syndromic surveillance for bioterrorism following the attacks on the World Trade Center—New York City, 2001. *MMWR* 2002;51 (Special Issue)13–5.

**FIGURE 3. Timeliness plots of outbreak-detection methods for respiratory infection with fever and gastroenteritis (diarrhea, vomit) without blood syndromes**



\* Autoregressive integrated moving average.

† Cumulative sum.

§ Temporal aberration detection.

¶ T1 = first day of outbreak; T2 = second day of outbreak; T3 = third day of outbreak; T4 = fourth day of outbreak.

4. Tsui F-C, 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;10:399–408.
5. Lazarus R, Kleinman KP, Dashevsky I, DeMaria A, Platt R. Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection. *BMC Public Health* 2001;1:9.
6. Liu Z, Tsui F-C, Zeng X. Cache table design for disease surveillance system [Abstract]. In: Kohane IS (ed). *Proceedings of the Annual Fall Symposium of the American Medical Informatics Association*, November 9–13, 2002, San Antonio, Texas. Philadelphia, Pennsylvania: Hanley & Belfus, Inc, 2002;1086.
7. Ackelsberg J, Layton M. Update #5: terrorist attack at the World Trade Center in New York City: medical and public health issues 2001. Available at <http://www.nyc.gov/html/doh/html/cd/wtcf.html>.
8. Lewis MD, Pavlin JA, Mansfield JL, et al. Disease outbreak detection system using syndromic data in the greater Washington, DC, area. *Am J Prev Med* 2002;23:180–6.
9. Department of Defense Global Emerging Infections System. Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE). Available at <http://www.geis.fhp.osd.mil/geis/surveillanceactivities/essence/essence.asp>.
10. Schafer K. LEADERS (Lightweight Epidemiology Advanced Detection & Emergency Response System), presentation W203. Available at <http://www.tricare.osd.mil/conferences/2001/agenda.cfm>.
11. Zelicoff A, Brillman J, Forslund DW, et al. The rapid syndrome validation project (RSVP). *Proc AMIA Symp* 2001:771–5.
12. Lober WB, Karras BT, Wagner MM, et al. Roundtable on bioterrorism detection: information system-based surveillance. *J Am Med Inform Assoc* 2002;9:105–15.
13. Mundorff MB, Gesteland PH, Rolfs RT. Syndromic surveillance using chief complaints from urgent care facilities during the Salt Lake 2002 Olympic Winter Games [Abstract]. *MMWR* 2004;53 (Suppl):254.
14. Gesteland PH, Wagner MM, Chapman WW, et al. Rapid deployment of an electronic disease surveillance system in the state of Utah for the 2002 Olympic Winter Games. *Proc AMIA Symp* 2002:285–9.
15. Tsui F-C, Espino JU, Wagner MM, et al. Data, network, and application: technical description of the Utah RODS Winter Olympic Biosurveillance System. *Proc AMIA Symp* 2002:815–9.
16. Jorm LR, Thackway S, Churches TR, Hills MW. Watching the Games: public health surveillance for the Sydney 2000 Olympic Games. *J Epidemiol Community Health* 2003;57:102–8.
17. Meehan P, Toomey KE, Drinnon J, Cunningham S, Anderson N, Baker E. Public health response for the 1996 Olympic Games. *JAMA* 1998; 280:1229–30.
18. Dafni U, Gkolfinopoulou K, Tsiodras S, Saroglou G. Planning syndromic surveillance for the Athens 2004 Olympic Games: a pilot study [Abstract]. *J Urban Health* 2003;80(2 Suppl 1):i125.
19. Rogerson PA. A cumulative sum approach to syndromic surveillance in geographic regions [Abstract]. *J Urban Health* 2003;80(2 Suppl 1):i130.
20. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. *Proc Natl Acad Sci U S A* 2003;100:1961–5.
21. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. *BMC Med Inform Decis Mak* 2003;3:2.
22. Hutwagner L. Centers for Disease Control, Early Aberration Reporting System (EARS). Presentation, National Syndromic Surveillance Conference, New York, NY, September 23–24, 2002.
23. Hutwagner L, Thompson W, Seaman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). *J Urban Health* 2003;80(2 Suppl 1):i89–96.
24. Lawson A. Temporal-spatial methodologies and data sources. Presentation, National Syndromic Surveillance Conference, New York, NY, September 23–24, 2002.
25. Merriam GR, Wachter KW. Algorithms for the study of episodic hormone secretion. *Am J Physiol* 1982;243:E310–18.

26. Nindl BC, Hymer WC, Deaver DR, Kraemer WJ. Growth hormone pulsatility profile characteristics following acute heavy resistance exercise. *J Appl Physiol* 2001; 91:163–72.
27. Lee WM, Meij BP, Bhatti SF, Mol JA, Rijnberk A, Kooistra HS. Pulsatile secretion pattern of growth hormone in dogs with pituitary-dependent hyperadrenocorticism. *Domest Anim Endocrinol* 2003;24:59–68.
28. 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;3:395–400.
29. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979;74:829–36.
30. Hurvich CM, and Simonoff JS. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J R Stat Soc B* 1998;60:271–93.
31. Heffernan R, Mostashari F, Das D, et al. New York City syndromic surveillance systems. *MMWR* 2004;53(Suppl):25–7.
32. SAS Institute, Inc. SAS,® version 8.2 [software]. Cary, NC: SAS Institute, Inc.
33. CDC. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. *MMWR* 2004;53(No. RR-5):1–11.
34. Sonesson C, Bock D. A review and discussion of prospective statistical surveillance in public health. *J R Stat Soc A* 2003;165:5–21.
35. Mostashari F, Hartman J. Syndromic surveillance: a local perspective. *J Urban Health* 2003;80(2 Suppl 1):i1–7.
36. Kulldorff M. Spatial scan statistics: models, calculations and applications. In: Glaz J and Balakrishnan N, eds. *Scan statistics and applications*. Boston, MA: Birkhäuser, 1999;303–22.
37. Mostashari F, Kulldorff M, Hartman JJ, Miller JR, Kulasekera V. Dead bird clusters as an early warning system for West Nile virus activity. *Emerg Infect Dis* 2003;9:641–6.
38. Lawson AB, Cressie N. Spatial statistical methods for environmental epidemiology. In: Sen PK, Rao CR, eds. *Handbook of statistics*. Amsterdam, Holland: Elsevier, 2000;18:357–96.
39. Clark AB, Lawson A. Change-point detection using directional derivatives [Abstract]. *MMWR* 2004;53(Suppl):233.
40. Wessman P. Some principles for surveillance adopted for multivariate processes with a common change point. *Communications in Statistics* 1998;27:1143–61.