

Comparison of Two Signal Detection Methods in a Coroner-Based System for Near Real-Time Mortality Surveillance

MATTHEW R. GROENEWOLD,
MSPH^a

SYNOPSIS

Objectives. This article describes and compares the performance characteristics of two approaches to outbreak detection in the context of a coroner-based mortality surveillance system using controlled feature set simulation.

Methods. The comparative capabilities of the outbreak detection methods—the Epidemic Threshold and Cusum methods—were assessed by introducing a series of simulated signals, configured as nonoverlapping, three-day outbreaks, into historic surveillance data and assessing their respective performances. Treating each calendar day as a separate observation, sensitivity, predictive value positive, and predictive value negative were calculated for both signal detection methods at various outbreak magnitudes. Their relative performances were also assessed in terms of the overall percentage of outbreaks detected.

Results. Both methods exhibited low sensitivity for small outbreaks and low to moderate sensitivity for larger ones. In terms of overall outbreak detection, large outbreaks were detected with moderate to high levels of reliability, while smaller ones were detected with low to moderate reliability for both methods. The Epidemic Threshold method performed significantly better than the Cusum method for overall outbreak detection.

Conclusions. The use of coroner data for mortality surveillance has both advantages and disadvantages, the chief advantage being the rapid availability of coroner data compared to vital statistics data, making near real-time mortality surveillance possible. Given the lack of sensitivity and limited outbreak detection reliability of the methods studied, the use of mortality surveillance for early outbreak detection appears to have limited usefulness. If it is used, it should be as an adjuvant in conjunction with other surveillance systems.

^aLouisville Metro Health Department, Louisville, KY

Address correspondence to: Matthew R. Groenewold, MSPH, Louisville Metro Health Department, 400 East Gray St., P.O. Box 1704, Louisville, KY 40201; tel. 502-574-5292; fax 502-547-5635; e-mail <matt.groenewold@louisvilleky.gov>.

©2007 Association of Schools of Public Health

Mortality surveillance based on the analysis of official death records has long been one of the basic methods used in epidemiology for planning and evaluating the delivery of health services, setting health policy and priorities, and generating new causal hypotheses.¹ More recently, the threat of bioterrorism and other emerging threats to public health have focused attention on the use of public health surveillance systems, including mortality surveillance systems, for the early detection of outbreaks.² While such activities may often take place at the state and national level, the use of local mortality statistics also plays an important role in monitoring and improving community health.³

Timely dissemination of results to those who are in a position to take action or who need to know is generally considered an important element of any surveillance system.⁴ With increasing emphasis being placed on the importance of surveillance for the early detection of disease outbreaks, including potential instances of bioterrorism, the standard of what is considered timely is shifting. The need for rapid intervention in response to outbreaks requires that surveillance systems, including mortality surveillance, operate within a framework approaching real time. Therefore, reliance upon state vital records for community mortality surveillance is problematic, as the complete availability of official records may lag by months or even years. This time lag requires an alternative data source to vital records for timely community mortality surveillance.

The use of data from medicolegal death investigations—those carried out by coroners and/or medical examiners—has been suggested for mortality surveillance, particularly for sudden or unexpected deaths.^{5,6} Because they investigate deaths among people who have not accessed the health-care system and who may have died without a confirmed diagnosis, the Centers for Disease Control and Prevention (CDC) has recommended that coroners and medical examiners “should be a key component of population-based surveillance for biologic terrorism.”^{6(p.17)} The CDC has further recommended that electronic information systems for sharing data between medicolegal death investigators and public health authorities be designed for the rapid recognition of excess mortality.⁶

Kentucky law (KRS Chapter 72) provides for a joint coroner/medical examiner system in which elected coroners in each county and their deputies investigate unattended, unexplained, violent, and suspicious deaths to establish the cause and manner of death. When requested, the Office of the State Medical Examiner assists county coroners—who are not required to have medical training—in their investigations by providing postmortem forensic medical examinations, including autopsies and toxicological tests. State medi-

cal examiners do not have independent jurisdiction to investigate deaths and are only involved in cases referred to them by a county coroner. Because, in many cases, the cause and manner of death can be reasonably inferred from evidence gathered at the scene or from the decedent’s known medical history, not all coroners’ cases are referred to the medical examiner. For this reason, and because most data from coroners’ investigations are available much sooner than those from medical examiners, coroner data are more suitable for a system of near real-time, population-based mortality surveillance in Kentucky.

The Louisville Metro Health Department (LMHD), Jefferson County Coroner’s Office, and the University of Louisville School of Public Health and Information Sciences have collaborated to develop and deploy such a system in the Louisville/Jefferson County Metro with funding from the school’s Center for the Deterrence of Biowarfare and Bioterrorism.

The widespread use of public health surveillance systems for early outbreak detection and response is a recent development, having been stimulated by the anthrax attacks of 2001 and an increased awareness of the threat of bioterrorism. The usefulness of such systems for this purpose has not been definitively established; measurement of the performance characteristics of early detection surveillance systems is required to determine the relative value of systems using differing approaches and designs.²

However, in the absence of historic data containing known examples of the type of event a surveillance system is designed to detect, evaluation of the system’s performance is only possible using simulation. This type of simulation can be based on datasets that are authentic, synthetic, or semisynthetic. One way to create such a semisynthetic dataset is to spike authentic (historic) background data with a simulated signal. The use of controlled feature sets allows for the systematic definition of simulated signals based on four variable parameters: outbreak magnitude, duration, spacing, and temporal progression.⁷

This article describes and compares the performance characteristics of two approaches to early event detection in the context of LMHD’s coroner-based mortality surveillance system using controlled feature set simulation.

METHODS

The comparative capabilities of the two outbreak detection methods were assessed by introducing a series of simulated signals, configured as temporal clusters, into historic surveillance data from 2004 and assessing their respective performances.

Surveillance system description

In this system, deputy coroners input data from death investigations in the field using laptop computers, and these data are uploaded to a master database upon completion of the investigation. Every 24 hours, an updated copy of the database is uploaded to a LAN folder that is shared with LMHD. The data is downloaded by LMHD each morning for daily analysis. Two different methods are used in the analysis of daily case counts to detect significant departures from baseline mortality: the Epidemic Threshold and Cusum methods.

When the system signals a significant increase in mortality, further epidemiologic analyses are performed using data already available from the coroner's database, including analyses of the presumed cause and manner of death and geographic as well as age, gender, and race stratified analyses. Such analyses are intended to determine whether further data collection and investigation are warranted. In some cases, these intermediate analyses may rule out the need for further follow-up; for example, identifying excess deaths as the result of a multiple-vehicle traffic crash. In cases in which further follow-up is required, they will define or narrow the focus of the epidemiologic field investigations to be undertaken.

Data

Daily counts of coroner cases were electronically available from the beginning of 2001. Data from 2001 through 2003 were used to model an expected baseline and simulations were performed using data from 2004, which was spiked with the simulated outbreaks. To account for seasonality and secular trend in the data, the baseline was constructed using a cyclical regression approach similar to that proposed by Serfling.⁸

The Serfling method normally aggregates deaths at the weekly level. In this case, however, the model was based upon daily aggregation of event data, requiring slight modification of the regression equation. The applied model took the following form: $Y_t = a + b \times t + c \times \sin(2 \times \pi \times t/365) + d \times \cos(2 \times \pi \times t/365)$ where Y_t is the predicted number of coroner's cases for day t ; t is the index for day of death; a is the mean number of daily cases for the period 2001 through 2003; $b \times t$ is a linear secular trend component; and c and d are harmonic coefficients used in the sine wave component to model seasonality. The confidence limits for the predicted values were computed as ± 1.96 standard error (SE) of the model. The results of the model, along with the observed data from 2004, are shown in Figure 1.

The Serfling method also calls for the removal of epidemic periods from the data upon which the model

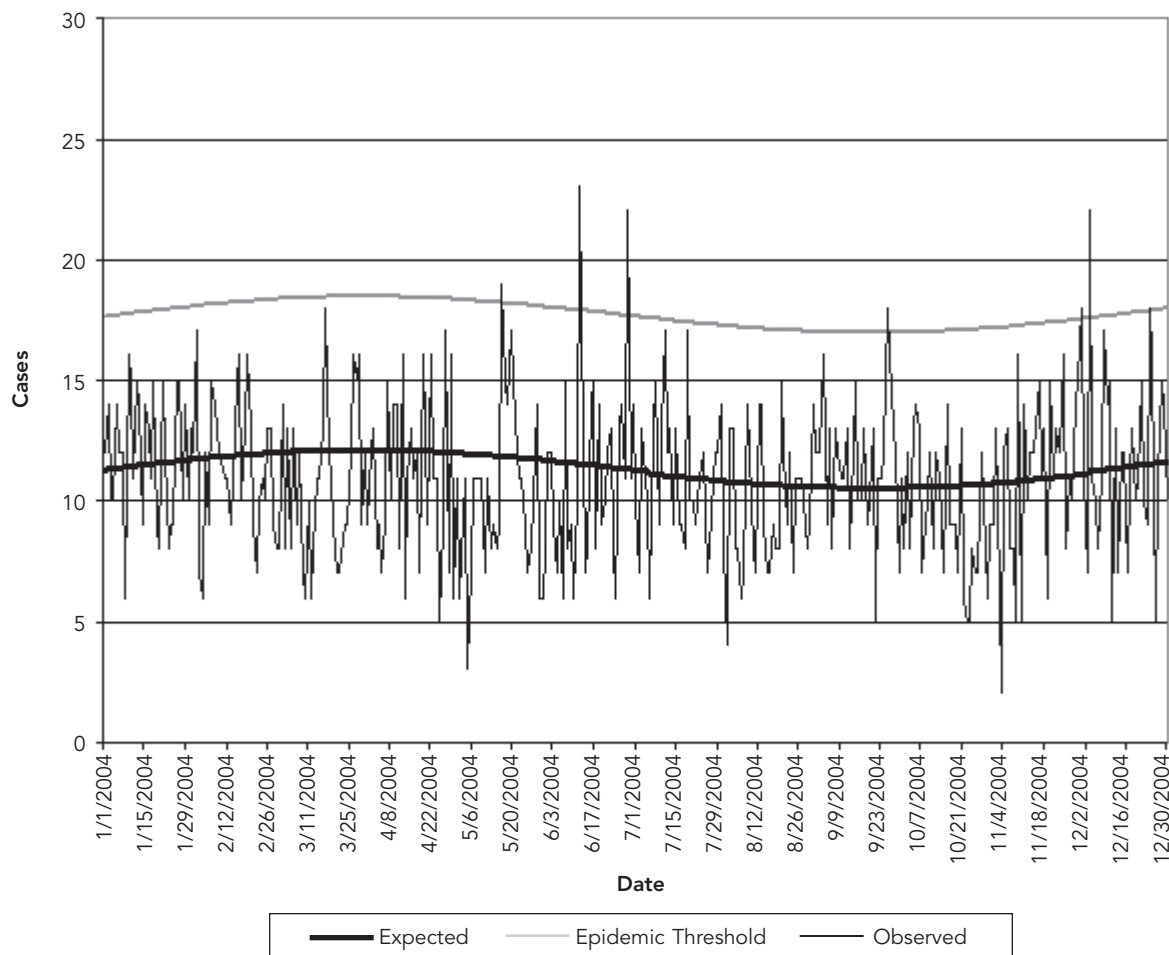
is based. However, because nothing in the data that could be considered a period of epidemic mortality was associated with any particular, identifiable event or process, data from the entire period were used.

Controlled feature sets

Semisynthetic datasets consisting of authentic background noise and controlled feature set-simulated outbreaks were created by injecting a series of identical, three-day-long outbreaks spaced three days apart into the 2004 surveillance data. A three-day lag was maintained between the end of each outbreak and the onset of the subsequent one, thus the first of these outbreaks began on January 1, the second on January 7, and so on. In this way, there were 183 outbreak days comprising 61 distinct outbreaks over the course of the 366 days of data (2004 was a leap year). These outbreaks were then removed and another, identical series was introduced but shifted forward by one day (i.e., the first outbreak began on January 2). This procedure was repeated two more times, introducing the controlled feature set on January 3 and 4, respectively. As a result, the entire dataset yielded 244 distinct, nonoverlapping outbreaks for analysis, each occurring in a unique temporal context.

The previously mentioned simulation was carried out nine times, using outbreaks of increasing magnitudes. Outbreak size was increased using multiples of the standard deviation of the error profile of the model. The error profile of the model is defined as the distribution of the daily forecast errors, which are defined as the expected minus the observed number of events for each day.⁷ The error profile can be thought of as the amount of background noise from which an outbreak must be distinguished in order to signal an alert. When the signal-to-noise ratio is near one, outbreak detection will typically be poor. As this ratio increases (i.e., with increasing outbreak magnitude), detection performance should improve.⁷

The smallest outbreak magnitude tested was two times the standard deviation of the error profile. Assuming that forecast error is random, the error profile should be normally distributed, and positive two standard deviations would correspond approximately with the upper 2.5% of the forecast error values. The outbreak size was increased by one standard deviation (rounded to the nearest whole number) up to a maximum of 10 times the standard deviation of the error profile. In all cases, the temporal progression of the outbreaks was configured as linear, with the number of additional cases increasing by the same amount each day of the outbreak.

Figure 1. Expected and observed daily numbers of coroner cases: Jefferson County, KY, 2004

Signal detection

Epidemic threshold. The Epidemic Threshold method was originally designed by Serfling⁸ and is currently used by the CDC, with some modifications, for pneumonia and influenza mortality surveillance.⁹ This method has also been applied to nontraditional mortality surveillance in Madrid, Spain, using weekly counts from the city's undertaker database.¹⁰ The epidemic threshold was established based upon the upper 95% confidence limit of the modeled baseline. Daily excursions above the epidemic threshold are counted as alerts. An example of the application of the Epidemic Threshold method to surveillance data is illustrated in Table 1.

Cusum. The Cusum method is a modification of a technique that was originally developed for quality-control monitoring of industrial processes that relies on the monitoring of cumulative differences between

observed and expected data in a time window when compared to a threshold.¹¹ The method was applied to public health surveillance beginning in the 1960s^{12,13} and, more recently, has been used in bioterrorism surveillance applications.¹⁴⁻¹⁶

In this method, for each time interval, a predetermined reference value, k , based on the expected number of cases, is subtracted from a second value, Y , derived from the observed number. The resulting sum, $Y - k$, is then added to the sum for the previous period, and so on. If the cumulative sum of these values, the cusum, exceeds a warning value, H , based on the expected number of cases, a significant departure from baseline is indicated. Because it is only necessary to detect significant increases in disease frequency, not decreases, the cusum is not allowed to fall below zero. If a negative value is obtained, the cusum is set to zero.

In this case, an adaptation of the Cusum method

Table 1. Detection of a simulated outbreak: Epidemic Threshold method with and without simulated outbreak

Date	Expected ^a	Observed ^b	Inject ^c	New Obs ^d	Epidemic Threshold	Status
Without simulated outbreak						
1/22/2004	11.55	12	0	12	17.92	Normal
1/23/2004	11.57	8	0	8	17.93	Normal
1/24/2004	11.58	9	0	9	17.95	Normal
1/25/2004	11.60	9	0	9	17.96	Normal
1/26/2004	11.61	15	0	15	17.97	Normal
1/27/2004	11.62	15	0	15	17.99	Normal
With simulated outbreak						
1/22/2004	11.55	12	0	12	17.92	Normal
1/23/2004	11.57	8	0	8	17.93	Normal
1/24/2004	11.58	9	0	9	17.95	Normal
1/25/2004	11.60	9	2	11	17.96	Normal
1/26/2004	11.61	15	4	19	17.97	Normal
1/27/2004	11.62	15	6	21	17.99	Alert

^aExpected number of coroner's cases, modeled from 2001–2003 data

^bActual observed number of coroner's cases, from 2004 surveillance data

^cNumber of simulated coroner's cases

^dObserved number of coroner's cases, including simulated cases

described by Aldrich and Drane¹⁷ was used. This methodology was adopted because the expected daily count of coroner's cases was larger than the design of the traditional Cusum method intended. In this adaptation, rather than using raw event counts, a transformed, *z*-like value—*Y*—is used. It is derived as:

$$Y = \frac{X - \mu}{\sqrt{\mu}}$$

X is the observed and μ the expected number of daily events. The value for *k* is set at one and the alert threshold for the cusum value, *H*, is set at two. Excursions by the cusum value above the alert threshold were counted as outbreak alerts. The expected daily number of coroner's cases was derived from the Serfling baseline.

Because the cusum value is dependent upon preceding time periods, it was necessary to temporally isolate each outbreak to ensure that any response from the previous outbreak was eliminated from the system before a new one was introduced. This isolation was accomplished by removing the simulated data for each outbreak before recording the response for the next. An example of the application of the Cusum method to surveillance data is illustrated in Table 2.

ANALYSIS

Sensitivity, predictive value positive (PVP), and predictive value negative (PVN) were calculated for both

signal detection methods for each outbreak magnitude. These calculations were based on the detection of specific outbreak days, with each day considered as a separate, independent case.

For both signal detection methods, a true positive result was recorded when the system signaled an alert on an outbreak day, and a true negative result was recorded if no alert was signaled on a nonoutbreak day. A false positive result was recorded when the system signaled an alert on a nonoutbreak day, and a false negative result was recorded if no alert was signaled on an outbreak day.

Sensitivity was calculated as the number of true positives divided by the number of outbreak days (true positives plus false negatives). PVP was calculated as the number of true positives divided by the total number of alert days (true positives plus false positives). PVN was calculated as the number of true negatives divided by the total number of non-alert days (true negatives plus false negatives). Specificity—calculated as the number of true negatives divided by the total number of nonoutbreak days (true negatives plus false positives)—was determined for both methods by simply applying each method to the baseline data, without injecting any signals, and was held constant for all simulations.

The performance of each signal detection strategy was also assessed in terms of the percentage of overall outbreaks detected, viewing each three-day outbreak as a single entity. Using this approach, the system was considered to have correctly identified an outbreak

Table 2. Detection of a simulated outbreak: Cusum method with and without simulated outbreak

Date	Exp ^a	Obs ^b	Inject ^c	New Obs ^d	Y	K	H	Y - K	Cusum	Status
Without simulated outbreak										
1/22/2004	11.55	12	0	12	0.13	1	2	-0.87	0	Normal
1/23/2004	11.57	8	0	8	-1.05	1	2	-2.05	0	Normal
1/24/2004	11.58	9	0	9	-0.76	1	2	-1.76	0	Normal
1/25/2004	11.60	9	0	9	-0.76	1	2	-1.76	0	Normal
1/26/2004	11.61	15	0	15	1.00	1	2	0.00	0	Normal
1/27/2004	11.62	15	0	15	1.00	1	2	0.00	0	Normal
With simulated outbreak										
1/22/2004	11.55	12	0	12	0.13	1	2	-0.87	0	Normal
1/23/2004	11.57	8	0	8	-1.05	1	2	-2.05	0	Normal
1/24/2004	11.58	9	0	9	-0.76	1	2	-1.76	0	Normal
1/25/2004	11.60	9	2	11	-0.17	1	2	-1.17	0	Normal
1/26/2004	11.61	15	4	19	2.17	1	2	1.17	1.17	Normal
1/27/2004	11.62	15	6	21	2.75	1	2	1.75	2.92	Alert

^aExpected number of coroner's cases, modeled from 2001–2003 data

^bActual observed number of coroner's cases, from 2004 surveillance data

^cNumber of simulated coroner's cases

^dObserved number of coroner's cases, including simulated cases

(a true positive) if an alert was signaled on any of the three days of the outbreak.

Because the signal detection capabilities of the Cusum method depend upon the accumulation of excess events over successive time periods, the comparative performance of the two methods was assessed using only the percentage-of-overall-outbreaks-detected approach. The significance of the difference between the percentages of overall outbreaks detected by each method at each outbreak magnitude was assessed using McNemar's test for paired proportions.

RESULTS

The standard deviation of the model's error profile was 3.272. Thus, the smallest outbreak magnitude tested was six excess cases. This number was increased in increments of three to a maximum outbreak size of 30. So that the shape of the epidemic curve would be approximately the same at each magnitude, outbreaks were not simply increased in size by one excess case per day. Instead, the number of additional cases per day was increased by one at every other size increment, while still increasing the overall outbreak size by three at each increment. This method resulted in outbreak progressions of 1-2-3, 2-3-4, 2-4-6, 3-5-7, 3-6-9, 4-7-10, and so on, rather than 1-2-3, 2-3-4, 3-4-5, 4-5-6, 5-6-7, and so on.

The sensitivities, PVP, and PVN of the two signal detection methods, in terms of the number of specific outbreak days detected at each outbreak magnitude,

are shown in Table 3. Analyzing the data without artificial signals resulted in the Epidemic Threshold method indicating five false positives and the Cusum method indicating three false positives. These results corresponded to specificities of 0.973 and 0.984, respectively.

The comparative percentages of overall outbreaks detected by the two methodologies are depicted in Figure 2. These results, as well as the significance of the difference in the proportions detected by the two methodologies, are also presented in Table 4.

DISCUSSION

The purpose of this article, apart from providing a description of a coroner-based system for near real-time community mortality surveillance, is to describe and compare the performance characteristics of two analytic methods for early event detection in the context of that system. More broadly, it is to examine the usefulness of controlled feature set simulation for evaluating surveillance systems and the analytic methodologies used in them for signal detection.

To be useful, signal detection methods must have acceptable operating characteristics. However, determining these characteristics in real-world data settings is difficult due to the limited availability of datasets containing bioterrorism attacks or attack-like events and for other reasons.^{18,19} Consequently, while early outbreak detection systems have been implemented in a variety of locales across the country as a result of fed-

Table 3. Sensitivity, PVP, and PVN of signal detection methods

Outbreak magnitude	N	Epidemic Threshold			Cusum				
		Outbreak days detected	Sens	PVP	PVN	Outbreak days detected	Sens	PVP	PVN
6	732	53	0.072	0.862	0.514	19	0.026	0.731	0.504
9	732	98	0.134	0.828	0.531	41	0.056	0.854	0.512
12	732	179	0.245	0.923	0.565	82	0.112	0.921	0.527
15	732	243	0.332	0.942	0.595	162	0.221	0.959	0.560
18	732	318	0.434	0.955	0.634	224	0.306	0.970	0.588
21	732	384	0.525	0.962	0.673	288	0.393	0.976	0.620
24	732	442	0.604	0.967	0.712	334	0.456	0.979	0.646
27	732	489	0.668	0.970	0.747	402	0.549	0.990	0.687
30	732	516	0.705	0.972	0.768	422	0.577	0.984	0.700

Sens = sensitivity
 PVP = predictive value positive
 PVN = predictive value negative

eral, state, and local initiatives, the number of rigorous assessments of their signal detection methods has been relatively limited,^{19,20} especially in terms of comparative studies that assess their relative strengths.¹⁸

The CDC has promulgated provisional recommendations for evaluating surveillance systems for

early outbreak detection, including guidelines for assessing the performance of analytical algorithms through simulation. The closely related metrics—sensitivity, PVP, and PVN—comprise a useful framework for these assessments. However, because acceptable levels of precision will likely vary from community to

Figure 2. Percent of outbreaks detected

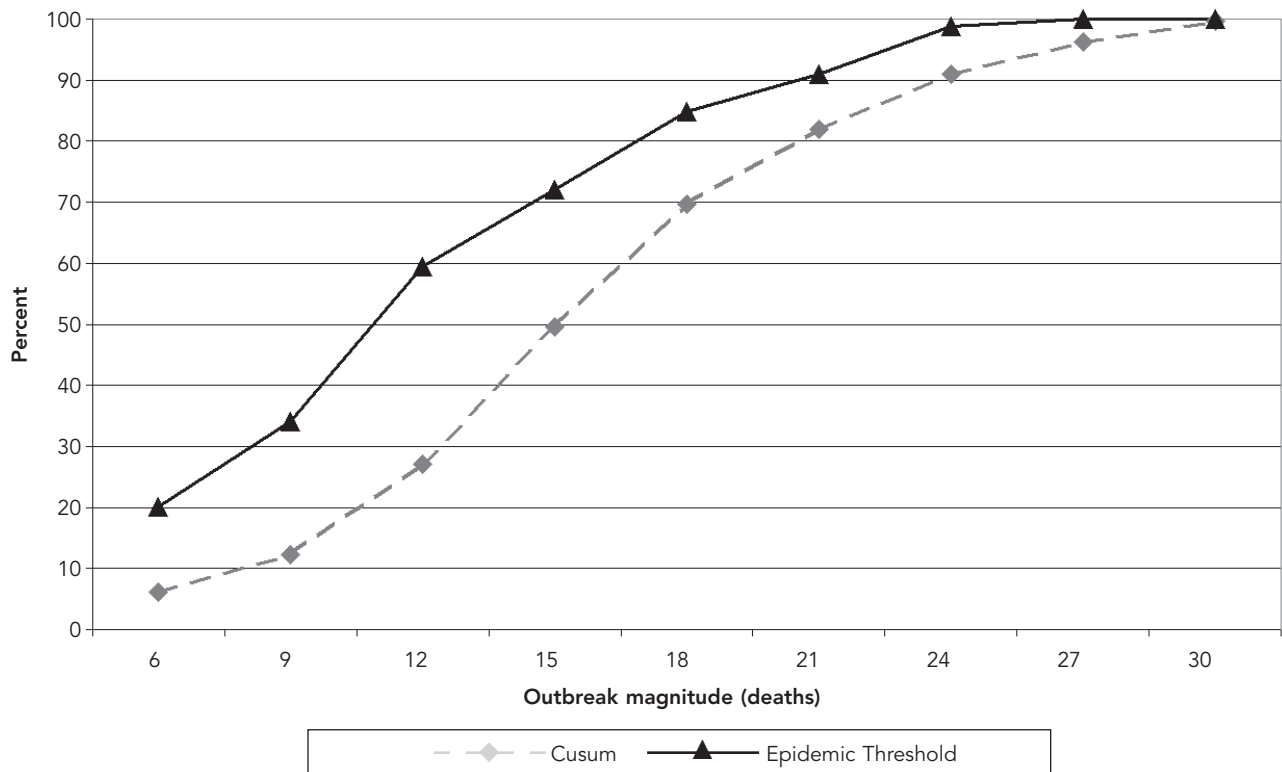


Table 4. Proportion of outbreaks detected

Outbreak magnitude	N	Outbreaks detected				
		Epidemic Threshold	Proportion	Cusum	Proportion	P-value ^a
6	244	49	0.201	15	0.061	<0.001
9	244	83	0.340	30	0.123	<0.001
12	244	145	0.594	66	0.270	<0.001
15	244	176	0.721	121	0.496	<0.001
18	244	207	0.848	170	0.697	<0.001
21	244	222	0.910	200	0.820	<0.00 ^b
24	244	241	0.988	222	0.910	<0.00 ^c
27	244	244	1.000	235	0.963	NA ^d
30	244	244	1.000	243	0.996	NA ^d

^aBased on McNemar's test for paired proportions, using the binomial distribution

^bDue to one small expected value, Fisher's Exact test also used. Fisher's Exact $p < 0.001$.

^cDue to two small expected values, Fisher's Exact test also used. Fisher's Exact $p = 0.001$.

^dNo statistics calculated. One variable (Epidemic Threshold) is constant.

NA = not available

community depending upon the perceived likelihood of an outbreak, benefits of early detection, and likely costs of investigating false alarms, specific targets for these metrics are not given.²

In this study, with specificity held constant, sensitivity, PVP, and PVN all increased with increasing outbreak magnitude, as would logically be expected. Following a generally linear increase, sensitivity ranged from single digits for the smallest outbreak to between 50% and 70% for the largest outbreak, depending on the analytic method. The PVP was high for both methods, even for small outbreaks.

Despite the fall 2001 anthrax attacks, bioterrorism is extremely rare in the United States and the likelihood of an attack must still be considered quite small in any given community. Surveillance systems designed for the early detection of such events, therefore, are sometimes justified on the basis of the reassurance they can provide that such events have not occurred or are not occurring, especially during perceived times of increased risk.² The credibility of such reassurance depends upon the validity of the system's negative results, which can be assessed in terms of PVN. In this study, PVN percentages ranged from the low 50s for the smallest outbreak size tested to the low-to-mid-70s for the largest, depending on the analytic method.

A side-by-side comparison based on the percentage of outbreaks detected suggested that the Epidemic Threshold method performs better than the Cusum method, at least under the conditions of this study. The Epidemic Threshold method detected significantly higher proportions of outbreaks than the Cusum method at all but the highest outbreak magnitudes,

with a virtually identical level of specificity. Outbreaks could be detected with reasonable reliability (probability of detection $\geq 80\%$) if they involved at least 18 excess deaths for the Epidemic Threshold method, or 21 for the Cusum method.

Limitations

The results of this study are subject to certain limitations that arise from both the data-simulation and analysis methodologies. In this study, the epidemic curves of the simulated outbreaks were linear and their durations quite short. However, there are a number of other canonical epidemic curve shapes, and irregular epidemic curves are possible as well.⁷ It is reasonable to expect that the outbreak detection methods used in this analysis would produce different operating characteristics for epidemics with different shapes or durations.

There are a number of statistical algorithms available for detecting aberrations in time-series data. Methodologies from the fields of epidemiology, statistical process control, signal processing, and data mining have been adapted for surveillance and early outbreak detection in recent years.²¹ It is not yet clear, however, which method or methods are best suited for early event detection in surveillance systems,¹⁸ and it is likely that certain methods will be better suited to particular data settings than others. This article focuses on two methods, one epidemiologic and one from statistical process control, which are used by the LMHD. Other methods could be expected to have different operating characteristics when applied to the same data.

Both methods rely upon the use of a baseline mod-

eled using a limited amount of historic data, from 2001 through 2003, as these were the only data available electronically. The use of more data to train the Serfling model could have produced a more accurate baseline, which could have affected the performance characteristics of either or both methods.

The use of coroner data has both advantages and disadvantages. On the one hand, these data are usually available much sooner than vital statistics data and often include detailed information on the circumstances surrounding the deaths, including the suspected cause and manner of death. This information not only allows for timely surveillance analysis, but also makes it possible to conduct the initial steps in the follow-up investigations of alerts without having to collect additional data. On the other hand, coroner and medical examiner databases do not include all deaths that occur in a community, and the subsets that they do contain are affected by both selection and information biases.⁵

The analysis of coroner data in near real time presents potential problems that are not normally encountered in the retrospective analysis of vital statistics data, including data lag issues and the requirement to preprocess and quality assure the data. While Jefferson County Coroner's Office policy requires that data from death investigations be entered in the database before the end of the shift on which the investigation took place, even relatively short delays in data entry (e.g., while waiting to receive or confirm certain data) can affect surveillance analyses. Also, coroner data available for near real-time analysis cannot undergo the extensive preprocessing and quality-assurance processes that vital statistics data normally do before they are officially made available for analysis. Both of these issues represent potential sources of error in surveillance analyses.

Implicit in the design of this study is the assumption that all of the deaths comprising the outbreaks would be investigated by the Coroner's Office and, therefore, reported to the surveillance system. In actuality, of course, this would not necessarily be the case, as deaths not appearing suspicious (e.g., from infectious disease) that occur in hospitals 24 hours or more after admission might not be reported to the Coroner's Office. It is for this reason that the coroner-based surveillance is considered useful primarily for surveilling community mortality, and other adjuvant systems should be relied upon for the surveillance of mortality in health-care facilities.

Additionally, while outbreak detection based on mortality surveillance alone may be feasible for outbreaks caused by highly lethal agents that produce

death quickly (e.g., nerve agents, other chemical or biological toxins) or for events normally recognized by the occurrence of excess deaths (e.g., heat waves), most of the likely bioterrorism agents have incubation periods that range from days to weeks, and acute illness is often preceded by a prodromal period of varying duration. In such cases, if an outbreak cannot be detected unless and until deaths occur, then a critical window of opportunity for early intervention will have been missed.

For these reasons, the LMHD uses coroner-based mortality surveillance, in conjunction with other forms of surveillance, including disease reporting, over-the-counter pharmaceutical sales, ambulance runs, and emergency department-based syndromic surveillance. In most cases, therefore, it is expected that the coroner-based mortality surveillance system would provide corroborating evidence of the existence of an outbreak rather than the initial alert.

Of course, extraordinarily large numbers of unexplained deaths are likely to arouse suspicion on the part of coroners even in the absence of a statistical signal. Given that, for both analytic methods, the probability of detecting an outbreak is acceptable only for relatively large outbreaks (≥ 18 cases over three days) (Table 4), this could be seen as calling into question the need for such a surveillance system. However, there are a number of reasons why, even in the face of limited sensitivity, the system remains useful.

First, without a surveillance system even relatively large outbreaks could conceivably go unrecognized if the circumstances of the deaths were not overtly suspicious, given that the outbreak could occur over several days and the individual cases distributed among two shifts per day and two deputy coroners per shift. Without the systematic aggregation and analysis of these data—a fundamental task of surveillance—temporal clusters could be overlooked. Second, statistical analysis of the data helps avoid false positives as well as false negatives, as reliance on coroners' subjective assessments of the occurrence of temporal clusters may trigger unnecessary investigations. Third, and perhaps most important, the existence of the system represents a formal channel of communication between the Coroner's Office and LHMD, which reinforces the need for and desirability of interagency information sharing.

CONCLUSION

Ultimately, the usefulness of community mortality surveillance systems for early outbreak detection must be evaluated prospectively and their true operating

characteristics determined using long-term studies that detect actual events. The high-mortality outbreaks such systems are designed to detect, however, are low-frequency events. Studies using real-world data, therefore, will require many years and much larger populations under surveillance to reach reliable conclusions. In the interim, controlled feature set simulation appears to be a useful method for describing and comparing the performance of the signal detection methods used in outbreak surveillance systems when real-world event data are lacking.

Within the context of the coroner-based mortality surveillance system used by LMHD, applications of both the Cusum and Epidemic Threshold methods exhibited low sensitivity for small outbreaks and low to moderate sensitivity for larger ones. In terms of overall outbreak detection, large outbreaks were detected with moderate to high levels of reliability while smaller ones were detected with low to moderate reliability for both methods. However, in this study, the Epidemic Threshold method performed significantly better than the Cusum method for overall outbreak detection.

The use of coroner data for mortality surveillance has both advantages and disadvantages. The chief advantage is the rapid availability of coroner data compared to vital statistics data, making near real-time mortality surveillance possible. The sole use of mortality surveillance for early outbreak detection has limited usefulness, however, and, given the lack of sensitivity and limited outbreak detection reliability of the methods employed in its coroner-based system, LMHD continues to use it as an adjuvant in conjunction with other surveillance systems.

The author gratefully acknowledges the generous contribution of funding and technical assistance provided by the University of Louisville School of Public Health and Information Sciences, Center for the Deterrence of Biowarfare and Bioterrorism, which made the coroner-based mortality surveillance system possible. The author would also like to thank the Jefferson County Coroner, Dr. Ronald Holmes, and the staff of his office, especially Chief Deputy Coroner Mark Handy. Without their patient cooperation and participation, the project would not have been possible.

Funding for the Louisville/Jefferson County coroner-based surveillance system was provided by the Center for the Deterrence of Biowarfare and Bioterrorism at the University of Louisville School of Public Health and Information Sciences.

REFERENCES

1. Denver A. *Epidemiology in health service management*. Fredrick (MD): Aspen; 1984.
2. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V. Framework for evaluating public health surveillance systems for early detection of outbreaks; recommendations from the CDC Working Group. *MMWR Recomm Rep* 2004;53(RR-5):1-11.
3. Durth JS, Bailey LA, Stoto MA, editors. *Improving health in the community: a role for performance monitoring*. Washington: National Academies Press; 1997.
4. Thacker SB, Berkelman RL. *Public health surveillance in the United States*. *Epidemiol Rev* 1988;10:164-90.
5. Graitcer PL, Williams WW, Finton RJ, Goodman RA, Thacker SB, Hanzlick R. An evaluation of the use of medical examiner data for epidemiologic surveillance. *Am J Public Health* 1987;77:1212-4.
6. Nolte KD, Hanzlick RL, Payne DC, Kroger AT, Oliver WR, Baker AM, et al. *Medical examiners, coroners, and biologic terrorism: a guidebook for surveillance and case management*. *MMWR Recomm Rep* 2004;53(RR-8):1-27.
7. Mandl KD, Reis B, Cassa C. Measuring outbreak-detection performance by using controlled feature set simulations. *MMWR Morb Mortal Wkly Rep* 2004;53(Suppl):130-6.
8. Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Rep* 1963;78:494-506.
9. Assad F, Cockburn WC, Sundaresan TK. Use of excess mortality from respiratory disease in the study of influenza. *Bul WHO* 1973; 49:219-33.
10. Alberdi JC, Ordoñas M, Quintana F. Construcción y evaluación de un sistema de detección rápida de mortalidad mediante análisis de Fourier, estudio de un valor con desviación máxima. *Rev Esp Salud Pública* 1995;69:207-17.
11. Ewan WD, Kemp KW. Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika* 1960; 47:363-80.
12. Hill GB, Spicer CC, Weatherall JA. The computer surveillance of congenital malformations. *Br Med Bull* 1968;24:215-8.
13. Weatherall JA, Haskey JC. Surveillance of malformations. *Br Med Bull* 1976;32:39-44.
14. Syndromic surveillance for bioterrorism following the attacks on the World Trade Center—New York City, 2001. *MMWR Morb Mortal Wkly Rep* 2002;51(Special issue):13-5.
15. Yuan CM, Love S, Wilson M; Centers for Disease Control and Prevention (US). Syndromic surveillance at hospital emergency departments—southeastern Virginia. *MMWR Morb Mortal Wkly Rep* 2004;53(Suppl):56-8.
16. Rogerson PA, Yamada I; Centers for Disease Control and Prevention (US). Approaches to syndromic surveillance when data consist of small regional counts. *MMWR Morb Mortal Wkly Rep* 2004;53(Suppl):79-85.
17. Agency for Toxic Substances and Disease Registry. *CLUSTER 3.1 software system for epidemiologic analysis instruction manual*. Atlanta: ATSDR; 1993.
18. Kleinman KP, Abrams A, Mandl K, Platt R. Simulation for assessing statistical methods of biologic terrorism surveillance. *MMWR Morb Mortal Wkly Rep* 2005;54(Suppl):101-8.
19. Rolka H, Bracy D, Russell C, Fram D, Ball R. Using simulation to assess the sensitivity and specificity of a signal detection tool for multidimensional public health surveillance data. *Stat Med* 2005; 24:551-62.
20. Buckeridge DL, Switzer P, Owens D, Siegrist D, Pavlin J, Musen M. An evaluation model for syndromic surveillance: assessing the performance of a temporal algorithm. *MMWR Morb Mortal Wkly Rep* 2005;54(Suppl):109-15.
21. Burkom HS. Development, adaptation, and assessment of alerting algorithms for biosurveillance. *Johns Hopkins APL Technical Digest* 2003;24:335-42.