
Can we lower the false positive rates of bioterrorism alarms enough to tip the cost-benefit scales to an acceptable balance?

Syndromic Surveillance: Is it Worth the Effort?

Michael A. Stoto, Matthias Schonlau, Louis T. Mariano

Unlike bombings, bioterrorism can be invisible, unapparent until people become ill, spreading silently as infected people interact with others. The flying bombs of September 11, 2001, the subsequent anthrax letters, and credible threats of future attacks have greatly heightened the need for a bioterrorism early warning system. The sooner public health officials know about a bioterrorist attack, the more effective their response can be. Early detection speeds quarantines, vaccinations, and treatments.

Since the terrorist attacks on September 11, a number of cities and states have worked to develop "syndromic surveillance" systems based on existing data. Syndromic surveillance involves the statistical analyses of data on individuals seeking care in emergency rooms or other health care settings for early symptoms of bioterrorist agents. Focusing on symptoms rather than confirmed diagnoses, syndromic surveillance aims to detect bioevents earlier than would be possible with traditional disease surveillance systems. Because potential bioterrorist agents such as anthrax, plague, brucellosis, tularemia, Q-fever, glanders, smallpox, and viral hemorrhagic fevers initially present a "flu-like illness," data suggesting a sudden increase of individuals with fever, headache, muscle pain, and

malaise could be the first indication of either a bioterrorist attack or a natural disease outbreak.

Researchers and public health experts are developing a wide variety of syndromic surveillance systems. These systems involve gathering data, automating statistical analysis of the data, and communicating results among physicians and public health officials. Some systems collect and analyze more than symptom data. They include data such as over-the-counter (OTC) pharmaceutical sales and absenteeism.

Syndromic surveillance systems are intended to raise an alarm, which then must be followed up by investigation and prevention. However, all alarm systems face trade-offs arising from imperfect knowledge about the occurrence of an attack. First, there is the common trade-off between true positives: sensitivity, sounding the alarm when an attack occurs, and false positives: one minus the specificity, sounding an alarm when there is no attack.

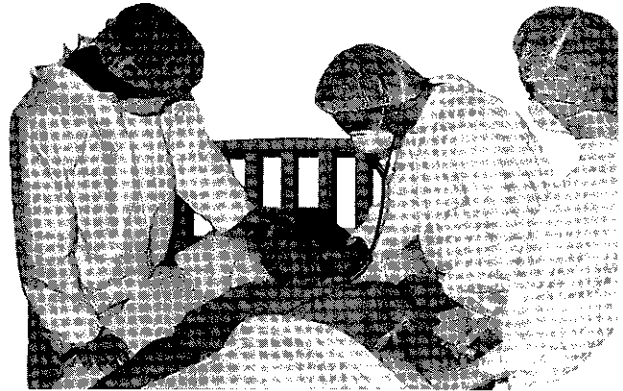


Photo courtesy of TimeLife Books

Timeliness is a key factor in the trade-off. Analyzing data for longer periods of time can help improve the accuracy of our alarm system, but too much time eliminates the value of the alarm.

Usually, it is possible to make the false positive rate tolerably small, but lowering the false positive rate generally decreases the true positive rate, slows timeliness, or both. Moreover, it is likely that thousands of syndromic surveillance systems will be running simultaneously in cities, counties, and states throughout the United States. Each of these jurisdictions might be looking at data in six to eight symptom categories, separately from each hospital in the area, and so on. As an example, suppose every county in the United States has an independent statistical

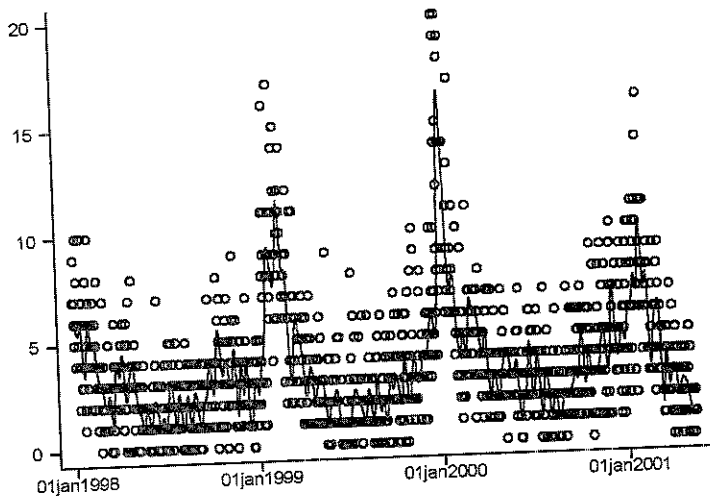


Figure 1. Number of "influenza-like illness" cases at the George Washington University Hospital Emergency Department, January 1998 – June 2001.

algorithm in place that is used daily and that has a 0.1% false-positive rate. Because there are approximately 3,000 counties, on average three counties a day would have a false-positive alarm. The impact of false alarms is financial and psychological. Response to phantom events costs money and false alarms desensitize responders to real ones. While any particular county would experience a false-positive alarm only about once every three years, this nationwide false-positive rate would be unacceptable.

Public health agencies around the country that are considering investing in syndromic surveillance systems must address a number of questions. What kinds of attacks are syndromic surveillance systems reasonably able to detect? Can more sophisticated statistical tools do better than simple methods? Should syndromic surveillance be integrated into public health practice? How can syndromic surveillance systems help meet other health goals? We present a simple simulation study to explore the first two questions.

A Simple Simulation Study

In order to explore the first two questions, we ran a simple experiment. Using data from the emergency department at the George Washington University Hospital in Washington, D.C., we simulated a bioterrorist attack and examined how different detection algorithms performed. While our analy-

sis focused on syndromic surveillance of emergency room data, the principles apply to other data types as well.

Figure 1 shows the number of cases of "influenza-like illness" (ILI) cases that presented each day between January 1998 and June 2001. Over the entire year the average is about three per day, but there is a clear annual pattern with up to 20 cases per day in the winter. We then "seeded" these data with two kinds of outbreaks, as illustrated in Figure 2. A "fast" outbreak was defined as 18 additional cases over three days – 3 on the first day, 6 on the second, and 9 on the third. Our simu-

lated "slow" outbreak involved the same total number of cases, but they were distributed over nine days as follows: 1, 1, 1, 2, 2, 2, 3, 3, 3. Each of these simulated outbreaks was added starting on each of the roughly 1,000 days in our database. We looked separately at the winter flu season and the rest of the year since background patterns were so different. Bioterrorist attacks and natural disease outbreaks can appear in many forms, but as we discuss further, the patterns that we tested were chosen to illustrate a middle ground in which statistical detection methods would have an advantage over traditional approaches.

We tried four statistical algorithms to determine when an alarm should be sounded. The first and simplest of these, the Shewhart algorithm, is familiar to epidemiologists, although not under statistician Walter Shewhart's name. This method triggers an alarm when the number of ILI cases on a single day t , y_t , exceeds a threshold, h . The threshold is set so that alarms are sounded with some small, prespecified probability, in our case 1%, even when there is no attack.

The other three statistical detection algorithms averaged cases over several days, so they are called integrative methods. The Exponentially Weighted Moving Average (EWMA) statistic for day t is defined recursively as

$$z_t = \lambda y_t + (1-\lambda)z_{t-1}$$

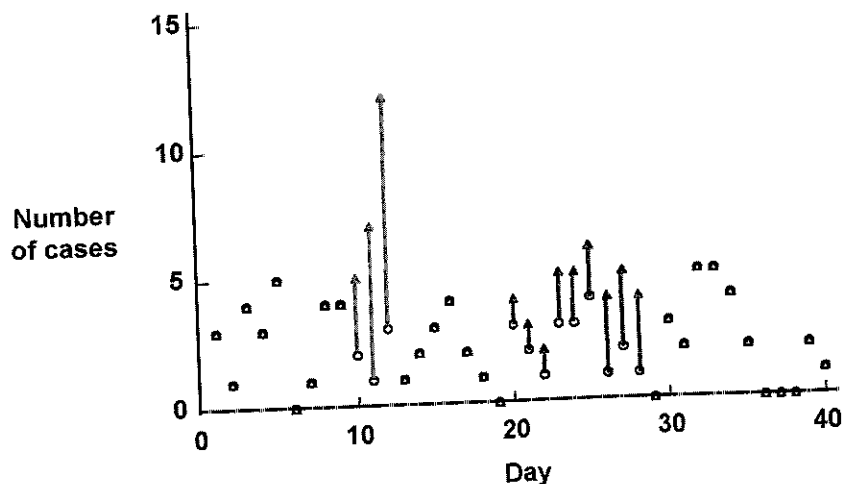


Figure 2. Simulated outbreaks superimposed on 40 days of data. A "fast" outbreak was defined as three, six, and nine cases above normal loads on three subsequent days. A "slow" outbreak was defined as 1, 1, 1, 2, 2, 2, 3, 3, 3 excess cases over nine days.

where λ is a parameter that determines how much relative weight should be given to recent observations. As with the Shewhart method, we choose an alarm threshold so that the false positive rate would be 1%.

The simple cumulative summation (CUSUM) statistic is defined as

$$C_t = \max(0, [y_t - \mu] - k + C_{t-1})$$

where μ is the daily average, and an alarm is sounded whenever C_t exceeds a preset threshold, h . This statistic cumulates positive deviations from the average, so has an improved chance of detecting a small but persistent increase in cases. Its performance may be limited, however, by the known annual patterns in ILI cases. We therefore also used a "mean-adjusted" CUSUM, in which the constant average is replaced with the exponentially weighted average over recent days, z_{t-1} :

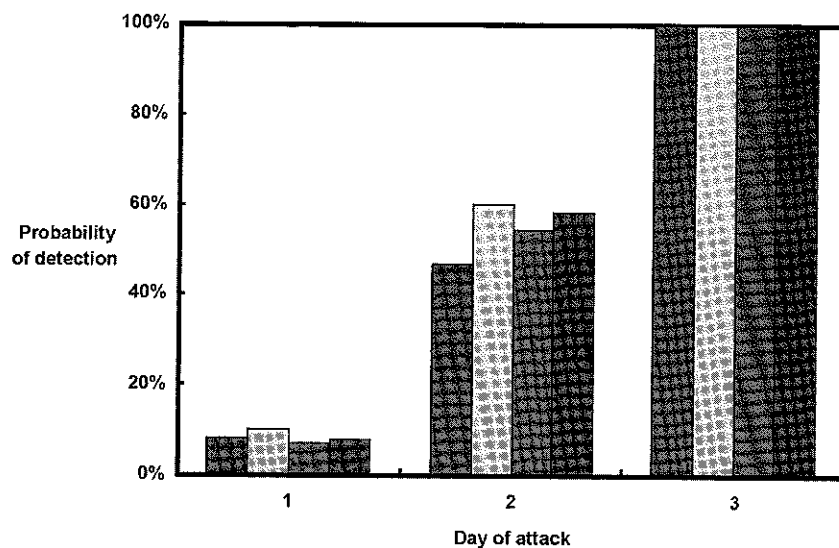
$$C_t = \max(0, [y_t - z_{t-1}] - k + C_{t-1})$$

As in the other methods, k , λ , and h are set to achieve the maximum true positive rate with a 1% false-positive rate over a three-year test period.

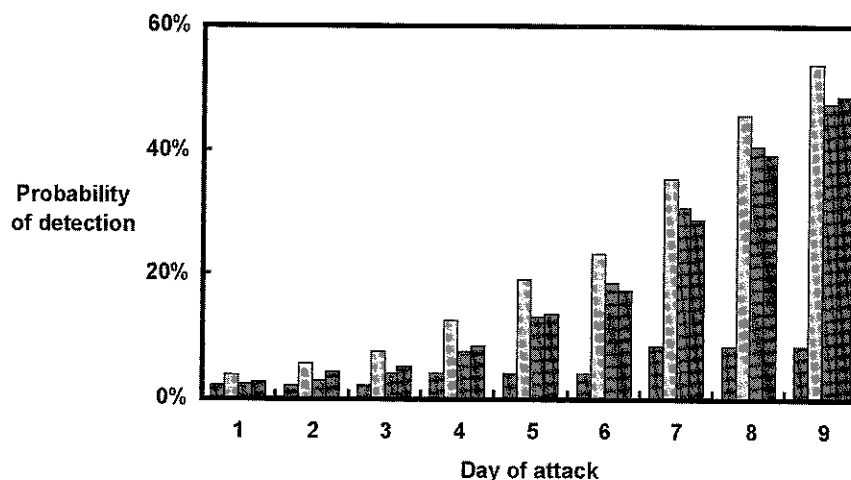
Figure 3a shows the results for each of the four methods on a fast attack outside of the winter flu season. In this situation all of the methods perform about equally well. Each was able to detect the outbreak about 50% of the time on day two, and 100% of the time on day three. In the flu season, however, the threshold has to be set so high that none of the detection algorithms sounds an alarm more than 20% of the time on the third day of the outbreak.

The detection algorithms do not work quite as well for the slow outbreak, as Figure 3b shows. For this type of outbreak, outside of the winter flu season, there is only a roughly 50% chance of detecting an outbreak on the ninth day. The three integrated methods are roughly equal, and all are superior to the Shewhart method that uses only one day's data for any one test. In the winter flu season, none of the methods can detect an outbreak more than 10% of the time, even on the ninth day.

These results tell us about the minimal size and tempo of outbreak that can be detected through syndromic surveillance of this type. The sudden appearance of large number of ILI cases—four or five times the daily average of three, especially in the sum-



A. Fast attack.



B. Slow attack.

Figure 3. Probability of detection of attack by day. Shaded bars correspond to the four detection algorithms: Shewhart, EWMA, CUSUM, and mean-adjusted CUSUM respectively.

mer—are clearly exceptional, and ER physicians do not require any sophisticated statistical algorithm to tell them so. Small outbreaks in which the number of excess cases on a given day is similar to the average number of ILI cases are simply not detectable by these means. Our analysis focused on outbreaks of medium size, outside of the winter flu season, and has identified the type of outbreak for which there is a 50% chance of detection. For a fast attack, each of the algorithms can "detect" an outbreak on the second day. For the slow outbreak, the integrated methods require nine days to achieve this level of accuracy.

Can this Performance be Improved?

There are a number of alternatives that could be considered, and actually are the subjects of current research. Most detection algorithms can be characterized in three respects: (1) what they assume as the background level and pattern of diseases or symptoms; (2) the type of departures from normal that they are tuned to detect (an exponential increase in the number of cases, a geographic cluster of cases, and so on), and (3) the statistical algorithm they use to determine when the data indicate a departure from normal, or an "alarm."

With respect to the background level and pattern, current syndromic surveillance methods typically compare current cases to the number in the previous day or week, the number in the same day in the previous year, or some average of past values. More sophisticated approaches use statistical models to "filter" or reduce the noise in the background data so that a signal will be more obvious. For instance, if a hospital emergency room typically sees more ILI patients on weekend days (when other facilities are not open), a statistical model can be developed to account for this effect. As a result any outbreak would be easier to detect. With a long enough data series, annual effects can also be incorporated. The flu season that appears most years, however, varies in its size and timing, so this approach would not be effective.

Better performance might also be obtained by carefully tuning the detection algorithm to detect specific types of outbreaks. For instance, one might choose to analyze a "syndrome" that is less common than ILI. We used the same methods, for instance, to analyze the data on the number of patients with "Viral NOS" (Not Otherwise Specified) symptoms, which averaged one per day. Outside of the flu season, we were able to detect a fast outbreak on day two 50–60% of the time, only a small improvement over ILI. With a slow outbreak, however, integrated methods had a 50% chance of detecting outbreaks from day five to day seven, compared to day nine for the same chance for ILI. This improved performance, however, has a cost—it is only sensitive to symptoms that ER physicians would classify as Viral NOS. The combination of fever and rash is rare and suggests the early stages of smallpox. A syndromic surveillance system set up to look at this combination would likely be more effective than the results above suggest, but would only be sensitive to smallpox and not terrorist agents that have other symptoms.

Data can also be analyzed geographically. If there were an extra 18 cases of ILI in a city, and all lived in the same neighborhood, that would surely be more informative than 18 cases scattered throughout the city; it would suggest a biological agent released at night in that area. This is only effective, however, for such a geographically focused

attack. It would not work if terrorists chose to expose people in an office building or at an airport, but the data were analyzed by home address.

Finally, more sophisticated detection algorithms could lead to better performance. The simplest detection algorithms focus on the number of excess cases on a given day (the actual number minus some baseline value). If this is above two or three standard deviations, an alarm is sounded. The methods described in the previous example go beyond this in two respects: they consider methods that integrate deviations over more than one day, and they have been set to fix the false-positive rate.

Current syndromic surveillance systems are typically set up to monitor eight or more separate sets of symptoms. Doing so increases sensitivity simply because more conditions are monitored. If each set of symptoms has a 1% false-positive rate, however, increasing the number monitored will also increase the number of false positives. Searching for patterns in these set of symptoms – fever up but rash down, for instance – might lead to better performing detection algorithms, but more research is needed to identify the most effective combination.

Another possibility is to pool data over multiple ERs, perhaps all hospitals in a metropolitan area or state, and that is what cities such as Boston and New York are doing. If this results in both the signal and the background increasing proportionally, it will result in a more effective system. If, for instance, there were 10 hospitals in the Washington area, each with an average of three ILI cases per day, and outbreaks were 10 times as large as those we simulated, the performance of detection algorithms would be substantially improved. If, however, there were 18 extra cases of ILI in the city, but they all appeared at the George Washington University Hospital (which is the closest to the White House, the World Bank, and other possible terrorist targets), this signal would be lost in the noise of the entire city's cases. One can analyze the data for the entire city and for each hospital individually, but with 10 separate analyses the number of false positives would also go up.

Ultimately, there really is no free

lunch. As is the case in other areas of statistics, there is an inherent trade-off between sensitivity and specificity, and the special need for timeliness makes it even more difficult in this application. Every approach to increasing sensitivity to one type of attack is less specific for some other scenario. To circumvent this trade-off, we would have to have some knowledge about how a terrorist may attack.

Integrating Syndromic Surveillance and Public Health

Beyond purely statistical issues, the value of syndromic surveillance depends on how well it is integrated into public health systems. Syndromic surveillance is intended as an alert system, and to be effective it must be linked to the public health agencies that our nation relies on for protection against all disease outbreaks, whatever their origin. Public health agencies also have access to health information that would otherwise be confidential and is needed for effective syndromic surveillance.

The detection of a sudden increase in cases of flu-like illness can mean many things. It could be a bioterrorist attack, but is more likely a natural occurrence, perhaps even the annual flu season. If the evidence is an increase in sales of flu medication, the cause may simply be that pharmacies are having a sale. A surge in absenteeism could reflect natural causes, or even a nice spring day!

When a health department is alerted to a possible disease outbreak, they begin an epidemiologic investigation. This involves identifying the people who have been infected, diagnosing their conditions, and interviewing them to identify common exposure patterns that might suggest the source and route of contagion. If there is reason to suspect terrorism, the police or the FBI might also join the investigation. A syndromic surveillance system that only says "there have been five excess cases of flu-like illness at hospital X" is not much use unless the five cases can be identified and reported to health officials.

An epidemiologic investigation would also include looking for evidence of an outbreak in other data systems. Is

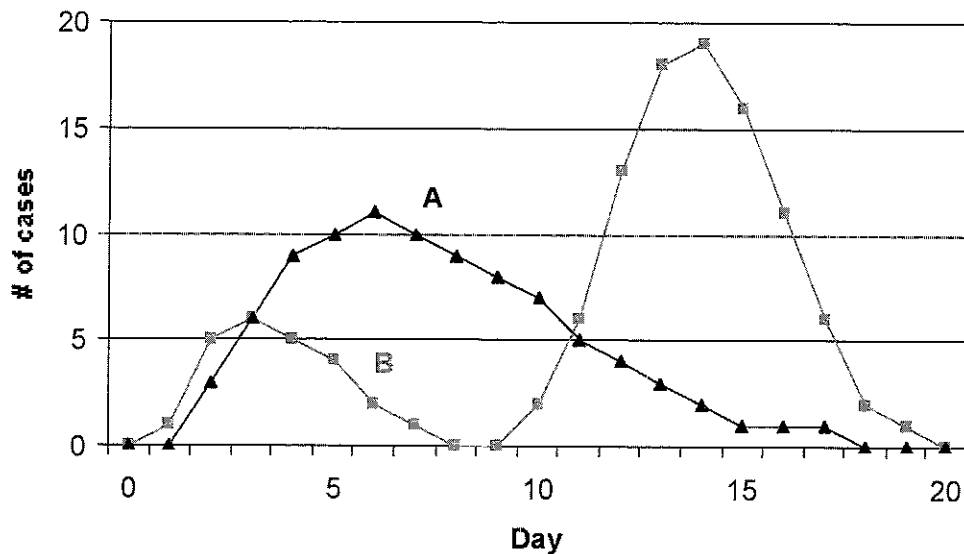


Figure 4. Two epidemic curves that are similar on days one through three, but then diverge. Line A (triangles) represents 90 cases drawn from a Gamma distribution with a mean of eight, a standard deviation of four, and 90 cases. Line B (squares) simulates an infectious agent with a mean latency of 10 days. It is made up of the sum of draws from two Gamma distributions, one with a mean of four and standard deviation of two (24 cases), and the second with a mean of 14 and a standard deviation of two (94 cases).

the increase in absenteeism matched by an increase in emergency room or doctor visits days later? Someone must have access to all of these data, and the health department is the most likely place

The new federal HIPAA regulations on confidentiality of health records (resulting from the Health Insurance Portability and Accountability Act) also come into play. Sharing health data on the scale contemplated by many syndromic surveillance systems is simply not allowed under the regulations. The data can be modified to reduce the risk of disclosing confidential patient information, but doing so decreases its value, especially for follow-up investigations. Public health agencies, however, are exempt from these constraints for officially sanctioned uses, including disease surveillance. This exemption applies to reportable diseases, but may not apply to the data needed for syndromic surveillance unless there is a public health mandate.

Another important component of an epidemiologic investigation is active surveillance, in which physicians are contacted and other means are used to identify additional cases. Some syndromic surveillance systems, however, have been set up to bypass physicians out of concern that their reporting would not be timely or complete. (This is a major concern with officially man-

dated reporting of certain conditions.) Writing physicians out of surveillance, though, could make it harder to gain their participation when it is needed in active surveillance and investigations, as well as epidemic control activities such as mass vaccination campaigns.

Health departments cannot act simply on the basis of a suspicion. Even when the cause and route of exposure are known, the available control strategies—quarantine of suspected cases, mass vaccination, and so on—are expensive and controversial, and often their efficacy is unknown. Coupled with the confusion that is likely during a terrorist attack or even a natural disease outbreak, making decisions could take days or weeks.

How much of an early warning can be expected from syndromic surveillance depends on the nature of the outbreak. First, as our calculations suggest, unless the number of people exposed in the initial attack is very large, it could be days before enough cases have accumulated to trigger an alert. Second, we must differentiate between attacks in which many people are exposed at the same time, and those in which a contagious agent might cause a large number of cases in multiple generations. Figure 4 gives two examples. Line A (triangles) illustrates what might be found if 90 people were exposed to a noncontagious agent

and symptoms first appeared, on average, eight days after exposure. Line B (squares) illustrates the impact of a smaller number of people (24) exposed to a contagious agent with an average incubation period of 10 days. Two waves of cases appear, the second larger and 10 days after the first. Because the two epidemic curves are similar on days one through three, it is difficult to know what can be expected, but if the agent were contagious (B), early intervention could save some or all of the second generation cases. In A, however, everyone would already have been exposed by the time the outbreak was detected.

Is It Worth the Effort?

Syndromic surveillance offers the potential for early warning of bioterrorist attacks and natural disease outbreaks, but this potential is limited and the cost can be high. Many cities and states are investing substantial amounts of federal bioterrorism funds in such systems. New York City's system, for instance, costs \$1 million per year in labor and computer expenses, and that doesn't count the cost of investigating alarms. But given our results, it is fair to say that the benefits of syndromic surveillance have not yet been established.

If hundreds or thousands of people in one city were infected simultaneous-

ly, no sophisticated systems would be necessary to detect it. Syndromic surveillance based on absenteeism and sales data might pick this up a day or so before patients came to emergency rooms, but that kind of advance warning is unlikely to make a difference. Syndromic surveillance was operating in both Washington and New York during the anthrax attack in 2001, but the number of cases was too small to set off an alarm. The negative reports from the syndromic surveillance systems during this period, it should be noted, were valuable in reassuring health officials – and the public – that the problem was not widespread.

Our simulation study gives some indication of the size and speed that outbreaks must attain before they are detectable, and the results are sobering. Even with an excess of nine cases over two days (the first two days of our ‘fast’ scenario), three times the daily average, there is only a 50% chance that the alarm would go off. If 18 cases are spread over nine days, it is not until the ninth day that there is a 50/50 chance that the alarm will sound. Moreover, this holds true only outside of the winter flu season. The sensitivity of the alarm can be improved, but only at the cost of more false positives or less timely results. Given these constraints, claims that syndromic surveillance can save thousands of lives seem very unrealistic.

The benefits of syndromic surveillance also depend on how well these systems are integrated into public health. Syndromic surveillance is designed to set off an alarm, nothing more. This must be investigated to determine the cause before any action can be taken, and deciding what to do can be very difficult. Better integration of syndromic surveillance with public health can shorten the time for decision-making, but only marginally.

So what should be done? City and state health departments should be cautious in investing in costly syndromic surveillance systems. Much is still to be learned about exactly what to monitor and how to analyze the data effectively, and government agencies could support research into better methods. A better choice of data series and detection algorithms may be able to improve the balance between sensitivity and specificity and maintain timeliness.

Evaluation tools, performance measures, and evidence-based standards for syndromic surveillance systems are sorely needed. Since bioterrorist attacks are rare, however, the performance of syndromic surveillance systems for this purpose cannot be directly assessed or evaluated, and some creativity is needed. One approach is to see how well existing systems have done in finding natural disease outbreaks, both new outbreaks of existing diseases and “new” diseases, such as West Nile Virus and SARS. Researchers should also extend the simulation studies that we have done to include the simultaneous analysis of multiple data streams, spatial analysis, and so on.

Another possibility is to consider “active syndromic surveillance.” The RSVP system developed at Sandia National Laboratories in New Mexico takes a more interactive approach to syndromic surveillance, focusing on the relationship between physicians and public health epidemiologists. RSVP is a web-based system that uses touch-sensitive computer screens to make it easy for physicians to report cases falling into one of six predefined syndromic categories without full clinical details or laboratory confirmation. The reports are transmitted electronically to the appropriate local health department, which may elect to follow up with the physician for more details. RSVP also includes analytical tools for state and local epidemiologists. To encourage participation, physicians get immediate feedback from the system on other similar cases in the region and practice guidelines for patients with this condition. The Voxiva Corporation has developed a similar system that allows physicians to call cases in with a touch-tone phone, and also to receive health alerts through voicemail. Systems of this sort are promising, but still have not been formally evaluated.

In addition, we should recognize that the potential of syndromic surveillance – and better public health data systems in general – might be greater for naturally occurring diseases than for bioterrorism. Every year, for instance, substantial effort goes into determining the beginning of the flu season, so that warnings and medical guidance can go out. Syndromic surveillance, or perhaps

simply the increased information sharing that is part of it, could help alert physicians and public health officials to increases in tuberculosis, sexually transmitted diseases, and many other problems that require attention. The potential beyond bioterrorism may make syndromic surveillance worth the effort. **E**

Additional Reading

- Goldenberg, A, Shmueli, G, Caruana R, and Fienberg S E 2002 “Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales.” *Proceedings of the National Academy of Sciences*: 99, 5237-5240
- Lazarus, R, Klienman, K, Dashevsky, I, Adams, C, Kludt, P, DeMaria, Jr, A, and Platt, R 2002 “Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events,” *Emerg Inf Dis*, 8, 753-760
- Lewis, M.D., Pavlin, J.A., Mansfield, J.L., O'Brien, S., Boomsma, L.G., Yevgeniy, E., and Kelley, P.W. 2002 “Disease outbreak detection system using syndromic data in the greater Washington, D.C. area,” *Am J Prev Med*, 23:180-186
- Reingold, A 2003 “If Syndromic Surveillance is the Answer, What is the Question?” *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, 1(2): 1-5 (Published online May 2, 2003, accessed May 31, 2003)
- Reis, B.Y., Pagano, M., and Mandl, K.D. 2003. “Using temporal context to improve biosurveillance,” *Proceedings of the National Academy of Sciences*, 100, 1961-1965
- Wagner, M., Tsui, F-C, and Espino, J, et al 2001. “The emerging science of very early detection of disease outbreaks,” *Journal of Public Health Management and Practice*, 7, 50-58
- Centers for Disease Control and Prevention. 2002 “Syndromic surveillance for bioterrorism following the attacks on the World Trade Center—New York City, 2001.” *MMWR*, S1, Special Issue, 13-15