

The Emerging Science of Very Early Detection of Disease Outbreaks

Michael M. Wagner, Fu-Chiang Tsui, Jeremy U. Espino, Virginia M. Dato, Dean F. Sittig, Richard A. Caruana, Laura F. McGinnis, David W. Deerfield, Marek J. Druzdzal, and Douglas B. Fridsma

A surge of development of new public health surveillance systems designed to provide more timely detection of outbreaks suggests that public health has a new requirement: extreme timeliness of detection. The authors review previous work relevant to measuring timeliness and to defining timeliness requirements. Using signal detection theory and decision theory, the authors identify strategies to improve timeliness of detection and position ongoing system development within that framework.

Key words: *decision theory, detection, epidemiology, population surveillance, public health*

Introduction

The early detection of outbreaks of disease long has been a concern of public health because of the potential to reduce morbidity and mortality. Recently, however, public health has witnessed a surge of development of new detection systems designed to provide more timely detection of outbreaks. This activity suggests that public health has a new requirement: extreme timeliness of detection.

The new detection systems—being developed by states,^{1–3} cities,⁴ the military,^{5,6} Centers for Disease Control and Prevention,⁷ and research groups^{8,9}—are designed to reduce time delays in reporting^{1,3,7} and to detect outbreaks when affected individuals first are experiencing early stages of illness.^{2,5,6,8–10} These systems employ different strategies including use of electronic data transfer (all), use of data routinely collected for other purposes,^{5,6,8} novel approaches to passive collection of data,² and novel approaches to active data collection.^{9,10} For convenient reference in this article, the authors refer to this set of systems collectively as *early warning systems*.

The applied work developing early warning systems is raising fundamental questions such as: Which data are useful for early detection? What are the timeliness requirements for outbreaks caused by different agents? How to measure timeliness of a detection system for a specific type of outbreak and especially for outbreaks such as large-scale inhalational anthrax that have not occurred in areas monitored by the new systems?

The work was supported in part by grants GO8 LM06625–01, and T15 LM/DE07059 from the National Library of Medicine; Contract No. 290–00–0009 from the Agency for Healthcare Research and Quality; and by Cooperative Agreement No. U90/CCU318753–01 from the Centers for Disease Control and Prevention (CDC). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC.

The combination of a new requirement of timeliness, a high level of applied work building early warning systems, and a set of research questions that need to be addressed to facilitate the applied work suggests that it may be profitable to view early detection of outbreaks as an area of new scientific inquiry and to ask what the theoretical underpinnings of this science are, what constitutes the relevant existing body of knowledge, and how the science can facilitate the applied work.

In this article, the authors identify and review the mathematical foundations of early detection and review the (limited) literature on defining timeliness

requirements for specific threats and on measuring the timeliness of specific detection systems for specific threats. The theory of early detection suggests strategies for improvement of timeliness that are relevant to the applied work and the authors indicate which strategies are being exploited by ongoing projects and which strategies remain to be exploited.

Mathematical Foundations of Early Detection

One of two mathematical foundations of early detection is signal detection theory,¹¹ the concepts of which are familiar to many researchers in public health. In public health, the objective of detection is to recognize from input data (signal) the occurrence of an event such as an epidemic (see Figure 1). A detection method processes the signal and produces as output a determination of whether an event is present or not. The detection method may be as simple as comparing the amplitude of the signal with a threshold. The accuracy of the detection method is reported using various ratios such as sensitivity, specificity, and positive predictive value.

Detection methods often can be adjusted to increase or decrease the sensitivity of the detection. For example, the threshold can be lowered. However, improvement in sensitivity usually occurs at the cost of loss of specificity. The optimal level of sensitivity relative to specificity depends on the consequences of false alarms and the benefits of true alarms. These consequences are not fundamental properties of the detection method itself, but are specific to the use to which the detection method is being applied. Therefore, researchers often report the specificity of a detection method over a range of sensitivities (by manipulating the threshold of detection) and report the results using a receiver-operating characteristic (ROC) curve, which plots sensitivity as a function of specificity.

Timeliness can be treated as a property of a detection method similar to the properties of sensitivity and specificity. Timeliness can be measured by subtracting the time of detection from the time of the event itself. It can be improved by adjustment of threshold and at the expense of the other parameters. Although it rarely is done, timeliness, sensitivity, and specificity can be plotted on a generalized ROC curve.

Michael M. Wagner, MD, PhD, is Assistant Professor of Medicine and Intelligent Systems, Center for Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania.

Fu-Chiang Tsui, PhD, is Research Assistant Professor, Center for Biomedical Informatics, University of Pittsburgh.

Jeremy U. Espino, MD, is a Postdoctoral Fellow, Center for Biomedical Informatics, University of Pittsburgh.

Virginia M. Dato, MD, MPH, is Senior Public Health Physician, Center for Public Health Practice at the Graduate School of Public Health, University of Pittsburgh.

Dean F. Sittig, PhD, is Senior Informatics Researcher, National Clinical Systems Development, Kaiser Permanente, Portland, Oregon.

Richard A. Caruana, PhD, is Research Scientist, Center for Automated Learning and Discovery, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Laura F. McGinnis, MSc, is Systems/Software Projects Coordinator, Pittsburgh Supercomputing Center, Pittsburgh, Pennsylvania.

David W. Deerfield II, PhD, is Assistant Director, Pittsburgh Supercomputing Center.

Marek J. Druzdzal, PhD, is Associate Professor of Information Science, University of Pittsburgh. He is currently on leave with ReasonEdge Technologies, Pte Ltd., Singapore.

Douglas B. Fridsma, MD, is Assistant Professor of Medicine, Center for Biomedical Informatics, University of Pittsburgh.

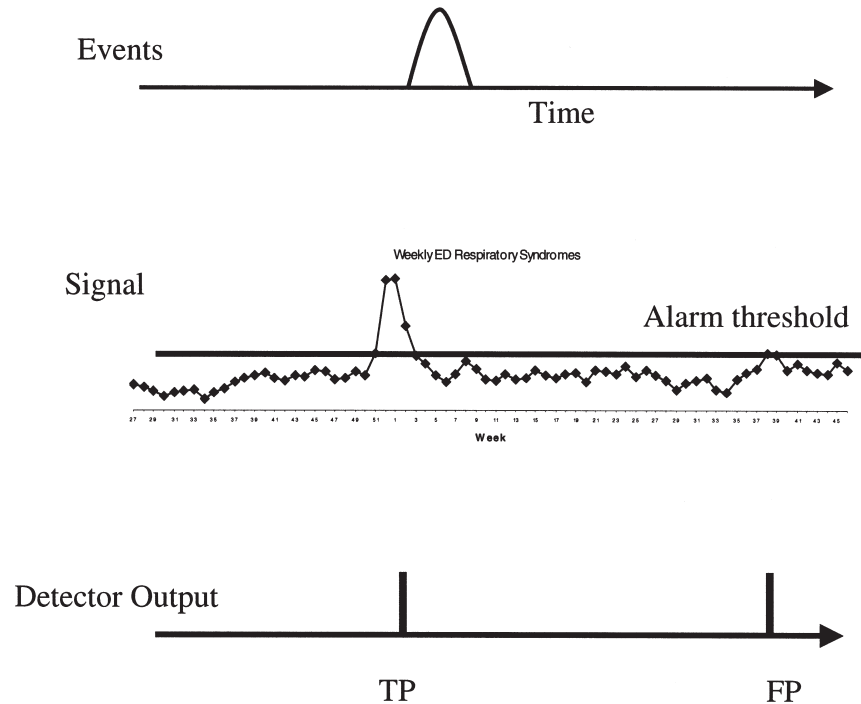


Figure 1. Principles of signal detection.

The second mathematical foundation of early detection is decision theory.¹² Decision theory is a mathematical formalism that can be used to identify an optimal sensitivity, specificity, and timeliness of detection for a specific application (such as detection of anthrax). Decision theory and the related field of decision analysis¹³ provide methods for estimating the benefit of true alarms and the costs of false alarms and for using these quantities analytically to identify an optimal sensitivity, specificity, and timeliness (and threshold) for that application.

Previous Work on Defining Timeliness Requirements

Timely detection is not a key requirement of every public health threat, but it is for many. Thus, it is surprising that quantitative (or even qualitative) analysis of timeliness requirements only has been performed for one condition: inhalational anthrax.¹⁴

That analysis was conducted for other purposes and was based on several simplifying assumptions, so the authors discuss it here as an example of the type of analysis that is needed to inform research and development of early warning systems, not as a proposed set of requirements for inhalational anthrax.

Figure 2 is a graph created from the data in that analysis. The graph shows the cumulative economic impact of a large-scale release of aerosolized *B. anthracis* as a function of time and as a function of delay in the response. The underlying model assumes that 100,000 individuals receive an LD50 dose. It uses a simple epidemiological model and estimates cost as the sum of the future earnings of those individuals who die, the costs of hospitalization, and the costs of prophylaxis of the community with antibiotics and vaccination. It also assumes that treatment with vaccination and antibiotics is 90 percent effective. The top-most curve corresponds to a no-treatment scenario and the bottom-most curve to the

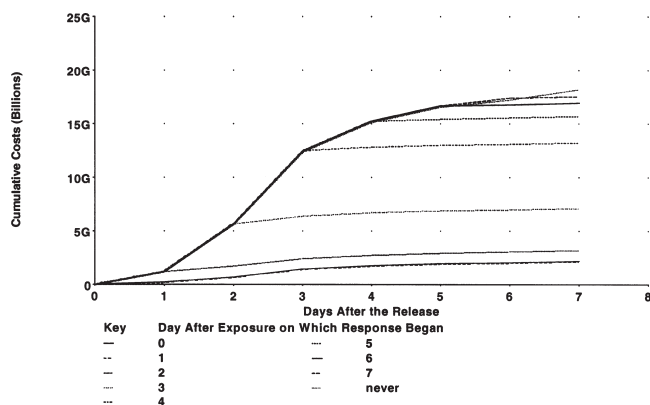


Figure 2. Cumulative economic impact with and without post-attack treatment following a hypothetical bioterrorist attack with *B. anthracis* on a population of 100,000. Cumulative cost is a measure of the economic impact of the epidemic. The bold curve is the cumulative cost without intervention and the other curves correspond to the cumulative cost under the condition that mass prophylaxis and treatment was begun on each of days 7 through 0, respectively. For simplicity, the underlying analysis assumes that after a decision to initiate treatment has been made, the response is instantaneous (all people treated immediately). Cumulative cost was computed using the following formula: Cumulative cost = (number of deaths x present value of expected future earnings) + (number of days of hospitalization x cost of hospitalization) + (number of outpatient visits x cost of outpatient visits) + cost of intervention.

earliest possible treatment scenario (treatment started concurrent with release of the pathogen into the air). The other curves correspond to treatment started on days one through six.

If the assumptions about the rapidity, lethality, and treatability of the epidemic are valid, this model helps define the timeliness requirement for detection of large-scale release of inhalational anthrax. In particular, the rate of cost accumulation during the steepest part of the curve is \$200 million per hour, which would be the cost of delaying a response decision by 1 hour under the assumptions of the model.

Previous Work on Measuring Timeliness of a Detection System

As the authors will discuss, researchers have measured the sensitivity and specificity of several outbreak detection systems; they rarely have measured the timeliness of outbreak detection systems. One of the few such studies was conducted by Quenel and colleagues,¹⁵ who used a Shewhart control chart to detect epidemics of Influenza A. Shewhart control charts commonly are used in manufacturing to ensure consistency and specification tolerance of a product. In epidemic detection, control charts work by signaling when the disease incidence (or disease indicator) is “out-of-control,” as defined by high and low boundaries set based on the standard deviation of the historical disease incidence. A 95 percent confidence threshold was computed from historical “non-epidemic” weeks. The method was applied separately to several types of data including emergency home medical visits, absenteeism from work, sentinel physician reports, and drug consumption (reported from a French pharmaceutical firm that provided 48 percent of the drugs distributed to Paris pharmacies). The gold standard was created by defining an epidemic period as two consecutive weeks during which > 1 percent of the specimens collected were positive for influenza type A, and by this method five gold standard epidemics were identified from 1984 to 1989. Sensitivity ranged from 32 percent (pediatrician activity due to influenza-like illness) to 81 percent (emergency visits), specificity ranged from 63 percent (sentinel physician overall activity) to 75 percent (emergency visits), and timeliness ranged from 5 weeks lag (pediatrician overall activity) to 4 weeks ahead (emergency visits) of the gold standard. Two other studies provide example methodologies for measuring timeliness.^{16,17}

Additional studies that measured sensitivity and specificity of outbreak detection systems but not timeliness could be reanalyzed to measure timeliness.^{18–27} Such analyses would provide insights about the timeliness of detection of those outbreaks studied (influenza, gastroenteritides, hepatitis, measles, mumps, meningitis, pertussis, rubella, and nosocomial outbreaks), a variety of detection methods (quality control charts,^{15,18–20} time-series analyses,^{17,21–23} naïve Bayes,²⁴ hidden Markov models,²⁵ and ad hoc algorithms^{26,27}), and a variety of signals

(microbiology cultures, emergency department visits, absenteeism from work, sentinel physician reports, consumption of drugs, and hospital fatalities).

Strategies To Improve Timeliness of Detection

Signal detection theory and decision theory suggest strategies for improving the timeliness of detection. As the authors will discuss, each of the early warning systems under development exploits a few of these strategies but none exploits all or even a majority of them. Additionally, none explicitly references detection theory or decision theory as a framework for design. The strategies suggested by theory are: (1) more complete sampling to improve the quality of existing signals, (2) identifying additional signals, especially those that occur earlier in the epidemic process, (3) improving the way detectors process signals, and (4) tuning the detection system for improved timeliness at the expense of specificity.

Improving quality of existing signals

If researchers have an existing signal, such as mandatory reporting of *Salmonella* cultures and other enteric organisms, they can improve the signal by more complete reporting from a region or increasing the size of the reporting region to match the scope of the largest possible outbreak. Researchers also can remove artificial time lags in signals due to delays in transmission. The National Electronic Disease Surveillance System project⁷ is an example, presently exploiting these two strategies through electronic reporting and through standardization of vocabularies, databases, and other measures needed to achieve regional data integration.

Signals also can be processed to remove non-epidemic temporal trends and known periodicities such as diurnal or weekly patterns. The detection and removal of such repetitive patterns can be achieved by time series analytical methods such as seasonal autoregressive integrated moving average,²⁸ fre-

quency analysis (e.g., Fourier transform²⁹), or wavelet transforms.³⁰

Adding new signals

A second strategy is to obtain additional signals that can augment or replace existing signals. Sources for new signals seem to fall into three categories that differ along the dimensions of cost and feasibility.

One approach is to facilitate *passive* reporting of new types of data (e.g., syndromes) by removing barriers to reporting and providing incentives to clinicians. An example is the RSVP system that encourages reporting directly from clinicians of patients with early presentations.² This approach removes barriers by making the reporting forms available on the Internet, and it provides information about similar cases as an incentive to clinicians. Passive reporting as a strategy, although inexpensive, usually suffers from underreporting.

A second approach is to develop new *active* data collection systems such as drop-in surveillance^{6,10} or new continuous sentinel systems for the types of data needed for timely detection.³¹ These approaches are feasible, but are costly to operate. An approach to active data collection that may address the limitation of cost is focused active data collection driven by value-of-information considerations.⁸ Value of information refers to an application of decision theory to the decision of whether to actively collect new information. A decision is made in favor of additional collection when the expected value of the additional information exceeds the expected cost of the collection.³²

Attractive from the perspective of both cost and feasibility is the use of data *collected routinely for other purposes*. For example, Figure 3 shows Pittsburgh regional data effects of the 1999 Influenza outbreak (the WebMD data shown are national). Examples of new early warning systems that exploit this strategy include systems that monitor emergency department diagnoses,^{8,33–36} microbiology results,^{3,8,37} pharmaceutical sales,⁴ and satellite image data.³⁸

Recent and ongoing research on the application of natural language processing (NLP) to detection of outbreaks and cases can be understood as an example of the use of data collected routinely for other purposes. Researchers have applied NLP to medical reports,^{39–42} and potentially it can be applied to other

Researchers also can remove artificial time lags in signals due to delays in transmission.

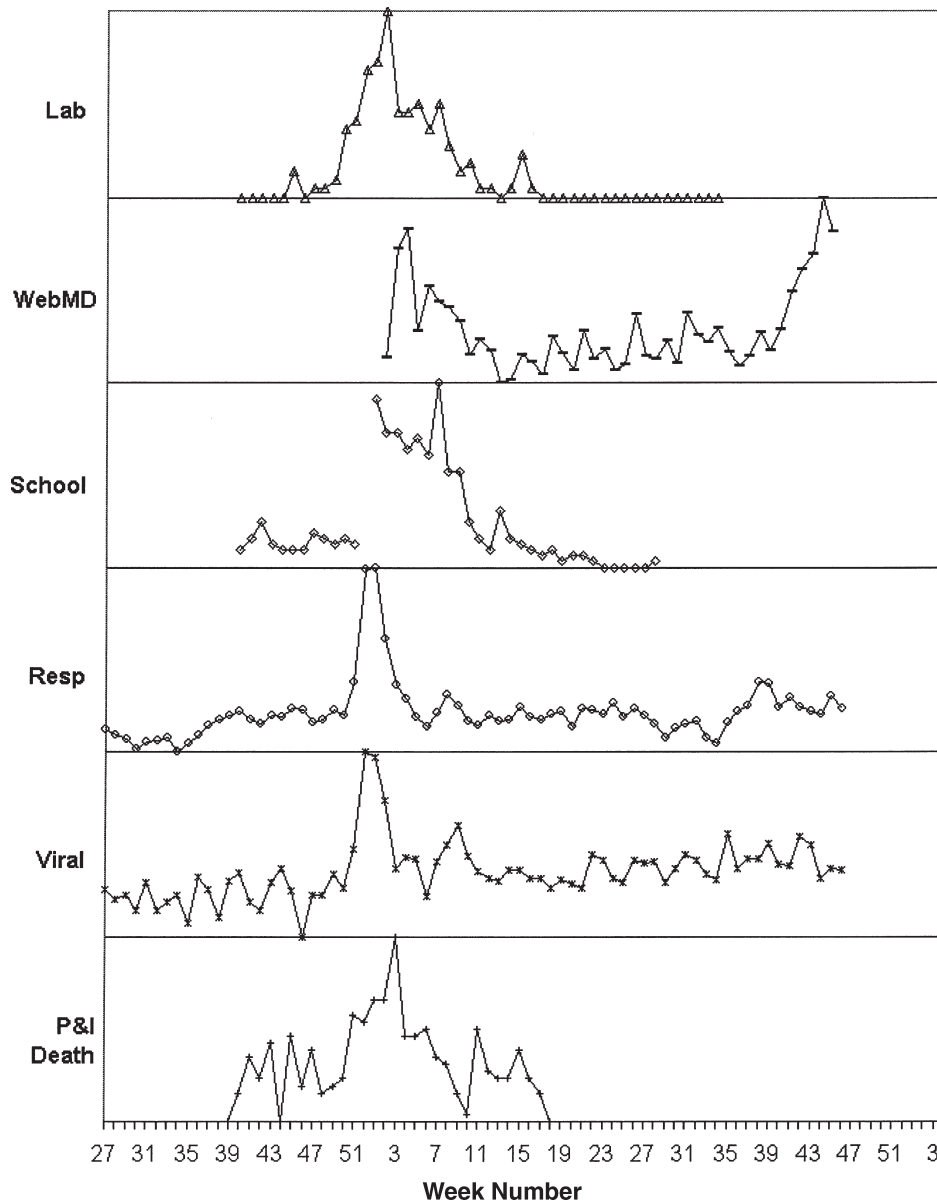


Figure 3. Weekly counts for several types of routinely collected data for different time periods around the December 1999 Influenza outbreak in Pittsburgh, Pennsylvania. Each data type is plotted on a normalized 0–1 scale. Legend: Lab, influenza cultures from the UPMC Health System; WebMD, counts of queries to a national Web health site using words such as cold and flu; School, school nurse influenza reporting; Resp. and Viral, categories of emergency department ICD-9-coded chief complaints.

types of data such as Web health-related queries, 911 emergency calls, and obituaries.

For most diseases, understanding of the value of data routinely collected for other purposes for early

detection is incomplete. An exception is Influenza, for which researchers have studied changes that appear in weekly or daily counts of numerous types of routinely collected data.^{15,17,43} For many other dis-

eases, public health experts know that routinely collected microbiology cultures are a useful source of data; however, for other types of routinely collected data, the potential value for early detection of the full set of diseases of interest to public health (a large matrix) is unknown. If such a matrix were available, it would be of potential relevance to designers of national public health surveillance systems because it would suggest priorities for data source development and integration.

Improving detection algorithms

The third strategy suggested by detection theory is to improve the detection method itself. The available signals may contain information that detection methods are not capable of utilizing fully. For example, algorithms have yet to be developed that can detect an outbreak of inhalational anthrax by searching for temporal and spatial associations in the affected population that are consistent with recent weather patterns and the incubation period of the disease.

Optimizing the detection threshold

A last strategy involves tuning the detection system for improved timeliness at the expense of specificity or sensitivity. Detection systems typically produce a numerical output (e.g., number of cases of X observed per day), and this output is compared against a threshold to determine whether to alarm. This last strategy involves the relatively simple-minded idea of lowering the threshold so that alarms are sounded earlier. A lower threshold, however, will result in more false alarms, creating a trade-off between the cost of the additional false alarms and the potential benefit of earlier detection. Decision theory suggests that the optimal threshold should be determined by an explicit consideration of this trade-off, and it provides analytical tools for this purpose.

Discussion

The recent surge of development of early warning systems for disease outbreaks suggests that extreme timeliness is a new requirement of public health. This requirement seems to be a response to emerging diseases and potential terrorist acts. Signal detection theory and decision theory provide a framework for analysis and measurement in this new area, and they

may be used to understand ongoing work and identify a set of strategies for improving detection.

Progress may be facilitated by viewing very early detection of outbreaks as a new focus area of scientific inquiry, as the authors have done in this article. This perspective suggests the importance of defining specific questions to be addressed, identifying prior relevant work, and identifying techniques from other fields that may be of value. The range of techniques identified in this article suggest that early detection will be an interdisciplinary field involving economic modelers, public health experts, informaticians, biostatisticians, and epidemiologists.

Progress also may be facilitated by the recognition that early detection as a scientific discipline has some characteristics of “big” science: Outbreaks of disease are not common and a surveillance population that is sufficiently large to have experience in disease outbreaks will be required to measure and validate early warning systems. There are many diseases with different detection requirements to survey, many potential signals to evaluate, and many methods to obtain and process the signals. Very few have been evaluated to date.

Although timeliness of detection is the subject of this article, timeliness is not the only or even the most important requirement for every disease. Sensitivity to small outbreaks may be of equal or greater importance in foodborne illnesses or more slowly progressing outbreaks. The principles and methods of signal detection theory and decision theory, fortunately, are not limited to the building of *early* warning systems, but apply equally to the problem of building systems that must be sensitive or specific.

REFERENCES

1. Georgia Division of Public Health. “Statewide Electronic Notifiable Disease Surveillance System” [online] 2001 [cited 2001 June 2]. Available from: <http://www.ph.dhr.state.ga.us/inpho/statewide/sendss.shtml>
2. New Mexico Department of Health. “RSVP” [online] 2001 [cited 2001 May 10]. Available from: <http://epi.health.state.nm.us/rsvpdesc/default.shtml>
3. Effler, P., Ching-Lee, M., Bogard, A., Man-Cheng, L., Nekomoto, T., and Jernigan, D. “Statewide system of electronic notifiable disease reporting from clinical laboratories: comparing automated reporting with conventional methods,” *JAMA* 282 (1999): 1845–50.

4. Mikol, Y., Miller, J., and Ashendorff, A. "Diarrheal disease surveillance programs: New York City's experience," *International Conference on Emerging Infectious Diseases*. Atlanta, Georgia: Centers for Disease Control, 2000.
5. DoD-GEIS. "Electronic surveillance system for early notification of community-based epidemics (ESSENCE)" [online] 2000 [cited 2001 June 2]. Available from: <http://www.geis.ha.osd.mil/getpage.asp?page=SyndromicSurveillance.htm&action=7&click=KeyPrograms>
6. Schafer, K. "LEADERS (Lightweight Epidemiology Advanced Detection & Emergency Response System)" [online] 2001 [cited 2001 May 10]. Available from: <http://www.tricare.osd.mil/conferences/2001/agenda.cfm>
7. "Supporting Public Health Surveillance through the National Electronic Disease Surveillance System (NEDSS)" [online] 2000 [cited 2001 April 30]. Available from: <http://www.cdc.gov/od/hissb/docs.htm#nedss>
8. Espino, J., Wagner, M., Tsui, F., and Wilson, T. "Real-time Outbreak and Disease Surveillance (RODS)" [online] 2001 [cited May 8, 2001]. Available from: <http://ultra.cbmi.upmc.edu/~jue/rods>
9. Talan, D.A., Moran, G.J., Mower, W.R., Newdow, M., Ong, S., Slutsker, L., Jarvis, W.R., Conn, L.A., and Pinner, R.W. "EMERGENCY ID NET: an emergency department-based emerging infections sentinel network. The EMERGENCY ID NET Study Group," *Ann Emerg Med* 32 (1998): 703-11.
10. Ernst and Young. "EYT Announces Medical Surveillance System, Early Warning System for Infectious Disease Outbreaks—Bio-terrorist Attacks" [online] 2001 [cited 2001 May 30]. Available from: http://www.eyt.com/news/pr_medsurv.asp
11. Kay, S. *Fundamentals of Statistical Signal Processing: Detection Theory*. New York: Prentice Hall, 1998.
12. von Neumann, J., and Morgenstern. *Theory of Games and Economic Behavior (2nd edition)*. Princeton, NJ: Princeton University Press, 1947.
13. Weinstein, M.C., and Fineberg, H.V. *Clinical Decision Analysis*. Philadelphia: Saunders, 1980.
14. Kaufmann, A., Meltzer, M., and Schmid, G. "The economic impact of a bioterrorist attack: Are prevention and postattack intervention programs justifiable?," *Emerging Infectious Diseases* 3 (1997): 83-94.
15. Quenel, P., Dab, W., Hannoun, C., and Cohen, J.M. "Sensitivity, specificity and predictive values of health service based indicators for the surveillance of influenza A epidemics," *International Journal of Epidemiology* 23 (1994): 849-55.
16. Hashimoto, S., Murakami, Y., Taniguchi, K., and Nagai, M. "Detection of epidemics in their early stage through infectious disease surveillance," *International Journal of Epidemiology* 29 (2000): 905-10.
17. Tsui, F.-C., Wagner, M.M., Dato, V., and Chang, C.-C.H. "Value of ICD-9-coded chief complaints for detection of epidemics." Proceedings of the AMIA Annual Symposium, 2001, in press.
18. Schiffman, R.B., and Palmer, R.A. "Surveillance of nosocomial infections by computer analysis of positive culture rates," *Journal of Clinical Microbiology* 21 (1985): 493-5.
19. Stern, L., and Lightfoot, D. "Automated outbreak detection: a quantitative retrospective analysis," *Epidemiology & Infection* 122 (1999): 103-10.
20. Hutwagner, L.C., Maloney, E.K., Bean, N.H., Slutsker, L., and Martin, S.M. "Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks," *Emerging Infectious Diseases* 3 (1997): 395-400.
21. Toubiana, L., and Flahault, A. "A space-time criterion for early detection of epidemics of influenza-like illness," *European Journal of Epidemiology* 14 (1998): 465-70.
22. Quenel, P., and Dab, W. "Influenza A and B epidemic criteria based on time-series analysis of health services surveillance data," *European Journal of Epidemiology* 14 (1998): 275-85.
23. Farrington, C.P., Andrew, N.J., Beale, A.D., and Catchpole, M.A. "A statistical algorithm for the early detection of outbreaks of infectious disease," *Journal of the Royal Statistical Society* 159 (1996): 547-563.
24. Zeng, G., Thacker, S.B., Hu, Z., Lai, X.J., and Wu, G.K. "An assessment of the use of Bayes' theorem for forecasting in public health: the case of epidemic meningitis in China," *International Journal of Epidemiology* 17 (1988): 673-9.
25. Le Strat, Y., and Carrat, F. "Monitoring epidemiologic surveillance data using hidden Markov models," *Stat Med* 18 (1999): 3463-78.
26. Stroup, D.F., Williamson, G.D., Herndon, J.L., and Karon, J.M. "Detection of aberrations in the occurrence of notifiable diseases surveillance data," *Stat Med* 8 (1989): 323-9; discussion 331-2.
27. Wharton, M., Price, W., Hoesly, F., Woolard, D., White, K., Greene, C., and McNabb, S. "Evaluation of a method for detecting outbreaks of diseases in six states," *American Journal of Preventive Medicine* 9 (1993): 45-9.
28. Box, G.E.P., and Jenkins, G.M. *Time series analysis: forecasting and control*. San Francisco: Holden-Day, 1976.
29. Oppenheim, A.V., Schafer, R.W., and Buck, J.R. *Discrete-time signal processing*. Upper Saddle River, NJ: Prentice Hall, 1998.
30. Tsui, F.-C. "Time-series prediction using a multiresolution dynamic predictor," Doctoral Thesis, Department of Electrical Engineering. Pittsburgh: University of Pittsburgh, 1996.
31. Jackson Foundation. "Jackson Foundation Research" [online] 2001 [cited 2001 June 2]. Available from: http://www.hjf.org/research/GEIS_1999/geis_1999.html
32. Heckerman, D., Horvitz, E., and Middleton, B. "An Approximate Nonmyopic Computation for Value of Information," *Seventh Conference on Uncertainty in Artificial Intelligence*. Los Angeles, 1991.
33. Lewis, M., Pavlin, J., O'Brien, S., Mansfield, J., Godlenbaum, M., Boomsma, L., and Kelley, P. "The use of an automated syndromic surveillance system for early detection of infectious disease outbreaks," *International Conference on Emerging Infectious Diseases*. Atlanta, Georgia: Centers for Disease Control, 2000.

34. Wagner, M., Espino, J., Tsui, F., Wilson, T., Tsai, T., Harrison, L., and Pasculle, W. "The role of clinical event monitors in public health surveillance." *Journal of the American Medical Informatics Association*, submitted for publication.
35. Idaho Technology. "Epidemiology Advanced Detection and Emergency Response (LEADERS)" [online] 2000 [cited 2001 May 8]. Available from: <http://www.army-technology.com/contractors/nbc/idaho>
36. Brinsfield, K., and Gunn, J. "Using volume-based surveillance for an outbreak early warning system," *Academy of Emergency Medicine* 8 (2001): 492.
37. Focus/MRL Inc. "The Rapid Antibiogram Reference-USA (RARTM-USA)" [online] 2001 [cited Available from <http://www.techmall.com/techdocs/TS990913-8.html>
38. GEIS. "GEIS Rift Valley Fever Monitor" [online] 2001 [cited 2001 May 10]. Available from: <http://www.geis.ha.osd.mil/riftvalleyfever/index.htm>
39. Hripcsak, G., Friedman, C., Alderson, P.O., DuMouchel, W., Johnson, S.B., and Clayton, P.D. "Unlocking clinical data from narrative reports: a study of natural language processing," *Annals of Internal Medicine* 122 (1995): 681–8.
40. Hripcsak, G., Knirsch, C., Jain, N., and Pablos-Mendez, A. "Automated tuberculosis detection," *Journal of the American Medical Informatics Association* 4 (1997): 376–81.
41. Hripcsak, G., Kuperman, G., Friedman, C., et al., "A reliability study for evaluating information extraction from radiology reports," *Journal of the American Medical Informatics Association* 6 (1999): 143–150.
42. Chapman, W., and Haug, P. "Comparing expert systems for identifying chest x-ray reports that support pneumonia," *Proceedings/AMIA Annual Symposium* (1999): 216–220.
43. Espino, J., and Wagner, M. "The accuracy of ICD-9 coded chief complaints for detection of acute respiratory illness," *Proceedings/AMIA Annual Symposium* (2001) In press.

