Evaluating Detection of an Inhalational Anthrax Outbreak

David L. Buckeridge,* Douglas K. Owens,†‡ Paul Switzer,‡ John Frank,§ and Mark A. Musen‡

Timely detection of an inhalational anthrax outbreak is critical for clinical and public health management. Syndromic surveillance has received considerable investment, but little is known about how it will perform relative to routine clinical case finding for detection of an inhalational anthrax outbreak. We conducted a simulation study to compare clinical case finding with syndromic surveillance for detection of an outbreak of inhalational anthrax. After simulated release of 1 kg of anthrax spores, the proportion of outbreaks detected first by syndromic surveillance was 0.59 at a specificity of 0.9 and 0.28 at a specificity of 0.975. The mean detection benefit of syndromic surveillance was 1.0 day at a specificity of 0.9 and 0.32 days at a specificity of 0.975. When syndromic surveillance was sufficiently sensitive to detect a substantial proportion of outbreaks before clinical case finding, it generated frequent false alarms.

In the early stage of an inhalational anthrax outbreak, a 1-day delay in the initiation of chemoprophylaxis and treatment of exposed persons can result in thousands of additional deaths and millions of dollars of additional expenditures (*1*,*2*). Thus, timely detection of an inhalational anthrax outbreak is critical. Rapid detection is also important for disease outbreaks that result from other bioterrorism agents and from emerging infectious diseases, such as severe acute respiratory syndrome or avian influenza (*3*).

To detect an epidemic such as inhalational anthrax, which is nonendemic and results in severe symptoms, public health authorities have relied traditionally on identification and rapid reporting of the sentinel clinical case. However, because the perceived likelihood of a bioterrorism attack has increased, public health authorities have sought novel approaches for rapid outbreak detection. One approach that has received considerable economic investment over the past 5 years is syndromic surveillance. This approach follows prediagnostic data sources in an attempt to detect an increase in the prevalence of nonspecific symptoms. For example, the BioSense system (*4*), developed by the Centers for Disease Control and Prevention (CDC) at a cost of >$75 million (*5*), follows records of outpatient visits, pharmaceutical prescriptions, and laboratory orders in an attempt to detect disease outbreaks rapidly. Hundreds of similar systems are maintained or are under development by various groups around the world (*6*). Other examples include systems operated by the Department of Homeland Security (*5*) and academic centers in partnership with state or county public health departments (*7*–*9*).

In addition to supporting outbreak detection, these syndromic surveillance systems provide situational awareness for public health authorities and may serve other purposes. Nevertheless, a major justification for these systems is outbreak detection. Despite substantial investment in syndromic surveillance and calls for further research from groups such as the Institute of Medicine (*3*), little evidence exists to suggest how syndromic surveillance will perform relative to clinical case finding for detection of an inhalational anthrax outbreak (*10*). The reason for this lack of evidence is that data from real outbreaks are not available to evaluate the performance of syndromic surveillance alone or in comparison to clinical case finding. Moreover, even if data were available from a large-scale outbreak, those data would allow only an evaluation of performance in 1 specific setting. CDC recently endorsed the use of simulated outbreaks to address the dearth of outbreak data (*11*), but existing simulation studies have not compared detection through clinical case finding with syndromic surveillance (*12*–*14*). Our aim was to develop a model for simulating use of healthcare services after a large-scale exposure to aerosol anthrax spores and then to use this

*McGill University, Montreal, Quebec, Canada; †Veterans Affairs Palo Alto Health Care System, Palo Alto, California, USA; ‡Stanford University, Stanford, California, USA; and §University of Toronto, Toronto, Ontario, Canada

model to estimate the detection benefit of syndromic surveillance compared with clinical case finding.

Methods

Study Design

We developed a model to simulate the dispersion of released anthrax spores; the infection of exposed persons; the progression of disease in infected persons; and symptomatic persons' use of the healthcare system, including blood culture testing in clinical settings. Using the simulation model, we generated outbreak signals and time until the first clinical diagnosis for 3 amounts of spores released. To incorporate into the model the uncertainty in parameter values, we used a Latin hypercube sampling design, which allows many parameter values to vary simultaneously (*15*). The 3,000 simulated signals generated with this sampling strategy were superimposed in turn onto baseline administrative records of ambulatory healthcare visits in the Norfolk, Virginia, area. These records are generated daily and similar types of records are used widely for syndromic surveillance (*4*,*7*,*9*). We assessed the usefulness of syndromic surveillance by modeling the healthcare system use that would occur after an anthrax attack and superimposing this use onto actual administrative data over 1-year period. Finally, we assessed, over a range of specificity, the sensitivity and timeliness of syndromic surveillance and the detection benefit of syndromic surveillance compared with clinical case finding for each simulated outbreak. We summarize our methods in the remainder of this section and refer readers to the online Technical Appendix (available from http://www.cdc.gov/ncidod/EID/vol12no12/06-0331_app.htm) for additional details.

Simulation Model

The simulation model builds on our previous work (*16*–*18*) and is composed of 4 components: dispersion, infection, disease, and healthcare system use. The dispersion model simulates the number of anthrax spores a person would inhale at locations throughout the region after release of aerosolized spores. We used the Hazard Prediction and Assessment Capability (HPAC) software developed by the Defense Threat Reduction Agency to simulate the dispersion of spores (*19*). The HPAC model accounts for factors such as atmospheric conditions and terrain. We simulated a point release of 3 amounts of anthrax spores: 1 kg, 0.1 kg, and 0.01 kg (Figure 1A).

The infection model simulates the number of persons infected, according to residential address and dispersion of spores (Figure 1B). The probability of infection given exposure to an amount of spores was modeled by using a probit regression model. The disease model uses a semi-Markov process to simulate the progression of infected persons through 3 discrete states of disease. Each infected person began in the incubation state and then progressed through the prodromal state and the fulminant state. The time in each state was sampled from a log normal distribution.

The healthcare use model uses a semi-Markov process to simulate the probability and timing of a symptomatic person seeking care and submission of blood for culture and culture results when care is sought. For persons in the prodromal or fulminant state of disease who sought care, the instantaneous probability of seeking care increased linearly over the duration of the state. For patients whose blood samples were cultured, the testing process was modeled as the transition through 2 discrete states: growth and isolation. The time spent in each of these states was modeled by using an exponential distribution.

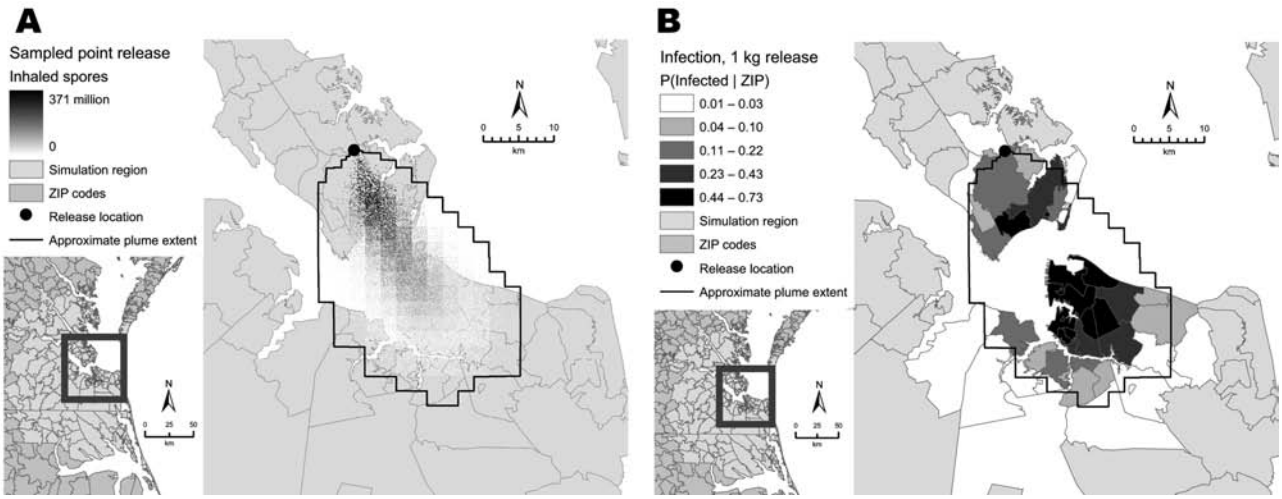


Figure 1. Maps showing output from dispersion (A) and infection (B) components of the simulation model. The dispersion component simulates geographic distribution of anthrax spores after an aerosol release. The infection component simulates infection of persons exposed to spores.

Data for Simulation Model

The infection model used an infection function corresponding to the data reported by Glassman (*20*). This is a probit model with a 50% lethal dose (LD_{50}) of 8,600 spores and a slope of 0.67. Uncertainty exists about the values for many of the parameters in the disease and healthcare use models. To incorporate this uncertainty into our estimates, we used a Latin hypercube sampling approach to sample parameter values for random variables in our simulation model (*15*). This approach requires specifying equal probability bins for parameter values. We specified 3 bins for each parameter value, a narrow bin around the most likely estimate, and wider bins on either side of the estimate. Table 1 shows the bins we used for each parameter value, the probability distribution that each value parameterizes, and the sources that we used to define the bins.

We used previous work modeling anthrax for the distribution of time periods in each disease state (*2*,*21*,*22*). For the probability of seeking care while in the prodromal disease state, cross-sectional surveys indicate that 14%–30% of persons visit a physician at some point during an episode of upper respiratory tract illness (*23*,*24*). For the fulminant disease state, we estimated the probability of seeking care before death as 90%–95%, given the severity of the symptoms in that state.

After a person made a healthcare visit, we simulated the syndrome assigned to the person by using probabilities that reflect the distribution of clinical presentations for inhalational anthrax reported in the literature (*25*,*26*). Because we considered only respiratory syndromes for surveillance, we varied directly only the probability of being assigned a respiratory syndrome to persons in the prodromal disease state.

For visits from persons in either symptomatic disease state, the estimate of sensitivity from published studies of blood culture testing was 0.8–0.9 (*27*). For a visit in the prodromal state, we estimated the probability of a physician ordering a blood culture as 0.01–0.015 on the basis of data from the National Ambulatory Medical Care Survey (*28*). For a visit in the fulminant state of disease, we estimated the probability of a blood culture test as 0.9–0.95. After gram-positive rods grew in the blood culture, we estimated the probability of isolating the organism to be 0.8–0.9 (*29*). We modeled the time until growth and isolation as exponential (*25*,*30*).

Baseline Data and Release Scenarios

We used records of ambulatory visits in the Norfolk, Virginia, region acquired from the TRICARE health maintenance organization as a baseline onto which we superimposed simulated outbreak records. The data covered the period 2001–2003, and the simulation region included 17 clinical facilities within an ≈160-km × 200-km area that encompasses 158 ZIP codes from 2 states. Over the 3 years of available data, 427,634 persons made >5 million visits. We classified the records into syndromes by using the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) to syndrome mapping defined by the ESSENCE system (*7*) and used only 351,749 records for which persons were classified as having a respiratory syndrome. The Human Subjects Panel at the Stanford School of Medicine approved the use of these data for this study. We examined 3 scenarios defined by the amount of spores released: 1 kg, 0.1 kg, and 0.01 kg. For each scenario, we performed 1,000 simulations.

Table 1. Sampling intervals for parameter values in the simulation model*

Parameter	Parameter value intervals	Probability distribution	Source†
Disease model			
Incubation duration, d; median	(5, 9) (9, 11) (11, 15)	Log normal	(*2*,*21*,*22*)
Incubation duration, dispersion	(1.5, 1.9) (1.9, 2.1) (2.1, 2.5)	Log normal	(*2*,*21*,*22*)
Prodromal duration, d; median	(1.5, 2.3) (2.3 ,2.7) (2.7, 3.5)	Log normal	(*2*,*22*)
Prodromal duration, dispersion	(1.2, 1.4) (1.4, 1.5) (1.5, 1.7)	Log normal	(*2*,*22*)
Healthcare use			
Probability of visit, prodromal state	(0.05, 0.25) (0.25, 0.35) (0.35, 0.55)	Bernoulli	(*23*,*24*)
Probability of visit, fulminant state	(0.7, 0.9) (0.9, 0.95) (0.95, 1)	Bernoulli	Estimate
Probability of respiratory syndrome, prodromal state	(0.5, 0.7) (0.7, 0.8) (0. 8,1)	Bernoulli	(*25*,*26*)
Blood culture test, prodromal state	(0.001, 0.01) (0.01, 0.015) (0.015, 0.025)	Bernoulli	(*28*)
Blood culture test, fulminant state	(0.7, 0.9) (0.9, 0.95) (0.95, 1)	Bernoulli	Estimate
Sensitivity of blood culture	(0.5, 0.8) (0.8, 0.9) (0.9, 1)	Bernoulli	(*27*)
Time until blood culture growth, d	(0.4, 0.8) (0.8, 1.0) (1.0, 1.4)	Exponential	(*30*)
Probability of isolation given growth	(0.5, 0.8) (0.8, 0.9) (0.9, 1)	Bernoulli	(*29*)
Time until blood culture isolation, d	(0.5, 0.6) (0.6, 0.9) (0.9, 1.5)	Exponential	(*25*)

*Using a Latin hypercube strategy, a value for each parameter was sampled by randomly selecting 1 of the 3 intervals for the parameter and randomly sampling a value on the selected interval. The sampled values parameterize probability distributions, which are sampled for the simulation model.
†References that support the parameter value intervals.

Outbreak Detection

The time to outbreak detection through clinical case finding for a simulated outbreak was calculated for each simulated outbreak as the time between exposure to spores and the first positive blood culture. To calculate time to outbreak detection through syndromic surveillance, we superimposed the simulated records for respiratory syndrome visits onto the authentic baseline data, beginning on a randomly selected date in 2003, and then applied the outbreak detection algorithm to the combined baseline and simulated data. The outbreak detection algorithm used a time-series model (*31*) to generate daily 1-step-ahead forecasts for the total number of respiratory syndrome visits (*13*) and then applied a cumulative sum (*32*) to the forecast residual. To vary the specificity of the detection algorithm, we varied the decision threshold of the cumulative sum.

Evaluation Metrics

To evaluate outbreak detection through syndromic surveillance, we calculated sensitivity, specificity, and timeliness at a range of decision thresholds. Timeliness is the duration between the release of anthrax spores and the first report of an outbreak. We also computed the detection benefit of syndromic surveillance relative to clinical case finding, and the proportion of runs with a detection benefit >0. Detection benefit is the potential time saved in detection from using syndromic surveillance compared with clinical case finding. The benefit is calculated as the difference in the timeliness between syndromic surveillance and clinical case finding in those simulations in which detection occurred first by syndromic surveillance. When an outbreak was not detected by syndromic surveillance, the detection benefit was 0. For a given release scenario, each of the 1,000 simulations integrated both randomness in the component model outputs as well as uncertainty in component model parameters. Each of the 1,000 simulations is a sample from the integrated distribution of possible outcomes. To indicate the spread of the integrated uncertainty distribution, we calculated the upper and lower deciles from the 1,000 simulations. For plots, we calculated 95% confidence intervals, which reflect finiteness of the simulation.

Results

Detection Performance of Clinical Case Finding

Because all outbreaks were ultimately detected by clinical case finding through routine blood culture, the sensitivity of this approach was 1.0 for the scenarios considered. Clinical case finding detected outbreaks from an average of 3.7 days to 4.1 days after release, with larger amounts of spores detected before smaller amounts (Table 2). Results from analyses of additional release scenarios (data not shown) suggested that the influence of amount released on time to detection was mediated, in part, through the number infected. Mean timeliness across the scenarios examined was associated with the mean number infected (Pearson's r –0.94, 95% confidence interval –0.98 to –0.79), and an increase of 10,000 infected persons resulted in a decrease in the time until detection of ≈4 hours.

Table 2. Average numbers of persons infected and average times to outbreak detection through clinical case finding for 3 release scenarios*

Amount released (kg)	Mean no. infected	Mean days to detection
1	49,000	3.7 (2.5, 5.0)
0.1	31,000	3.9 (2.7, 5.3)
0.01	15,000	4.1 (2.9, 5.5)

*Values in parentheses are 10th and 90th percentiles of the distribution.

Detection Performance of Syndromic Surveillance

The sensitivity and timeliness of syndromic surveillance were influenced by the release amount and by specificity. Table 3 shows this relationship over the release scenarios examined and 2 levels of specificity. At a specificity of 0.90, a 1-kg release was detected in 100% of our simulations (sensitivity 1.0) at a mean detection time of 3.1 days. For a release that was much smaller, 0.01 kg, sensitivity was 0.94, and the mean detection time increased to 3.6 days. Although the sensitivity of syndromic surveillance was high when we set specificity to 0.90, this specificity resulted in a false alarm (false-positive detection) ≈1 every 10 days. By increasing specificity to 0.975, we reduced the false alarm rate to ≈1 every 40 days (Table 3). However, with increased specificity, the sensitivity of syndromic surveillance decreased (from 0.98 to 0.82 depending on the size of the release) and the mean time until detection lengthened to 4.3 days for a 1-kg release and to 5.1 days for a 0.01-kg release (Table 3).

Results from analyses of additional release scenarios (data not shown) indicated that the trends in sensitivity and timeliness across release amount were mediated to some extent by the number infected. Sensitivity was a nonlinear function of the number of persons infected, with sensitivity increasing more quickly when fewer persons were infected. At a specificity of 0.975, an increase of 10,000 infected persons resulted in a decrease in time to detection of ≈6 hours.

Detection Benefit of Syndromic Surveillance Compared with Clinical Case Finding

The detection benefit of syndromic surveillance compared with clinical case finding was influenced by specificity and the release amount. Table 3 shows this relationship for the release amounts examined and 2 levels

Table 3. Sensitivity, time to outbreak detection (timeliness), proportion of outbreaks detected through syndromic surveillance before clinical case finding, and mean detection benefit of syndromic surveillance compared with clinical case finding for 3 release scenarios and 2 levels of specificity*

Amount released (kg)	Specificity 0.900 (1 false alarm every 10 d)				Specificity 0.975 (1 false alarm every 40 d)			
	Sensitivity per outbreak	Mean timeliness, d	Proportion with detection benefit	Mean detection benefit, d	Sensitivity per outbreak	Mean timeliness, d	Proportion with detection benefit	Mean detection benefit, d
1	1.00	3.1 (0, 5)	0.59	1.0 (0, 3.3)	0.98	4.3 (2, 7)	0.28	0.32 (0, 1.0)
0.1	0.99	3.3 (0, 6)	0.55	1.0 (0, 3.5)	0.95	4.7 (2, 7)	0.24	0.33 (0, 1.1)
0.01	0.94	3.6 (0, 7)	0.51	1.1 (0, 3.7)	0.82	5.1 (2, 8)	0.19	0.33 (0, 1.3)

*Values in parentheses are 10th and 90th percentiles of the distribution.

of specificity. When the specificity was 0.9, syndromic surveillance detected from 51% to 59% of outbreaks before clinical case finding, and the mean detection benefit was 1.0–1.1 days, but this specificity resulted in a false alarm every 10 days. At a specificity of 0.975, which reduced false alarms to 1 every 40 days, syndromic surveillance detected 19%–28% of outbreaks before clinical case finding and the mean detection benefit was 0.32–0.33 days, or ≈8 hours. Figure 2 shows that for the 0.01-kg and 1-kg release scenarios (results for the 0.1-kg release are similar, but are not shown), the proportion of outbreaks detected first by syndromic surveillance and the mean detection benefit of surveillance each increased as specificity decreased. Figure 2 also shows that the release amount had a strong effect on the proportion of outbreaks detected first by syndromic surveillance but that it did not have a strong effect on the mean detection benefit.

At a set specificity, syndromic surveillance tended to detect a higher proportion of outbreaks before clinical case finding with increasing release amount. The mean detection benefit, in contrast, tended to decrease when the amount of spores released increased. This decrease in average detection benefit occurred because even though syndromic surveillance detected more outbreaks before clinical case finding as the release amount increased, the detection benefit for the additional outbreaks was small, and the average detection benefit thus decreased.

Discussion

When we compared the performance of clinical case finding with that of syndromic surveillance for detecting an inhalational anthrax outbreak, we found that clinical case finding detected outbreaks on average 3.7–4.6 days after release of spores. The ability of syndromic surveillance to detect an outbreak before clinical case finding was influenced by both specificity and release size, with specificity being the predominant factor. Our results suggest that syndromic surveillance could detect an inhalational anthrax outbreak before clinical case finding. However, we regularly observed a detection benefit only when syndromic surveillance operated at a specificity in the range of 0.9, which corresponds to 1 false alarm every 10 days. When operating at this relatively low specificity with a concomitant high sensitivity, syndromic surveillance detected outbreaks, on average, 1 day before clinical case finding did.

One of the most useful findings of our study was the tradeoff between sensitivity and specificity of syndromic surveillance. To reduce the false alarm rate, specificity must be high. However, as specificity increased in our study, the sensitivity of syndromic surveillance decreased, and the proportion of outbreaks that was detected first by syndromic surveillance decreased more substantially. If the response to a result from syndromic surveillance is resource intensive and includes follow-up investigations in multiple healthcare settings, then a false alarm rate of 1 every 10 days may be too high for such a system to be useful. Alternatively, if public health personnel can rule out false-positive results with minimal investment, then a higher rate of false alarms may be acceptable.

The detection benefit of syndromic surveillance might be an important lead, depending on the action triggered by a surveillance alarm. Because many clinical and public health departments have defined protocols for actions after clinical confirmation of an inhalational anthrax case (*33*), the action after detection of a clinical case is fairly well defined in many jurisdictions. In contrast, the appropriate action after a result from syndromic surveillance system is not well-defined (*34*). For example, some public health departments routinely wait 1 day for a second alarm before taking action (*35*). This strategy could eliminate the potential detection benefit of syndromic surveillance. Another concern is the relatively low specificity at which syndromic surveillance must operate to consistently result in a detection benefit. A system producing this many false alarms may result in excessive costs, and users may minimize the importance of these results.

To be useful, however, syndromic surveillance does not necessarily have to detect all outbreaks, or even most outbreaks, before a clinician detects the first case. The additional lead in detection offered by syndromic surveillance in some outbreaks may result in enough benefit to support the use of syndromic surveillance. Syndromic surveillance may also be useful for applications other than detecting an outbreak caused by bioterrorism; e.g., for detecting other types of disease outbreaks (*36*), for ruling out the existence

of an outbreak, or for evaluating the effect of a public health intervention. Assessment of the question of the utility of syndromic surveillance in general would require consideration of a broader range of costs and benefits than we included in our study.

Our methods are an advance over those used in previous studies because we were able to examine rigorously, within a single modeling framework, the ability of clinical case finding and syndromic surveillance to detect anthrax outbreaks. The nature of our model allowed us to vary some outbreak characteristics directly (e.g., release amount) and to incorporate the uncertainty in parameter values into our final estimates of detection performance and detection benefit. Although our sampling approach did allow us to vary many parameter values simultaneously, it did not clarify how the results vary in relation to changes in the value of a single parameter. Our estimate of detection performance through syndromic surveillance is comparable to estimates observed through studies that used simulation models (*12*,*37*), but those studies did not allow direct comparison of detection through syndromic surveillance with detection through clinical case finding. Our estimate of the time until detection through clinical case finding is longer than the estimate used by the authors of a study aimed at modeling response strategies to an anthrax outbreak (*2*), but those authors did not provide a clear rationale for the value they chose. An initial presumptive diagnosis may occur earlier than the first positive blood culture result (e.g., through clinical symptoms and chest radiographs), but a decision for large-scale intervention would likely not be made until at least after the first definitive diagnosis was made.

In our study, we considered 1 approach to syndromic surveillance for an outbreak resulting from 1 type of organism, and we considered clinical case finding through 1 type of routinely applied diagnostic test. There are many different approaches to syndromic surveillance; e.g., different types of data and different detection algorithms. Although different approaches to surveillance might produce different results, the choice of the infectious organism is likely to have a greater effect on results. Anthrax is relatively unique among bioterrorism agents in that a routinely used diagnostic test (i.e., blood culture) will identify the organism definitively. The benefit of syndromic surveillance relative to clinical case finding may therefore be greater for outbreaks caused by other organisms, and an anthrax outbreak may be a worst-case scenario for syndromic surveillance.

Syndromic surveillance detected an inhalation anthrax outbreak before the first clinical case was diagnosed in as many as half of simulated outbreaks. However, the potential detection benefit of syndromic surveillance compared with clinical case finding depended critically on the specificity and sensitivity at which a surveillance system operated and on the size of the outbreak. When syndromic surveillance was sufficiently sensitive to detect a substantial proportion of outbreaks, it generated frequent false alarms. Public health authorities should be aware that the

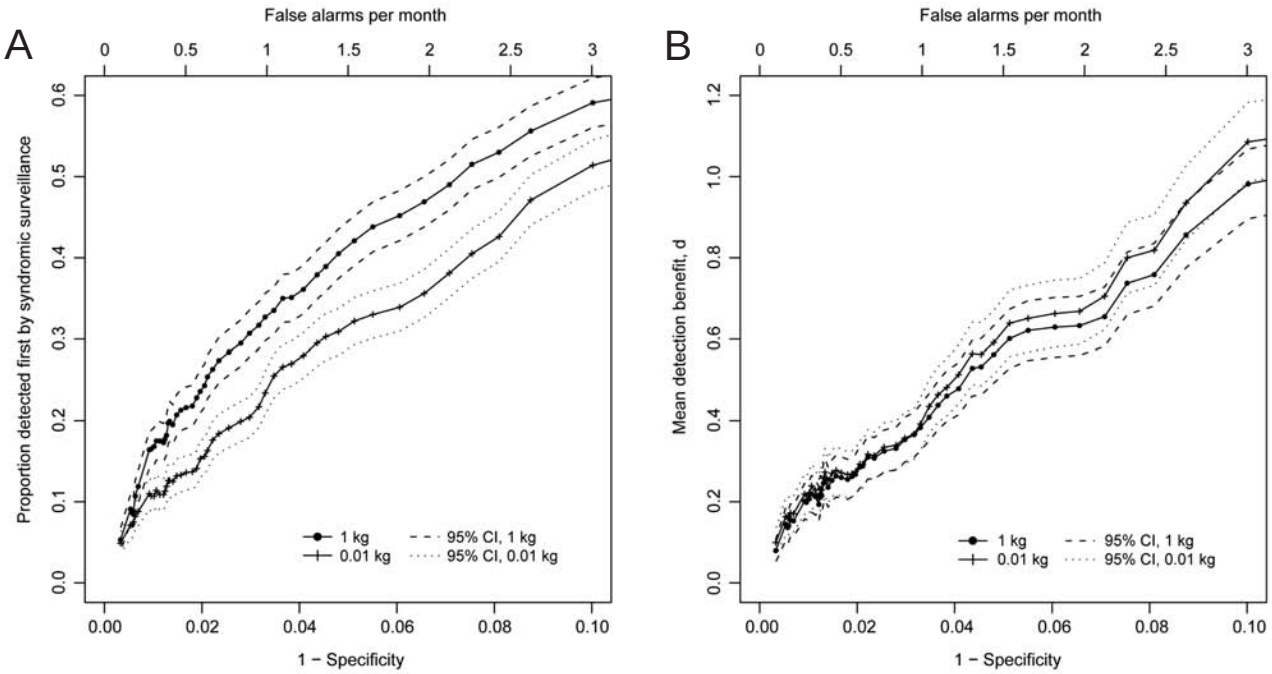


Figure 2. Proportion of inhalational anthrax outbreaks detected by syndromic surveillance before clinical case finding (A) and mean detection benefit of syndromic surveillance compared with clinical case finding as a function of specificity (and false-alarm rate)(B) for 3 release scenarios. CI, confidence interval.

potential detection benefit of syndromic surveillance compared with clinical case finding is influenced strongly by the specificity at which a surveillance system operates. To help detect outbreaks more rapidly, future research should examine the cost-effectiveness of syndromic surveillance and explore approaches to linking syndromic surveillance and clinical case finding more closely.

Acknowledgments

We thank Julie Pavlin for her help in obtaining the data used in this study and for comments on earlier versions of the simulation model. Part of this work was performed while David L. Buckeridge was a Department of Veterans Affairs postdoctoral informatics fellow.

The research of Dr Buckeridge is supported by a Canada research chair in public health informatics.

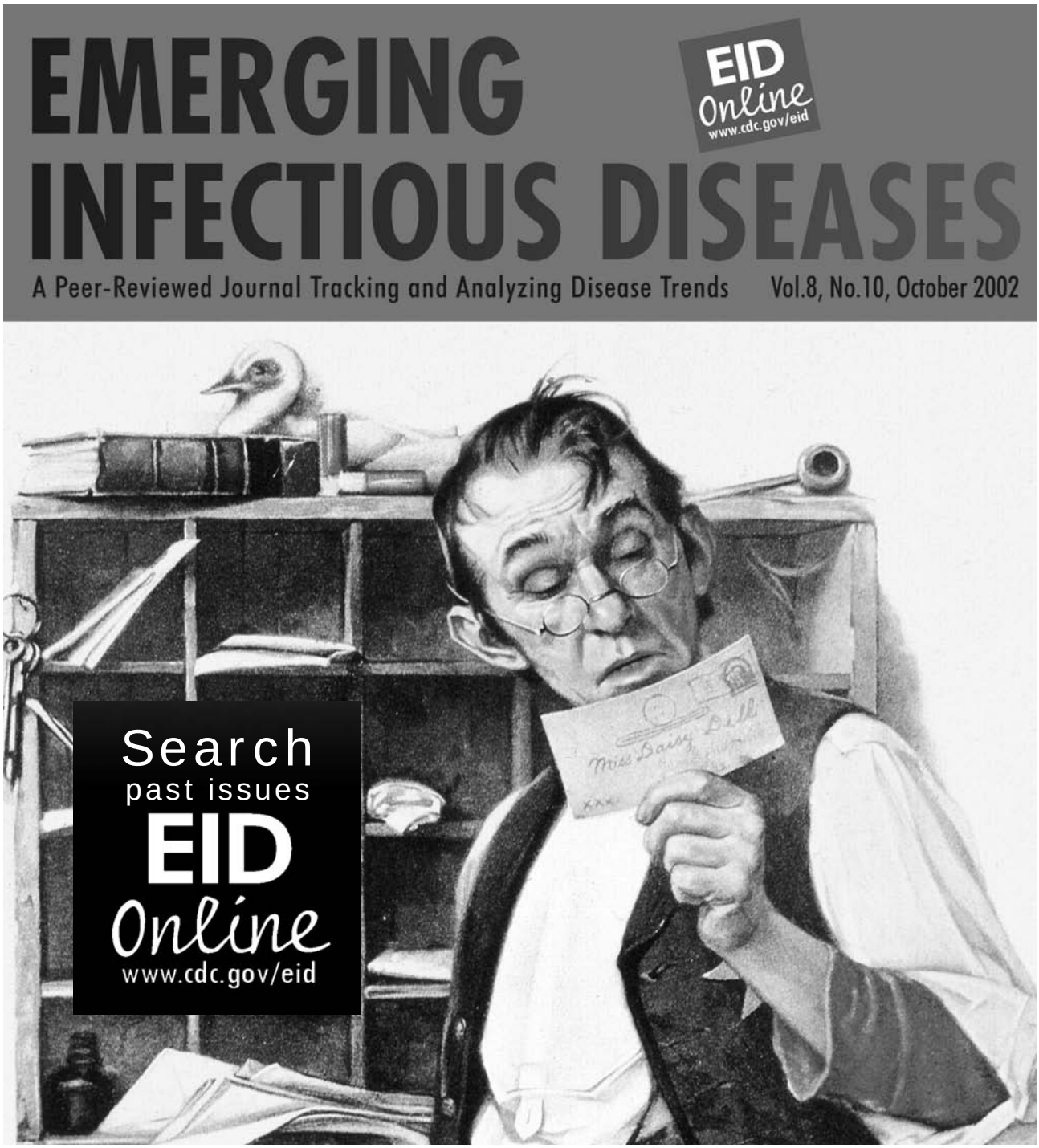
Dr Buckeridge is assistant professor in the Department of Epidemiology, Biostatistics, and Occupational Health at McGill University and a medical consultant with the Montreal Department of Public Health and the Quebec Institute of Public Health. His research interests include public health informatics with a particular focus on the informatics of public health surveillance.

References

1. Kaufmann AF, Meltzer MI, Schmid GP. The economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable? Emerg Infect Dis. 1997;3:83–94.
2. Wein LM, Craft DL, Kaplan EH. Emergency response to an anthrax attack. Proc Natl Acad Sci U S A. 2003;100:4346–51.
3. Smolinski M, Hamburg ML. Lederberg J, editors. Microbial threats to health: emergence, detection, and response. Washington: National Academy Press; 2003.
4. Loonsk JW. BioSense: a national initiative for early detection and quantification of public health emergencies. MMWR Morb Mortal Wkly Rep. 2004;53(Suppl):53–5.
5. United States Government Accountability Office. Information technology. Federal agencies face challenges in implementing initiatives to improve public health infrastructure. Volume GAO-05–308. Washington: The Office; 2005.
6. Buehler JW, Berkelman RL, Hartley DM, Peters CJ. Syndromic surveillance and bioterrorism-related epidemics. Emerg Infect Dis. 2003;9:1197–204.
7. Lombardo J, Burkom H, Elbert E, Magruder S, Lewis SH, Loschen W, et al. A systems overview of the electronic surveillance system for the early notification of community-based epidemics (ESSENCE II). J Urban Health. 2003;80(Suppl 1):i32–42.
8. Wagner MM, Robinson JM, Tsui F-C, Espino JU, Hogan WR. Design of a national retail data monitor for public health surveillance. J Am Med Inform Assoc. 2003;10:409–18.
9. Yih WK, Caldwell B, Harmon R, Kleinman K, Lazarus R, Nelson A, et al. National Bioterrorism Syndromic Surveillance Demonstration Program. MMWR Morb Mortal Wkly Rep. 2004;53(Suppl):43–9.
10. Bravata DM, McDonald KM, Smith WM, Rydzak C, Szeto H, Buckeridge DL, et al. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. Ann Intern Med. 2004;140:910–22.
11. 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;53:1–11.
12. Nordin JD, Goodman M, Kuldorff M, Ritzwoller DP, Abrams AM, Kleinman K, et al. Simulated anthrax attacks and syndromic surveillance. Emerg Infect Dis. 2005;11:1394–8.
13. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. Proc Natl Acad Sci U S A. 2003;100:1961–5.
14. Stoto M, Schonlau M, Mariano L. Syndromic surveillance: is it worth the effort? Chance. 2004;17:19–24.
15. McKay M, Beckman R, Conover W. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics. 1979;21:239–45.
16. Buckeridge DL. A method for evaluating outbreak detection in public health surveillance systems that use adminstrative data. [Doctoral dissertation]. Stanford (CA): Stanford University; 2005.
17. Buckeridge DL, Burkom H, Moore A, Pavlin J, Cutchis P, Hogan W. Evaluation of syndromic surveillance systems: design of an epidemic simulation model. MMWR Morb Mortal Wkly Rep. 2004;53(Suppl):137–43.
18. 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.
19. Defense Threat Reduction Agency. The HPAC user's guide. Hazard prediction and assessment capability, version 4.0. Fort Belvior (VA): The Agency; 2001.
20. Glassman HN. Industrial inhalation anthrax. Discussion. Bacteriol Rev. 1965;30:657–9.
21. Brookmeyer R, Blades N, Hugh-Jones M, Henderson DA. The statistical analysis of truncated data: application to the Sverdlovsk anthrax outbreak. Biostatistics. 2001;2:233–47.
22. Webb GF, Blaser MJ. Mailborne transmission of anthrax: modeling and implications. Proc Natl Acad Sci U S A. 2002;99:7027–32.
23. McIsaac WJ, Levine N, Goel V. Visits by adults to family physicians for the common cold. J Fam Pract. 1998;47:366–9.
24. Metzger KB, Hajat A, Crawford M, Mostashari F. How many illnesses does one emergency department visit represent? Using a population-based telephone survey to estimate the syndromic multiplier. MMWR Morb Mortal Wkly Rep. 2004;53(Suppl):106–11.
25. Inglesby TV, O'Toole T, Henderson DA, Bartlett JG, Ascher MS, Eitzen E, et al. Anthrax as a biological weapon, 2002: updated recommendations for management. JAMA. 2002;287:2236–52.
26. Jernigan JA, Stephens DS, Ashford DA, Omenaca C, Topiel MS, Galbraith M, et al. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. Emerg Infect Dis. 2001;7:933–44.
27. Reimer LG, Wilson ML, Weinstein MP. Update on detection of bacteremia and fungemia. Clin Microbiol Rev. 1997;10:444–65.
28. National Center for Health Statistics. Public use data tapes: National Ambulatory Medical Care Survey 2003. Hyattsville (MD): The Center; 2003.
29. Begier EM, Barrett N, Mshar P, Johnson D, Hadler J; Connecticut Bioterrorism Field Epidemiology Response Team. Gram-positive rod surveillance for early anthrax detection. Emerg Infect Dis. 2005;11:1483–6.
30. Beekmann SE, Diekema DJ, Chapin KC, Doern GV. Effects of rapid detection of bloodstream infections on length of hospitalization and hospital charges. J Clin Microbiol. 2003;41:3119–25.
31. Box GEP, Jenkins GM. Time series analysis, forecasting and control. Englewood Cliffs (NJ): Prentice Hall; 1976.
32. Page E. Continuous inspection schemes. Biometrika. 1954;41:100–15.
33. Tengelsen L, Hudson R, Barnes S, Hahn C. Coordinated response to reports of possible anthrax contamination, Idaho, 2001. Emerg Infect Dis. 2002;8:1093–5.

34. Pavlin JA. Investigation of disease outbreaks detected by "syndromic" surveillance systems. J Urban Health. 2003;80(Suppl 1):i107–14.
35. Heffernan R, Mostashari F, Das D, Karpati A, Kuldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. Emerg Infect Dis. 2004;10:858–64.
36. Marx MA, Rodriguez CV, Greenko J, Das D, Heffernan R, Karpati AM, et al. Diarrheal illness detected through syndromic surveillance after a massive power outage: New York City, August 2003. Am J Public Health. 2006;96:547–53.
37. Miller B, Kassenborg H, Dunsmuir W, Griffith J, Hadidi M, Nordin JD, et al. Syndromic surveillance for influenzalike illness in ambulatory care network. Emerg Infect Dis. 2004;10:1806–11.

Address for correspondence: David L. Buckeridge, McGill Clinical and Health Informatics, Department of Epidemiology and Biostatistics, McGill University, 1140 Pine Ave West, Montreal, Quebec, H3A 1A3, Canada; email: david.buckeridge@mcgill.ca



This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

Technical Appendix

In this document we provide additional information about the methods used in the simulation study that we describe in the main paper. In section 1 we describe the simulation model and in section 2 we describe the study design.

1 Simulation Model

The simulation model used in this study builds on our earlier work [5–7].

1.1 Dispersion

The dispersion model takes as input the amount of anthrax spores released and returns as output the number of spores an individual would inhale at locations on a regular grid overlaid upon the study region. To model the dispersion of anthrax spores, we used the Hazard Prediction and Assessment Capability (HPAC) software, which is developed by the US Defence Threat Reduction Agency (DTRA) to allow modeling of threats from weapons of mass destruction [8]. This software uses meteorological and terrain data to interpolate time-integrated exposure concentrations from a release scenario onto a spatial grid, relying on the second-order cluster integrated puff model [24].

The HPAC software outputs a mean exposure plume for a release scenario. The mean plume is determined within HPAC by simulating many individual plumes and then calculating the mean and standard deviation of exposure at each location on an exposure grid. We did not use the mean plume directly because it smooths out the variation in spore concentration within a plume and because the mean plume tends to overestimate exposure at the margins of the plume. To address these limitations of the mean plume, we re-sampled randomly from the mean plume.

The release scenario we used for all release amounts was a stationary release in the area of the Langley Air Force Base (37.1N, 76.4W), using weather data for Veteran's Day (November 11, 5 am EST, 2 m s^{-1} NNW wind, clear skies), with a 2 m release height, and a 10 s duration. The HPAC software was set to calculate a boundary layer and large-scale variability, and to include terrain effects. We assumed that 1 kg of spores contained 10^{15} spores and we modeled the dispersion of 1 kg, 0.1 kg, and 0.01 kg of spores. We calculated a mean plume at each release amount and then re-sampled 1,000 random plumes from each mean plume. The HPAC model provides the time-integrated concentration

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

at each location and we assumed that individuals breathed at a rate of 0.0005 $m^3 s^{-1}$ to determine the number of spores an individual would inhale at locations throughout the study region.

1.2 Infection

The infection model takes as input the number of spores an individual would inhale at locations on a regular grid (output from the dispersion model) and the population by ZIP code. The model returns the number of infected individuals by residential ZIP code. We assumed that all individuals were exposed at their home address and that each individual had a uniform probability of being exposed at any location within their home ZIP code. To determine the probability of infection following exposure to a given number of spores, we used a function corresponding to the data reported by Glassman following experimental exposure of primates [10]. This is a probit model with an LD50 of 8,600 spores and a slope of 0.67. Figure 1 shows a plot of the infection function. This function allows for infection at low dose exposure. For example, approximately 2.5% of individuals exposed to 10 spores will become infected.

To determine the number infected by home ZIP code, we first determined the probability of infection for each ZIP code Z_i as the average probability of infection across the ZIP code:

$$p_{Z_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} f(c_j),$$

where there are n_i cells from the exposure grid within ZIP code Z_i and $f(c_j)$ is the infection function that returns the probability of infection given the number of spores inhaled in cell c_j. The number of individual infected within a ZIP code I_{Z_i} was then sampled from a binomial distribution:

$$I_{Z_i} \sim binomial(N_{Z_i}, p_{Z_i}),$$

where N_{Z_i} is the population of ZIP code Z_i.

1.3 Disease

The disease model takes as input the number of infected individuals and returns a disease path for each individual. The disease path describes the amount of time spent in each discrete disease state. We used a semi-Markov process to model progression of an individual with inhalational anthrax through three disease states: incubation, prodromal, and fulminant (Figure 2) [26].

The definition of the semi-Markov process requires identification of the states, including the holding-time functions, and specification of the transition probabilities between states. The initial state in the model was incubation, followed by certain transition to the prodromal state, and then the fulminant state. For holding-time functions, we used the lognormal distribution, which appears to describe the duration of incubation for many diseases [20, 23], including inhalation

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

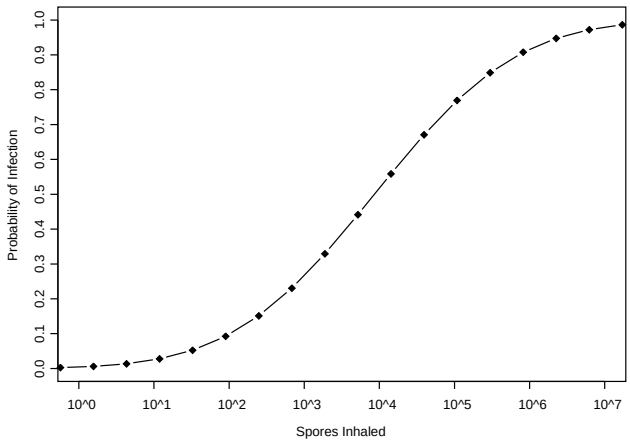


Figure 1: The infection function for inhaled anthrax determined mainly from primate research [10].

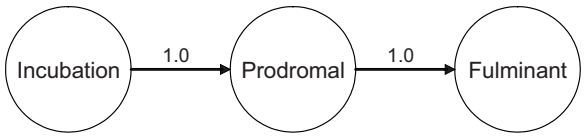


Figure 2: The disease model for inhalational anthrax. Infected individuals all pass through three disease states. In the incubation state individuals have no symptoms. In the prodromal state individuals experience an influenza-like illness. Finally, in the fulminant state individuals experience severe symptoms, such as shock. The holding-time function in each state is a lognormal distribution with parameters shown in Table 1.

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

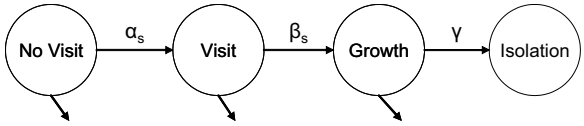


Figure 3: The health-care utilization model. When individuals enter the prodromal or fulminant disease state, they enter the health-care utilization model. The probability of making a visit (α_s) varies by disease state (Table 1). The probability of a positive blood culture (β_s) is the product of the probability of ordering a test (which varies by disease state) and the sensitivity of the test (which does not vary by disease state). The probability of isolating the organism does not vary by disease state. The holding-time function for the 'No Visit' state is triangular, with a duration equivalent to the length of the disease state. Holding-time functions for the 'Visit' and 'Growth' states are exponential, with parameters shown in Table 1.

anthrax [4, 17]. The values used to parameterize the holding-time functions are taken from observational studies of human exposure [4, 17] and other modeling studies [25, 26], and are shown in Table 1.

1.4 Health-Care Utilization

The health-care utilization model takes as input a set of disease paths and for each path performs three tasks: (1) it identifies if and when individuals seek care in each disease state, (2) it determines the presenting syndrome for individuals that seek care, and (3) it identifies the timing and results of blood culture testing once care is sought.

We used a semi-Markov process to model health-care utilization (Figure 3). A separate process was used to describe health-care utilization in each of the prodromal and fulminant disease states. Both processes had the same states and transitions (Figure 3), but some values for holding-time functions and transition probabilities differed between the disease states (i.e., those states with a s subscript in Figure 3) and the values used in the simulation study are shown in Table 1.

The transition from 'No Visit' to 'Visit' represents an individual seeking care at an ambulatory clinic or emergency department. The probability of this transition occurring (α_s) differs between disease states (s). We set the probability of a visit in the prodromal disease state (α_p), to approximately 0.30 because cross-sectional surveys suggest that this proportion of individuals visit a physician at some point during an episode of upper respiratory tract illness [15, 18]. For the fulminant disease state (α_f), we estimated the probability of seeking care as approximately 95% given the severity of the symptoms in that state.

The transition from 'Visit' to 'Growth' represents an individual having a positive blood culture test after making a visit. The probability of this transition (β_s) is the product of the probability of performing a blood culture test (β_s^1) and

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

the sensitivity of the test (β^2, i.e., $\beta_s = \beta_s^1 \times \beta^2$). The probability of performing a test in the prodromal state (β_p^1), was estimated from the National Ambulatory Health Care Survey as approximately 0.0125 [9]. In the fulminant state, we assumed that the probability of a blood culture test (β_f^1) was approximately 0.95. We relied on published studies of blood-culture testing to estimate the sensitivity of blood-culture testing in both symptomatic disease states (β^2) as approximately 0.85 [21].

The final transition, from 'Growth' to 'Isolation', represents the decision to isolate the organism from a blood culture bottle growing gram-positive rods. We relied on data from a recent survey to estimate this value (γ) as approximately 0.9 [2].

In addition to a transition probability, each of the first three states in the health-care utilization model also requires a holding-time function. The holding-time function for the 'No Visit' state models the distribution of time to seeking care, given that care is sought. We used a right triangular distribution fit to the time spent in the disease state. So, for example, if an individual had a prodromal disease state duration of 10 days, then the probability of seeking care at the instant of entering the disease state would be zero, and the instantaneous probability of seeking care would increase linearly to 0.2 at ten days, with a mean time to seeking care of 6.7 days.[1] This approach to modeling visits effectively limits individuals to a single visit in each disease state. The selection of a triangular distribution reflects the lack of published evidence about the timing of health-care utilization following the onset of symptoms.

The holding time function for 'Visit' reflects the distribution of times until growth occurs given that the test is positive. The holding time function for 'Growth' is the distribution of times until the organism is isolated given than a decision is made to isolate a specific organism. We modeled both these holding times as exponential with means obtained from published reviews of blood-culture testing [1, 11].

Finally, for individuals that made a health care visit, we simulated the syndrome assigned to the individual using probabilities that reflect the distribution of clinical presentations for inhalational anthrax reported in the literature [11, 12]. As we consider only respiratory syndromes for surveillance, we modeled only the probability of being assigned a respiratory syndrome in the prodromal disease state, which we estimated at approximately 0.75 (Table 1).

2 Study Design

2.1 Generation of Simulated Signals

We generated 1,000 simulated outbreak signals at each release amount for a total of 3,000 signals. To generate the simulated outbreaks for each release amount,

[1] These values result from the properties of the triangular distribution, which is defined by three parameters: a, b, and c. In a right triangular distribution, b = c. The maximum point density is 2 / (b - a) and the mean is (a + b + c) / 3.

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

we first sampled an exposure grid and calculated the number infected by home ZIP code. The next step was to select a set of disease and health-care utilization parameters. We then generated a disease path for each infected individual, the timing of visits to physicians for symptomatic individuals, the administrative codes generated through visits, and the occurrence, timing and results of blood culture testing.

Due to the large number of parameter values in the disease and health-care utilization components of the simulation model, we used Latin hypercube sampling (LHS) to select parameter values for each simulation run of these components. LHS is an approach to sampling parameter values from a high-dimensional parameter space in order to obtain estimates of output variables that are more efficient and precise than would be obtained with simple random sampling [16]. When specifying a simulation model there are K parameters. A given run of the simulation model requires a value for each parameter, or a set of parameter values $\mathbf{X} = \{X_1, \ldots, X_K\}$. Each parameter has a space of possible values $\mathbf{S} = \{S_1, \ldots, S_K\}$. In LHS the space for each parameter is partitioned into N intervals of probability size $1/N$. The Cartesian product of these intervals partitions S into N^K cells, which form a hypercube. To obtain $\mathbf{X}$ for a simulation run requires randomly sampling a partition for each parameter, and then sampling a parameter value from within that partition, assuming that values are uniformly distributed within a partition. Table 1 shows the parameter value intervals used in the simulation study.

In order to reduce the variance between scenarios, we used the same 1,000 sets of disease and health-care utilization parameters, selected through LHS, for each scenario. In other words, the first runs for each of the 3 release amounts all used the first set of sampled parameters, the second runs all used the second set of sampled parameters, and so on. Similarly, we sampled 1,000 exposure grids and all 3 scenarios used the same 1,000 re-sample exposure plumes.[2] Finally, the same random number generator with the same seed value was used for each scenario. We used a combined multiple recursive generator as proposed and implemented by L'Ecuyer with the default initial seed [14]. This sampling strategy was intended to improve the efficiency of the simulation and reduce the variance of the output variables [13]. The net result is to facilitate comparison of the results across the different scenarios.

2.2 Combination of Simulated Data with Baseline Data

We first defined a 330 day interval on the baseline data from January 19, 2003 until December 15, 2003 as possible starting dates for a simulated outbreak. The gap at the beginning of the year was to allow a period for the detection algorithm to initialize using test data, and the gap at the end of the year was to ensure that injected outbreaks did not run off the end of the test data. We then selected 1,000 random dates for injecting outbreaks. Each of the 330 dates

[2]The pattern of dispersion is constant and the number of spores scales linearly with amount release. We therefore simulated a 1 kg release and scaled the results for the two other release amounts.

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

Parameter	Distribution	Parameter Value Intervals	Source
Holding-Time Functions			
Disease Model			
Incubation (days), median	lognormal	[5, 9] [9, 11] [11, 15]	[4]
Incubation, dispersion∗	lognormal	[1.5, 1.9] [1.9, 2.1] [2.1, 2.5]	[4]
Prodromal (days), median	lognormal	[1.5, 2.3] [2.3, 2.7] [2.7, 3.5]	[26]
Prodromal, dispersion	lognormal	[1.2, 1.4] [1.4, 1.5] [1.5, 1.7]	[26]
Fulminant (days), median†	lognormal	[1.5]	[26]
Fulminant dispersion†	lognormal	[1.44]	[26]
Health-Care Utilization Model			
Time until visit	right triangular	disease-state duration‡	Estimate
Time until blood culture growth (days)	exponential	[0.4, 0.8] [0.8, 1.0] [1.0, 1.4]	[1]
Time until blood culture isolation (days)	exponential	[0.5, 0.6] [0.6, 0.9] [0.9, 1.5]	[11]
Transition Probabilities			
No Visit to Visit (α_s)			
Prodromal state visit (α_p)	Bernoulli	[0.05, 0.25] [0.25, 0.35] [0.35, 0.55]	[18]
Fulminant state visit (α_f)	Bernoulli	[0.7, 0.9] [0.9, 0.95] [0.95, 1]	Estimate
Visit to Growth ($\beta_s = \beta_s^1 \times \beta^2$)			
Blood culture test, prodromal state (β_p^1)	Bernoulli	[0.001, 0.01] [0.01, 0.015] [0.015, 0.025]	[9]
Blood culture test, fulminant state (β_f^1)	Bernoulli	[0.7, 0.9] [0.9, 0.95] [0.95, 1]	Estimate
Sensitivity of blood culture (β^2)	Bernoulli	[0.5, 0.8] [0.8, 0.9] [0.9, 1]	[21]
Growth to Isolation (γ)			
Prodromal and fulminant states	Bernoulli	[0.5, 0.8] [0.8, 0.9] [0.9, 1]	Estimate
Assignment of Presenting Syndrome			
Probability of respiratory syndrome, prodromal state	Bernoulli	[0.5, 0.7] [0.7, 0.8] [0.8, 1]	[12]

Table 1: Ranges of parameter values used in the Latin hypercube sampling. The indicated source was used to identify the best estimate of the parameter value interval and inform the range of the intervals. ∗Following Sartwell [23], the parameter $d = e^s$, where s^2 is the variance, is referred to as the *dispersion factor*. †We did not model uncertainty about the fulminant disease state because the there is general agreement that this state has a short duration. ‡The length of disease state was used to parameterize the triangular visit-time distribution.

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

was used 3 times and ten dates were sampled randomly from the 330 dates to be used a fourth time. All 100 dates were then shuffled randomly.

To inject the simulated outbreak signals for a release amount, we used the following method. Each release amount had set of simulated outbreaks, $O = \{O_1, \ldots, O_{1000}\}$, and the set of randomly ordered dates, $D = \{D_1, \ldots, D_{1000}\}$. For outbreak j, $j \in \{1, \ldots, 1000\}$, we selected O_j and D_j. The outbreak O_j is a time series of counts, lasting n days and representing the visits for respiratory conditions from the day of the simulated release until the day of the peak incidence of cases.

The outbreaks series O_j is an ordered set of values, $O_j = \{o(1), \ldots, o(n)\}$, which we define to run from D_j until $D_j + n - 1$, or $O_j = \{o(D_j), \ldots, o(D_j + n - 1)\}$. The background time series is also an ordered set of values, $B = \{b(1), \ldots, b(m)\}$. For inject j, we extract a subset of the background time series $B_j = \{b(D_j - g), \ldots, b(D_j), \ldots, b(D_j + n - 1)\}$, where g is the length of the lead-in gap, which is the amount of time in the inject series before the beginning of the outbreak. We then define the inject series $I_j = \{i(D_j - g), \ldots, i(D_j + n - 1)\}$, with the entries in the series defined as,

$$i(k) = \begin{cases} b(k) + o(k) & \text{if } \; k \geq D_j \text{ and } k < D_j + n \\ b(k) & \text{otherwise} \end{cases}$$

In other words, the inject series is formed by adding the values of the outbreak series to the values of the background series, after aligning the two series so that the first day of the outbreak series is added to day D_j in the background series. We then applied the outbreak detection algorithm to each day in the inject series to generate alarm values from day $D_j - g$ to day $D_j + n - 1$. The lead-in gap $g = 16 + (3 \times 28) = 100$ days, and the first 16 values were excluded from the alarm series as they were required to initialize the temporal forecast algorithm.[3] This left an 84 day (approximately 3 months) initilization period.

2.3 Outbreak Detection Algorithm

We used an autoregressive seasonal integrated moving average (SARIMA) model [3] to calculate one-step-ahead daily forecasts of respiratory syndrome counts and then used a cumulative sum [19] to detect positive deviations in the forecast residuals. Other researchers have used a similar approach to outbreak detection in a surveillance setting [27].

To fit the SARIMA model, we used two years of data for respiratory syndrome visit counts (2001 to 2002, inclusive) and followed a procedure similar to that described elsewhere [22]. This entailed subtracting the overall mean, day-of-week means, month means, and holiday means from the original count data to give a series centered on zero. We used trimmed means (alpha = 0.1) for both day-of-week and month to minimize the influence of outliers. We then assessed the temporal autocorrelation in this series and fit a SARIMA model using a standard approach to model specification [3].

[3]The number of days excluded is determined by the number and order or terms in the time-series model

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

We evaluated the fit of the SARIMA model to the training (2001 to 2002) and test data (2003) using the mean absolute percentage error (MAPE), defined as

$$MAPE = \frac{1}{j}\sum_{j=1}^{m}\frac{|\mu_j - x_j|}{x_j},$$

where there are m days in the training interval, μ_j is the forecast value on day j and x_j is the observed value. After subtracting the overall mean and means for day-of-week, month and holiday, the zero-centered series exhibited temporal autocorrelation at short lags on the order of days, and cyclical lags of order seven. We found that a SARIMA model (2,0,1) x (2,0,1)7 had the best fit to the zero-centered series. One-step-ahead forecasts from the SARIMA model resulted in a mean absolute percentage (MAPE) of 14.9% on the training data, which implies that the forecast values were, on average, within 14.9% of the true value. This fit is similar to or better than the fit reported by others using the same algorithm and similar data [22].

To detect temporal aberrancies in the observed counts, we applied a cumulative sum to the standardized forecast residual and declared an aberrancy when the cumulative sum exceeded a threshold. We calculated the standardized residual for each day as the observed total respiratory visit count minus the one-step-ahead forecast from the SARIMA model, divided by the standard error of the forecast. Standardization, or dividing the residual by the standard error of the forecast, resulted in residuals with a standard normal distribution.

We then applied a one-sided cumulative sum to the standardized residuals to detect a positive shift in the mean of the residual series:

$$S_t = max(0, S_{t-1} + ((X_t - (\mu_0 + k\sigma_x))/\sigma_x)).$$

The cumulative sum requires four parameters: the series mean μ_0, the series standard deviation σ_x, the shift k, and the decision threshold h. The shift specifies, in standard deviations, the minimum detectable change in the mean. We set the shift at 0.5 standard deviations and declared an aberrancy for a given day, j, when the cumulative sum crossed the decision threshold, or

$$A(h)_j = \begin{cases} 1 & \text{if } S_j > h \\ 0 & \text{otherwise.} \end{cases}$$

We varied h between 0 and 10 to determine the outbreak detection performance over a range of thresholds.

2.4 Evaluation of Outbreak Detection Performance

2.4.1 Sensitivity

Sensitivity is the probability of an alarm given an outbreak, or

$$Sensitivity = P(A|O) = \frac{n(A,O)}{n(O)},$$

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

where $n(O)$ is the number of outbreaks and $n(A, O)$ is the number of outbreaks during which at least one alarm sounded. We calculated sensitivity for detecting an outbreak following a given release amount at a decision threshold h as:

$$Se(h) = \frac{1}{n} \sum_{i=1}^{n} \min(1, \sum_{j=1}^{m_i} A(h)_{ij}),$$

where there are n simulation runs and simulated outbreak i has m_i days from onset to peak incidence. Note that sensitivity is computed using only the outbreak intervals and no distinction is made as to when in the course of an outbreak alarms occur.

2.4.2 Specificity

Specificity is the probability of no alarm given that there is no outbreak, or

$$Specificity = P(\overline{A}|\overline{O}) = \frac{n(\overline{A}, \overline{O})}{n(\overline{O})},$$

where $n(\overline{O})$ is the number of days in the test data and $n(\overline{A}, \overline{O})$ is the number of alarms when the algorithm is applied to the test data without any superimposed outbreaks. We calculated specificity at a decision threshold h as:

$$Sp(h) = \frac{1}{m} \sum_{j=1}^{m} A(h)_j,$$

where there are m days in the test data set. Note that specificity is calculated using only non-outbreak data. We therefore assume that any alarm in the test data without a superimposed outbreak is a false alarm.

2.4.3 Timeliness

Timeliness is the time between the onset of the outbreak and the first alarm sounded during an outbreak. We calculated timeliness for a single simulated outbreak as:

$$T(h) = \min_j (j : A(h)_i = 1),$$

where there are m_i days from the onset until the peak of an outbreak i and timeliness is not defined if $\sum_{j=1}^{m_i} A(h)_j = 0$. Note that the mean timeliness for a set of outbreaks is not necessarily monotonically increasing or decreasing with the threshold. The reason for this potential variation in timeliness is that as the thresholds changes, new outbreaks may be detected at the changed threshold. The mean is re-calculated incorporating the timeliness from these additional outbreaks, but there is no guarantee that the timeliness for these new outbreak will be less than (or greater than) the mean timeliness before their inclusion.

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

The effect of newly incorporated outbreaks is less at lower specificity, where there are already a considerable number of outbreaks detected. At higher specificity, though, the derivative of the mean timeliness can change from negative to positive as the threshold changes.

Timeliness for detection through clinical case-finding in a simulation run was calculated as the minimum of the times to a positive blood culture diagnosis for all positive blood cultures ordered for visits occurring in either disease state which were assigned any syndrome.

2.4.4 Detection Benefit

Detection benefit is the potential gain in time to detection from using a new detection method relative to a standard or existing method. For a new detection method A and an existing method B, the benefit of A over B for a single outbreak is calculated as the difference in the timeliness using the two methods, or

$$D_{AB}(h) = \max(0, T_B(h) - T_A(h)).$$

The detection benefit is always greater than or equal to zero. Note that, as with timeliness, the mean detection benefit is not necessarily monotonically increasing or decreasing with the threshold.

Another measure of detection benefit is the proportion of times that method A detects an outbreak before method B. For a single outbreak, this is equivalent to calculating a binary measure in place of the continuous detection benefit, or

$$P_{AB}(h) = \begin{cases} 0 & \text{if } D_{AB}(h) = 0 \\ 1 & \text{if } D_{AB}(h) > 0. \end{cases}$$

References

[1] SE Beekmann, DJ Diekema, KC Chapin, and GV Doern. Effects of rapid detection of bloodstream infections on length of hospitalization and hospital charges. *Journal of Clinical Microbiology*, 41(7):3119–25, 2003.

[2] EM Begier. Gram-positive Rod Surveillance for Early Anthrax Detection. *Emerg Infect Dis*, 11(9):1483–1486, Sep 2005.

[3] GP Box and GM Jenkins. *Time Series Analysis, Forecasting and Control.* Prentice Hall, Englewood Cliffs, Revised edition, 1976.

[4] R Brookmeyer, N Blades, M Hugh-Jones, and DA Henderson. The statistical analysis of truncated data: application to the Sverdlovsk anthrax outbreak. *Biostatistics*, 2(2):233–247, 2001.

[5] DL Buckeridge. *A Method for Evaluating Outbreak Detection in Public Health Surveillance Systems that Use Adminstrative Data.* PhD thesis, Stanford University, 2005.

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

[6] DL Buckeridge, H Burkom, A Moore, J Pavlin, P Cutchis, and W Hogan. Evaluation of syndromic surveillance systems–design of an epidemic simulation model. *MMWR Morb Mortal Wkly Rep*, 53 Suppl:137–43, 2004.

[7] DL Buckeridge, P Switzer, D Owens, D Siegrist, J Pavlin, and M Musen. An evaluation model for syndromic surveillance: assessing the performance of a temporal algorithm. *MMWR Morb Mortal Wkly Rep*, 54 Suppl:109–15, 2005. 1545-861X (Electronic) Journal Article.

[8] The Defense Threat Reduction Agency. *The HPAC User's Guide. Hazard Prediction and Assessment Capability, Version 4.0*, 2001.

[9] National Center for Health Statistics. Public use data tapes: National ambulatory medical care survey 2003. Technical report, National Center for Health Statistics, 2003.

[10] HN Glassman. Industrial inhalation anthrax, discussion. *Bacteriological Review*, 30:657–659, 1965.

[11] TV Inglesby, T O'Toole, DA Henderson, JG Bartlett, MS Ascher, E Eitzen, AM Friedlander, J Gerberding, J Hauer, J Hughes, J McDade, MT Osterholm, G Parker, T M Perl, P K Russell, and K Tonat. Anthrax as a biological weapon, 2002: updated recommendations for management. *Journal of the American Medical Association*, 287(17):2236–52, 2002.

[12] JA Jernigan, DS Stephens, DA Ashford, C Omenaca, MS Topiel, M Galbraith, M Tapper, TL Fisk, S Zaki, T Popovic, RF Meyer, CP Quinn, SA Harper, SK Fridkin, JJ Sejvar, CW Shepard, M McConnell, J Guarner, WJ Shieh, JM Malecki, JL Gerberding, JM Hughes, BA Perkins, and Anthrax Bioterrorism Investigation Team. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. *Emerging Infectious Diseases*, 7(6):933–44, 2001.

[13] AM Law and WD Kelton. *Simulation modeling and analysis*. McGraw-Hill, Third edition, 2000.

[14] P L'Ecuyer. Random number generation. In JE Gentle, E Härdle, and Y Mori, editors, *Handbook of Computational Statistics*, pages 35–70. Springer-Verlag, 2004.

[15] WJ McIsaac, N Levine, and V Goel. Visits by adults to family physicians for the common cold. *The Journal of Family Practice*, 47(5):366–9, 1998.

[16] MD McKay, RJ Beckman, and WJ Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245, 1979.

[17] M Meselson, J Guillemin, M Hugh-Jones, A Langmuir, I Popova, A Shelokov, and O Yampolskaya. The Sverdlovsk anthrax outbreak of 1979. *Science*, 266(5188):1202–8, 1994.

This material, provided by the author(s) as a supplement to Evaluating Detection of an Inhalational Anthrax Outbreak, is not part of Emerging Infectious Diseases contents.

[18] KB Metzger, A Hajat, M Crawford, and F Mostashari. How many illnesses does one emergency department visit represent? Using a population-based telephone survey to estimate the syndromic multiplier. *MMWR Morbidity and Mortality Weekly Report*, 53 Suppl:106–11, 2004.

[19] E Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954.

[20] P Philippe. Sartwell's incubation period model revisited in the light of dynamic modeling. *Journal of Clinical Epidemiology*, 47(4):419–33, 1994.

[21] LG Reimer, ML Wilson, and MP Weinstein. Update on detection of bacteremia and fungemia. *Clinical Microbiology Reviews*, 10(3):444–65, 1997.

[22] BY Reis, M Pagano, and KD Mandl. Using temporal context to improve biosurveillance. *Proceedings of the National Academy of Sciences USA*, 100(4):1961–5, 2003.

[23] PE Sartwell. The distribution of incubation periods of infectious diseases. *American Journal of Hygiene*, 51:310–318, 1950.

[24] RI Sykes, SF Parker, DS Henn, and WS Lewellen. Numerical simulation of ANATEX tracer data using a turbulence closure model for long-range dispersion. *Journal of Applied Meteorology*, 32:929–947, 1993.

[25] GF Webb and MJ Blaser. Mailborne transmission of anthrax: Modeling and implications. *Proceedings of the National Academy of Sciences USA*, 99(10):7027–32, 2002.

[26] LM Wein, DL Craft, and EH Kaplan. Emergency response to an anthrax attack. *Proceedings of the National Academy of Sciences USA*, 100(7):4346–4351, 2003.

[27] GD Williamson and HG Weatherby. A monitoring system for detecting aberrations in public health surveillance reports. *Statistics in Medicine*, 18(23):3283–98, 1999.