

SITUATIONAL AWARENESS IN A BIOTERROR ATTACK VIA PROBABILITY MODELING

Edward H. Kaplan
William N and Marie A Beach Professor of Management Sciences
Yale School of Management
Professor of Public Health
Yale School of Medicine
Professor of Engineering
Yale Faculty of Engineering

Johan Walden
Assistant Professor of Finance
Haas School of Business
University of California, Berkeley

1. Introduction

Following the events of September 11, 2001 and the subsequent anthrax mailings in the United States, the threat posed by bioterrorism has received renewed attention by governments across the globe. Though no large-scale attacks have occurred as of the date of this writing, several scenarios have been studied in detail, including deliberate releases of smallpox [1, 2], anthrax [3], plague [4], and botulinum [5], among others. While the specifics of these scenarios differ in terms of the dispersion mechanisms of the infectious agents and the resulting casualties that could result, all such analyses emphasize the importance of early detection and rapid response in mitigating the consequences of a bioterror attack.

There have been many different suggestions to detect a bioterror attack, ranging from the use of biosensors [6] to syndromic surveillance [7], to screening blood donors [8], to case identification. Once an attack has been detected and the agent has been identified, however, the key concerns become establishing the scale of the attack and forecasting how casualties will emerge over time. Such forecasts can guide the allocation of response resources; for example, is it necessary to add “surge capacity” to hospitals in the impacted area or are existing facilities adequate for expected caseloads. The real-time aspect of this approach enables rapid reassessment

of the situation as new caseload information arrives over time, and also provides useful guidance for deciding when the outbreak has ended (in the sense that few cases remain to be uncovered). We refer to such real-time monitoring, assessment, and forecasting activities as *situational awareness*.

This chapter illustrates the basic situational awareness problem with the aid of a recently developed model applied to a hypothetical anthrax bioterror attack [9]. The intent is to introduce the sorts of analyses that are possible using probability models that are simple enough to evaluate in an ordinary spreadsheet program such as Excel. Though some specific numerical assumptions will be made to provide a working example, the reader should focus on the nature of the information one could obtain from such methods rather than the specific numerical results that follow from the details of this example.

2. The Situational Awareness Problem

Suppose that the first human case in a bioterror attack has just been observed, and that the agent involved has been determined. As more human cases are reported, it is critical to ascertain answers to two questions: how big was the attack (i.e., how many persons were infected); and when did it happen. The answers to these questions can then be used to create forecasts of future cases, which in turn can be used for emergency planning (e.g., is there sufficient hospital capacity to handle expected future cases, or is there a need for adding surge capacity). Estimates of the time and size of a bioterror attack also figure prominently in determining when the resulting outbreak has ended (i.e., when the expected number of future cases is very small compared to the estimated size of the outbreak). Additional situational awareness questions include establishing the location and spatial spread of an outbreak. Also, while this chapter will focus on techniques derived from observing human cases over time, one certainly could extend these ideas to attacks that are detected via biosensors, syndromic surveillance, or other means.

3. An Anthrax Example and Attendant Assumptions

In the example to be developed, an anthrax attack infecting 100 persons has occurred. How will those infected in the attack reveal themselves over time? The key assumption employed is that the probability distribution of the *incubation time* from infection through the onset of symptoms is known.

Cases will thus appear in a manner consistent with this probability distribution. For anthrax, Brookmeyer et al. [10] studied the incubation time data associated with the Swerdlovsk accidental anthrax release and via maximum likelihood were able to fit a lognormal distribution with median 11 days and dispersion 2 days to the incubation time data associated with that outbreak. Figure 1 shows the lognormal probability density with the parameters stated above.

It is assumed that no interventions were launched prior to detecting the attack, which is reasonable: in the absence of any evidence of an attack, on what basis would one launch an intervention? However, while interventions might be launched once an attack has been detected (e.g., distribution of antibiotics to persons in the vicinity of the attack), it is further assumed that such interventions will not have any impact on the occurrence of cases during the first several days of the outbreak (as these cases would be among persons with sufficiently advanced infections to render intervention ineffective). Anthrax being a noncontagious agent, there will be no need to account for secondary transmission in our illustrative example. More complicated models allowing for secondary and continuing transmission (e.g., as would be the case with smallpox) can also be constructed, though such extensions will not be pursued in this chapter.

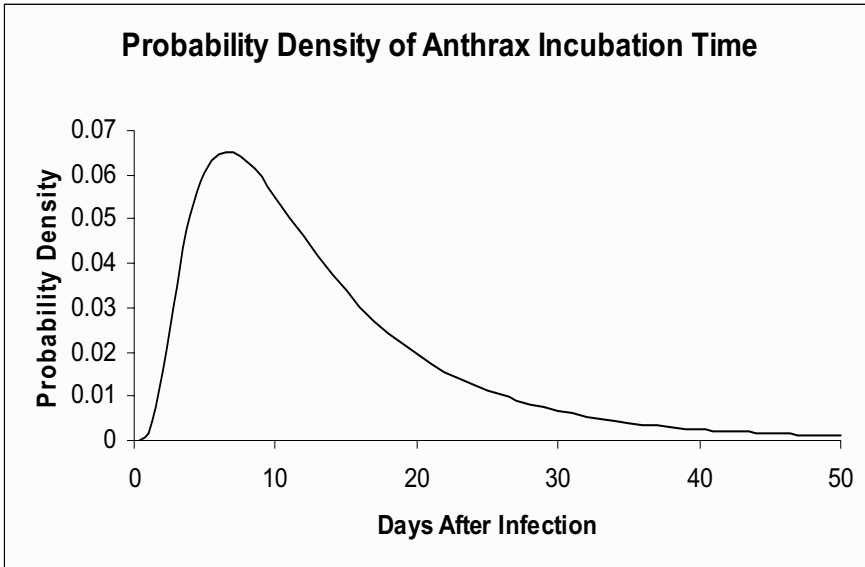


Figure 1. Probability density of anthrax incubation time.

4. Basic Ideas

The key insight that enables one to estimate the time and size of a bio-terror attack from case reporting data is this: if n persons were infected in the attack, then the time from the attack until the 1st, 2nd, 3rd, ..., n th cases occur would be the *order statistics* from the incubation time distribution [11]. In particular, the time from the attack to the first observed case would correspond to the *minimum* of n independent incubation times (where the independence reasonably assumes that the progression of infection in any one individual has no influence on the progression of infection in any other individual). Thus, if one knew the size of the attack, one could estimate the time of the attack by working backwards from the time of the first observed case.

Of course, the size of the attack is not known, but it is possible to quantify prior beliefs regarding the likelihood that attacks infecting different numbers of individuals could occur. A Bayesian approach formally accomplishes this by postulating a prior probability distribution governing the size of the attack. Any such distribution should be diffuse to express our considerable ignorance over the likelihood that any one of a wide range of different attack sizes would occur. An attractive choice that satisfies this criterion and is also easy to work with is given by the log-uniform distribution – that is, a distribution where the logarithm of the attack size is taken to be uniform between zero (an attack size of unity) and the largest order-of-magnitude considered reasonable for an attack of this sort (for recall that it is assumed that the agent used has been determined; e.g., anthrax). This amounts to order-of-magnitude uncertainty: the chance that an attack infects a number in the tens is the same as the chance that an attack infects a number in the hundreds, thousands, or ten thousands.

Once a prior distribution for the initial attack size has been assigned, it then becomes a straightforward probability exercise to, in real time as new case reports arrive, estimate the conditional (or posterior) probability distributions of the attack size and time via Bayes' rule, given the number and timing of observed cases to date. Once armed with these conditional distributions, further probabilistic calculations enable a forecast of the number of cases expected in the future, as well as computation of the probability that at least a certain percentage of all persons infected in the attack have already been observed (which is useful in deciding when the outbreak has essentially ended).

The formal probability arguments involved in this approach were presented in [9] and are summarized in the appendix of this chapter for the technical reader. To show how such a model could be used in an actual

bioterror attack, we return to our anthrax example and consider what might happen over the course of a single outbreak. Note that all of the calculations involved in this example were performed using an Excel spreadsheet, indicating the technical ease with which the models can be implemented [9].

5. A Simulated Example

Imagine that an anthrax attack occurs that infects 100 persons. The time from infection to symptoms follows the lognormal incubation time density of Figure 1. To obtain a specific example, 100 incubation times were randomly selected from this probability distribution and ordered to create the “actual” occurrence of cases over time (as reported on a daily basis) shown in Figure 2. Using the probability models proposed, it is possible to estimate the conditional probability distribution of the number infected in the attack given the cases observed to date since the detection of the attack. Figures 3–7 show this distribution having observed all cases through days 2, 5, 10, 20, and 50, respectively. From Figure 2, one can see that a total of 8, 25, 59, 80, and 100 cases were observed as of these dates. Note how the uncertainty in the attack size distribution disappears over time as more cases are reported, that is, how the probability densities increasingly place most of their weight on attack sizes in the vicinity of (the true value of) 100 as more information is collected. This illustrates how the model is able to “learn” from the data (and via the incubation time distribution) over time.

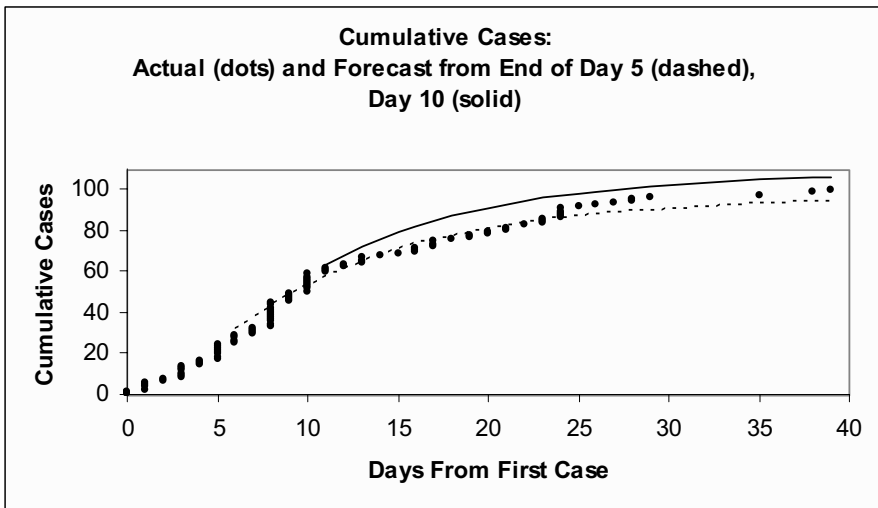


Figure 2. Cumulative actual and forecasted anthrax cases over time in the simulated outbreak.

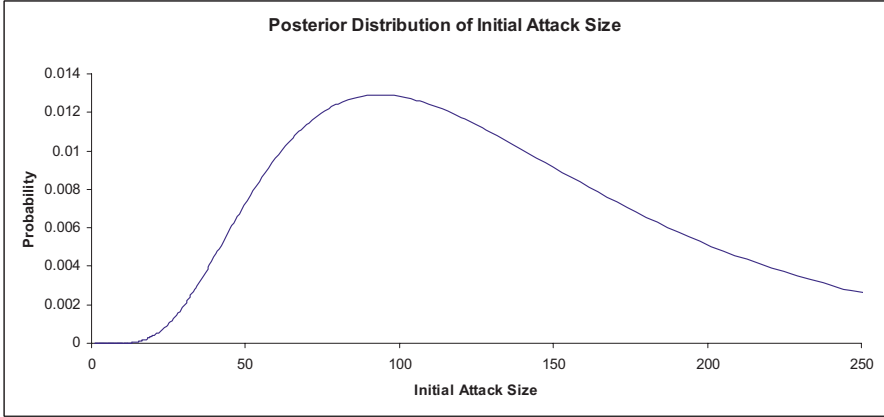


Figure 3. Attack size probability distribution 2 days after seeing the first case (Eight cases observed in total).

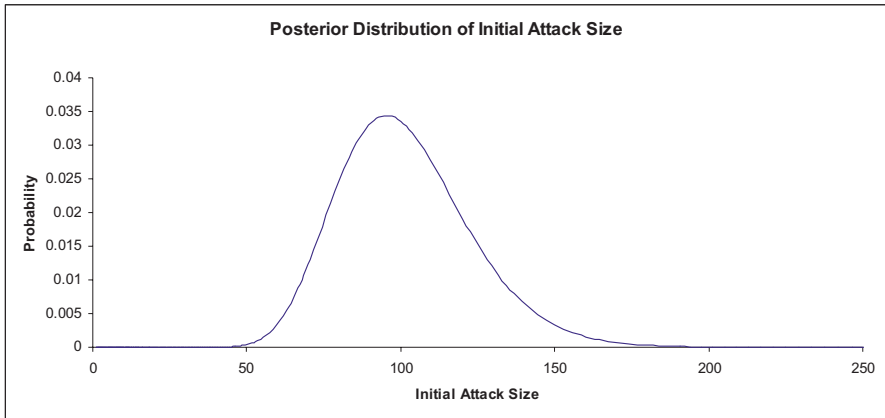


Figure 4. Attack size probability distribution 5 days after seeing the first case (25 cases observed in total).

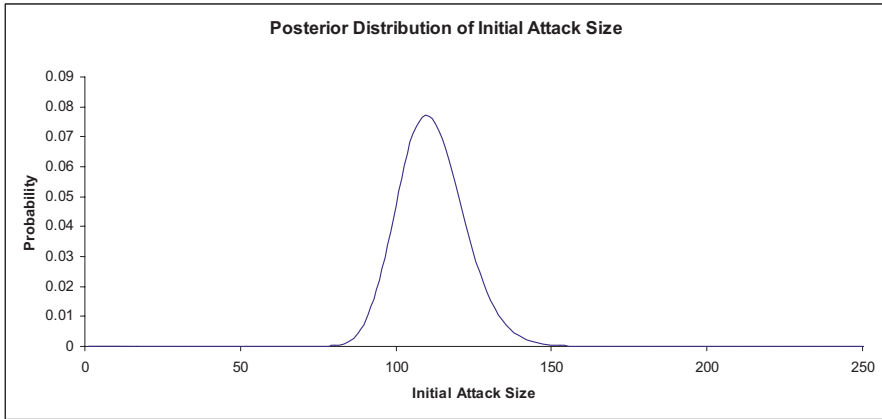


Figure 5. Attack size probability distribution 10 days after seeing the first case (59 cases observed in total).

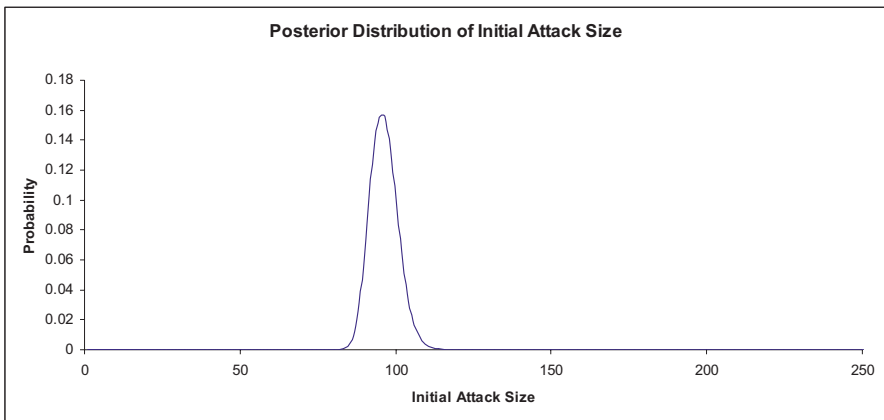


Figure 6. Attack size probability distribution 20 days after seeing the first case (80 cases observed in total).

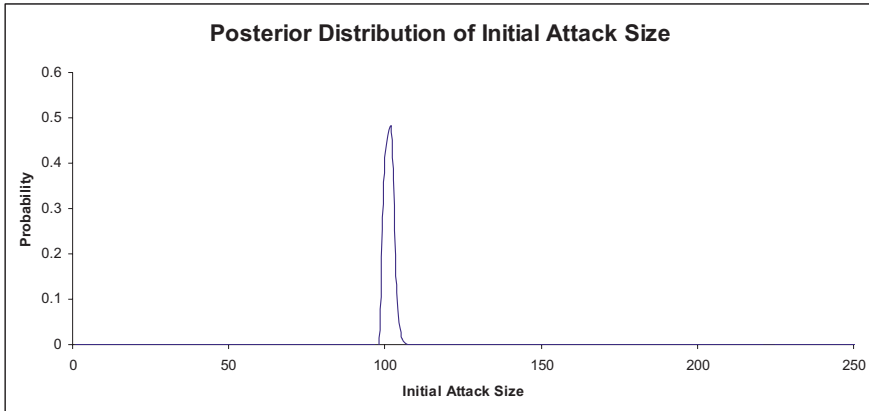


Figure 7. Attack size probability distribution 50 days after seeing the first case (100 cases observed in total).

While Figures 3–7 reveal that considerable “learning” occurs over the course of this simulated outbreak, it is interesting that the mean attack size (a point estimate as it were) computed from these distributions as new information arrives over time is quite stable. Figure 8 plots the mean attack size as estimated over time from the updated probability distributions, and shows that the point estimates obtained do indeed hover closely about the true value of 100. One would expect accurate estimates once most of those infected in the attack have progressed to symptoms, but Figure 8 shows that good results are obtained much earlier.

Considerably less information is gained regarding the timing of the attack. Figures 9 and 10 report the probability distribution of the time of the attack as measured in days prior to observing the first case after days 2 and 50 of the outbreak. In both cases the bulk of the probability locates the attack to have occurred between 1 and 3 days before the first case was observed. The information gained from day 2 to day 50 essentially rules out the possibility that the attack occurred more than 3 days prior to the date of the first case.

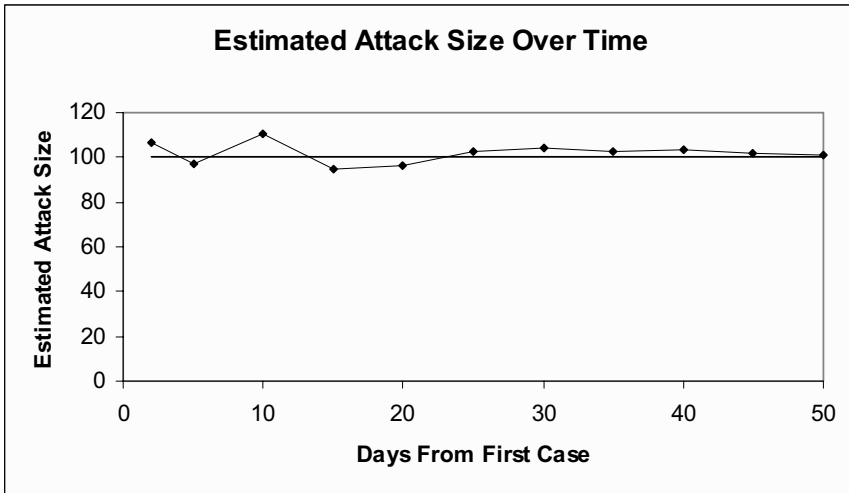


Figure 8. Expected (mean) attack size as estimated over time from detection of the outbreak from the first observed case.

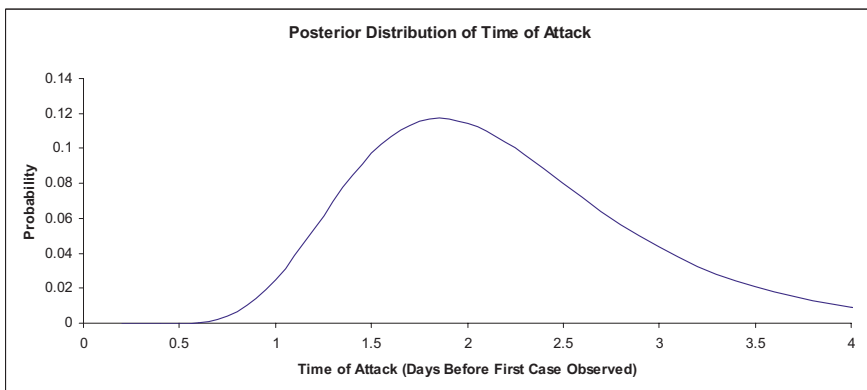


Figure 9. Probability distribution of the time of attack (measured as days before observing the first case) 2 days after seeing the first case (Eight cases observed in total).

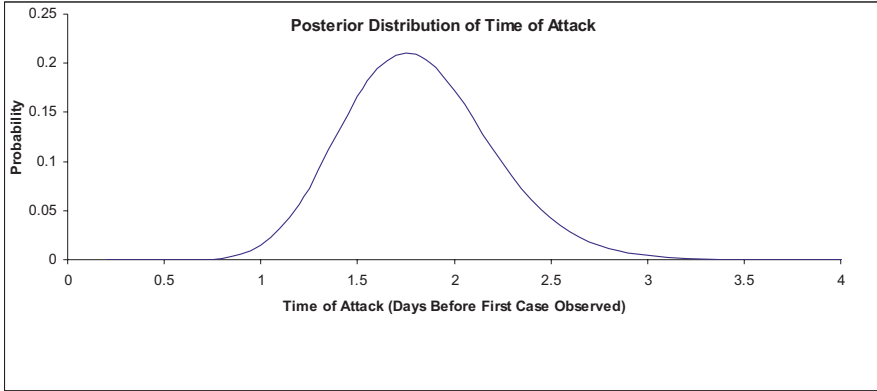


Figure 10. Probability distribution of the time of attack (measured as days before observing the first case) 50 days after seeing the first case (100 cases observed in total).

6. Forecasting Future Cases

Armed with the updated probability distributions of the attack size and time at any point during the outbreak, it is straightforward to forecast the occurrence of future cases. Figure 2 reports two such forecasts along with the simulated number of cases observed over time. The forecasts created 5 days after the attack appeared as a dashed curve, while the solid curve presents forecasts created 10 days post-detection. Both these forecasts provide sufficiently accurate pictures of downstream cases of infection to be useful for planning purposes. That the forecasts from day 5 are more accurate than the forecasts from day 10 (at least through the first 24 days of the outbreak) is specific to the example simulated and should not be interpreted as meaning that earlier forecasts will outperform later ones in general. The key point, however, is that the forecasts provide an early view of how those infected in the attack will reveal themselves over time, enabling appropriate planning for care and treatment to ensue early in the outbreak.

7. Is the Outbreak Over?

Thus far the focus has been on estimating the time and size of a bioterror attack in order to predict future caseloads. However, whether or not to declare that an outbreak has ended is also a crucial decision. Brookmeyer and You [12] proposed a statistical hypothesis test based on the spacing between observed cases for deciding whether an outbreak has ended. An alternative approach follows directly from the Bayesian principles employed

in the present chapter. Specifically, one can calculate the probability that at least $100\alpha\%$ of *all* cases have been observed at any point during the outbreak (and for any value of α). For example, one can compute the probability that at least half of all cases have already occurred as of the date of computation, or if at least 80%, 95%, or even 100% of all cases have occurred. The probability that at least $100\alpha\%$ of all cases have been observed is referred to as the α -level outbreak completion probability.

Figure 11 plots α -level outbreak completion probabilities for $\alpha = 0.5, 0.75, 0.95,$ and 1 for the data in our simulated anthrax attack. Note that one is about 80% certain that at least half of all cases have occurred after day 10 of the outbreak (and indeed, 59 cases already occurred by day 10, which is more than 50% of 100, the true number infected in the outbreak). It of course takes longer for the outbreak completion probabilities to approach unity as α increases; it requires 80 days and 100 observed cases (which in the example constitute *all* of the cases that could occur) to be 95% certain that the outbreak is over (i.e., the α -level outbreak completion probability equals 95% for $\alpha = 1$ only after 80 days). The method is thus appropriately conservative when asked if the outbreak is truly over.

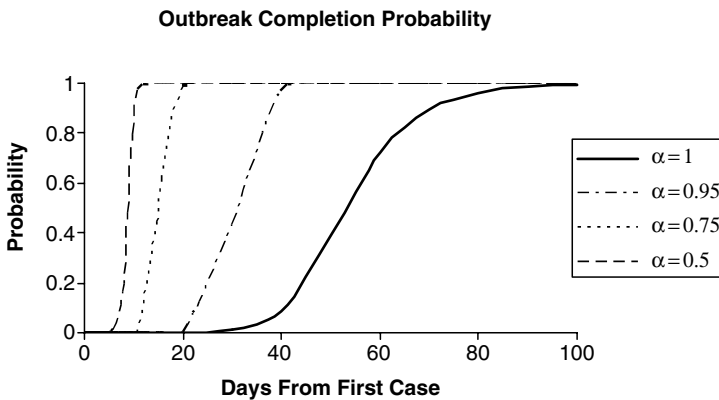


Figure 11. α -level outbreak completion probabilities over time.

8. Summary

This chapter has provided a simple example showing how one can use the probability distribution governing the incubation time from infection through symptoms for a bioterror agent with observed cases in real time to estimate the time and size of a bioterror attack, forecast the number of downstream cases for planning purposes, and assess over time the extent to

which the outbreak is totally or partially complete. The computations required can be carried out in an Excel spreadsheet [9], as was all of the computing for the example presented.

While it is hoped that that this chapter serves to convince the reader of the value of probability modeling in bioterror preparedness and response, the methods and example presented are not the final word on this subject. This chapter has only concerned itself with a noncontagious bioterror agent such as anthrax. Attack with a contagious agent such as smallpox, plague, or tularemia would require additional modeling to account for continuing transmission from those initially infected; deconvolution methods similar to the back calculation of human immunodeficiency virus (HIV) incidence from acquired immunodeficiency syndrome (AIDS) reporting data would be required [13]. The approach has also assumed that the incubation time probability distribution is known, but while information is available for many feared bioterror agents [14], it might be that for either new or modified viral or bacterial infections, the data observed in the outbreak itself becomes the best information describing incubation times (as was the case with the accidental release of anthrax in Swerdlovsk [10]). In this instance, it would be possible to augment the modeling approach suggested to account for uncertainty in the incubation time distribution, and to statistically update this distribution and hence learn in real time as those infected present themselves as cases. However, although it is not difficult in principle, loss of precision governing the incubation time distribution would translate to considerable loss in precision regarding the forecasts of downstream cases.

Finally, the model proposed based all inferences over time on observed cases of infection, yet as noted earlier there are several proposals, such as the use of biosensors and/or syndromic surveillance systems meant to speed the detection and general assessment of bioterror events. It remains to integrate models of the form reported in this chapter with those other more elaborate detection and reporting systems; doing so would be a major advance in the provision of decision support to those concerned with managing the consequences of a bioterror attack.

References

1. Kaplan EH, Craft DL, Wein LM. Emergency response to a smallpox attack: the case for mass vaccination. *Proc Natl Acad Sci USA* 2002; 99:10935–10940.
2. Halloran ME, Longini IM, Nizam A, Yang Y. Containing bioterrorist smallpox. *Science* 2003; 298:1428–1432.
3. Wein LM, Craft DL, Kaplan EH. Emergency response to an anthrax attack. *Proc Natl Acad Sci USA* 2003; 100:4346–4351.

4. Gani R, Leach S. Epidemiologic determinants for modeling pneumonic plague outbreaks. *Emerg Infect Dis* 2004; 10:608–614.
5. Wein LM, Liu Y. Analyzing a bioterror attack on the food supply: the case of botulinum toxin in milk. *Proc Natl Acad Sci USA* 2005; 102:9984–9989.
6. Estacio, PL. Bio-Watch Overview. Washington, DC: Department of Homeland Security, 2004. Available at: <https://www.nleetc.org/training/nij2004/estacio.pdf>
7. Das D, Metzger K, Heffernan R, Balter S, Weiss D, Mostashari F. New York City Department of Health and Mental Hygiene. Monitoring over-the-counter medication sales for early detection of disease outbreaks – New York City. *MMWR Morb Mortal Wkly Rep* 2005; 54(Suppl):41–46.
8. Kaplan EH, Patton CA, FitzGerald WP, Wein LM. Detecting bioterror attacks by screening blood donors: a best-case analysis. *Emerg Infect Dis* 2003; 9:909–914.
9. Walden J, Kaplan EH. Estimating time and size of bioterror attack. *Emerg Infect Dis* 2004; 10:1202–1205.
10. Brookmeyer R, Blades N, Hugh-Jones M, Henderson DA. The statistical analysis of truncated data: application to the Swerdlovsk anthrax outbreak. *Biostatistics* 2001; 2:233–247.
11. David HA. Order statistics. New York: Wiley, 1970.
12. Brookmeyer R, You X. A hypothesis test for the end of a common source outbreak. *Biometrics* 2005 (online early, doi:10.1111/j.1541-0420.2005.00421.x).
13. Brookmeyer R, Gail MH. *AIDS Epidemiology: A Quantitative Approach*. New York: Oxford University Press, 1994.
14. Centers for Disease Control and Prevention. *Biological agents/diseases: category A*. Atlanta (GA): The Centers, 2003. Available at: <http://www.bt.cdc.gov/agent/agentlist-category.asp#a>

Appendix

This appendix summarizes the formulas used in the models of this chapter. For greater detail please consult [9]. The analysis centers on two random variables, the number infected in the attack N , and the time of the attack (measured as time before the first observed case) A . The probability distribution of the incubation time from infection through symptoms is assumed known, and in particular the probability that an incubation time exceeds x is denoted by the survivor distribution $S(x)$.

Suppose that the attack size $N = n$. Since the attack is detected by the first observed case, the probability density of the time of attack given $N = n$ is given by

$$f(a | n) = nS(a)^{n-1} f(a) \text{ for } a > 0, n = 1, 2, 3, \dots \quad (1)$$

If the prior probability distribution for the attack size is given by $p(n) = \Pr\{N = n\}$, then the joint probability distribution of the attack time and size is given by

$$f(a, n) = p(n) \times nS(a)^{n-1} f(a) \text{ for } a > 0, n = 1, 2, 3, \dots \quad (2)$$

Now, suppose that by time τ , an additional $k-1$ cases have been observed beyond the initial case observed at time 0 that signaled the outbreak. Conditional on the attack size $N = n$ and the attack time $A = a$ (days before observing the first case), the probability of observing these additional cases at times $t_2, t_3 \dots t_k \leq \tau$ (where $t_1 = 0$ denotes the time at which the first case is observed) is given by

$$\mathbf{L}(\mathbf{t} | a, n, \tau) = \frac{(n-1)!}{(n-k)!} \left\{ \frac{S(a+\tau)}{S(a)} \right\}^{n-k} \prod_{j=2}^k \left\{ \frac{f(a+t_j)}{S(a)} \right\} \quad (3)$$

The joint probability of observing $N = n$, $A = a$, and the data actually seen then equals

$$J(a, n, \mathbf{t} | \tau) = p(n) \frac{n!}{(n-k)!} S(a+\tau)^{n-k} \prod_{j=1}^k f(a+t_j) \text{ for } a > 0, n = k, k+1, k+2, \dots \quad (4)$$

Bayes' rule then implies that the conditional probability distribution of the attack time and size given the data observed up to time τ after the attack was detected is given by

$$f(a, n | \mathbf{t}, \tau) = \frac{p(n) \frac{n!}{(n-k)!} S(a+\tau)^{n-k} \prod_{j=1}^k f(a+t_j)}{\sum_{i=k}^{\infty} \int_{u=0}^{\infty} p(i) \frac{i!}{(i-k)!} S(u+\tau)^{i-k} \prod_{j=1}^k f(u+t_j) du} \text{ for } a > 0, n = k, k+1, \dots \quad (5)$$

All of the quantities of interest can be computed from Equation (5). For example the α -outbreak completion probabilities can be computed as

$$\Pr\left\{\frac{k}{N} \geq \alpha \mid \mathbf{t}, \tau\right\} = \sum_{n=k}^{\lfloor k/\alpha \rfloor} \int_{a=0}^{\infty} f(a, n | \mathbf{t}, \tau) da \text{ for } 0 \leq \alpha < 1 \quad (6)$$

Details for computing forecasts of future cases and other quantities can be found in [9].