SANDIA REPORT

SAND2006-ABCD
Unlimited Release
Printed February 2006

A Bayesian method for characterizing distributed micro-releases: I. The single-source case for non-contagious diseases

J. Ray, Y. Marzouk and H. N. Najm,
Sandia National Laboratories, CA

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online ordering:	http://www.ntis.gov/ordering.htm

SAND2006-ABCD
Unlimited Release
Printed February 2006

A Bayesian method for characterizing distributed micro-releases: I. The single-source case for non-contagious diseases

J. Ray
Advanced Software R. & D.,
Sandia National Laboratories
P. O. Box 969, Livermore CA 94550
jairay@ca.sandia.gov

Y. Marzouk and H. N. Najm
Reacting Flow Research
Sandia National Laboratories
P. O. Box 969, Livermore CA 94550
{ymarzou,hnnajm}@sandia.gov

M. Kraus
Analysis Directorate, NORAD-NORTHCOM
250 S. Peterson Blvd, Peterson AFB, CO 80914-3180
Mark.Kraus.ctr@northcom.mil

P. Fast
Center for Advanced Scientific Computing
Lawrence Livermore National Laboratory, Livermore CA 94550
pfast1@llnl.gov

Abstract

We present results from a preliminary investigation into the use of Bayesian inference methods to characterize the genesis of epidemics, caused by natural and unnatural means, based solely on a short (3-4 days) time-series of diagnosed patients exhibiting symptoms. In particular, we infer the number of index cases, the time of infection and the dosage of the pathogen, under the following assumptions: (1) the epidemic has a single-focus epidemic (2) all the infected people receive the same dosage and (3) the effect of contact spread of the disease is negligible in the data used for the inference. The estimates of the parameters are developed as PDFs (probability density functions). Smallpox and anthrax are used as the pathogens in this study. The methods are tested against simulated epidemics as well as the anthrax outbreak of Sverdlovsk in 1979.

Acknowledgment

This work was supported Sandia National Laboratories' LDRD (Laboratory Directed Research and Development) funds, sponsored by the Computational and Information Sciences IAT.

Contents

Appendix

Figures

Tables

A Bayesian method for characterizing distributed micro-releases: I. The single-source case for non-contagious diseases

1 Introduction

Pathogens may be introduced into a population via natural or unnatural means. Naturally, they may infect a person via contact between an infected human or, in some cases, domesticated animals. Alternatively, they may be introduced via an atmospheric release of an aerosol preparation in a military or criminal attack - large bio-warfare programs existed in both the Soviet and NATO countries before the 1972 Biological Warfare Convention [1, 2] . In many cases, such introductions may not be detected via environmental sensors - examples include small releases which do not travel far, releases where the formulation is coarse (and heavy) enough to precipitate quickly and releases in areas which are not well instrumented with sensors. In such cases, the first intimation of the attack will be the definitive diagnosis of the first patient - but by then the pathogen may have established itself in the population. Thus the ability to infer the characteristics of the introduction plays an important part in formulating a medical response. The characteristics can also serve as the initial condition for various epidemic models which can then be used to predict the evolution and spread of the disease in a population as well as its ramifications on the society.

The two cases – natural and unnatural introduction – place slightly different requirements of the inference process. In case of a natural epidemic (of a communicable disease), the initial infecteds (the index cases) are expected to be few. If the disease has a contagious asymptomatic phase (e.g. H5N1, "bird flu"), the relevant question, when a definitive diagnosis is made, is how many infected people (very few of them would be index cases) exist in the population. In case of an aggressive attack (which we will refer to henceforth as a bioterror attack or BT attack), the relevant questions involve estimating the number of index cases, the time of the attack as well as the doses received by the infected people. The reasons for these different requirements stem from the pathogens of choice for BT attacks as well as the mechanisms of release. BT attacks are meant to infect as large a population as possible, and the ratio of index cases to people who "caught" the disease by contact will not be less than one, at least in the early days of the epidemic when such inferences are of operational importance. Also most BT pathogens (e.g. smallpox, anthrax, plague etc [1]) are either non-contagious, act slowly or have symptomatic contagious phases whereby the contagious person can be quarantined. The time of the attack is critical to determining the location of the attack by tracing the movement of the diagnosed patients back to the estimated attack time. The distribution of doses among the infected population is determined by the mode of release (indoor versus outdoor) - it varies inversely with distance from the point of release, and in an indoor context, is affected by circulation. Further, BT attacks could be carried out as a set of small releases, staggered in time and space - the small size reduces detectability and allows low-quality formulations to be used. In such a case, one would have to infer the number of distributed attacks

and the size, time and dose-distributions corresponding to each of them.

Drawing these inferences can be challenging - the only two phenomena that can be exploited for this purpose are (1) the distribution of incubation period of the disease, which in some cases is dependent on the dose received and (2) the "mixture" of index and non-index cases in the diagnosed patients. To be relevant in an operational, consequence management sense (as opposed to forensics), these inferences have to be drawn early in the epidemic; thus a time-series of patient data, 3-4 days long, should be considered the norm. Thus the inferences are expected to be rather inaccurate / uncertain and quantifying this uncertainty becomes an inherent requirement of the inference process. This uncertainty can be reduced if one has prior beliefs (gathered via processes that are independent of the epidemic) regarding the values of the quantities being evaluated. Including them into the inference approach in a mathematically rigorous way contributes to its robustness.

In this report, we explore how such BT attacks may be characterized. This preliminary study targets single-focus attacks (i.e. single release) where we also assume that (1) the disease is non-contagious or the delay in the outbreak of secondary cases is long enough that only the index cases are observed in the period of interest and (2) the dose received by the index cases is a constant. We study how the inferences of the size, time and dosage behave with the size of the attack and the time-resolution of the observed data i.e. if the patient data were to be collected in 6-hour intervals instead of 24 hours, would the inferences be substantially more accurate? Smallpox and anthrax are used in the study. We adopt a Bayesian inference approach since (1) it allows us to develop the inferred quantities as PDFs (probability density functions), thus quantifying the uncertainty and (2) it allows a straightforward accommodation of prior beliefs using Bayesian *priors*. We will essentially study the suitability of Bayesian methods to deliver estimates within our self-imposed limit of 4 days (of the patient time-series) and investigate how these may be improved e.g. with better resolved data as well as possible causes of inaccuracy e.g. noisy data. The results of this study will indicate whether more detailed questions (e.g. dose-distributions, multiple attacks etc) can be answered satisfactorily by such an approach as well as help identify a few promising approaches to enhance the method's efficiency.

2 Prior Work

Prior work on the topic of inferring the characteristics of an attack are best described under two headings

1. the problem of inference and
2. the model of the disease used for inference. In our case, we will limit the discussion to smallpox and anthrax.

The exact question of estimating the size and time of an attack from a time-series of patient data seems to have been studied little. Walden and Kaplan [3] developed a Bayesian formulation which they tested on a low dose anthrax attack corresponding, roughly, to the Sverdlovsk outbreak [4]

of 1979, using an incubation period model by Brookmeyer et al [5]. The formulation is also valid for communicable diseases as long as they have a long, non-contagious incubation period e.g. smallpox. They also demonstrated the use of Bayesian *priors* - prior belief in the size of the attack N - to develop a smooth PDF for N, even for a small infected population ($N = 100$) and a time-series, 5 days long, with data collected on a day-by-day basis.

An alternative (likelihood maximization) method was employed by Brookmeyer and Blades [6] to infer the size of attacks in the 2001 anthrax attacks in the US [7], preparatory to estimating the reduction of casualties by the timely administration of antibiotics. The inference process was difficult given the small number of symptomatic patients (10 infectees in 3 separate attacks). They also used the incubation period model from [5].

All inference studies need a model of the disease in question - in the studies described above, a model of the incubation period sufficed since contact spread of the disease was not an issue (anthrax is not contagious) or that the long, non-contagious incubation period guaranteed that the second generation of infectees would appear long after the first symptomatic patient (e.g. smallpox). A significant amount of work has been done on the incubation period of anthrax. Brookmeyer et al [5] developed a low-dose incubation period model applicable to the Sverdlovsk outbreak; their more recent work, based on a competing risks formulation, includes dose-dependence [8]. A more empirical approach, but based on significantly more data, was done recently by Wilkening [9], where he also compared four different models, including the dose-dependent one by Brookmeyer [8] (referred to as Model D). While Wilkening's Model A agreed with Model D at the high-dose limit, their low-dose behavior was different.

The question of whether a person exposed to a number of spores will actually contract the disease is a separate question that will not be addressed in the report. The problem of estimating the probability of infection from S spores was addressed in [8] as well as by Glassman [10].

The modeling of smallpox poses very different challenges because of the lack of a good non-human primate model for experimental studies. Recently, Jahrling et al [11] have reproduced human-like response in cynomolgus macaques using immensely high doses of variola major; previous studies [12] were largely qualitative or statistically insufficient. In the absence of such studies, almost all our information comes from the records collected during the WHO Eradication Program [13, 14]. These are strictly low-dose behaviors. Meltzer et al [15] have collated this data into a form amenable to modeling. Eubank's detailed simulations of a smallpox epidemic [16] makes use of the data in [13, 14] and is probably the most sophisticated to date. For the purposes of this study, we will construct a "consensus" model in Appendix A based largely on [15, 11]. We will not model the various clinical forms of smallpox [12] in the interest of simplicity; further, there is no data based on which we can segregate and separately model the various clinical forms.

3 Formulation of the problem

Consider an attack at time τ where N people are infected, with each of the N people receiving the same dose D of the pathogen. N, D and τ are unknown. Assume, too, that the attack is not caught on environmental sensors. Also assume that none of the data being used in the inference process was generated due to contact spread of the disease i.e. all the patients are index cases.

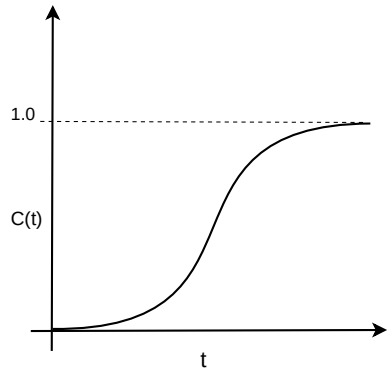

Figure 1. CDF of an incubation period distribution.

Phenomenologically, the N people will incubate the disease. Incubation period follows a distribution that is dose-dependent and after a period of time, the outliers in the population will become symptomatic. They will be diagnosed in the hospitals and the date of showing symptoms (usually a day) will be recorded. So for a few days M (say 3 - 5 days) we can expect

1. a series $t_i, i = 0 \ldots M$ of times, perhaps the end points 24-hr intervals, when patients' symptoms are observed
2. the time-series $n_i, i = 0 \ldots M$ of new patients who turned symptomatic between t_{i-1} and t_i

The incubation period of a disease follows a strict distribution, often log-normal [17] (see Fig. 1). Some people come out of incubation and show symptoms early, others later. If N people are infected, the number of people showing symptoms will vary over time, approximately similarly to the CDF (exactly in the limit as $N \to \infty$). So, on a person-by-person basis, this can be modeled as

1. each person, after infection, has a probability of showing symptoms, $C(t)$, which asymptotically approaches 1 as $t \to \infty$. $C(t)$ is the cumulative distribution function or CDF.
2. its derivative $P(t) = \frac{dC}{dt}$ is the probability density function, PDF. It can be derived (e.g. from an experimental study) by plotting the histogram of the number of new symptomatics per a given time interval (e.g. 24 hours).

Thus,

- A person infected at time $t = 0$ with dose D will show symptoms by time t with probability $C(t, D)$.
- the corresponding survival probability is $P_{surv}(t, D) = 1 - C(t, D)$.

We can state the problem as such: Given a time series $(t_i, n_i), i = 0 \ldots M$ of patients showing symptoms over a few days M, estimate (N, τ, D) from the data. n_i patients are assumed to have developed symptoms over the time interval between t_{i-1} and t_i.

Let a total of L patients develop symptoms over M days i.e. $L = \sum_{i=0}^{M} n_i$.

Thus the number of infectees yet to show symptoms is $N - L$

The probability that $(N-L)$ infectees will not exhibit symptoms by day M is $\{P_{surv}(t_M - \tau, D)\}^{N-L}$

The probability that n_i patients will exhibit symptoms between time $t_i - \Delta t$ and t_i, (where $\Delta t = t_i - t_{i-1}, i \neq 0$ is the length of the data collection interval) is $\{C(t_i - \tau, D) - C(t_i - \Delta t - \tau, D)\}^{n_i}$

The number of ways L people can be chosen out of a total of N infectees is

$$\frac{N!}{L!(N-L)!}$$

The number of ways L people can be divided into the sequence $n_0, n_1, \ldots n_M$ is

$$\frac{L!}{n_0! n_1! \ldots n_M!}$$

Therefore, the likelihood of observing a $\{t_i, n_i\}, i = 0 \ldots M$ series given a (N, τ, D) attack is

$$\begin{aligned}
\mathcal{L}(\{t_i, n_i\}, i = 0, \ldots M | \{N, \tau, D\}) &= \frac{N!}{L!(N-L)!} \times \frac{L!}{n_0! n_1! \ldots n_M!} \times \{P_{surv}(t_M - \tau, D)\}^{N-L} \times \\
&\quad \prod_{i=0}^{M} \{C(t_i - \tau, D) - C(t_{i-1} - \tau, D)\}^{n_i} \\
&= \frac{N!}{(N-L)! \prod_{i=0}^{M} n_i!} \times \{P_{surv}(t_M - \tau, D)\}^{N-L} \times \\
&\quad \prod_{i=0}^{M} \{C(t_i - \tau, D) - C(t_{i-1} - \tau, D)\}^{n_i} \qquad (1)
\end{aligned}$$

The likelihood in Eq. 1 now needs to be incorporated into an expression that allows inference. We start with Bayes' rule [18]. Assume that there exists a stochastic forward model that, given an input $\mathbf{x}$, produces an output $\mathbf{y}$. Since the model is stochastic and the output $\mathbf{y}$ is not guaranteed, one considers the probability of observing $\mathbf{y}$ given $\mathbf{x}$ i.e $\Pi(\mathbf{y}|\mathbf{x})$. Let the probability of having $\mathbf{x}$ as an input itself be $\Pi(\mathbf{x})$ and $\mathbf{y}$ as an output be $\Pi(\mathbf{y})$. Then by Bayes rule

$$\Pi(\mathbf{x}|\mathbf{y})\Pi(\mathbf{y}) = \Pi(\mathbf{y}|\mathbf{x})\Pi(\mathbf{x}) \qquad (2)$$

In our case the time series (t_i, n_i) forms the output/observables $\mathbf{y}$ while $\mathbf{x} = (N, \tau, D)$. We are interested in guaging $\Pi(\mathbf{x}|\mathbf{y})$. Eq. 1 gives $\Pi(\mathbf{y}|\mathbf{x})$. We construct $\Pi(\mathbf{x})$ by exploiting the fact that N, τ and D are uncorrelated with each other and the PDFs expressing their distributions can simply be multiplied to give

$$\Pi(\mathbf{x}) = \pi_N(N)\pi_\tau(\tau)\pi_D(D)$$

where $\pi_N(N)$, $\pi_\tau(\tau)$ and $\pi_D(D)$ are PDFs encompassing prior beliefs in N, τ and D. Such an approach is flexible - a near-certainty regarding the value of a certain parameter e.g. $N = N_0$ can be incorporated using $\pi_N(N) = \delta(N - N_0)$ while a belief that $N_S \leq N \leq N_S$ can be incorporated as a uniform distribution $\pi_N(N) = \mathcal{U}(N_S, N_E)$. The expression for the inferred quantities (N, τ, D), given the observed data $\{t_i, n_i\}, i = 0, \ldots M$, is

$$\begin{aligned} \Pi(\{N,\tau,D\}|\{t_i,n_i\}, i=0,\ldots M) &= \frac{\mathcal{L}(\{t_i,n_i\}, i=0,\ldots M|\{N,\tau,D\})\pi_N(N)\pi_\tau(\tau)\pi_D(D)}{\Pi(\{t_i,n_i\}, i=0,\ldots M)} \\ &= \Pi(\mathbf{x}|\mathbf{y}) \end{aligned} \tag{3}$$

For the rest of the proposal, we will use uniform distributions with finite supports for (N, τ, D), and probe the entire (N, τ, D) space within the supports to determine promising values for (N, τ, D). We do this in a brute-force manner by specifying the starting and ending values for (N, τ, D) as well as increments and exhaustively evaluating $\Pi(N, \tau, D|\{t_i, n_i\}, i = 0, \ldots M)$ in the space. These values are then normalized to get $\bar{P}(N, \tau, D)$

$$\bar{P}(N,\tau,D) = \frac{\Pi(N,\tau,D|\{t_i,n_i\}, i=0,\ldots M)}{\int_{N=N_S}^{N=N_E}\int_{\tau=\tau_S}^{\tau=\tau_E}\int_{D=D_S}^{D=D_E}\Pi(N,\tau,D|\{t_i,n_i\}, i=0,\ldots M)dNd\tau dD} \tag{4}$$

The subscripts S and E refer to the starting and ending values of the various attack characteristics. The normalized probability $\bar{\Pi}$ can then be marginalized to provide PDFs for the attack characteristics:

$$\begin{aligned} g(N) &= \int_\tau\int_D \bar{P}(N,\tau,D)dDd\tau \\ g(\tau) &= \int_N\int_D \bar{P}(N,\tau,D)dDdN, \\ g(D) &= \int_\tau\int_N \bar{P}(N,\tau,D)dNd\tau \end{aligned} \tag{5}$$

In order to complete the formulation, we need to state $C(t, D)$ for smallpox and anthrax. Anthrax is modeled with a log-normal incubation period with median and standard deviation dependent on the dose [9]. Using $C_A(t, D)$ as the CDF for anthrax

$$C_A(t,D) = \frac{1}{2}\left[1 + erf\left(\frac{\ln(t/t_0)}{\sqrt{2}S}\right)\right], \;\; t_0 = 10.3 - 1.35\log_{10}(D), \;\; S = 0.804 - 0.079\log_{10}(D) \tag{6}$$

The CDF $C_S(t, D)$ for smallpox, which is modeled as a normal distribution and does not include dose-dependence (Appendix A), is

$$C_S(t,D) = \frac{1}{2}\left[1 + erf\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)\right], \;\; \mu = 15, \;\; \sigma = 2 \tag{7}$$

In the next section, we will evaluate the utility of Eq. 5 for inferring the characteristics of an attack given a small time-series. We will evaluate how it behaves for small and large N and D and how long a time-series is required for meaningful inferences. It is expected that large N and D should help inferences (large statistical populations and reduced variability). We will also investigate the possibility of lengthening the time-series by collecting data (symptomatic patients) on a finer-grained time resolution e.g. 6 hour intervals instead of 24 hours.

We will simulate BT attacks using the incubation period in Eq. 6 and the smallpox model in Appendix A. For small smallpox attacks, we can expect a few second generation infectees to be included in the data for inferences, thus invalidating one of our assumptions. This can be used to test the robustness of the inference approach.

4 Test cases

In this section we will investigate the utility of the formulation in Sec. 3 in inferring attack characteristics. We will investigate the effect of the size of the infected population and the effect of having a fine-grained time-series, while following our restriction of a small observation duration of 3-4 days. We will investigate smallpox and anthrax (via simulations) and address the Sverdlovsk outbreak [4] of 1979.

4.1 Smallpox attacks

A model of smallpox is described in Appendix A. Smallpox has a long incubation time. For a large attack, the first 3-4 days of the epidemic will yield symptomatic patients who are almost certainly the index cases; in small attacks, a few second generation patients may also appear. Since small attacks are unlikely to yield copious fresh cases every day, even a few second generation cases can seriously derail the inference process. This is a weakness of the current model which depends entirely on the incubation period and ignores all processes related to contact spread of the disease.

In Table 1 we present data on new patients appearing on a day by day basis. Case C, the smallest attack, has some second generation cases on Days 4 and 5. These results were generated using the "non-deterministic" method described in Appendix A. These stochastic runs will generate different time-series if run with a different stream of random numbers; however, these differences *should not change the inference*. This independence of the inferences from the particular realization we consider is critical to its applicability in real-life scenarios.

In Fig. 2 we plot $\bar{P}(N,\tau)$ as given by Eq. 4 on Day 3 for the Case A in Table 1. Marginalization in each direction provides the PDFs for N and τ. The development of the PDFs over time is also shown in Fig. 2 (below); by day 2 we get a multimodal distribution for N which, had a good prior been available, could have significantly simplified the inference - the peaks are quite distinct. PDFs for day 3 and 4 still have an ambiguity (multiple peaks) though by day 5, the PDF is quite unambiguous. However, by day 3 we are within an order of magnitude of the correct answer. No

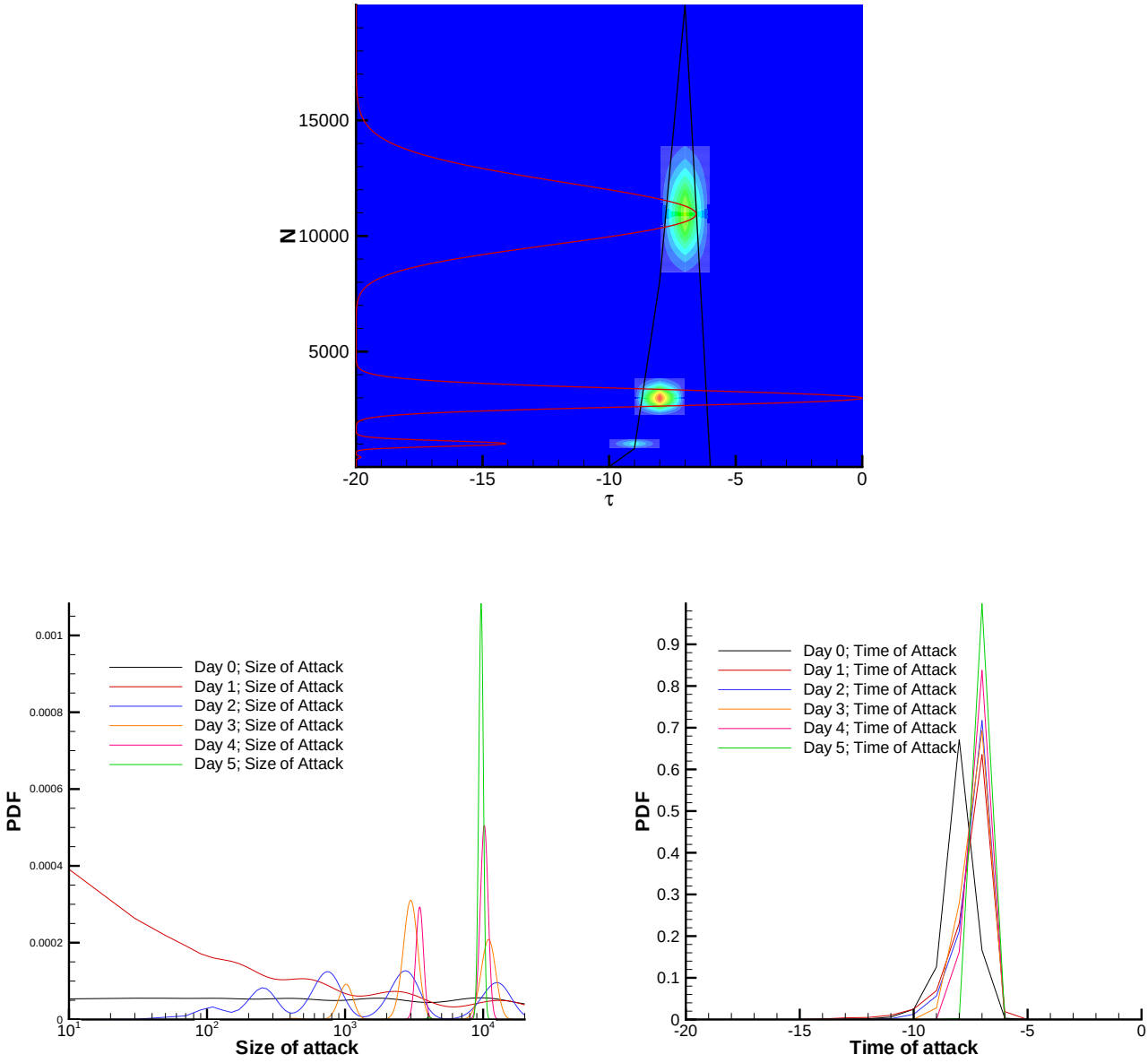

Figure 2. Above: The plot of $\bar{P}(N,\tau)$ as given by Eq. 4 on day 3 for the Case A in Table 1. Blue indicates $P = 0$; red maps to large numbers (though, of course, less than 1). The inlaid plots of PDFs of τ and N show the marginalization at work. Below: The PDFs of N, the size of the attack and τ, the time of attack developing as more data becomes available. The domain $-20 \leq \tau \leq 0$ and $10 \leq N \leq 20000$ was probed.

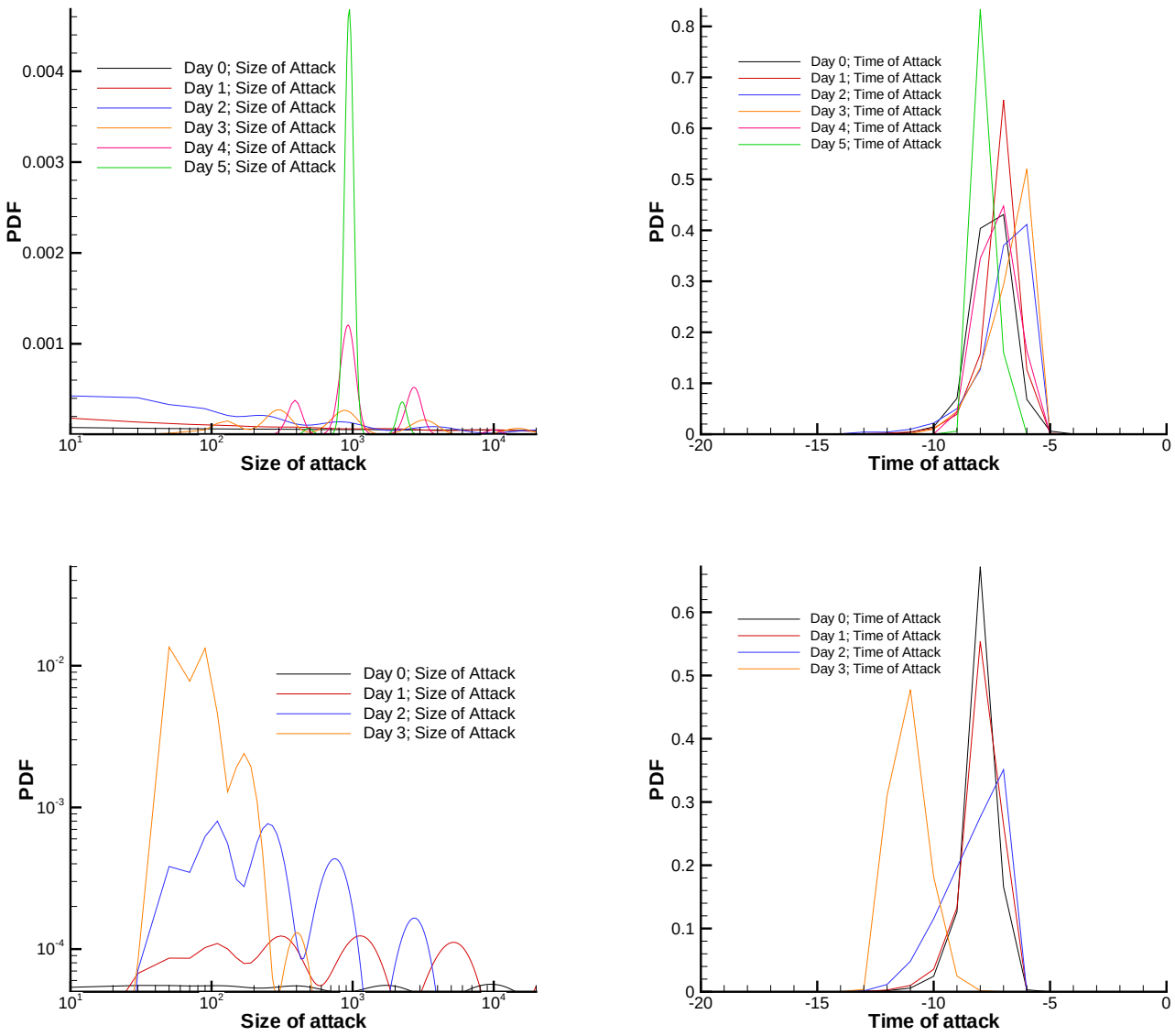

Figure 3. The PDFs for N and τ for Case B (above) and Case C (below). The plots for Case C are not shown for Days 4 and 5 since the data some second generation infectees which invalidated the inference process.

such ambiguity is seen in the PDF for τ; by day 1 the PDF peaks around -7 (seven days before the first symptomatic patient), which agrees with the results in Table 1. In Fig. 3 we plot the PDFs for Case B and C. The behavior seen in Case A holds true for Case B, though τ converges slightly slowly. Ambiguity, in the form of multimodal PDFs are seen for N; the data indicates that uncertainty in the size of N given the uncertainty in τ. An early attack of a smaller size can produce a signature similar to a larger, though recent attack. This can also be seen in Table 1, where the time-series from the three attacks vary very little from each other. The behavior for Case C, which has a small population of infected people, is different. Note that the vertical axis for the characterization of N is logarithmic; the characterization of the attack by day 3 is quite clear. However, the characterization of τ is largely incorrect till day 3. The results for day 4 and 5 are not plotted since *all* the values of $\bar{P}(N,\tau)$ were close to zero. This is because the data was corrupted by a few second generation infectees and consequently did not fit into the model of an incubation-based inference. While the quantification of the ratio of the index cases to second generation cases has not been done in the study, it is clear that this inference process is sensitive to such contamination.

Given that the smallest attack (Case C) has the most difficult inference task, we redo the inference with data collected on a 6 hour rather than daily basis. The correct values for (N,τ) are $(100,-10.75)$. The time-series is $\{2,0,0,1,1,3,3,3,1,3,0,4,5,1,7,6,5,3,6,4,8,3,3\}$. The PDFs for N and τ are in Fig. 4. Comparing them to Fig. 3 we see that the PDFs are generally smoother, and the variation of the PDFs with increasing amounts of data is more predictable. Also, τ is estimated more accurately. Thus, even though we do not use more data (i.e. we do not go beyond day 3), the Bayesian method exploits the *structure* in the high-resolution (i.e. 6 hour collection interval) data to arrive at "crisper" inferences. This relates directly to efforts (using syndromic surveillance and other techniques) to collect and process data quickly - if the data is collected at a finer resolution, inferences can be greatly enhanced without necessarily asking for *more* data.

Day	Case A	Case B	Case C
0	2	1	2
1	1	1	5
2	14	3	10
3	51	15	10
4	164	42	31
5	414	88	21
N	10,000	1,000	100
τ	-7	-8	-11

Table 1. Time series from 3 scenarios. The actual (N,τ) values are also provided.

4.2 Anthrax attacks

The incubation period for anthrax shows a distinct dependence on the dose (see Eq. 6) - thus an anthrax attack presents us with the possibility of inferring D the dose, along with N and τ. This leads to a more difficult problem - while the data is limited (3-4 days), the number of "degrees of

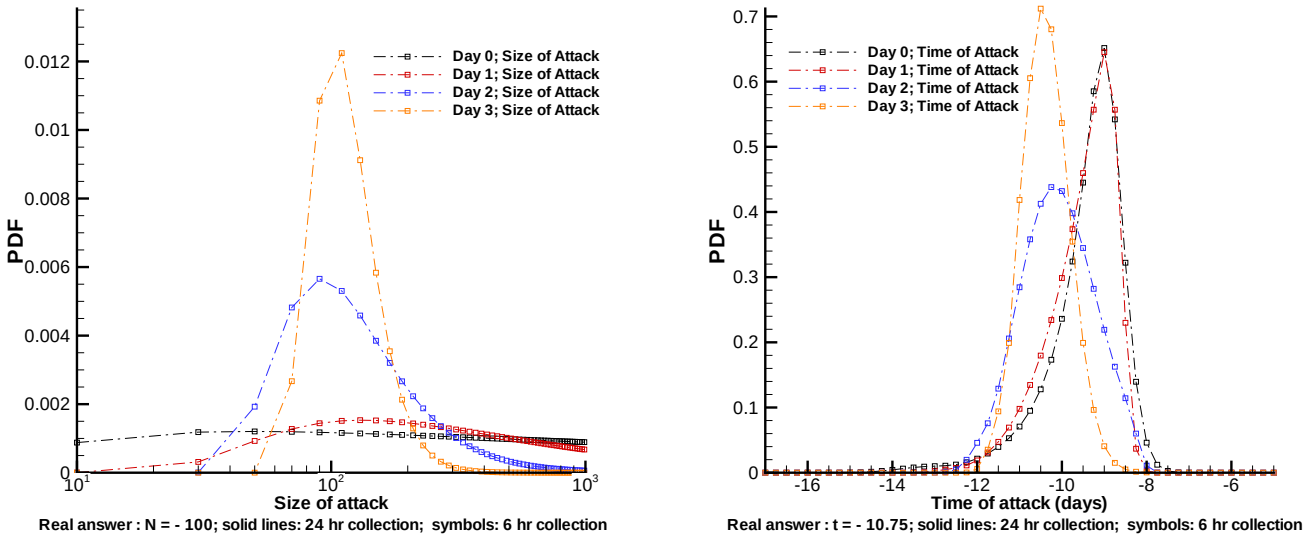

Figure 4. The PDFs for N and τ for Case C with data collected on a 6 hour interval.

freedom" increases from 2 to 3. As in Sec. 4.1 we simulate a number of attacks and collect the time-series; these are then used to estimate (N,τ,D). We make the simplifying assumption that all infected people get the same dose of D spores each.

Table 2 contains the time-series for 6 different attacks. These are also stochastic runs, based on the incubation period distribution (Eq. 6). The evaluation of $\bar{P}(N,\tau,D)$ is done in a 3D space and a figure akin to Fig. 2(top) would involve visualizing a 3D space or viewing isosurfaces of $\bar{P}$. We marginalize the distribution to provide PDFs of N, τ and D. These are plotted in Fig. 5 and 6. From Fig. 5 it is clear that it is difficult to infer the dosage for small attacks; even at high-dose attacks, the dosage PDF is too broad for much use. The observation also holds true for τ for the low-dose $D = 1$ attack. Only the characterization of N is seen to be somewhat better. Matters are different for large attacks (Fig. 6). Most characterizations are clean, though not necessarily unimodal. In case of Case F, we see a distinct possibility of either $N \approx 10,000$ or $N \approx 16,000$ corresponding to a $D = 10^4$ or $D = 10^3$ attack; the data up to day 3 supports both hypotheses quite well. Thus Bayesian methods can provide one with ambiguous characterizations, dependent upon data.

We now reanalyze Case A and B by collecting data on a 6 hour basis. The time series for Case A is $\{1,0,0,0,2,1,0,0,1,1,4,1,1,4,1,3,2,0,1,2,3,2,2\}$ and for Case B is $\{1,0,1,0,2,1,2,2,2,2,6,4,1,6,1,3,0,2,1,3,5,3,1\}$. The correct value for τ is -2.25 days in both cases. The PDFs are plotted in Fig. 7. We do not see any marked improvement in the inferences. Thus the extra structure in the data was simply not sufficient to overcome the uncertainty introduced by the small size of the population and the low dose.

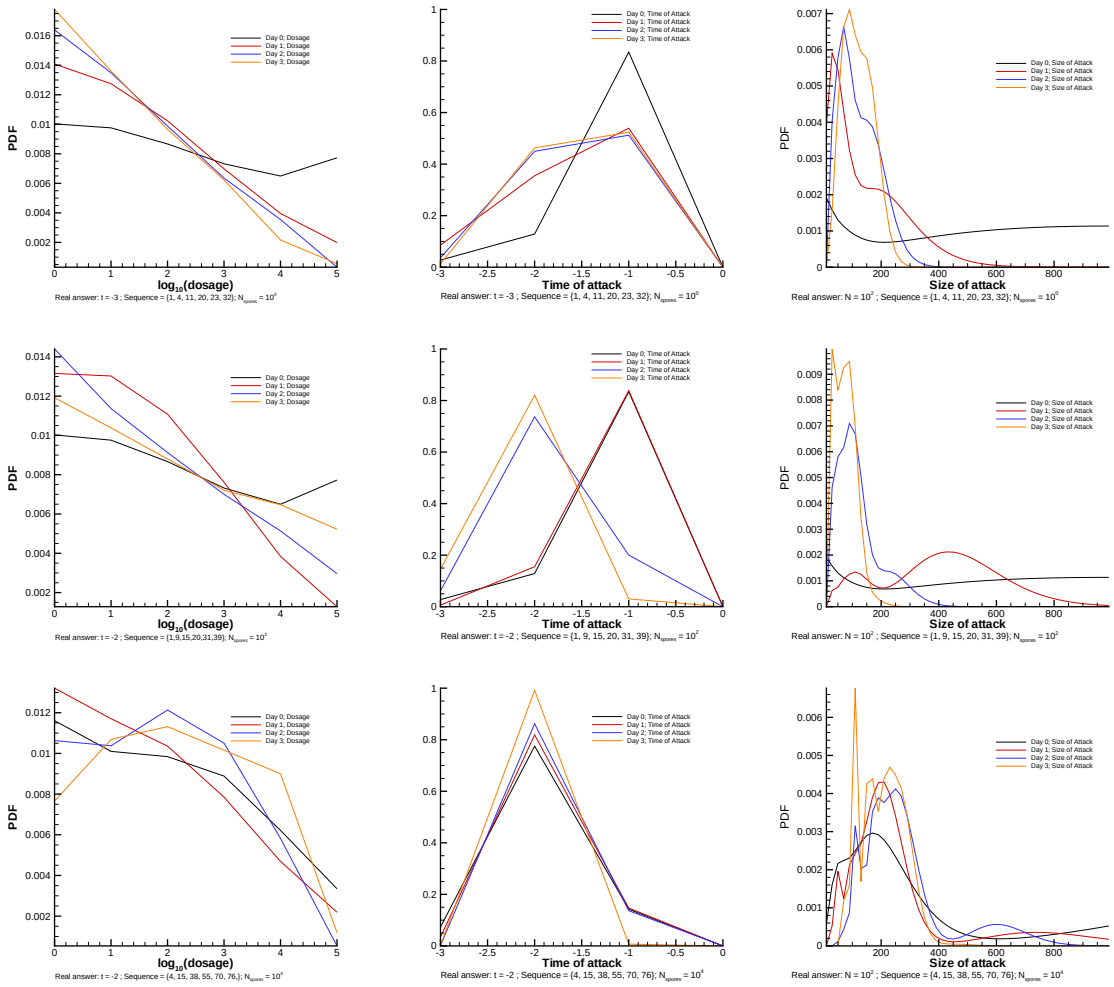

Figure 5. PDFs for the $\log_{10}(\text{dose})$, τ and N, for cases A, B and C. These are all small attacks $(N = 100)$. Black lines include day 0 data only, red lines include day 1, blue lines include day 2 in addition and orange lines involve day 0, 1, 2 and 3. The "sequence" in the figures is the total number of symptomatic as a function of time.

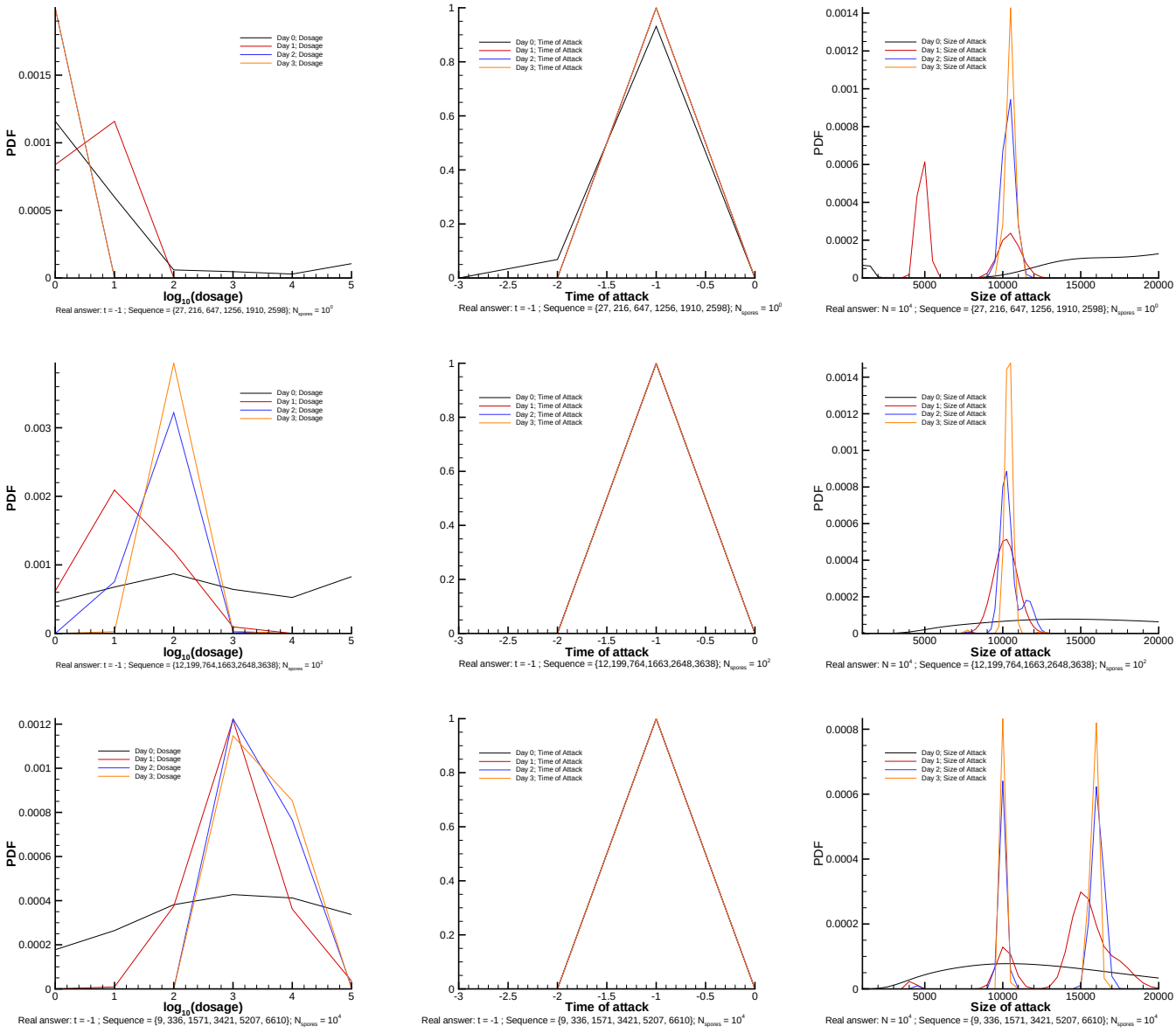

Figure 6. PDFs for the $\log_{10}(\text{dose})$, τ and N, for cases A, B and C. These are all large attacks $(N = 10{,}000)$. Black lines include day 0 data only, red lines include day 1, blue lines include day 2 in addition and orange lines involve day 0, 1, 2 and 3. The "sequence" in the figures is the total number of symptomatic as a function of time.

R9cm

Day	Case A	Case B	Case C	Case D	Case E	Case F
0	1	1	4	27	12	9
1	3	8	11	189	187	327
2	7	6	23	431	565	1235
3	9	5	17	609	899	1850
4	3	11	15	654	985	1786
5	9	8	6	688	990	1403
N	100	100	100	10,000	10,000	10,000
τ	-3	-2	-2	-1	-1	-1
D	1	100	10,000	1	100	10,000

Table 2. Time series from 6 anthrax attack scenarios. The actual (N, τ, D) values are also provided.

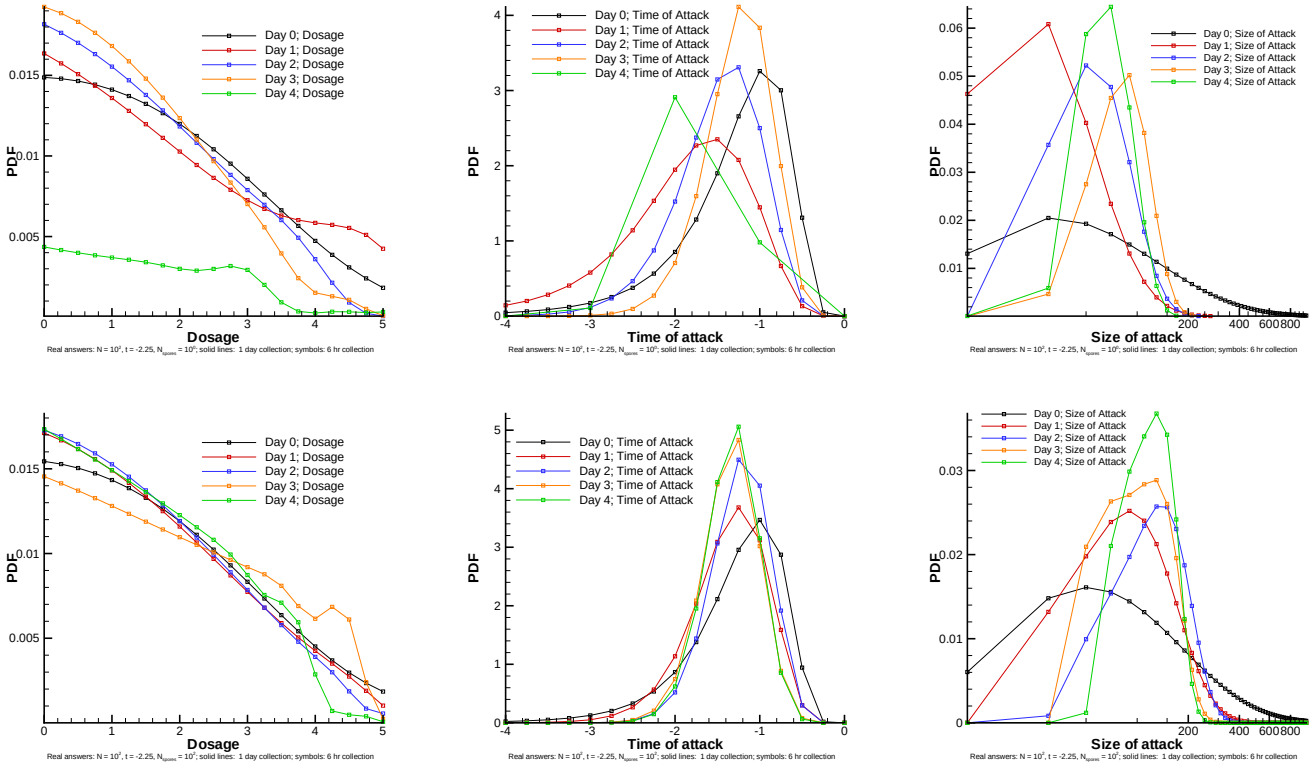

Figure 7. PDFs for the $\log_{10}(\text{dose})$, τ and N, for cases A and B with data collected at 6 hour intervals. These are all small attacks $(N = 100)$. Black lines include day 0 data only, red lines include day 1, blue lines include day 2 in addition, orange lines involve day 0, 1, 2 and 3. Green lines include day 4, in addition.

4.3 The Sverdlovsk outbreak

It is suspected that on April 2nd, 1979, there was an accidental release of a high-grade anthrax formulation from a military facility in Sverdlovsk, Russia [4]. 70 people are believed to have died [4, 5] and it has been estimated that 80 were infected [5]. This estimation was done under the assumption that all the fatalities were due to inhalational anthrax. This outbreak provides a good test case for our inference procedure. Wilkening [9] estimates that the average dosage was either around 2-3 spores, based on his Model A, or around 300 spores based on his Model D, which is the competing risks model of Brookmeyer [8]. Meselson [4] estimates 100-2000 spores as the likely dosage.

The Sverdlovsk case presents significant challenges. It was a low-dose attack infecting fewer than a hundred people. The first patient was detected on April 4th, 1979. The time-series of symptom onset is available on a day-by-day basis in [19]. Around April 12th, tetracycline was administered around Sverdlovsk; around the middle of April people were vaccinated. These prophylactic measures probably cured a few and increased the incubation period in others. Further, the data we work with almost certainly contains some recording errors. Also, the data was reconstructed from a variety of sources; public health records had been confiscated by the KGB [4]. Noisiness of the data, the effect of prophylaxis (which is not modeled in our inference process), and the small size of the infected population are expected to stress our inference process.

In Fig. 8 we plot the inferences for $\log_{10}(\text{dose})$, τ and N, based on the data in [19]. The time of release τ was easy to infer. The PDFs for dosage are unclear, though after 19 days, it is clear that it is a low-dose attack. The inference for N centers around 55 consistently though the inference with 5 days of data (black line) severely underestimates N. The data in [19] shows a long tail after April 18th, roughly after 5-6 days of the start of prophylaxis, when the antibiotics might be expected to affect the progress of the disease, with a single person showing symptoms everyday. Thus, inclusion of any data beyond Day 14 does not add much information but rather destabilizes the inference process as the (unmodeled) effect of prophylaxis begins to matter more. However, we are certainly within a factor of two of the correct value of N even with 10 days of data.

5 Conclusions

We have developed a promising, but still immature approach to inferring the characteristics of a BT attack. Based on Bayesian inference, we have inferred the characteristics of attacks with very little data (3-4 days). Large attacks ($O(10^3)$ infectees) are easy to infer for size, time and dosage; small attacks pose a harder problem because of the stochastic nature of the observed data. Even with these weaknesses, the basic method was able to infer characteristics of the Sverdlovsk outbreak with commendable accuracy. The method is lightweight and runs in seconds on a regular workstation. Its simplicity allows it to be improved, enhanced and modified in innumerable ways; as in most Bayesian approaches, these are achieved, most simply, by exploiting priors.

In the smallpox cases that we presented in Sec. 4.1, the inferences were done using the same

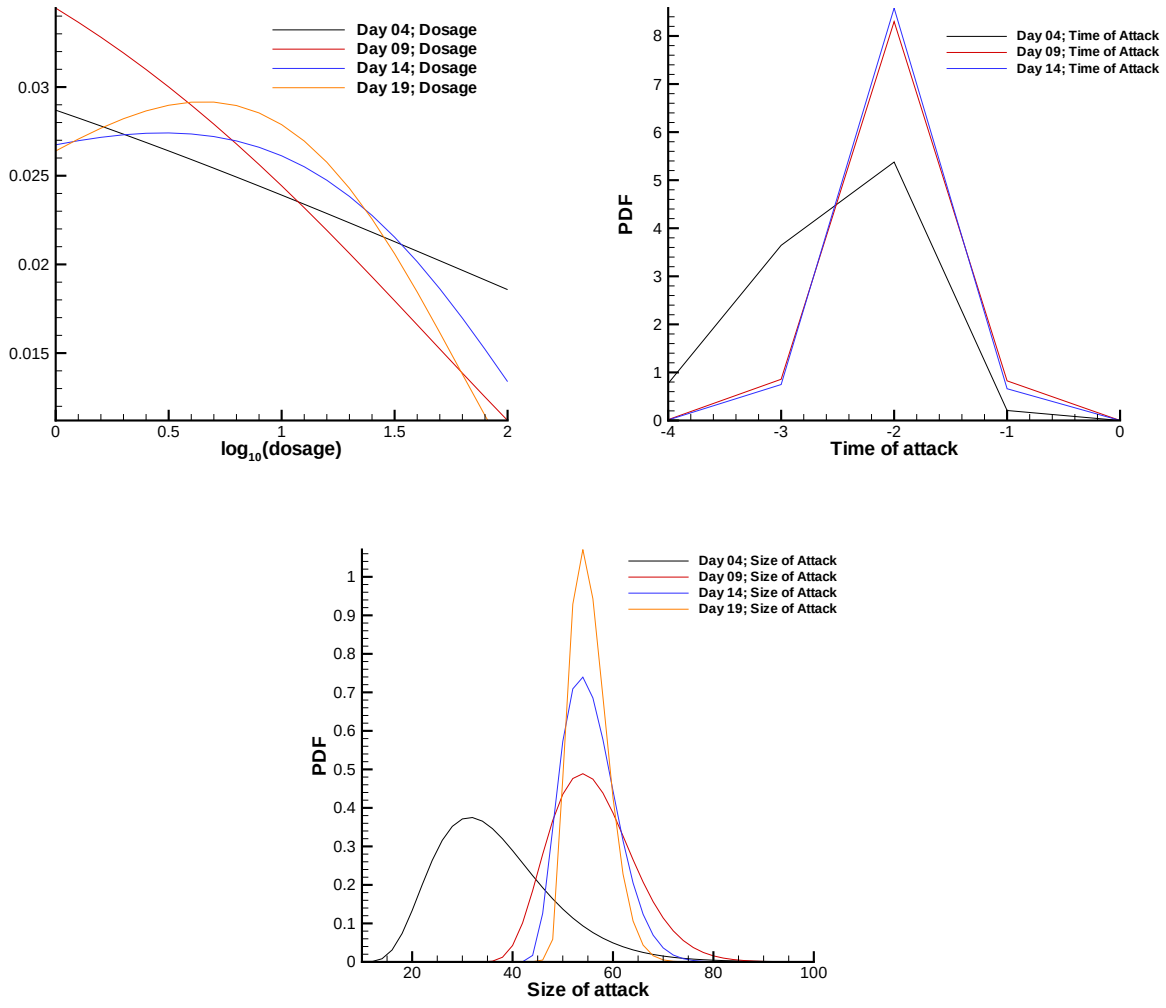

Figure 8. PDFs for the $\log_{10}(\text{dose})$, τ (above) and N(below), for the Sverdlovsk attack. Black, red, blue and orange lines denote inferences that use 5, 9, 14 and 19 days of data.

model of smallpox as the attacks. However, the attack simulation being stochastic, the inferences were drawn from only one realization of the attack and thus, strictly speaking, this was not an *inverse crime* [18]. However, it does raise the question of what the inferences would be if the characteristics of the pathogen used in the attack were markedly different from the model used in the inference. While this study did not address this question, it is a valid problem and can be approached in two ways:

1. perform a sensitivity study of the effect of discrepancy in the incubation period mean (or median) and the standard deviation of the pathogen used for inference and in the attack. The "distance" between the true PDF and the one developed when a discrepancy exists is one of the quantities to measure.

2. Instead of evaluating sensitivities, we infer (N, τ, μ, σ) from the observed data. However, we note that a good inference of four parameters will need a lot of data (or very helpful characteristics like a large N) given our experience in estimating three parameters in the anthrax tests in Sec. 4.2.

It is clear that many cases exist when fine-grained data (e.g. collected over 6 hour intervals rather than on a daily basis) is helpful. However, in reality, finer-grained data will suffer more from noise that more coarser grained collection, where measurement errors (reporting errors) will tend to cancel each other out. While Bayesian methods are good at navigating random noise, it remains to be seen if noisier, but finer grained data offer any advantage over cleaner coarse-grained data. Alternatively, if this measurement noise can be modeled, we may be able to include it in the inference process itself. This "filtering" idea might bear fruit if the time-series were long. Further, we also need to conduct a verification study regarding the effect of realizations (of the stochastic BT attack) on the quality of the inference.

One of the planned extensions of this model that is one where we try to infer the dose-distribution following an attack, strictly from the time-series (t_i, n_i). Conceptually, under the constraint of a single attack, this may be done by estimating the population N_j in each of $j = 0 \ldots J$ dosage bins which cover, in a non-overlapping manner, the entire dosage range. It is unclear, though, how one chooses a representative dosage for a given bin and what the sensitivity of inferences are to this choice. Being able to characterize the distribution can identify important signatures of the release process - a low-dose distribution with large variations is indicative of an outdoor attack with significant dispersion while a high-dose attack would indicate an indoors release. However it introduces J parameters $(N_j, j = 0 \ldots J)$ into the inference process, and given our experience in Sec. 4.2, J will probably need to be a small number.

Another planned extension of this approach is that of discriminating between a single versus multiple distributed attacks. In some cases, if the attacks are displaced only slightly (in time and space) it will be impossible to discriminate between them. Conceptually, one could create the equivalent of Eq. 4 dependent on $(N_k, \tau_k, D_k), k = 0 \ldots K$, where K is the number of attacks being considered. However this dramatically increases the number of parameters to be estimated, and will require significant amounts of data unless K is small. This approach is expected to be of relevance in small, covert attacks.

The most glaring shortcoming in this inference model is the lack of contact-spread processes in the inference procedure. In order to augment Eq. 4 to deal with contagious diseases, we will need to incorporate the probability of observing second generation infectees as a function of t. This PDF can only be developed by running ensembles of stochastic epidemic simulations and projecting the data on a suitable basis set so that the expression (which will involve at least N, R_0 and the total susceptible population) can be compactly represented. This is currently underway.

References

[1] Luther E. Lindler, Frank J. Lebeda, and George W. Korch. *Biological Weapons Defense*. Humana Press, Totowa, NJ, USA, 2005.

[2] Ken Alibek and Stephen Handelman. *Biohazard*. Delta, New York, NY, USA, 2000.

[3] J. Walden and E. H. Kaplan. Estimating time and size of bioterror attack. *Emerging Infectious Diseases*, 10(7):1202–1205, 2004.

[4] Matthew Meselson, Jeanne Guillemin, Martin Hugh-Jones, Alexander Langmuir, Ilone Popova, Alexis Shelokov, and Olga Yampolskaya. The Sverdlovsk anthrax outbreak of 1979. *Science*, 266:1202–1208, 1994.

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

[6] Ron Brookmeyer and Natalie Blades. Statistical models and bioterrorism : Application to the U.S. anthrax attacks. *Journal of the American Statistical Association*, 98(464):781–788, 2003.

[7] John A. Jernigan et al. Bioterrorism-related innhalational anthrax: The first 10 cases reported in the United States. *Emerging Infectious Diseases*, 7(6):933–944, 2001.

[8] R. Brookmeyer, E. Johnson, and S. Barry. Modelling the incubation period of anthrax. *Statistics in Medicine*, 24:531–542, 2005.

[9] D. Wilkening. Sverdlovsk revisted : Understanding human inhalational anthrax. *Science*. Submitted.

[10] H. N. Glassman. Discussion of Industrial Inhalational Anthrax. *Bacteriological Review*, 30:657–659, 1966.

[11] Peter B. Jahrling, Lisa E. Hensley, Mark J. Martinez, James W. LeDuc, Kathleen H. Rubins, David A. Relman, and John H. Huggins. Exploring the potential of variola virus infection of cynomolgus macaques as a model for human smallpox. *Proceedings of the National Academy of Science*, 101(42), 2004.

[12] A. R. Rao, M. Savithri Sukumar, S. Kamalakshi, T. V. Paramasivam, Thiru A. R. Parasuraman, and M. Shantha. Experimental variola in monkeys Part I: Studies on disease enhancing property of cortisone in smallpox. *Indian Journal of Medical Research*, 56(12):1855–1865.

[13] F. Fenner, D. A. Henderson, I. Arita, Z. Jezek, and I. D. Ladnyi. *Smallpox and Its Eradication*. World Health Organization, Geneva, 1988.

[14] A. R. Rao. *Smallpox*. The Kothari Book Depot, Bombay, 1972.

[15] M. I. Meltzer, I. Damon, J. W. LeDuc, and J. D. Millar. Modeling potential responses to smallpox as a biological weapon. *Emerging Infectious Diseases*, 7(6), November 2001.

[16] Stephen Eubank, Hasan Guclu, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429:180–184, 2004.

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

[18] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, New York, 2004.

[19] Thomas V. Inglesby, Donald A. Henderson, John G. Barteltt, Michael S. Ascher, Edward Eitzen, Arthur M. Friedlander, Jerome Hauer, Joseph McDade, Michael T. Osterholm, Tara O'Toole, Gerald Parker, Trish M. Perl, Philip K. Russel, and Kevin Tomcat. Anthrax as a biological weapon - Medical and public health management. *J. Am. Med. Assoc.*, 281(18):1735–1745, 1999.

[20] T. M. Mack. Smallpox in Europe, 1951-71. *Journal of Infectious Diseases*, 125:161–169, 1972.

[21] A. K. Joarder, D. Tarantola, and J. Tulloch. *Eradication of Smallpox from Bangladesh*. WHO Regional Publications; South-East Asia Series No. 8, New Delhi, 1980.

[22] M. Kortepeter, G. Christopher, T. Cieslak, R. Culpepper, R. Darling, J. Pavlin, J. Rowe, K. McKee, and E. Eitzen. *Medical Management of Biological Casualties Handbook*. U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Maryland 21702 - 5011, 2001.

[23] E. H. Kaplan, D. L. Craft, and L. M. Wein. Emergency response to a smallpox attack : The case for mass vaccination. *Proceedings of the National Academy of Science*, 99(16), August 2002.

[24] A. Brown, R. Czigan, J. Garcia, and A. Straubinger. Modeling a smallpox outbreak. http://www.czigan.net/pg2.htm. San Diego State University.

[25] A. R. Rao. Infected Inanimate Objects (Fomites) and Their Role In Transmission of Smallpox. Technical Report 72.40, World Health Organization, 1972. South East Asia Series.

[26] R. Gani and S. Leach. Transmission potential of smallpox in contemporary populations. *Nature*, 414, December 2001.

[27] R. M. Anderson, R. M. May, and B. Anderson. *Infectious Diseases of Humans : Dynamics and Control*. Oxford University Press, New York, 1992.

A A mathematical model of smallpox

A.1 Introduction

This document lays out a mathematical model for the infection and evolution of variola major in a well-mixed, homogeneous population. The index cases are generated via inhalation of aerosolized virons while the subsequent transmission between humans is via contacts. Data for the pathogenesis (development of the disease in a host) shows large variations and consequently we will derive some "consensus" values. Pathogenetic characteristics are strongly dependent on a variety of factors (health, sanitation and age being the most important); however data showing the functional dependence of these characteristics on the factors are scarce or non-existent. Thus the pathogenetic characteristics we "derive" via consensus are necessarily somewhat approximate. Note that these numbers apply neither to *Variola minor* nor to the somewhat benign (!) version of variola major that existed in Sahelian Africa. The data refers almost entirely to the virulent strain of variola major found in the Indian subcontinent and draws heavily from the data of Rao [14], Mack et al. [20] and Joarder et al. [21]

A.2 Epidemiology and pathogenesis

Smallpox (or more precisely variola major) is an *Orthopox* virus. Its principal mode of ingress into a host is via the respiratory tract, though cutaneous invasion (as during inoculation via scarification or variolation) is an oft-exploited (though inefficient) mode of infection. Inhalational infection requires that the virons be agglomerated into particles $< 1\mu$m in size; larger particles are either filtered out by the cilia in the respiratory tract or trapped and swallowed - an action that effectively deactivates the virons. Sub-micron particles are generated by the aerosolizing effect of coughing and sneezing of infected patients, provided that the virons exist in oral secretions. Once inside the host, the virons establish themselves in the respiratory tract and incubate.

The *incubation* period varies (see Section A.4 for details) but its length is around 12 days. The infectee displays no symptoms and is not contagious. The incubation period leads into the *prodromal* period of around 3 days where the patient undergoes high fever (40-41°C) and suffers from body aches; however none of these symptoms are unique to smallpox. The patient is immobilized (due to the fever) and does not shed the virus - hence the stage is non-contagious. The prodromal period leads to the *contagious* period.

During the early epoch of the contagious period, the fever abates somewhat and "spots" appear in the oral cavity. Over approximately 4 days, they evolve into sores and release virons. Coughing and sneezing during this epoch aerosolizes the virons, causing this to be the most contagious period of the disease evolution. At the end of this "sore-in-mouth" phase, rashes start appearing over the body, a process that continues for a further 4 days. The fever returns. At the end of "sore-in-mouth" and "rash-over-body" phases, the rashes evolve into bumps, pustules and finally scabs. Due to the overt nature of the rashes (and the recurrence of the fever) the patient has little contact (except with

immediate caregivers) and the "rash-over-body" phase is only somewhat contagious (the pustules contain virons and can release them if they break). The process of development of rashes into pustules and scabs occurs over a period 4-5 days from the end of the "rash-over-body" and leads into what we term the *scab* period. The contagious period lasts (very approximately) 7 days and the entire "rash and early scab" period lasts 13 days.

The scab period sees the drying and shedding of the scabs. The scabs contain high concentrations of virons encased in a tough dry covering which prevents them from escaping and thus infecting others. This scabbing stage is non-contagious and leads either to an abatement of fever and recovery (with possible sequelae of blindness, pockmark disfigurations etc) or death. Mortality rates (Chapter 4 in [13]) show a clear structure, with the very young (< 4 years) and the old (> 40 years) succumbing to the disease in large percentages ($\geq 30\%$). Vaccination reduces the mortality rate by a factor of 4. Overall, a mortality rate δ of 30 % was observed amongst the unvaccinated populace. The entire duration of the disease, from infection to recovery is estimated to be around 36 days. Recovery engenders long-term immunity.

A.3 Inhalational smallpox

In theory, a *single* viron "taken" by the host is sufficient to cause smallpox. However, "taking" efficiency is not 100 % and nothing is known regarding the LD_{50} for smallpox. USAMRIID estimates that 10-100 inhaled organisms are sufficient for infection (Appendix C in [22]). Experiments with cynomolgus macaques have shown that extremely high concentrations (10^8) can cause smallpox in non-human primates who are otherwise immune to it.

We assume that the human ID_{50} (infectious dose, 50 % probability) is 5 PFUs (plaque forming units). We will also assume that 10^2 ID_{50} will provide an almost 100 % infection probability. Assuming a log-normal probability distribution function, the cumulative probability distribution function (CDF) $D(x)$ (where x is the dose in PFUs) is

$$D(x) = 0.5 \left[1 + erf \left(\frac{\ln(x/x_0)}{\sqrt{2}S} \right) \right] \tag{8}$$

where x_0 is 5 PFUs and $S = 1.628$. The assumption here is that $erf(2.0) \approx 1$. Note, that the mean μ and variance σ^2 of a log-normal distribution is given by

$$\begin{aligned} \mu &= x_0 \, e^{\frac{S^2}{2}} \\ \sigma^2 &= x_0^2 \, e^{S^2} \left(e^{S^2} - 1 \right) \end{aligned}$$

A.4 Pathogenesis

The evolution of smallpox shows distinct phases but a great deal of uncertainty exists regarding the duration of the phases. This uncertainty regarding the duration of the contagious stage is critical to obtaining credible estimates of casualties. The various stages are:

1. **Incubation :** This period (the incubation) is asymptomatic and non-contagious with a length of approximately 12 days.

2. **Prodromal :** This period is marked by high fever (40-41°C) and aches. In the absence of any sneezing and coughing there is no virus shedding, but it seems to have been considered both contagious [23] and non-contagious [15]. We will assume that this period is non-contagious. It lasts about 3 days. The characteristics of the fever are not unique to smallpox.

3. **Symptomatic, contagious :** This follows the prodromal period and shows various symptoms unique to smallpox. This can be divided into 2 sub-phases.

   - "Sore-in-mouth" sub-phase : Following the prodromal period, this lasts 4 days. The fever drops but coughing and sneezing persist. Small "spots" appear on the tongue and in the mouth, which over the 4 days, develop into sores. Towards the end of the sub-phase, the sores burst releasing high quantities of virons into the oropharyngeal cavity, which are then promptly aerosolized via coughing and sneezing and so shed into the environment. This is the most contagious period of the disease.
   - "Rash-on-body" sub-phase : The 4-day "sore-in-mouth" sub-phase is followed by a 1-day transition after which rashes start appearing on the body. The fever returns and this process continues for around 4 days too. Towards the end of the 4-day period, the rashes start developing into bumps, pustules and finally scabs. While the pustules contain virons and may release them if ruptured, the high fever and overt symptoms effectively isolate the patients from everyone except the immediate caregivers. We will consider the first two days of the "rash-on-body" as being in the contagious symptomatic phase and the last two days in the noncontagious symptomatic phase, mentioned below.

   The contagious phase thus lasts about 7 days.

4. **Symptomatic, non-contagious :** Once the scabs appear after the "rash-on-body" sub-phase, they dry and are shed. These scabs contain large amounts of virons encased in a tough shell which rarely ruptures. Thus this phase is non-contagious. The high fever persists and then abates as the person recovers or the phase ends in the patient's death. The cause of death is usually toxemia; organ failure has not been observed. The duration of this phase is around 14 days - it contains the last 2 days of "rash-on-body", the 4 days during which the rashes mature into pustules and scabs and 8 days of shedding of the scabs.

5. **Recovery :** The duration of the disease (from infection to recovery/death) is around 36 days. Recovery leads to long-term immunity.

Ref.	[23]	[24]	[15]
τ_I	11	13 (7-17)	12 (P_5 = 8, P_{95} = 17)
τ_P	3	3 (2-4)	3
τ_{I_1}	12	9	10
τ_{I_2}	-	11 (4-16)	8

Table A.1. Durations of the Incubation, Prodromal, Symptomatic (contagious) and Symptomatic (non-contagious) phases from 3 recent studies. [24] listed bounds on the duration of the various phases. All numbers are in days. [15] provided the 5^{th} and 95^{th} percentiles for the Incubation period. They also assumed that the disease terminated in 33 days while [24] assumed 36. [23] did not assume a duration for the disease since it was not required for their modeling study.

We now need to arrive at a "consensus" set of values for the duration of these periods, which we denote as $\tau_I, \tau_P, \tau_{I_1}$ and τ_{I_2} respectively. We tabulate data from some recent calculations in Table A.1. Using the data in Table A.1 we arrive at the "consensus" values in Table A.2.

Periods	Values
τ_I	$\mu = 12, \sigma = 2$
τ_P	3
τ_{I_1}	$\mu_{\mathrm{I}_1} = 7, \sigma_{\mathrm{I}_1} = 1$
τ_{I_2}	14

Table A.2. We assume that the incubation and symptomatic (contagious) phases are normally distributed. The σs here take into account the variation in the data in Table A.1. τ_{I_2} is calculated from the μ of the other periods and a disease lifetime $\mu_{\mathrm{R}} = 36$ days, $\sigma_R = 3$. All numbers are in days.

A.5 Transmission

Transmission of the virons occurs via inhalation of aerosols generated either by the patients coughing and sneezing or when handling his bedclothes and personal effects. The latter, called transmission by fomites, seems to be ineffective [25] and we will concentrate on the first.

Smallpox deactivates quickly in air and so transmission is usually short-range - i.e. by contact. The classical measure of "contagiousness" is the basic reproductive number R_0, the number of susceptibles a single infected individual infects during the contagious period. Since this is spread over τ_{I_1}, the rate of infection is R_0/τ_{I_1}. R_0 has been estimated for a number of smallpox outbreaks

[26]. Numbers varied between 3.5 and 6.0. [23], [24] and [15] all assumed 3.0 and we will adopt this number too.

The actual modeling of transmission via contact in a homogeneous, well-mixed population will be done via the law of mass action[27]. The probability κ that a susceptible will be infected during a time period Δt is

$$\kappa = \Delta t \beta I_1 = \Delta t \frac{R_0}{\tau_{I_1}} \frac{I_1}{N}$$

where I_1 is the number of symptomatic contagious people and N is the total population. Thus β, the effective infection rate is

$$\beta = \frac{R_0}{\tau_{I_1} N}$$

A.6 Evolution models

Evolution models are of two types - deterministic, which are based on ODEs (ordinary differential equations) and probabilistic which exploit probability distribution functions (as defined in Table A.2) to evolve a population of entities. Both models divide the population into 5 compartments

1. Susceptibles (S), the uninfected people who can be infected i.e. people without immunity
2. Latent (L) which includes people in the Incubation and Prodromal stages. τ_L is characterized by $\mu_L = 15, \sigma_L = 2$.
3. Contagious (I_1), which includes people in the Symptomatic, contagious phase
4. Advanced (I_2), which includes people in the Symptomatic, non-contagious phase
5. Removed (R), which includes people who have recovered or died.

Both the models transition people through these compartments, but via different means.

A.7 Deterministic

In the equations below we will assume that vaccinations do not occur as the disease progresses in the populace. We also assume that people are infected at a constant rate i.e. given a random infected person, the probability of transitioning to the next compartment is not a function of when the person was chosen. This does not quite hold true in a bio-terrorist scenario - the index cases

are infected simultaneously and transition together via the various compartments. Thus the rate of transition for the entire compartment is a function of time, at least in the initial generations of the infected entities. The equations are similar to the SIR and SEIR models [27].

$$\begin{aligned}
\dot{S} &= -\beta I_1 S \\
\dot{L} &= \beta I_1 S - \frac{1}{\tau_L} L \\
\dot{I_1} &= \frac{1}{\tau_L} L - \frac{1}{\tau_{I_1}} I_1 \\
\dot{I_2} &= \frac{1}{\tau_{I_1}} I_1 - \frac{1}{\tau_{I_2}} I_2 \\
\dot{R} &= \frac{1}{\tau_{I_2}} I_2
\end{aligned} \tag{9}$$

The number of dead people is δR. The τ used here are the means listed in Table. A.2.

A.8 Non-deterministic

Non-deterministic simulations evolve a set of entities over a period of time in steps of Δt. In the following, we will assume a normal distribution, characterized by a mean μ and standard deviation σ. Thus the probability of an event happening between t and $t+\Delta t$ is

$$\frac{1}{\sigma\sqrt{2\pi}} \int_t^{t+\Delta t} e^{-\frac{1}{2}(\frac{q-\mu}{\sigma})^2} dq = \Phi(t+\Delta t;\mathbf{p}) - \Phi(t;\mathbf{p})$$

where

$$\begin{aligned}
\Phi(t;\mathbf{p}) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}(\frac{q-\mu}{\sigma})^2} dq \\
&= \frac{1}{2}\left(1 + erf\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)\right)
\end{aligned}$$

where $\mathbf{p} = \{\mu, \sigma\}$. Thus, the probability $P_{i\to i+1}$ of a person transitioning from stage i to stage $i+1$ is related to $\frac{1}{\sigma\sqrt{2\pi}} \int_t^{t+\Delta t} e^{-\frac{1}{2}(\frac{q-\mu}{\sigma})^2} dq$ by the expression

$$(1-\Phi(t;\mathbf{p}))P_{i\to i+1} = \frac{1}{\sigma\sqrt{2\pi}} \int_t^{t+\Delta t} e^{-\frac{1}{2}(\frac{q-\mu}{\sigma})^2} dq \tag{10}$$

or

$$P_{i\to i+1}(t;\mathbf{p}) = \frac{\Phi(t+\Delta t;\mathbf{p}) - \Phi(t;\mathbf{p})}{1-\Phi(t;\mathbf{p})} \tag{11}$$

Thus the transition probabilities are:

1. A susceptible can either stay in compartment S or transition to L, on being infected, in time Δt with a probability

$$P_{S\to L} = \beta I_1 \Delta t$$

2. A latent, who *has been in the latent state for time t* can transition to I_1 over a time period Δt with a probability $P_{L\to I_1}$ given by Eqn. 11 with $\mathbf{p} = \{\mu_L, \sigma_L\}$.

3. A person who *has been in compartment I_1 for time t* will transition over to I_2 in time period Δt with a probability $P_{I_1\to I_2}$ given by Eqn. 11 with $\mathbf{p} = \{\mu_{I_1}, \sigma_{I_1}\}$.

4. A person in I_2 who *has been infected for time T* will transition over to R in time period Δt with a probability $P_{I_2\to R}$ given by Eqn. 11 with $P_{I_2\to R}$ with $\mathbf{p} = \{\mu_R, \sigma_R\}$.

DISTRIBUTION:

4 M9159
Jaideep Ray, 08964

2 M9056
Y. Marzouk and Habib Najm, 08351

3 MS 9018
Central Technical Files, 8940-1

1 MS 0899
Technical Library, 9616

1 MS 9021
Classification Office, 8511, for Technical Library, MS 0899, 9616
DOE/OSTI via URL