Journal of
Applied Microbiology

Journal of Applied Microbiology ISSN 1364-5072

ORIGINAL ARTICLE

Variance in *Bacillus anthracis* virulence assessed through Bayesian hierarchical dose–response modelling

J. Mitchell-Blackwood[1,2], P.L. Gurian[1], R. Lee[3,4] and B. Thran[4]

1 Department of Civil, Architectural and Environmental Engineering, Drexel University, Philadelphia, PA, USA
2 US Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC, USA
3 US Army Medical Research Institute of Chemical Defense, Aberdeen Proving Ground, MD, USA
4 Environmental Health Risk Assessment Program, Health Risk Management Portfolio, US Army Public Health Command Army Institute of Public Health, Aberdeen Proving Ground, MD, USA

Keywords
Bacillus anthracis, dose-response modelling, environmental, spores, virulence.

Correspondence
Jade Mitchell-Blackwood, US Department of Agriculture, Food Safety Inspection Service, 1400 Independence Avenue, Stop 3766, Washington, DC 20250, USA.
E-mail: mitchell-blackwood.jade@epa.gov

2011/1916: received 9 November 2011, revised 7 April 2012 and accepted 10 April 2012

doi:10.1111/j.1365-2672.2012.05311.x

Abstract

Aims: To develop a predictive dose–response model for describing the survival of animals exposed to *Bacillus anthracis* to support risk management options.
Methods and Results: Dose–response curves were generated from a large dose–mortality data set (>11 000 data points) consisting of guinea pigs exposed via the inhalation route to 76 different product preparations of *B. anthracis*. Because of the predictive nature of the Bayesian hierarchical approach (BHA), this method was used. The utility of this method in planning for a variety of scenarios from best case to worst case was demonstrated.
Conclusions: A wide range of expected virulence was observed across products. Median estimates of virulence match well with previously published statistical estimates, but upper bound values of virulence are much greater than previous statistical estimates.
Significance and Impact of the Study: This study is the first meta-analysis in open literature to estimate the dose–response relationship for *B. anthracis* from a very large data set, generally a rare occurrence for highly infectious pathogens. The results are also the first to suggest the extent of variability, which is contributed by product preparation and/or dissemination methods, information needed for health-based risk management decisions in response to a deliberate release. A set of possible benchmark values produced through this analysis can be tied to the risk tolerance of the decision-maker or available intelligence. Further, the substantial size of the data set led to the ability to assess the appropriateness of the assumed distributional form of the prior, a common limitation in Bayesian analysis.

Introduction

The need to establish health-based exposure guidelines for agents of bioterrorism and many other micro-organisms is relevant in a variety of contexts, including the development of first responder guidelines, medical response plans (Mitchell-Blackwood *et al.* 2011), remediation standards (Hong *et al.* 2010) and sensor system design (Huang *et al.* 2010). A key element in the development of these guidelines is an understanding of the dose–response relationship, which predicts the consequence of a microbiological exposure to a given dose of organisms (Haas *et al.* 1999). The impact of exposure guidelines and remediation goals on resource management and the associated expenditures are significant. Therefore, an analysis which transparently presents the science supporting the derivation and selection of dose–response curves, which in turn support the establishment of health-based exposure guidelines, is needed.

The ability of dose–response models to aid in making predictions of potential consequences of future events is impeded by both aleatory and epistemic uncertainty

© 2012 No claim to US Government works/J. Mitchell-Blackwell & P.L. Gurian

because of inherent variability and lack of full knowledge about both biological and physical processes, respectively. As is the case with most environmental stressors, chemical or biological, a primary source of variability arises from the need to extrapolate experimental data from animal surrogate hosts to humans, although data driven methods for describing interspecies differences are rapidly advancing (Jones *et al.* 2009). Unfortunately, even with the development of more complex models, the fact remains that biosafety and ethics concerns preclude the ability to perform large *in vivo* studies in animals and humans. Hence, both the limited availability and small sample sizes of historic experimental data sets for bioterrorism agents further prevent complete reduction of uncertainty by this source. Additionally, there are other relevant sources of uncertainty primarily because of variability to consider in the microbial risk assessment of these agents. These include (i) strain differences when surrogate strains of the same organism are used to assess virulence experimentally and (ii) the use of different experimental protocols and preparation methods. Because micro-organisms are uniquely subject to evolution and bioengineering, even if a perfect model could be developed based on all available observations to describe the sample-to-sample variations from all relevant inputs and assumptions (e.g. host species, strains, isolates and dissemination techniques), it would be limited in its ability to predict the virulence of a pathogen in a future bioterrorist attack (National Research Council, 2010). Therefore, in this paper, we consider the importance of uncertainty in our characterization of variability by fitting a Bayesian hierarchical dose–response model to a historic US Army data set. By using this approach to parameterize the dose–response curve, estimates of the virulence of different preparations or products of *Bacillus anthracis* were produced for use in the formulation of information that supports health-based risk management decisions.

Virulence, a term used to describe the relative ability of a pathogen to cause harm in a host organism, can also be used to compare different formulations or products generated from the same pathogen. Generally, virulence is described using a lethal dose (LD) estimate for a given percentage of the population (e.g. LD_{50}). By comparing LD estimates, one can rank virulence for a given set of pathogenic test products. Because preparation of the test products may potentially impact delivery (e.g. aerosol source strength and decay rates), environmental and respiratory tract deposition, the viability of the spores themselves, which is directly related to infectivity, the consideration of the variability related to this parameter is important in assessing and managing microbial risk.

Several previous estimates of *Bacillus anthracis* virulence are available. However, questions remain about both the amount and relevance of the evidence used to inform these estimates (Coleman *et al.* 2008). The most quoted LD_{50} for *B. anthracis* is 8000–10 000 spores (Defense Intelligence Agency, 1986). However, there are no experimental details to support the reported LD_{50} in the Defense Intelligence Agency document (DIA). Another commonly cited LD_{50} is 2500–55 000 inhaled anthrax spores (Inglesby *et al.*, 2002), which references the former mentioned DIA document; therefore, it does not provide the improved confidence needed to use as a basis for exposure guidelines. A more recent estimate of infectivity was reported by Bartrand *et al.* (2008) – 94 320 inhaled spores (with 95% confidence interval bounds equal to 74 100–125 060). This much larger LD_{50} estimate was the result of statistical modelling of pooled data sets from guinea pigs exposed via the intranasal route and rhesus monkeys exposed via the inhalation route. The reported inconsistencies in virulence for *B. anthracis* are attributable to many possible explanations, including but not limited to variability in the sensitivity of the host organisms, strain differences, delivery mechanism of the experimental test product, related environmental conditions and/or test product preparation.

While the utility of statistical dose–response modelling is clear, classical or frequentist dose–response approaches do not capture the variability from experiment to experiment (US EPA, 2008). Whereas the Bayesian approach allows for hierarchical models, in which results from one experiment are seen as deriving from an overall distribution of potential experimental outcomes. In the classical approach, it is generally necessary to assume the results of different experiments are either produced by identical or related underlying processes (so that the results of different experiments can be pooled) or are produced by completely independent processes (in which case the results must be analysed separately and results of one experiment cannot inform the analysis of another experiment). In addition to overcoming these limitations, the dose–response curves generated by the Bayesian hierarchical approach allow predictions to be made for unobserved pathogenic agents (i.e. future agents that may be used in an attack), based on the observed data available for analysis. Furthermore, because this method is fully probabilistic, a distribution of values for each dose–response parameter makes it possible to quantify uncertainty associated with the estimate.

The authors recognize that because of the enormous amount of variability from the sources discussed herein, the selection of a model and/or parameter estimate for use in downstream risk management decisions becomes complex and much less straightforward than acceptance of a particular statistical confidence level or margin of error. The consideration of *B. anthracis* products with

© 2012 No claim to US Government works/J. Mitchell-Blackwell & P.L. Gurian

different virulence-class (e.g. high, medium and low) characteristics provide insights into the potential public health outcome given a product used. For example, high virulence-class products would be good representatives of 'worst-case scenarios'. Therefore, a set of possible benchmark values, which may be tied to the risk tolerance of the decision-maker or available intelligence, are provided. Communication of uncertainty through a range of plausible values, which can be explored in scenario assessments made by decision-makers, is an efficient and effective means of informing downstream decisions (National Research Council, 2007). This study approaches the output of the analysis from the perspective of an emergency manager who prior to an event and in all likelihood throughout the initial stages of an event lacks information on the exact type of product that has been released, but instead must prepare for the full range of possible scenarios. This perspective allows one to communicate information required by risk analysts for the development of health-based exposure guidelines, without providing information as to which characteristics result in higher (or lower) virulence that could be exploited by adversaries.

Data and methods

Data set

The data set used for this analysis is a compilation of unclassified or declassified historical US Army data generated in the 1950s and 1960s. Original reports were obtained and mined for dose–response data as part of a larger ongoing dose–response project at the US Army Public Health Command – formerly the Center for Health Promotion and Preventive Medicine. The data set is extensive with more than 11 000 observations of dose and mortality in Hartley strain guinea pigs challenged with inhalational exposure to a virulent strain of *B. anthracis*. Based on the range of dates, these studies were conducted and papers published by USAMID in the 1960s the Vollum strain are most probable. An analysis of such a large body of evidence is rare.

This data set was obtained from a number of different tests performed by researchers over the test period. Each test was conducted using the same *B. anthracis* strain under several experimental variables: product preparation, chamber size, animal weight, temperature, relative humidity, season, collection device, fans and delivery mechanism. A total of 76 experimental products with a range of physical properties (e.g. water content, pH, dry weight and particle size), storage times, storage temperatures and virulence were used in these tests representing variability in product preparation. Each product, the actual aerosol to which the animals were exposed, represents a particular preparation under a particular set of experimental conditions as described above. However, the details of such preparations, delivery and experimental conditions are noted elsewhere (Thran *et al.* 2010) and are not deemed to be important to the understanding of the current analysis.

Analysis method

Dose–response relationships were fit to data using a three-stage Bayesian hierarchical structure (Gilks *et al.* 1996) to describe the observations of death of individual animals exposed to *B. anthracis*. In the first stage, an exponential dose–response function (Haas *et al.* 1999) given by Eqn 1 is used to describe the probability of mortality $P(D)$ at a given dose (d).

$$P(D) = 1 - \exp(-kd) \tag{1}$$

The parameter, k, is the probability that a single organism can survive and cause infection. In the second stage of the model, the *values of k* for each product are described by a lognormal hyperdistribution given by Eqn 2:

$$\ln k \sim \mathrm{N}(\mu_{\ln k}, \sigma^2) \tag{2}$$

where $\ln k$ is the natural log of k, N is a normal distribution, with parameters $\mu_{\ln k}$, the mean of the distribution, and σ^2, the variance of the distribution. In the third stage of the model, the parameters of $\mu_{\ln k}$ and σ^2 from the second stage hyperdistribution are described using diffuse priors with fixed parameters given by Eqns 3 and 4:

$$\ln\mu_{\ln k} \sim \mathrm{N}\left(-11{\cdot}9, 2^2\right) \tag{3}$$

$$-2\ln\sigma^2 \sim \mathrm{N}\left(0{\cdot}67, 0{\cdot}84^2\right) \tag{4}$$

The posterior or predictive distribution for each observed product is given mathematically by Eqn 5:

$$_{\sigma lnk}\mu_{lnk} f\left(k_j|x_j\mu_{lnk}, \sigma^2_{lnk}\right) f\left(\mu_{lnk}|x\right) f\left(\sigma_{lnk}|x\right) d\mu_{lnk}\, d\sigma_{lnk} \tag{5}$$

where:

$f(k_j|x_j, \mu_{lnk}, \sigma^2_{lnk})$is the posterior distribution of the *values of k* for a product number, j, given the observed experiment using that product number and the hyperparameters of the prior distribution.

$f(\mu_{lnk}|x)$ is the posterior distribution of the hyperparameter, μ_{lnk}, given the observed data.

$f(\sigma_{lnk}|x)$ is the posterior distribution of the hyperparameter, σ_{lnk}, given the observed data.

Monte Carlo Markov Chain (MCMC) methods were used to develop a sample from the posterior distribution

© 2012 No claim to US Government works/J. Mitchell-Blackwell & P.L. Gurian

of the model parameters. The algorithm was implemented in 'R' (http://www.r-project.org), a free software environment for statistical computing. A minimum of 10 000 samples were obtained for each posterior distribution.

The Bayesian approach allows for the estimation of dose–response model parameters and the associated uncertainty of unobserved agents (e.g. a product that has not been used in an experiment or terrorist attack). Mathematically, the posterior predictive distribution of an unobserved species is represented by Eqn 6:

$$_{\sigma lnk}\mu lnkf(k|\mu_{lnk},\sigma^2_{lnk})f(\mu_{lnk}|x)f(\sigma_{lnk}|x)\ d\mu_{lnk}\ d\sigma_{lnk} \qquad (6)$$

where:

$f(k|\mu_{lnk},\sigma^2_{lnk})$is the posterior distribution of k given μ_{lnk} and σ_{lnk}. This distribution is a simulated k based on the posterior distribution of the hyperparameters.

$f(\mu_{lnk}|x)$ is the posterior distribution of the hyperparameter, μ_{lnk}, given the observed data.

$f(\sigma_{lnk}|x)$ *is* the posterior distribution of the hyperparameter, σ_{lnk}, given the observed data.

An average predictive dose–response curve for an unobserved species can also be determined by integrating over the full range of variability and uncertainty in the parameters, as described in Eqn 7:

$$\begin{aligned} E[\text{risk}|\text{dose}] &= {}_{\text{k}}P(D|k,\text{dose})f(k|x)dk \approx 1/NP(D|k) \\ &= \text{Average}P(D) \end{aligned} \qquad (7)$$

where:

E [*risk*|*dose*] = The expected value of risk defined as $P(D)$, the probability of mortality at any given the dose

P (D|k,*dose*) is the probability of death given the parameter value and any given dose.

$f(k|x)$ is the posterior distribution of the parameter after taking the observed data into account; This represents the distribution of *values of k* that are plausible after updating the prior distribution with the observed data.

N is the number of MCMC samples from the posterior distribution of k.

$P(D|k)$ is the marginal probability of death integrated across all possible values of the parameter.

Model verification

To verify the internal validity of the model, a simple comparison with *values of k* estimated by the maximum likelihood estimation (MLE) method (Haas *et al.* 1999) was conducted. These were also computed in 'R' along with confidence intervals via bootstrap re-sampling. Additionally, goodness of fit was determined by likelihood ratio test where the fit of the exponential dose–response model to the individual product data sets was determined. The lognormal distribution which was specified as the prior distribution represents an assumption that the 76 *values of k* for the test products are log normally distributed. This assumption was made in the absence of observable data. The actual distribution of the *values of k* was obtained when the exponential dose–response model was fit to the experimental data using maximum likelihood estimation. Finally, the abundance of estimates available from the analysis of this data set made it possible to generate a quantile–quantile (Q–Q) plot to assess how well the assumed hierarchical distribution (the specified prior) fit the actual distribution of *values of k* obtained from MLE estimates for the different experiments.

Results

Dose–response parameters (*k*) for individual products

The box plots in Fig. 1 illustrate the posterior distributions of the ln$k$ for each product. The lack of overlap among many of the boxes indicates that different products have distinct dose–response behaviour. In fact, the range of variability among products is huge. While large numbers of products cluster around ln$k$ of −12, several products have *values of lnk* of −10, a factor of e^2 or 7·7 above the central tendency. Values below the central

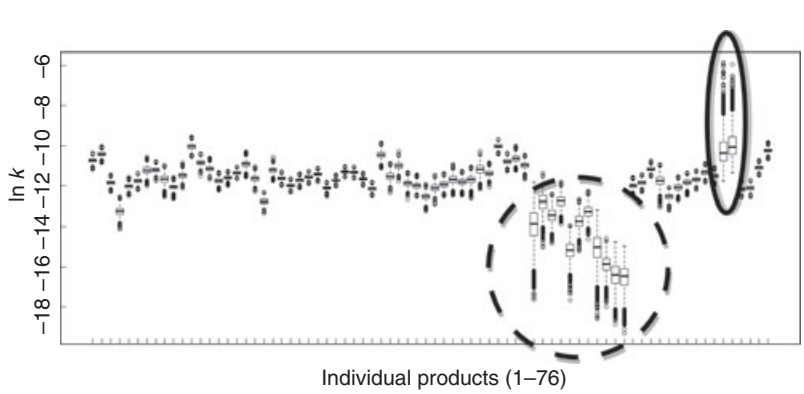

Figure 1 Box plots of individual test product posterior distributions. (⬚) The dashed oval indicated the lower bound of product variability, while (⬭) the solid oval shows the upper bound of observed product variability.

© 2012 No claim to US Government works/J. Mitchell-Blackwell & P.L. Gurian

tendency, as low as ln*k* of −16 are observed as well. This range of −16 to −10 indicates a factor of 400 (e^6) variability.

Predictive distribution for unobserved product

By observing the posterior distribution of the grand mean and variance describing the observed experiments, we derive a distribution from which we can make predictions about unobserved test products. The mean and median values of the distribution, which are very similar, were used to plot a predictive risk curve in Fig. 2 with associated confidence intervals. Figure 2 also includes the average predictive dose–response curve that is determined by integrating over the full range of variability and uncertainty in the parameters. The curve produced illustrates more conservative estimates of risks at lower doses, where data are lacking. The risk curves generated from parameter estimates of the most virulent product are included as a reference point for the upper bound risk predictions. These values fall within the credible bounds of risk from the predictive distribution approaching the upper limit.

Model verification

An internal check of validity consisted of comparing *values of k* estimated by both the BHA and MLE approach (Fig. 3). The MLE *values of k* and confidence intervals are compared to the posterior mean and credible intervals obtained by the BHA in Fig. 4. Values were not able to be estimated for six products using the MLE approach because of singularity in response. For the 70 products that could be compared, the *values of k* fall along the 1 : 1 correspondence line suggesting a nearly exact match (Fig. 3) between approaches. There are two exceptions for products at the lower end of the range where the BHA estimated *values of k* are higher than the MLE values.

Estimates of model uncertainty, as represented by the width of confidence intervals for the two approaches, are also largely convergent. A plot of 95% credible intervals from the BHA against 95% confidence intervals for the MLE (Fig. 4) shows that most products lie along the 1 : 1 correspondence line indicating very similar estimates of uncertainty from the two approaches. For the larger confidence intervals (i.e. when uncertainty is greatest), the BHA produces narrower credible intervals (points fall below 1 : 1 line) and this also corresponds to when there are fewer observations informing and estimate.

Goodness of fit

For the exponential model, the fit of 70 products was evaluated using likelihood ratio tests. As noted above, two (2) products had no observed fatalities and four (4) products had 100% fatalities, so they could not be estimated by maximum likelihood methods. The selection criterion of $P < 0{\cdot}01$ for rejection was used for the analysis. For 42 of the 70 products, the likelihood ratio test indicated acceptable fit. In consistency with Occam's razor, the most parsimonious model – the exponential model – was fit initially in this analysis. The beta-Poisson model (Haas *et al.* 1999) – a two parameter dose–response model – was explored as an alternative to the exponential dose–response model in the light of the goodness-of-fit test

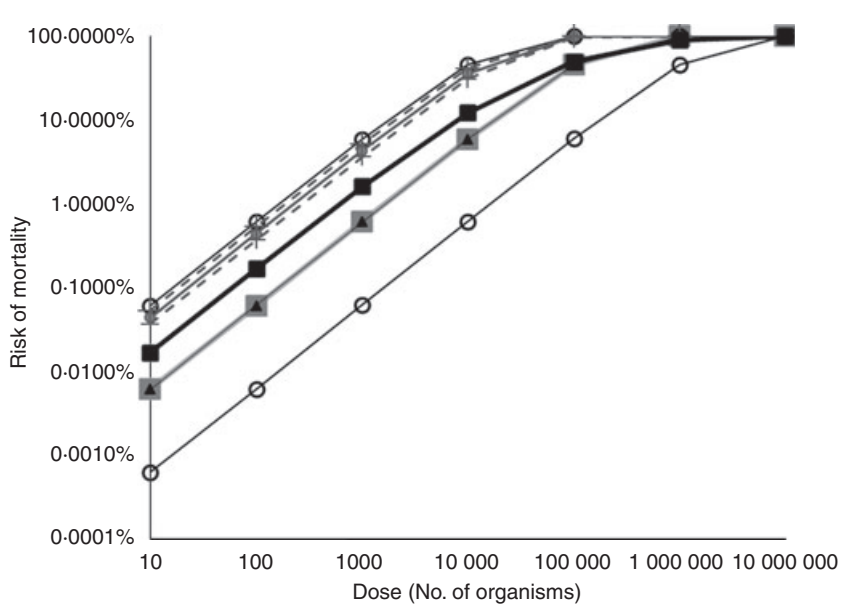

Figure 2 Predictive Risk Curve for Unobserved Products. Unobserved Products Mean (■); Unobserved Products Median (▲); Unobserved Products 5th percentile (○); Unobserved Products 95th percentile (○); Most Virulent Product Mean (●); Most Virulent Product 5th percentile (+); Most Virulent Product 95th percentile (+); Average Predictive Dose–response (■).

© 2012 No claim to US Government works/J. Mitchell-Blackwell & P.L. Gurian

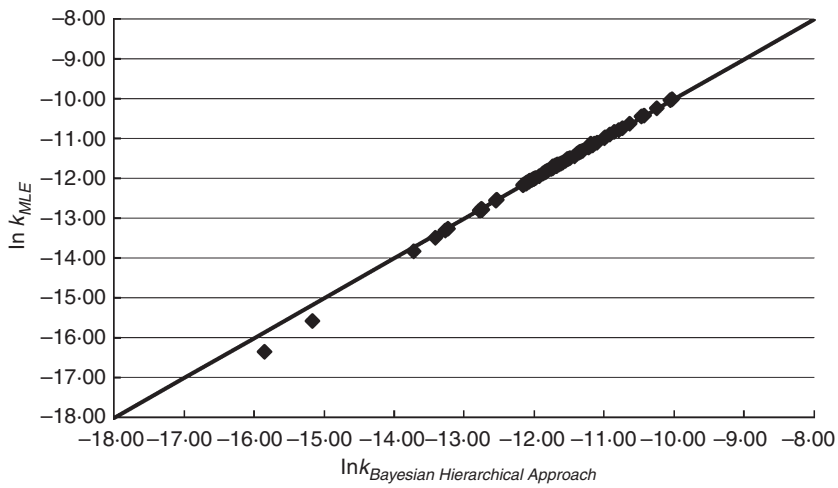

Figure 3 1 : 1 Correspondence between the Bayesian Hierarchical Approach estimate of dose–response parameter (ln*k*) and Maximum Likelihood Estimate (MLE) estimate of dose–response parameter (lnk_{MLE}). ln*k* (◆).

results. For seven products, the beta-Poisson provided acceptable fit when the exponential did not. While this does indicate that the exponential model was appropriate for most products, at least a third was not fit well by the exponential model. In interpreting this result, it is important to note that the test compares the proposed model with a perfectly fitting model (i.e. a model with a separate parameter for each dose group). Therefore, this is a very stringent test. A model may still have some predictive ability, and hence be useful, despite having some divergence from the data. A summary of these results for each product is provided in Table 1. The beta-Poisson model was recommended over the exponential model when it provided a statistically significant improvement in fit determined by minimized deviance.

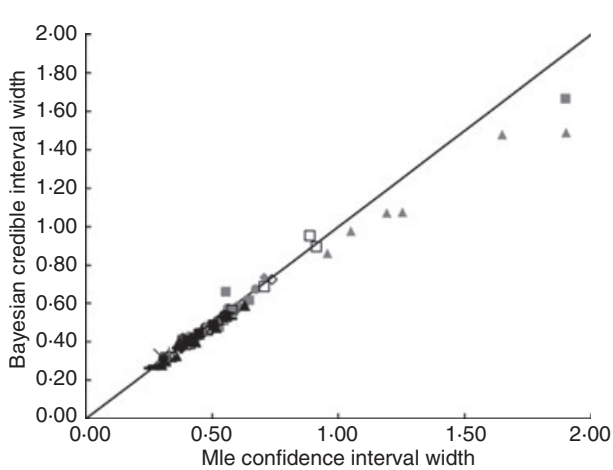

Figure 4 1 : 1 Correspondence Ratio of the 95th% Bayesian Credible Intervals and Maximum Likelihood Estimation (MLE) Confidence Intervals. ▲ 2 Observations (6 Products); ■ 8 Observations (6 Products); ● 9 Observations (6 Products); Δ 10 Observations (1 Product); □ 12 Observations (8 Products); ○ 15 Observations (6 Products); ◇ 16 Observations (4 Products); ◆ 17 Observations (2 Products); ▲ 18 Observations (9 Products); ■ 20 Observations (5 Products); ● 21 Observations (1 Product); ◆ 22 Observations (2 Products); - 24 Observations (11 Products); + 30 Observations (1 Product); × 32 Observations (1 Product) (Observations are the number of data points for each product.)

A second issue of model fit is whether the distribution of product virulence matches the proposed prior distribution, which was initially a subjective estimate. While the lognormal distribution is unbound (between 0 and 1), a requirement to represent *k* as a probability, the values observed in this study were not expected to violate this requirement. Figure 5 shows that there are some deviations between the distribution of *values of lnk* (for each product obtained by MLE) and a lognormal distribution (the assumed distribution for these values specified by the prior.) The observed values do not perfectly match theoretical quantiles of the lognormal distribution depicted by the solid line.

Benchmark estimates of virulence

A summary of possible parameter values are tabulated in Table 2. The approach produces a wide range of estimates of virulence, which can be used in a number of decision scenarios as suggested in the table. Lethal doses (LD10 and LD50) are also provided for comparative analysis. Both the central tendency and upper bound estimates of the Bayesian and MLE methods produce similar results. Highly conservative estimates can be drawn by using the 95th percentile of unobserved product predictive distribution or by observing the statistics from the posterior of most virulent product tested in the data set (with a similar MLE result). The low lethal dose values produced are in the proximity of values previously published with inadequate detail (Defense Intelligence Agency, 1986; Inglesby *et al.*, 2002).

© 2012 No claim to US Government works/J. Mitchell-Blackwell & P.L. Gurian

Table 1 Goodness-of-fit test summary

Product no.	Exponential model fits	beta-Poisson model fits	beta-Poisson model recommended	Product no.	Exponential model fits	beta-poisson model fits	beta-Poisson model recommended
1	Yes	Yes		39	Yes	Yes	
2	Yes	Yes		40	Yes	Yes	
3	Yes	Yes		41	Yes	Yes	
4	Yes	Yes		42	Yes	Yes	
5	Yes	Yes		43	No	No	
6	Yes	Yes		44	Yes	Yes	
7	No	No		45	No	Yes	Yes
8	Yes	Yes		46	Yes	Yes	Yes
9	Yes	Yes		47	No	No	
10	No	No		48	No	Yes	Yes
11	No	No		49	No	Yes	Yes
12	Yes	Yes		50		No animals died	
13	No	No		51	Yes	NA	
14	Yes	Yes		52	Yes	NA	
15	No	No		53	Yes	NA	
16	No	No		54	Yes	NA	
17	No	No		55	Yes	NA	
18	Yes	Yes		56	Yes	NA	
19	Yes	Yes		57		No animals died	
20	No	No		58	Yes	Yes	
21	Yes	Yes		59		No animals died	
22	Yes	Yes		60		No animals died	
23	No	Yes	Yes	61	Yes	Yes	
24	No	Yes	Yes	62	No	No	
25	Yes	Yes		63	No	No	
26	No	No		64	No	No	
27	Yes	Yes		65	Yes	Yes	
28	No	No		66	Yes	Yes	
29	No	No		67	Yes	Yes	
30	No	Yes	Yes	68	Yes	Yes	
31	No	No		69	Yes	Yes	
32	No	No		70	No	Yes	Yes
33	Yes	Yes	Yes	71		All animals died	
34	Yes	Yes		72		All animals died	
35	Yes	Yes		73	No	No	
36	Yes	Yes		74	No	No	
37	No	No		75	Yes	Yes	
38	Yes	Yes		76	Yes	Yes	

Discussion

Observations

A very wide range of virulence was observed across the *B. anthracis* experimental test product preparations analysed herein. Virulence was previously defined as a term used to describe the relative ability of a pathogen to cause harm in a host organism. Virulence can also be described as the degree of pathogenicity within a group of microorganisms as indicated by case fatality rates and/or the ability of the organism to invade the tissues of the host. The pathogenic capacity of an organism is known to be determined by its virulence factors (US National Library of Medicine, 2010). These factors can be described as structures or chemical constituents that contribute to pathophysiology and include toxins, biological structures like endospores and surface adhesion molecules. While the biological reasons were unknown or simply not reported at the time the experimental data were collected, it was established that varying both preparation and environmental conditions clearly contributed to the resistance and virulence properties of the endospores formed. In the present context, it should be noted that the spore

© 2012 No claim to US Government works/J. Mitchell-Blackwell & P.L. Gurian

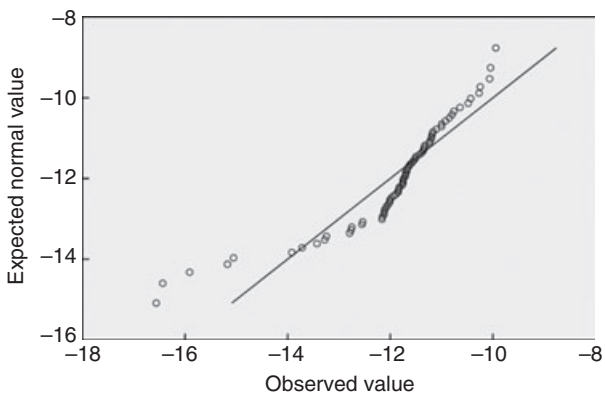

Figure 5 Distribution of Bayesian Hierarchical Approach mean *values of lnk* of each product. Mean of Posterior Distribution of ln*k* for each product (○).

adaptation of the micro-organism for dispersal and survival for extended periods of time also make it favourable for environmental transmission, persistence and proliferation. However, the present study is the first meta-analysis across experimental variables (associated with preparation and administration of the aerosol) to be conducted to determine the variability associated with these variables on virulence. The observed range of dose–response parameters (ln k) from −16 to −10 indicated a factor of 400 (e^6) variability, which compared to the variability associated with previously assessed sources, such as different host species (Mitchell-Blackwood 2010), microbial strains (Bartrand *et al.* 2008; Mitchell-Blackwood 2010) and aerosol sizes (Bartrand *et al.* 2008; Mitchell-Blackwood 2010), represents much larger differences.

The parameter estimates could have been produced by classical statistical methods, MLE, but this method only captures variability. As stated in the introduction, it is not only important to capture variability when assessing risks from exposure to *B. anthracis*, but also our uncertainty in this estimate because such a high hazard is associated with *B. anthracis* and we want to make decisions about future events using the most information available. In comparison with the MLE, it is reasonable to anticipate similarities between the central tendencies for *values of k* of individual products and MLE estimates given the use of a diffuse prior in the Bayesian hierarchical model structure. However, the confidence intervals can be expected to be smaller (tighter) by the BHA, as the BHA allows for information to be shared among different experiments. In fact, the tendency of the BHA to pull estimates for products towards the central tendency of other products is because of the sharing of information. Thus, products observed in Fig. 1 with very low *values of k* are pulled upward towards the overall mean. This is appropriate if the hyperdistribution accurately reflects the relatedness of different parameters. However, in this case, there may be aspects of the relationship among the products that are not captured by a single hyperdistribution. For example, some products may lack certain features, which would result in a markedly decreased virulence, or even cause them to be avirulent (not able to cause harm). In such a case, the assumption of a continuous distribution of virulence embodied in the hyperdistribution used herein would not be appropriate. However, it should be emphasized that the effect here is modest because the deviations are not large enough to produce an undesirable result. While the hyperdistribution used in this study is imperfect, it is health protective. The observed values are lower than the modelled (expected) values for both the upper and lower tails. As illustrated in Fig. 5, the points for both the upper and lower values of the observed range are above their corresponding modelled (expected) values, which are indicated by the solid line. Hence, the model produces conservative and therefore health-protective estimates.

Because limited information is available about what preparation might be used in a future event, the predictive curve for unobserved species (Fig. 2) was developed from the lower tier parameters of the Bayesian hierarchical model. The confidence intervals for the predictive distribution were much larger than those for individual experiments but this was expected because these predictions include variation across product types as well as uncertainty in the effects of a single product type. Intuitively, we would expect to have less confidence about the prediction for an unobserved product. The average predictive curve which integrated over the uncertainty in each estimate of risk made it clear that extrapolated values in the lower dose range have a greater degree of uncertainty than average parameter values consistent with the lack of observations in the experiments at this range of the curve.

Implications

Because of the intended applied nature of this work and the complexity of the associated downstream management responses, several selections of estimates were summarized in Table 2. The following discussion communicates how different estimates might be used as benchmarks in risk management decisions including exposure guideline development. While these values were derived from a small animal model, the application of the data to humans for risk management decisions is supported by the work published by Bartrand *et al.* (2008), which demonstrated statistically the ability to pool data sets between different host species. Additionally, the Bayesian

© 2012 No claim to US Government works/J. Mitchell-Blackwell & P.L. Gurian

Table 2 Benchmark dose estimates (with three significant figures) for two decision scenarios

Decision scenario	Model result	Parameter (ln *k*)	Benchmark Doses (organisms)	
			LD_{10}	LD_{50}
MLE – Central Tendency	Median of MLE estimates	−11·7	12 300	80 700
MLE – Conservative	Highest of observed products (95th percentile upper bound)	−10·0 (−9·81)	2320 (1920)	15 300 (12 600)
BHA – Central Tendency	Median (mean is similar to two significant digits) of predictive distribution	−12·0	17 100	113 000
	Average predictive distribution	n/a	7770	101 000
BHA – Conservative	Median of posterior distribution of most virulent product (95th percentile upper bound)	−10·03 (−9·85)	2360 (1970)	15 700 (13 100)
	95th percentile of predictive distribution (unobserved product)	−9·71	1720	11 500

BHA, Bayesian Hierarchical Approach; MLE, Maximum Likelihood Estimation; LD, lethal dose.

hierarchical framework has been used to conduct a meta-analysis of *Bacillus anthracis* challenge studies using different host species and strains of the organisms (Mitchell-Blackwood 2010). These results indicated that a data driven uncertainty factor of 10 may be appropriate for extrapolating between host species in such studies. This information should be weighed by a decision-maker utilizing these results.

There are several reasons why multiple values are required; (i) Benefit-cost analyses used to inform risk management decisions are generally based on mean or median estimates (Office of Management and Budget (OMB) (1992); (ii) In reality, responders need to be prepared for the worst-case scenario, and so they must have health-protective upper bound estimates of virulence; (iii) In some cases, a less virulent agent may be used. In such cases, less extensive remediation efforts may be appropriate, and some estimates of the lower bound of virulence are needed to inform preparations for such situations (Mitchell-Blackwood *et al.* 2011).

MLE approaches are often used for regulatory and guidance purposes as they are well established and understood to be asymptotically unbiased and efficient, making MLE approaches very favourable for large samples. Given existing guidance to base benefit-cost estimates on central tendencies, one might take the median of the observed *values of lnk* (Table 2), which in this case would be −11·7, corresponding to an LD_{10} of 12 300 and an LD_{50} of 80 700. However, this approach may not be appropriate for all circumstances. Health-protective, worst-case estimates may be needed initially to protect first responders and exposed individuals, before the details of the product are known. One straightforward manner to obtain a health-protective estimate of virulence is to select the most virulent product and base estimates on this product. This approach could be justified by assuming that an upper bound on virulence was reached during the course of developing the products included in this data set. This view is supported by the few points at the upper tail of the ln*k* estimates (Fig. 5). While several products have *values of lnk* of approx. −10, none exceed this value (despite higher expected values based on a normal distribution of *values of lnk*). This is consistent with the idea that there is some physical limitation on how virulent the products can be, and that this data set contains several products that reached this upper bound (i.e. the two in the upper right of Fig. 1 with observed *values of lnk* of approx. −10). Choosing the most virulent of the products dramatically changes the dose–response estimates. The most virulent product has an ln*k* estimate of −10·0, corresponding to an LD_{10} of 6320 and an LD_{50} of 15 300. If, as an additional measure to conservatism, the 95th percentile of the bootstrap MLE estimates is used, then an ln*k* of −9·81 is obtained corresponding to an LD_{10} of 1920 and an LD_{50} of 12 600.

The BHA offers several alternatives for estimates as well. The central tendency can be estimated by taking the posterior median value of the hyperdistribution, that is, the posterior median of the *values of lnk* predicted by the fitted model. This results in an ln*k* estimate of −12·0, corresponding to an LD_{10} of 17 100 and an LD_{50} of 113 000. As noted above, the hyperdistribution is not an exact fit and hence observed and modelled median values do not match exactly, but this approach is in at least rough agreement with the median of the MLE estimate (ln*k* = −11·7). An alternative measure of central tendency is to take the average of the risk predicted by different posterior samples of ln*k* (shown in Fig. 2). This average predictive distribution is often appropriate in microbial risk assessment because it integrates across different potential parameter values to estimate the central

© 2012 No claim to US Government works/J. Mitchell-Blackwell & P.L. Gurian

tendency of risk (Englehardt 2004; US EPA 2008). While there is no single ln*k* value to compare with other approaches, this method produces an LD_{10} of 7770 and an LD_{50} of 101 000, values somewhat lower than either the median of the MLE estimates and of the median of the Bayesian posterior distribution of ln*k*. This method accounts for the fact that risk is an exponential function of ln*k*. As benchmark values should be based on central tendencies of risk, not central tendencies of model inputs, this approach is proposed as being more appropriate for central tendencies of *B. anthracis* virulence than the other two methods.

The BHA also allows for more conservative estimates of virulence. Estimates of upper bounds for the most virulent of the observed products may be used, in which case very similar values to those obtained for this product by MLE are obtained (Table 2). An alternate approach is to take an upper bound of the posterior predictive distribution of ln*k*. If the 95th percentile is taken as an upper bound, then a value of −9·71 is obtained, which corresponds to an LD_{10} of 1720 and an LD_{50} of 11 500. These values indicate a greater virulence than any actually observed, exceeding even the 95th percentile of the most virulent product observed here. This may be seen as unrealistic or appropriately conservative, depending on how reliable the hyperdistribution is deemed. Given the lack of fit of the hyperdistribution used here, a great deal of confidence cannot be placed in these values, but they do at least allow for the reality that the sample of products observed here may not be indicative of the universe of potential products.

The central tendencies estimated here (given in Table 2) are in very rough alignment with the estimate of Bartrand *et al.* (2008) of an LD_{50} of 94 320 spores based on pooled guinea pig and monkey data. However, the conservative, health-protective values estimated here are considerably lower than the estimates of Bartrand *et al.* (2008) and more closely resemble the Defense Intelligence Agency (1986) values of 8000 to 10 000 spores (the 95th percentile posterior predictive value of 11 500 comes closest to this range).

It is acknowledged that the fit of both the individual product dose–response models and the overall hyperdistribution used by the BHA require further work. However, the use of LD_{50} values on the order of 10 000 spores can be justified by this work. The most lethal product was fit adequately by the exponential dose–response model and gives an LD_{50} value on this order.

Conclusions

This study evaluated a large data set to estimate *B. anthracis* virulence. A great deal of variability was found as estimates of the virulence of different products spanned a factor of 400. Both MLE and BHA approaches were found to generally provide similar estimates. The BHA approach produces narrower confidence intervals, particularly for sparsely sampled products, because the hierarchical model structure allowed sharing of information among products. Central tendencies of virulence observed here are in accordance with those of the study of Bartrand *et al.* (2008) rather than the values reported by Defense Intelligence Agency (1986) and are informed by a large body of evidence. However, the conservative, health-protective estimated here (LD_{50} = 11 500 organisms) are close to the range reported by Defense Intelligence Agency (1986). This work provides a statistical backing for the use of these more conservative, health-protective estimates of virulence. These values may be used to represent worst-case situations, in which the characteristics of the product are unknown and a highly health-protective value is desired. They could also be used to characterize the situation in which an intelligent adversary knowingly selects the most dangerous product from a large array of potential agents. Further work that considers alternative model structures to assess whether model fit can be improved is warranted.

Acknowledgements

The authors would like to thank Drs Tonya Nichols and Sarah Taft (US Environmental Protection Agency, National Homeland Security Research Center) for joint financial support, with the US Army, of the Pathogen Information Catalog (PICAT), which houses the data set, and their technical contributions to the project. Stephanie Hines (Battelle) also contributed during technical discussions. Ms. Erin Curran (US Army Public Health Command was instrumental in the gathering of the original research reports, and Peg Coleman, Stacey Massulik, Michele Stephenson, and Lisa Lowe (Syracuse Research Corporation) were instrumental in the population of the PICAT. Portions of this research were supported in part by an appointment to the Student Research Participation Program at the US Army Center for Health Promotion and Preventive Medicine (USACHPPM) administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and USACHPPM. Additional support was provided by the Center for Advancing Microbial Risk Assessment, which is supported by the US Department of Homeland Security and the US Environmental Protection Agency under the Science to Achieve Results (STAR) grant program (grant no. R83236201).

© 2012 No claim to US Government works/J. Mitchell-Blackwell & P.L. Gurian

References

Bartrand, T.A., Weir, M.A. and Haas, C.N. (2008) Dose-response models for inhalation of *Bacillus anthracis* spores: interspecies comparisons. *Risk Anal* **28**, 1115–1124.

Coleman, M., Thran, B., Morse, S., Hugh-Jones, M. and Massulik, S. (2008) Inhalation anthrax: dose response and risk analysis. *Biosecur Bioterror* **6**, 147–160.

Defense Intelligence Agency (1986) Soviet biological warfare threat. DST-161-OF-057-86.

Englehardt, J.D. (2004) Predictive Bayesian dose-response assessment for appraising absolute health risk from available information. *Human and Ecological Risk Assessment* **10**, 69–78.

Gilks, W., Richardson, S. and Spiegelhalter, D.J. (eds.) (1996) *Markov Chain Monte Carlo in Practice*. London, UK: Chapman and Hall.

Haas, C.N., Rose, J.B. and Gerba, C.P. (1999) *Quantitative Microbial Risk Assessment*. New York, NY, USA: John Wiley and Sons.

Hong, T., Gurian, P.L. and Ward, N.F.D. (2010) Setting risk-informed environmental standards for *Bacillus anthracis* spores. *Risk Anal* **30**, 1602–1622.

Huang, Y., Hong, T., Bartrand, T., Gurian, P.L., Haas, C.N., Liu, R. and Tamarakar, S. (2010) How sensitive is safe? Risk-based targets for ambient monitoring of pathogens *IEEE Sens J* **10**, 668–673.

Inglesby, T.V., O'toole, T., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Friedlander, A.M., Gerberding, J. *et al.* (2002) Anthrax as a biological weapon, 2002: updated recommendations for management. *JAMA* **287**, 2236–2252.

Jones, D.R., Peters, J.L., Rushton, L., Sutton, A.J. and Abrams, K.R. (2009) Interspecies extrapolation in environmental exposure standard setting: a Bayesian synthesis approach. *Regul Toxicol Pharmacol* **53**, 217–225.

Mitchell-Blackwood, J. (2010). *Using analytic Models for Risk-Based Responses to Pathogenic Agents in the Environment*. Ph.D. Doctoral, Philadelphia, PA: Drexel University.

Mitchell-Blackwood, J., Gurian, P. and O'donnell, C. (2011) Finding risk-based switchover points for response decisions for environmental exposure to *Bacillus anthracis*. *Human and Ecological Risk Assessment* **17**, 489–509.

National Research Council (2007) *Models in Environmental Regulatory Decision Making*. Washington, DC: National Academies Press.

National Research Council (2010) Review of the department of homeland security's approach to risk analysis. Committee to Review the Department of Homeland Security's Approach to Risk Analysis, ed.

Office of Management and Budget (OMB). (1992) Guidelines and discount rates for benefit-cost analysis of federal programs [Online]. Available: http://purl.access.gpo.gov/GPO/LPS46031, [Accessed 25 May 2012].

Thran, B., Lee, R., Mitchell-Blackwood, J. and Gurian, P.L. (2010) *Technical Guide 316 Supplement C2: Dose–Response Assessment for Inhalation Anthrax in Guinea Pigs Using Historical Army Data and Classical and Bayesian Statistical Methods*. Environmental Health Risk Assessment Program (MCHB-TS-REH) (ed). Aberdeen Proving Ground, MD: Limited Distribution through the US Army Public Health Command.

US EPA (2008) Development and evaluation of novel dose-response models for use in microbial risk assessment. In: National Center for Environmental Assessment (ed). Cincinnati, OH.

US National Library of Medicine. (2010) *MeSH – Medical Subject Headings* [Online]. Stockholm, Sweden: National Institutes of Health. Available: http://mesh.kib.ki.se/swemesh/show.swemeshtree.cfm?Mesh_No=G06.930andtool=karolinska [Accessed 2011].

© 2012 No claim to US Government works/J. Mitchell-Blackwell & P.L. Gurian