Dose–Response Modeling: Extrapolating From Experimental Data to Real-World Populations

Adrian Pratt,1 Emma Bennett,1 Joseph Gillard,2 Steve Leach,1 and Ian Hall1,3,*

Dose–response modeling of biological agents has traditionally focused on describing laboratory-derived experimental data. Limited consideration has been given to understanding those factors that are controlled in a laboratory, but are likely to occur in real-world scenarios. In this study, a probabilistic framework is developed that extends Brookmeyer’s competing-risks dose–response model to allow for variation in factors such as dose-dispersion, dose-deposition, and other within-host parameters. With data sets drawn from dose–response experiments of inhalational anthrax, plague, and tularemia, we illustrate how for certain cases, there is the potential for overestimation of infection numbers arising from models that consider only the experimental data in isolation.

KEY WORDS: Competing-risks framework; dose–response modeling; quantitative microbial risk assessment

1. INTRODUCTION

Mathematical models have long been used to help interpret experimental data arising from challenges of infectious disease agents to animals. Any model is an abstraction of reality and necessarily simplifies the underlying processes, often to arrive at a solution that is mathematically tractable. However, models of any particular phenomenon should be carefully constructed and the associated assumptions, where possible, fully evidence-based and clearly stated, especially if chosen for convenience in absence of such an evidence base.

One concept often used in challenge studies is that members of a population have a variety of inherent tolerance levels that, if exceeded by exposure level, result in response. Irwin was the first to coin the expression “individual effective dose” that has been widely used since (Irwin, 1937). In contrast to this concept which assumes “maximum synergism” between bacteria (Meynell & Stocker, 1957), models may alternatively assume the independent action of organisms. In such models, the variability with respect to host is replaced with stochasticity with respect to bacteria, whereby each individual organism has a probability of alone invoking response (through survival and sufficient proliferation). This is often termed a “hit” and one or more of these could be required to initiate response (Teunis & Havelaar, 2000). A third concept that may be considered is that of a bacterial threshold, which when reached results in host response. Here, the host starts with an initial load and through bacterial reproduction and death, the number of organisms eventually reaches zero (resolution) or a threshold (response).

Recent work on Bacillus anthracis has attempted to review dose–response models (both individual effective dose and single-hit) against a set of criteria (Toth et al., 2013). These criteria are that the model...
is fitted to data (rather than expert elicitation), the resulting fit is consistent with the pattern of infection arising from the Sverdlovsk incident (Meselson et al., 1994; Wilkening, 2006), the model is biologically motivated, and finally that the incubation period is inferred. Models more complicated than the exponential model (one form of single-hit type) are eventually rejected in Toth et al. (2013) because the extra complexity (as given by additional parameters) is not supported by the model fits (i.e., the fit is not statistically better with additional complexity and biological rationale in the model). These are sound conditions when interpreting a data set in isolation, however, no consideration was given to variation in say, individual susceptibility or clustered dosages when extrapolating away from experimental data to a human population. The purpose of this study is then to illustrate, with plausibly motivated but parsimonious models, the potential impact of extrapolating inferences based on experimental data to wider populations. In deriving models, we use language relevant to anthrax infection but will also discuss model fits to other biological organisms.

2. METHODS

To model the dose–response for an individual in a population to a challenge of B. anthracis, we assume that there are two key events for each bacterium: germination of the spore, after which point infection is inevitable in the host, or the removal of the spore from the system after which infection (arising from that spore) is impossible; this is the approach taken by Brookmeyer, Johnson, and Barry (2005). There is effectively a competition between the two processes and so if the removal of the spore were to occur at time t and germination at time u then if t ≤ u infection is stopped but if t > u infection occurs. This is conceptually often called a competing-risks framework: occurrence of one event precludes occurrence of the other.

There are a number of natural questions arising for any probabilistic interpretation of this framework:

(1) What distributions/processes govern σ and τ within an individual?
(2) What impact does inhalation of multiple spores have on the response?
(3) What differences are there in σ and τ (or related values) between individuals?

Here we attempt to provide a framework to address these questions for infection with B. anthracis. We will develop a general model below, then make use of three experimental data sets to illustrate the potential impact of extrapolating away from carefully calibrated experimental settings to wider scenarios. In the absence of data for these wider scenarios, we illustrate on an order of magnitude basis how changes in the additional parameters can influence inferred infection rates. Limitations in deriving and parameterizing the models are discussed.

2.1. The Competing-Risks Framework

One may derive the single-hit models from the competing-risks framework, and this derivation is repeated here for transparency and to allow clarity in assumptions made and extensions possible. Let us assume that the germination time for each spore in each individual is modeled by some probability density function \( f_G(t, \Theta_G) \) with associated parameter set \( \Theta_G \) and mean \( \mu_G \) and similarly the removal time is given by \( f_R(t, \Theta_R) \) and mean \( \mu_R \), with corresponding cumulative distribution functions, \( F_G \) and \( F_R \), respectively. We assume that germination and removal are independent processes.

The probability \( \beta \) that \( \sigma \geq \tau \) for any given spore in each exposed individual (and thus infection subsequently occurring) is then

\[
\beta = \int_0^\infty \int_0^\infty f_G(t) f_R(s) ds dt
\]

\[
= \int_0^\infty h_G(t) \exp \left( - \int_0^t h_G(u) du + \int_0^t h_R(u) du \right) dt, \quad (1)
\]

where \( h_G \) and \( h_R \) represent the hazard functions of germination and removal, respectively. If the hazard function over time for germination is proportional to the hazard of removal then \( h_R = \gamma h_G \) and

\[
\beta = \frac{1}{1 + \gamma}, \quad (2)
\]

and so a single parameter describes the underlying process, the ratio of the removal time to germination time. In the situation where the hazards of germination and removal are constant so the removal and germination times are Exponentially distributed with means \( \mu_R \) and \( \mu_G \), (and it follows that \( \gamma = \mu_R/\mu_G \), the assumption used from this point forward in article), this is equivalent to the form for \( \beta \) observed by Brookmeyer et al. (2005).

For an individual challenged with a dose \( D \) of spores, some fraction \( K \) (where \( K \sim \text{Binomial}(D, \phi) \))
with number of trials \( D \), and probability of success \( \phi \) will be deposited in the lung space. From this point, the competing-risks framework (as discussed above) may be applied. Alternatively one could construct a form of \( h_c \) and \( h_R \) in (1) that includes deposition directly, but for simplicity we continue with the former formulation explicitly coupling this faster-acting event. The parameter \( \phi \) may vary depending on activity level of the host and particle size inhaled, but will be assumed not to vary between individual spores, while \( D \) will depend on (atmospheric) dispersion of the agent and so will depend on the individual’s position relative to the source. Each inhaled spore \( i \) will then be subject to a Bernoulli test with probability \( \phi \beta \) of causing infection in the host. The probability \( Q \) of removal of all spores without infection occurring is then

\[
Q = (1 - \phi/(1 + \gamma))^D, \tag{3}
\]

which can be evaluated, provided with some estimate of the parameters \( D, \gamma \), and \( \phi \), for an individual receiving the challenge. At a population level these values will vary, here we propose possible options for modeling this variation.

### 2.2. Aerosol Dispersion

Ordinarily one assumes that the number of spores in a unit volume of inhaled air is modeled as a Poisson process (Haas, Rose, & Gerba, 2014), that is we assume that spores are randomly distributed in the atmosphere local to the inhaler and with expected dose proportional to the amount of air inhaled, \( D \sim \text{Pois}(d) \). In this case, we find the dose-response function takes the form of an Exponential function

\[
Q = \exp(-\alpha d), \tag{4}
\]

where \( \alpha = \phi/(1 + \gamma) \). If the expected inhaled dose \( d \) is overdispersed then the uncertainty in dose may be reasonably modeled by a \( \Gamma \)-Poisson process (more regularly called a Negative-Binomial distribution), and the resulting form of \( Q \) is given by

\[
Q = \left[1 + \frac{\phi}{\xi(1 + \gamma)}\right]^{-\xi d}, \tag{5}
\]

where the parameter \( \xi \in (0, \infty) \) (arising from the negative-binomial distribution) controls the level of overdispersion. This equation can be rewritten in the form of an exponential

\[
Q = \exp(-\hat{\alpha} d), \tag{6}
\]

where \( \hat{\alpha} = \xi \ln(1 + \phi/(\xi(1 + \gamma))) \). Given that Equations (4) and (6) have the same functional form, fitted to a single experimental data set we clearly cannot tell the amount of overdispersion from the resulting value of the superparameter, without further contextual information.

### 2.3. Variation in Germination and Removal

We note from Equation (2) that \( \gamma \) is a ratio and so if we assume that the mean germination and mean removal times, \( \mu_G \) and \( \mu_R \), respectively, are both \( \Gamma \) distributed in a particular host such that

\[
\mu_G \sim \Gamma(g; \alpha_G, \beta_G), \quad \mu_R \sim \Gamma(r; \alpha_R, \beta_R)
\]

then the ratio \( \gamma = \mu_R/\mu_G \) can be shown to have the probability density function

\[
f(\gamma; \alpha_G, \beta_G, \alpha_R, \beta_R) = \frac{\Gamma(\alpha_G + \alpha_R)}{\Gamma(\alpha_G)\Gamma(\alpha_R)} \left(\frac{\beta_R}{\beta_G}\right)^{\alpha_R} \gamma^{\alpha_R - 1} \left(1 + \frac{\beta_R\gamma}{\beta_G}\right)^{-\alpha_R - \alpha_G}.
\]

The parameter \( \beta_X \) (where \( X \) represents either \( G \) or \( R \)) is the index of dispersion and we define \( \kappa = \beta_G/\beta_R \) for future notational brevity.

If the index of dispersion is the same for germination and removal, so \( \kappa = 1 \), Equation (4) becomes

\[
rdQ = \gamma F_1 \left( \frac{1}{1 + \lambda}; \rho; -\phi d \right), \tag{7}
\]

where \( \gamma F_1 \) represents the generalized hypergeometric function (Slater, 1966) given by

\[
pFq(\alpha_1, \ldots, \alpha_p; \beta_1, \ldots, \beta_q; z) = \sum_{i=0}^{\infty} \frac{(\alpha_1)_n \ldots (\alpha_q)_n}{(\beta_1)_n \ldots (\beta_p)_n} \frac{z^n}{n!}, \tag{8}
\]

and \((x)_n\) denotes a rising factorial or Pochhammer symbol. Note, the generalized hypergeometric function will be used throughout this article. In Equations (4) and (6), the impact of overdispersion compared to a Poisson-distributed dose could not be determined as the models had the same functional form, however, if overdispersion in dose is included within (7), the dose-response relationship becomes

\[
Q = \gamma F_1 \left( \frac{\xi d}{1 + \lambda}; \rho; -\phi/\xi \right). \tag{9}
\]

The assumption that the dispersion indices for germination and removal are equal is returned to later
in this article. It is worth noting that Equation (7) is called the “Beta-Poisson” model (Teunis & Havelaar, 2000) and when \( \rho \rightarrow \infty \) but \( \lambda \sim O(1) \) then

\[
Q \sim \left(1 + \frac{\phi d}{\rho}\right)^{-\rho/(1+\lambda)}.
\]

This approximate dose–response relationship is commonly applied in the literature and confusingly also called the Beta-Poisson model (Haas et al., 2014; Huang, Bartrand, Haas, & Weir, 2009; Tamrakar, Haluska, Haas, & Bartrand, 2011; Teunis & Havelaar, 2000).

2.4. Deposition

The parameter \( \phi \) models the between-host variability in the dose that actually is retained following exposure. As this is a probability, it is sensible to assume this follows a beta-distribution (with associated parameters \( \alpha_\phi \) and \( \beta_\phi \)). Indeed, Heppell, Egan, and Hall (2017) find outputs from the multiple particle dosimetry (MPPD) model (Price, Asgharian, Miller, Cassee, & de Winter-Sorkina, 2002) can be approximated well with a beta-distribution. Incorporation of this additional source of variability can considerably increase the complexity of the dose–response model, depending on what other factors are considered and upon what restrictions are placed on these assumptions. With an overdispersed dose, and variation in germination and removal rates (albeit with equal index of dispersion), the dose–response relationship is defined by

\[
Q = 3F_2 \left( \zeta d, \phi \omega, \frac{\rho}{1+\lambda}; \rho, \omega; -d \right).
\]

The parameters have been rewritten as \( \phi = \alpha_\phi / (\alpha_\phi + \beta_\phi) \) which is the mean of the given beta-distribution and \( \omega = \alpha_\phi + \beta_\phi \) which controls variation (note, the true variance can then be written as \( \phi(1-\phi)/(1+\omega) \)). A slightly simpler relation is arrived at if the dose is assumed to be Poisson-distributed (i.e., not overdispersed):

\[
Q = 3F_2 \left( \phi \omega, \frac{\rho}{1+\lambda}; \rho, \omega; -d \right).
\]

2.5. Additional Models

All previous models have assumed that the index of dispersion is equal in germination and removal rates. If we remove this constraint, the models can quickly become intractable. However, where the dose is overdispersed and variation in deposition is not permitted (as in Equation (9)), allowing for an index of dispersion not equal to unity gives

\[
Q = k^{\alpha_R} \frac{\Gamma(\alpha_R + \alpha_G)}{\Gamma(\alpha_R) \Gamma(\alpha_G)} \int_0^\infty \frac{\gamma^{\alpha_R-1}}{(1+\gamma)^{\alpha_R+\alpha_G}} \left[1 + \frac{\phi}{\xi(1+\gamma)}\right]^{-\zeta d} dy
\]

\[
= k^{\alpha_R} \frac{\Gamma(\alpha_R + \alpha_G)}{\Gamma(\alpha_R) \Gamma(\alpha_G)} \int_0^1 \frac{u^{\alpha_R-1}(1-u)^{\alpha_R-1}}{(1-(1-\kappa)u^{\alpha_R+\alpha_G}(1+\phi(1-u)/\xi))} du
\]

\[
= k^{\alpha_R} \frac{\Gamma(\alpha_R + \alpha_G)}{\Gamma(\alpha_R) \Gamma(\alpha_G)} \sum_{i=0}^\infty \frac{1}{\Gamma(i+1)} \left(-\phi \right)^i \frac{\Gamma(i+\zeta d) \Gamma(\alpha_G)}{\Gamma(\zeta d) \Gamma(\alpha_R)} \int_0^1 \frac{u^{\alpha_R-1}(1-u)^{\alpha_R-1}}{(1-(1-\kappa)u)^{\alpha_R+\alpha_G}} du
\]

\[
= k^{\alpha_R} \sum_{i=0}^\infty \frac{1}{\Gamma(i+1)} \left(-\phi \right)^i \frac{\Gamma(i+\zeta d) \Gamma(\alpha_G)}{\Gamma(\zeta d) \Gamma(\alpha_R)} \frac{\Gamma(\rho)}{\Gamma(\rho+i) F(\alpha_R, \rho; \rho+i; 1-\kappa)}
\]

\[
= k^{\alpha_R} \sum_{i=0}^\infty \sum_{j=0}^\infty \frac{(\zeta d)(\alpha_G)(\alpha_R)}{\Gamma(i+1) \Gamma(j+1) \Gamma(\rho) \Gamma(\rho+i+j)} \left(1-\kappa \right)^i \frac{(-\phi)^j}{\xi}.
\]

where \( F_3 \) is the Third Appell hypergeometric function (Appell, 1925) and we have defined \( \rho = \alpha_R + \alpha_G \) and \( \lambda = \alpha_R/(\alpha_G+\kappa) \) to be the ratio of the underlying distribution mean values.

Allowing for a beta-distribution to describe the variation in deposition modifies the dose–response
function from Equation (13) to

\[
Q = \kappa a R \sum_{i=0}^{1} \sum_{j=0}^{\infty} \frac{\Gamma(i+1)\Gamma(j+1)(\rho)^{i+j}}{\Gamma(i+1) \Gamma(j+1)} \left( \frac{-\phi}{\zeta} \right)^i \frac{\phi^{a \phi-1} (1-\phi)^{\beta \phi-1}}{B(a, \beta)} d\phi
\]

\[
= \kappa a R \sum_{i=0}^{\infty} \frac{\Gamma(i+1)}{\Gamma(i+1)} \left( \frac{-1}{\zeta} \right)^i
\]

(14)

where \( \phi \) is the mean of the beta-distribution that models \( \omega = \alpha_0 + \beta_0 \) controls the associated variance of the distribution. This can be rewritten in the notation of a Kampé de Fériet function (Weisstein, Weisstein) given by

\[
Q = \kappa a R \text{Kampé de Fériet}_{1:1:0}^{0:3:2} \left( \rho \left| \begin{array}{c} \zeta d, \alpha_G, \varphi \omega \end{array} \right. \alpha_R, \rho \left| \begin{array}{c} (1-\kappa), -\frac{1}{\zeta} \end{array} \right. \right).
\]

(15)

This formula captures variability around individuals at roughly the same geographic location, though some of these parameters could be expected to change over greater distances. Note, both Equations (13) and (15) can be challenging to evaluate numerically.

A further two cases are included in this section for completeness. These are the dose–response models generated by assuming that no variation is present in removal and germination rates, though allowing for variation in deposition rates. These models are given by

\[
Q = _1F_1(\varphi \omega; \omega; -\frac{d}{(1+\lambda)}).
\]

(16)

\[
Q = _2F_1(\zeta d, \varphi \omega; \omega; -\frac{1}{(1+\lambda)})
\]

(17)

for Poisson-distributed, and \( \Gamma \)-Poisson-distributed doses, respectively. There is little reason to expect these \textit{a priori}, though it is interesting to note that they take the same functional form as Equations (7) and (9), albeit with a different interpretation of the parameters.

### 2.6. Summary and Implementation

A complete list of dose–response models discussed is given in Table I and associated parameters are given in Table II. In this article, the standard R functions are used for simpler models (Equations (4) and (6)), the \textit{gsl} package in R is used for evaluating \( _1F_1 \) and \( _2F_1 \) functions (Equations (7), (9), (16), and (17)). When \( \phi > \zeta \) the final argument is larger than one and convergence may be an issue, in this situation formula 15.3.5 from

| Table I. List of Models for Different Sources of Variation |
|---------------------------------|-----------------|-------------|
| Dispersion Type | Variance in Deposition | Variance in Germination & Removal | Form of Q | Eq. No. |
|-----------------|-----------------|-------------|-----------------|--------|
| Poisson | × | × | \( \exp \left( -\frac{\phi d}{(1+\lambda)} \right) \) | 4 |
| \( \Gamma \)-Poisson | × | × | \( \exp(-\zeta \ln \left( 1 + \frac{d}{(1+\lambda)} \right)) d \) | 6 |
| Poisson | ✓ | × | \( _1F_1(\varphi \omega; \omega; -\frac{d}{(1+\lambda)}) \) | 16 |
| \( \Gamma \)-Poisson | ✓ | × | \( _2F_1(\zeta d, \varphi \omega; \omega; -\frac{d}{(1+\lambda)}) \) | 17 |
| Poisson | × | \( k=1 \) | \( _1F_1(\frac{\zeta d}{\Gamma(\zeta)}; \rho; -\phi d) \) | 7 |
| \( \Gamma \)-Poisson | ✓ | \( k=1 \) | \( _2F_1(\zeta d, \varphi \omega; \omega; -\frac{d}{\gamma}) \) | 9 |
| Poisson | ✓ | \( k=1 \) | \( _2F_2(\varphi \omega; \frac{d}{\Gamma(\zeta)}; \rho, \omega; -d) \) | 12 |
| \( \Gamma \)-Poisson | ✓ | \( k=1 \) | \( _3F_2(\zeta d, \varphi \omega; \frac{d}{\Gamma(\zeta)}; \rho, \omega; -\frac{d}{\gamma}) \) | 11 |
| \( \Gamma \)-Poisson | × | \( k \neq 1 \) | \( \kappa a R F_3(\zeta d, \frac{d}{\Gamma(\zeta)}, \frac{d}{\Gamma(\zeta)}; \rho; \gamma; \rho; -\phi \gamma(1-\kappa)) \) | 13 |
| \( \Gamma \)-Poisson | ✓ | \( k \neq 1 \) | \( \kappa a R F_3^{0:3:2}(\zeta d, \alpha_G, \varphi \omega; \alpha_R, \rho; (1-\kappa), -\frac{1}{\zeta}) \) | 15 |
Table II. Parameter Summary

| Parameter | Description                                      | Notes                          |
|-----------|--------------------------------------------------|--------------------------------|
| $\gamma$ | Ratio of mean germination and removal times      |                                |
| $D$       | Inhaled dose                                     | $D \sim \text{NB}(\Delta, \delta)$ |
| $\Delta$ | Negative-Binomial distribution shape parameter   |                                |
| $\delta$ | Negative-Binomial distribution scale parameter   |                                |
| $d$       | Mean inhaled dose                                | $d = \Delta \delta$            |
| $\zeta$  | Overdispersion in dose                           | $\zeta = 1 + \delta$           |
| $\phi$   | Lung deposition fraction                         | $\phi \sim \text{Beta}(\alpha \phi, \beta \phi)$ |
| $\alpha_\phi$ | Beta-distribution parameter                     |                                |
| $\beta_\phi$ | Beta-distribution parameter                     |                                |
| $\psi$   | Superparameter—Mean lung deposition              | $\psi = \alpha_\phi / (\alpha_\phi + \beta_\phi)$ |
| $\omega$ | Superparameter—Controls variance of $\phi$      | $\omega = \alpha_\phi + \beta_\phi$ |
| $\gamma$ | Ratio of mean rates of removal and germination   | $\gamma = \mu_R / \mu_G$        |
| $\mu_R$  | Mean rate of spore removal                       | $\mu_R \sim \text{Gamma}(\alpha_R, \beta_R)$ |
| $\mu_G$  | Mean rate of spore removal                       | $\mu_G \sim \text{Gamma}(\alpha_G, \beta_G)$ |
| $\rho$   | Superparameter                                  | $\rho = \alpha_R + \alpha_G$    |
| $\lambda$| Superparameter—ratio of mean rates               | $\lambda = \alpha_R \beta_R / \alpha_G \beta_G$ |
| $\kappa$ | Ratio of dispersion indices for germination and removal | $\kappa = \beta_G / \beta_R$ |

Abramowitz and Stegun (1972) are used to improve convergence and compared with formula 15.3.4 and most stable result for dose regime considered is chosen. There are no further identified issue with numerical convergence of these functions for parameterizations inferred once one of these transforms is made. We call a Python library for 2F2 function (12) which suffers convergence for large parameter choices, in this study this is an issue for the anthrax data set below. We do not numerically evaluate the more complex functions derived for additional models given by (11), (13), or (15) in this study.

3. RESULTS

3.1. Application of Theory to Dose–Response Data Sets

In the previous section, equations were derived from the competing-risks framework that allow for variation in dose-dispersion, dose-deposition, and in the germination and removal processes. In typical dose–response experimental data, some of these factors may naturally vary and hence in attempting to describe the data, models should include them. However, other factors are controlled for and any extrapolation to real-world scenarios necessitates an understanding of these parameters. In this section, we investigate these additional factors.

The language used throughout this article has focused on the sporulating bacterium *B. anthracis*. We make use of data from two additional bacteria: *Yersinia pestis*, the causative agent of plague; and *Francisella tularensis*, the causative agent of tularemia. Descriptions of the mode of infection and the suitability of both agents for applying the modeling framework discussed in this study are given below. Data sets are identified for all three agents.

The pathogenesis of *Y. pestis* infection (by inhalation) is complicated and not fully understood. Several factors have been elucidated and center on the ability of *Y. pestis* to suppress and evade the host immune system in an initial anti-inflammatory phase (Li & Yang, 2008). To initiate infection, *Y. pestis* is initially phagocytozed by host leucocytes. If taken up by neutrophils bacteria are eliminated, whereas in macrophages they may resist cell-mediated destruction, express a number of virulence factors, then escape host cells with a capacity to proliferate with reduced risk of further phagocytosis (Cavanaugh & Randall, 1959; Lukaszewski et al., 2005). Thus success or failure in surviving macrophage destruction may determine whether infection persists or the bacterium is eliminated. This model of infection is analogous to that in the competing-risks model discussed throughout this article, and is thus an appropriate mathematical model to describe inhalational plague infection. Here we make use of two data sets that, to our knowledge, have not been used in published dose–response models before. Both data sets involve inhalational exposures of CO92 strain *Y. pestis* to *Cynomolgus* macaques. These data sets were obtained from Van Andel et al. (2008) and Warren et al. (2011); from this point forward we refer to the combined data sets simply as the plague data set. Fit to an
exponential model (Equation (4)), the ID$_{50}$ (dose required to cause infection in 50% of the exposed, susceptible population) for this combined data set is 88 organisms. Data were collected by exposing 40 participants to individual doses.

In a previous study, Wood, Egan, and Hall (2014) demonstrate that a mechanistic birth–death-survival model for inhalational tularemia infection is approximated well by an exponential model and can be interpreted as a hit model. The hit could be the failed elimination of a bacterium following phagocytosis: an unsuccessful attempt to kill allows the bacterium to multiply to number several hundred within the macrophage, before rupture and release into the extracellular environment for the process to be repeated. Similarly to Y. pestis, this describes a competing-risks type framework (clearance vs. survival and proliferation) and thus we can use the models developed in this article. The data sets featured in Wood et al. are human inhalational experiments using the virulent SCHU-S4 strain bacteria (Saslaw, Eigelsbach, Prior, Wilson, and Carhart, 1961; Sawyer, Dangerfield, Hogge, and Crozier, 1966; Sawyer, Jemski, et al., 1966). Fit to an exponential model, the ID$_{50}$ for this combined data set is 10 organisms. These data only include a relatively small range of doses and so are unlikely to support more complicated models when making inference directly but in this context the effect of structure in the models is being shown. The Saslaw study exposed 20 volunteers to a range of doses, with doses of 10, 20, 23, and 46 organisms being received by two participants each (Saslaw et al., 1961).

A third comparison data set is obtained from the experiment of Altboum et al. (2002). Here, guinea pigs were exposed inhalationally to Vollum strain B. anthracis. Fit to an exponential model, the ID$_{50}$ for this data set is $1.5 \times 10^5$ organisms. In total, 51 guinea pigs were used in study with 4, 8, 12, 12, 8, and 7 receiving each dose in increasing order.

### 3.2. Considering Overdispersion in Dose

In typical laboratory-controlled dose–response experiments (including the data sets we have chosen to use), we may assume that dosages are unstructured, i.e., they are Poisson-distributed. It is also reasonable to assume that deposition rates show little variation between hosts since activity levels will be similar. On this basis, the dose–response relationship given by Equation (7) may be an appropriate choice to describe our data. We thus begin by fitting our three experimental data sets to this model. Best-fit parameters obtained for each data set and deposition value are given in Table III. To investigate the impact of overdispersion in the dose, we substitute these best-fit parameters into the model given by Equation (9), and vary the unknown overdispersion parameter $\zeta$, on an order of magnitude basis (recalling that $\zeta \in (0, \infty)$). It is important to notice that changing $\zeta$ in this way shows the impact of such overdispersion rather than fitting the overdispersed model to the data, having the mechanistic structures allows prediction to be made of the effect of overdispersion.

Three values of $\zeta$ are considered ($\zeta = 10, 1, 0.1$). Fig. 1 illustrates for the three featured data sets (left to right: anthrax, plague, and tularemia) and a deposition rate of 0.1, the effects of varying the dispersion parameter $\zeta$. Figures for the other deposition–dispersion combinations are near identical to their parent models and hence omitted for brevity, however Table IV lists the response rates at the ID$_{50}$ of the original (parent) model for all parameter combinations.

For the anthrax data set, there is no discernible difference between dose–response relationships that allow the overdispersion parameter $\zeta$ to vary; this is true regardless of the level of deposition assumed. For plague and tularemia (which have lower ID$_{50}$s), allowing for overdispersion yields dose–response relationships that predict lower infection rates across all doses though the magnitude of the effect for most cases is modest (see Table IV). The impact of the overdispersion appears to increase as each of the dispersion parameter, deposition rate, and ID$_{50}$ of the data set shrinks. Only in the extreme, i.e., for the tularemia data set with a deposition rate of 0.1, and dispersion parameter $\zeta = 0.1$ do we see a noteworthy difference; the response rate at the previous ID$_{50}$ has reduced to 38.64% (rounded to four significant figures).

In most cases then, the added complexity of including overdispersion is not warranted, and the simpler parent model given by Equation (7) is valid, however when dealing with a highly infective organism, or a very low deposition rate, it may be important to consider the level of overdispersion.

It is interesting to note that in graph C (deposition value of 0.1, tularemia data set) the response rate appears not to tend to unity as the dose increases. This is not a numerical artifact but instead an artifact that the maximum-likelihood estimate for $\rho < 1$ in which case (7) tends to $\lambda/(1 + \lambda)$. Exploration of the likelihood function in the parameter-space suggests
### Table III. List of Parameter Fits for Various Deposition Values

| Deposition | Anthrax | Plague | Tularemia |
|------------|---------|--------|-----------|
| = 0.9      | ρ = 10.115, λ = 18.417 | ρ = 5.622, λ = 10.234 | ρ = 1.124, λ = 2.046 |
| = 0.5      | ρ = 64.47, λ = 52.65  | ρ = 35.79, λ = 28.96  | ρ = 7.02, λ = 5.24  |
| = 0.1      | ρ = 4.77, λ = 4.17   | ρ = 2.65, λ = 2.09   | ρ = 2.94 × 10⁻⁴, λ = 9.69 × 10⁻² |

---

![Fig 1](image)

**Fig 1.** Dose–response fits for (left to right columns) anthrax, plague, and tularemia data sets, deposition level 0.1, for various overdispersion parameter values. Data points on plot may represent multiple responses as detailed for each study in Section 3.1.

### Table IV. Response Rate for Different Levels of Overdispersion at Original ID₅₀

| Deposition | Anthrax | Plague | Tularemia |
|------------|---------|--------|-----------|
| = 0.9      | ζ = 10  | ζ = 1  | ζ = 0.1  |
|            | 50.00   | 50.00  | 50.00    |
| = 0.5      | ζ = 10  | ζ = 1  | ζ = 0.1  |
|            | 49.99   | 49.97  | 47.78    |
| = 0.1      | ζ = 10  | ζ = 1  | ζ = 0.1  |
|            | 49.72   | 47.65  | 34.9     |

---

that the given parameter values do indeed maximize the likelihood, though an alternative parameter with smaller likelihood tend to unity in a reasonable dose-dependent manner.

### 3.3. Considering Variation in Deposition

The impact of variation on deposition can also be considered independent of overdispersion. Equation (12) was generated by considering a beta-distribution on the deposition parameter ϕ. The parameter ϕ is fixed by the mean deposition rate, however the remaining parameter ω that controls the variance, is free to vary.

The variation in deposition parameter ω is considered for three values: 2, 20, and 200, where increasing values of ω indicate reduced variability in the deposition rate. Note that these values of ω were chosen to be at least 2 to prevent the underlying beta-distribution from being bimodal with local maxima at zero and unity; this would not be expected biologically for these causative agents, but in generality is this framework is applied to other infections where some inherent protection is extant in population the limit on ω may be relaxed.

**Table V.** Impact of incorporating variation into the deposition parameter on the dose–response relationship, and Table V details the impact
Dose–response fits for plague and tularemia data sets (top and bottom rows, respectively), deposition levels 0.9, 0.5, 0.1 (left to right, columns), for various levels of variability in deposition. Data points on plot may represent multiple responses as detailed for each study in Section 3.1.

**Table V.** Response Rate for Different Levels of Variation in Deposition at Original ID<sub>50</sub>

| Deposition | ω = 200 | ω = 20 | ω = 2 | ω = 200 | ω = 20 | ω = 2 | ω = 200 | ω = 20 | ω = 2 |
|------------|---------|--------|-------|---------|--------|-------|---------|--------|-------|
| Plague     | 50.02   | 49.95  | 49.41 | 49.95   | 49.39  | 45.12 | 49.43   | 44.96  | 27.48 |
| Tularemia   | 50.00   | 49.94  | 49.38 | 49.93   | 49.34  | 44.98 | 49.42   | 45.06  | 27.83 |

The differences observed when considering uncertainty on the deposition parameter are visually evident in both data sets from a deposition level of 0.5; they are however still relatively modest, for instance, the new response rate for the plague data set with a deposition rate of 0.5 and ω = 2 is 45.12%. A
significant reduction in response is observed for a deposition rate of 0.1 and $\omega = 2$ for both the plague and tularemia data sets with response rates of 27.43% and 27.83%, respectively. These results suggest that as the deposition rate decreases, the uncertainty around that deposition plays a larger role in determining the probability of infection.

For the purposes of this article, we have provided examples for deposition and dispersion independently. It is possible to consider the combined effects of variation using Equation (11), however the visualization becomes increasingly complex as the dimensionality increases.

4. DISCUSSION

A central paradigm in mathematical modelling is to build the simplest model that describes the data. While this may generally be true, if the simplest dose–response models fit solely to laboratory-controlled data are used to make inferences regarding real-world exposures, there is the possibility for an overestimation of the number of infected cases.

In this work, we have extended Brookmeyer’s competing-risks model of inhalational anthrax infection to consider how variation or uncertainty may impact upon the predicted response rate. Candidate models have been generated by considering general distributions for prescribed sources of variation. Parent models have been parameterized from real inhalational data sets for a variety of organisms, and the parameters used to inform the more complex child model. From this, we have been able to speculate on the impacts that population heterogeneity or parameter uncertainty may have on infection probabilities.

In the examples we have shown, allowing for variation in overdispersion has had minimal impact on the dose–response relationship in all but the most extreme case. The results suggest that the impact increases as the ID$_{50}$ for the data set, the deposition rate, and the overdispersion parameter $\omega$ decrease. This could be regarded as a high-infectivity effect, and thus, consideration of whether this model provides any added benefit over the parent model depends upon the organism being modelled. In contrast to dispersion, the impact of variation on deposition appears to be important at higher deposition levels, and potentially independent of the ID$_{50}$ of the data set. As the deposition rate decreases, increasing variability in the deposition rate results in a smaller response rate. Again, assessing whether this model should be used may be dependent on the deposition level, and hence may be organism- and host-dependent.

While the framework initially described is formulated for inhalational anthrax infection, we have shown examples in tularemia and plague for which the same model could be applied. The central concept binding all three examples is that in the case of independent action, organisms may behave in a manner such that a single bacterium has two possible fates: removal (from the system, by any means) or survival with subsequent proliferation resulting in infection. The focus on inhalation as route of infection in this study is not necessarily a barrier to translation to other modes of transmission, such as ingestion or cutaneous infection. The methods may have a broader applicability worthy of further study.

To our knowledge, there are no published dose–response models with application to human inhalational plague (i.e., applied to data of inhalational or intranasal exposure of human or nonhuman primate hosts). Using the same concept as the competing-risks model for inhalational anthrax infection, a model for inhalational plague parameterized from primate dose–response data has been posited. Further work is required to validate the modeling assumptions made, however in the absence of other models, we recommend this model be used for inhalational plague risk assessment. Note that the estimates made by this model are subject to the same concerns regarding all models that attempt to explain only the experimental data; they may overpredict infection estimates when variation is not taken into account.

Throughout the model generation process, different realizations of variation have yielded models with the same functional form. For example, assuming Poisson-distributed dispersion of organisms, a beta-distribution on the deposition parameter would produce a model with the same functional form as a model allowing for gamma distributions on the variation in germination and removal rates (with equal dispersion index). Furthermore, some authors (Teunis & Havelaar, 2000) consider an exponential model for dose–response where the unknown parameter is simply the probability of an individual bacterium causing infection, and a beta-distribution on this parameter allows for population heterogeneity at the level of either host or bacterium. Each interpretation produces models with the same functional form, and hence model fit alone is insufficient in
providing evidence that a specified set of assumptions gives rise to observed behavior. Clearly, some understanding of the experimental conditions (and therefore model selection) is required if model users wish to make inferences about the processes that dictate infection events.

Mathematical tractability has been a problem throughout this article. For instance, the hypergeometric functions are often computed as a power series, and subject to computational overflows when certain parameters are sufficiently large. One measure of whether \( \frac{1}{2}F_2 \) may be simulated for a data set is in the \( ID_{50} \). For large \( ID_{50} \), the infection generating event rate must be small (relative to the removal rate), resulting in at least one large parameter. This was the case for the anthrax data set, and hence the model was omitted in Fig. 2. There is therefore a tradeoff in terms of model complexity versus practicality. This may not be the case if alternative descriptions of variation are used that do not result in a Gaussian hypergeometric function and numerical and mathematical representation is subject to further work.

The modeling assumptions we have made to generate a set of candidate models require further work to validate their appropriateness and so this analysis can be considered as illustrative only. However, the assumptions made have used general functions and so serve some purpose beyond simple illustration. Further work is required to consider if the distributions we have considered capture real-world variability, or if alternatives need to be investigated.

In the context of emergency response and planning, often the most risk-averse scenario is considered. Here, this equates to selecting the competing-risks model (with the same functional form as an exponential model), as it gives a higher number of infections than other candidate models discussed in this article. However, the uncertainty needs to be considered when assigning reasonability to the worst-case scenario. If the assumption that deposition rates can be described by a nonbimodal beta-distribution is valid, then (as previously discussed) simulations for which \( w = 2 \) may provide a bound on this uncertainty. In addition, the analysis presented here suggests when uncertainty and variation should be considered, that is, for highly infective organisms, and for low deposition rates. The work presented here examining uncertainty and variation in the absence of data therefore contributes to the overall risk assessment.

5. CONCLUSION

The competing-risks model of bacterial infection can be extended to incorporate sources of variation that are often controlled for in laboratories yet likely to arise in real-world scenarios. Model realizations involving additional variational parameters can be partially parameterized from their simpler parent models fitted to real data. Exploration of remaining unknown parameters has shown that model predictions may overestimate infection numbers when simpler models fit to laboratory-controlled data are used in isolation.

ACKNOWLEDGMENTS

This study was funded by the UK Home Office. The research was also partially funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emergency Preparedness and Response at King’s College London and in Modelling Methodology at Imperial, both in partnership with Public Health England (PHE). IH is Principal Investigator of the NIHR Policy Research Programme in Operational Research for Emergency Response Analysis (OPERA, PR-R17-0916-21001). IH and SL are also Members of NIHR Health Protection Research Units in Emerging and Zoonotic Infections and IH is a Member of the NIHR HPRU in Gastrointestinal Infections, both at Liverpool. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health, Home Office, or Public Health England. The authors declare that they have no competing interests.

REFERENCES

Abramowitz, M., & Stegun, I. A. (1972). Handbook of mathematical functions. New York, N.Y.: Dover Publications Inc.
Allbourn, Z., Gozes, Y., Barnea, A., Pass, A., White, M., & Kobiler, D. (2002). Postexposure prophylaxis against anthrax: Evaluation of various treatment regimens in intranasally infected guinea pigs. Infection and Immunity, 70(11), 6231–6241.
Appell, P. (1925). Sur les fonctions hypergéométriques de plusieurs variables, les polynomes d’hermite et autres fonctions sphériques dans l’hyperespace. Mémorial des sciences mathématiques, 3, 1–75.
Brookmeyer, R., Johnson, E., & Barry, S. (2005). Modelling the incubation period of anthrax. Statistics in Medicine, 24(4), 531–542.
Cavanaugh, D. C., & Randall, R. (1959). The role of multiplication of Pasteurella pestis in mononuclear phagocytes in the pathogenesis of flea-borne plague. Journal of Immunology, 83(4), 348–363.
Haas, C. N., Rose, J. B., & Gerba, C. P. (2014). Quantitative Microbial Risk Assessment. Hoboken, NJ: Wiley.
Heppell, C. W., Egan, J. R., & Hall, I. (2017). A human time dose response model for Q fever. *Epidemics, 21*, 30–38.

Huang, Y., Bartrand, T., Haas, C., & Weir, M. (2009). Incorporating time postinoculation into a dose–response model of *Yersinia pestis* in mice. *Journal of Applied Microbiology, 107*(3), 727–735.

Irwin, J. (1937). Statistical method applied to biological assays. *Supplement to the Journal of the Royal Statistical Society, 4*(1), 1–60.

Li, B., & Yang, R. (2008). Interaction between *Yersinia pestis* and the host immune system. *Infection and Immunity, 76*(5), 1804–1811.

Lukaszewski, R. A., Kenny, D. J., Taylor, R., Rees, D. C., Hartley, M. G., & Oyiston, P. C. (2005). Pathogenesis of *Yersinia pestis* infection in BALB/c mice: Effects on host macrophages and neutrophils. *Infection and Immunity, 73*(11), 7142–7150.

Meselson, M., Guillemin, J., Hugh-Jones, M., Langmuir, A., Popova, I., Shelokov, A., & Yampolskaya, O. (1994). The Sverdlovsk anthrax outbreak of 1979. *Science, 266*(5188), 1202–1208.

Meynell, G., & Stocker, B. (1957). Some hypotheses on the aetiology of fatal infections in partially resistant hosts and their application to mice challenged with *Salmonella paratyphi-B* or *Salmonella typhimurium* by intraperitoneal injection. *Microbiology, 16*(1), 38–58.

Price, O., Asgharian, B., Miller, F., Cassee, F., & de Winter-Sorkina, R. (2002). Multiple path particle dosimetry model (MPPD v1.0): A model for human and rat airway particle dosimetry. RIVM rapport 650010030.

Saslaw, S., Eigelsbach, H. T., Prior, J. A., Wilson, H. E., & Carhart, S. (1961). Tularemia vaccine study: II. Respiratory challenge. *Archives of Internal Medicine, 107*(5), 702–714.

Sawyer, W. D., Jemski, J. V., Hogge, A. L., Eigelsbach, H. T., Wolfe, E. K., Dangerfield, H. G., ..., Crozier, D. (1966). Effect of aerosol age on the infectivity of airborne *Pasteurella tularensis* for *Macaca mulatta* and man. *Journal of Bacteriology, 91*(6), 2180–2184.

Slater, L. J. (1966). Generalized hypergeometric functions. Cambridge: Cambridge University Press.

Tamrakar, S. B., Haluska, A., Haas, C. N., & Bartrand, T. A. (2011). Dose–response model of *Coxiella burnetii* (Q fever). *Risk Analysis, 31*(1), 120–128.

Teunis, P., & Havelaar, A. (2000). The Beta Poisson dose-response model is not a single-hit model. *Risk Analysis, 20*(4), 513–520.

Toth, D. J., Gundlapalli, A. V., Schell, W. A., Bulmahn, K., Walton, T. E., Woods, C. W., ..., Adler, F. R. (2013). Quantitative models of the dose–response and time course of inhalational anthrax in humans. *PLoS Pathogens, 9*(8), e1003555.

van Andel, R., Sherwood, R., Gennings, C., Lyons, C., Hutt, J., Gigiotti, A., & Barr, E. (2008). Clinical and pathologic features of cynomolgus macaques (Macaca fascicularis) infected with aerosolized *Yersinia pestis*. *Comparative Medicine, 58*(1), 68–75.

Warren, R., Lockman, H., Barnewall, R., Krile, R., Blanco, O. B., Vasconcelos, D., ..., Fellows, P. (2011). Cynomolgus macaque model for pneumonic plague. *Microbial Pathogenesis, 50*(1), 12–22.

Weisstein, Eric W. Kampé de Fériet Function. MathWorld–A Wolfram Web Resource. Retrieved from https://mathworld.wolfram.com/KampedeFerietFunction.html

Wilkening, D. A. (2006). Sverdlovsk revisited: Modeling human inhalation anthrax. *Proceedings of the National Academy of Sciences, 103*(20), 7589–7594.

Wood, R., Egan, J., & Hall, I. (2014). A dose and time response Markov model for the in-host dynamics of infection with intracellular bacteria following inhalation: With application to *Francisella tularensis*. *Journal of the Royal Society Interface, 11*(95), 201401192–10.