A method for incorporating a time-dose-response model into a *Giardia lamblia* outbreak

Bidya Prasad, Michael O. Ryan and Charles N. Haas

ABSTRACT

Experimental time-to-infection data is a useful, but often underutilized, material for examining the mechanics of *in vivo* pathogen growth. In this paper, the authors attempt to incorporate a time-dose-response (TDR) equation into a model which predicts the number of ill persons per day in a *Giardia lamblia* epidemic using data collected from a Pittsfield, Massachusetts outbreak. To this end, dose-response and TDR models were generated for *Giardia* exposure to beaver and human volunteers, and a maximum likelihood estimation approach was used to ensure that the models provided acceptable fits. The TDR equation that best-fit the human data was the beta-Poisson with exponential-reciprocal dependency model, and this was chosen to be incorporated into the outbreak model. The outbreak model is an expanded probability model that convolutes an assumed incubation distribution of the infectious agent with an exposure distribution. Since the beta-Poisson with exponential-reciprocal dependency models the time-to-infection density distribution, it is input as the incubation distribution. Several density functions, including the Weibull, lognormal, gamma, and uniform functions served as exposure distributions. The convolution of the time-dependent probability distribution with the lognormal distribution yielded the best-fit for the outbreak model.

Key words | dose-response, epidemiology, *Giardia lamblia*, outbreak model, time-dose-response

Bidya Prasad (corresponding author)
Michael O. Ryan
Charles N. Haas
Department of Civil, Architectural and
    Environmental Engineering,
Drexel University,
3141 Chestnut Street, Curtis Hall, Room 251,
Philadelphia,
PA 19104,
USA
E-mail: *bp348@drexel.edu*

INTRODUCTION

Giardia lamblia (syn. *duodenalis* or *intestinalis*), referred to herein as *Giardia*, is a zoonotic flagellated parasite that infects a wide variety of mammalian hosts through the fecal-oral route. Common transmissions occur either through direct contact with the fecal matter of an infected person, or through cyst ingestion, which often occurs through the vehicle of contaminated food or water (Erlandsen & Meyer 2013). *Giardia* cysts are immediately infectious once they are shed, and are protected by a hard outer shell that allows them to survive outside the body for several weeks or even months, and makes them tolerant to chlorine disinfection (Olson *et al.* 1999). Human dosing studies have shown that the median infectious dose for *Giardia* via the oral route is between 10 and 100 cysts, and an infected person might shed $1-10^9$ cysts daily for several months

(Rendtorff 1954). Infection may be acute or chronic, with susceptible populations such as children and the elderly often having difficulty in clearing the cysts. Although carriers can be asymptomatic, gastroenteritis normally begins 1 to 3 weeks after infection, which can lead to severe diarrhea and life-threatening sickness (Heyman 2004).

Due to its stability and widespread distribution, *Giardia* is considered to be one of the most common causes of protozoan diarrhea in the world. In a 2007 review of worldwide waterborne outbreaks caused by parasitic protozoa in developed countries, Karanis *et al.* found that *Giardia* was the etiological agent responsible for 32% of total outbreaks (Karanis *et al.* 2007). In the United States, the total number of reported cases of *Giardia* was 19,140 in 2008 (Yoder *et al.* 2010), but since infected populations may be

asymptomatic or lack access to sufficient health care, some experts estimate that the actual number of giardiasis in the United States is in the order of 2×10^6 (Mead *et al.* 1999). Annual global incidences of giardiasis are estimated to be at 2.8×10^8 infections per year, and in 2004, *Giardia* was added to the World Health Organization (WHO) Neglected Diseases Initiative (Savioli *et al.* 2006).

Despite such prevalence, there have been few outbreak models for *Giardia* that forecast transmission of the pathogen into susceptible populations. The susceptible–infected–recovered (SIR) compartment model proposed by Kermack–McKendrick in 1927 remains the cornerstone of disease modeling (Kermack & McKendrick 1927), and has been expanded upon numerous times since its inception (Anderson & May 1991; Hethcote 1996). In 1999, Gupta reworked the compartment model so that density distributions were assumed for parameters that had historically been point estimates (Gupta 1999). Gupta's method could be further improved upon by assuming a density distribution that closely models the incubation time (the period between infection and clinical disease) for *Giardia*. The incubation density distribution would be assumed from time-dose-response (TDR) modeling.

This TDR model, which quantifies a relationship between pathogen kinetics and host response, provides the means by which dosing study equations can be incorporated into outbreak models (Huang & Haas 2009). Between 2009 and 2011, Huang *et al.* published a series of papers that aimed to quantify the onset time for clinical disease by building incubation time into the framework of traditional exponential and beta-Poisson dose-response equations (Huang *et al.* 2009; Huang & Haas 2009, 2011). These TDR equations can be incorporated into the SIR compartment model. To date, a TDR model has not been generated for *Giardia*, nor has an incubation distribution been incorporated into a disease model for this pathogen.

The TDR outbreak model quantifies the fate and transport dynamics of infectious agents, and such a model can be used to determine if the host-pathogen dynamic can adequately describe a public health risk. Predictive modeling of disease transmission and risk plays a crucial role in public health responses during an outbreak. The results of this work can be used to improve risk assessments for *Giardia*, and can enable health authorities to make prompt decisions in controlling the spread of infectious diseases.

The purpose of this study is to: (i) generate and discuss the results stemming from a non-time-dependent DR model for *Giardia*; (ii) determine the best-fit TDR model for *Giardia*; (iii) incorporate this model into the incubation distribution of an outbreak model; and (iv) compare the results of the time-dependent model with other non-time-dependent models. Criteria and parameter values are also investigated.

METHODS

This study was broken down into four parts, each of which built upon the previous step. The following tasks were undertaken, as summarized in Figure 1: (i) a literature review of dose-response and outbreak datasets; (ii) modeling of time-dependent dose-response using animal and human studies; and (iii) the development and utilization of an outbreak model which incorporates a selected TDR equation into the model. A literature search for usable datasets for the dose-response and outbreak models was performed first. The dose-response datasets from the literature search were then modeled using the TDR equations, and the best-fit model among them was chosen. From the initial literature search, a usable *Giardia* outbreak dataset was found, and this dataset was modeled by convoluting an exposure density function with an incubation density function. The exposure density functions that were explored were the Weibull, gamma, lognormal and uniform. The best-fit TDR (in the form of a density function) was used to model the incubation density function in the outbreak model. The work builds upon the TDR models developed by Huang *et al.* (Huang *et al.* 2009; Huang & Haas 2009, 2011), and the outbreak model developed by Gupta (1999).

Datasets for TDR and outbreak models

Dose-response dataset

A literature search was first performed to find datasets from which a dose-response model could be generated. The inclusion criteria for choosing the data were that the article included: (i) a clear description of dosing methods; (ii) reported mode of exposure; (iii) the dose the subjects were given; (iv) the number of subjects that experienced adverse
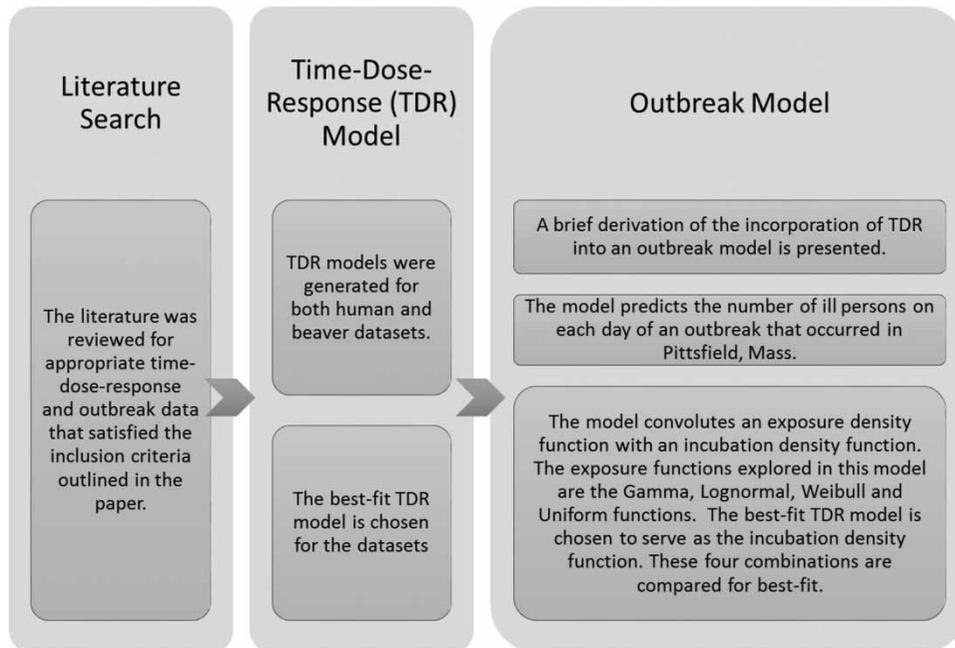
Figure 1 | Outline of methods. The methods can be divided into three parts: (i) the literature search; (ii) the time-dose-response (TDR) model; and (iii) the outbreak model.

effects on each respective day; (v) the criteria used to define a positive endpoint; (vi) a detailed description of the pathogen (i.e. source and strain); (vii) at least one observation 0 < p < 1; and (viii) the days to adverse response for each individual subject (Haas *et al.* 1999).

Outbreak dataset

The literature was then searched again to find an appropriate outbreak that could be fit to our model. The inclusion criteria for selecting an appropriate outbreak for the *Giardia* disease model included: (i) that the size of population susceptible to infection was known or could be estimated; (ii) members of this susceptible population who became symptomatically ill were confirmed cases; (iii) the beginning and end of the pathogen contamination is known; (iv) there is incubation data for the pathogen; (v) the outbreak occurred over a defined period of time; (vi) the exposure period is short compared to the outbreak; (vii) each confirmed case comes from the original source and there is no secondary spread of disease; or, if there is some secondary spread, there is a clear distinction between the primary and secondary spread; and (viii) that all outbreak cases

that are identified have the same case definition. The final criterion is necessary because case definitions have evolved over the years, and a case of infection defined decades ago may not be the same as today. A laboratory confirmed case of giardiasis is defined as the detection of *Giardia* organisms, antigen, or DNA in stool or other fluid and tissue sample (Wharton *et al.* 1990). A probable case occurs when a person has the same symptoms as the confirmed cases, which are specified by the epidemiologist in the outbreak investigation; these cases may be associated with a common source (exposure) or another case, if secondary transmission is appropriate (Wharton *et al.* 1990).

Dose-response models

A Cochran-Armitage test of trend was applied to determine whether there was an association between increasing dosages and adverse effects (Haas *et al.* 1999; Neuhäuser & Hothorn 1999). Written formally, the Z-statistic is calculated by

$$Z_{CA} = \frac{\sum_{i=1}^{k}(x_i - \bar{x})p_i}{\sqrt{\bar{p}(1 - \bar{p})\sum_{i=1}^{k} n_i(x_i - \bar{x})^2}} \tag{1}$$

where

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} \quad \text{and} \quad \bar{p} = \frac{\sum_{i=1}^{k} p_i}{\sum_{i=1}^{k} n_i} \qquad (2)$$

The natural log of the mean dose in group i is $x_i = \ln(d_i)$, with d_i total subjects and p_i positive subjects. The null hypothesis of lack of trend is rejected if Z_{CA} is above the upper 5th percentile of the normal distribution, which is 1.64 for a one-tailed test. If such an association is determined, the data can be modeled.

Traditional dose response predicts the risk of adverse consequences (endpoints of illness or death) from infection through a given dose (Haas *et al.* 1999). The most commonly used dose-response models are the Exponential or approximate beta-Poisson equation, respectively (Haas *et al.* 1999):

$$P(d) = 1 - e^{-kd} \qquad (3)$$

$$P(d) = 1 - \left[1 + \frac{d}{N_{50}} \times \left(2^{1/\alpha} - 1 \right) \right]^{-\alpha} \qquad (4)$$

The approximate beta-Poisson (Equation (4)) was developed from the Exponential (Equation (3)) by replacing the parameter k with a beta distribution, and has historically been used in dose-response modeling (Haas *et al.* 2014). Both expressions assume that: (i) a single surviving pathogen is sufficient to initiate infection; (ii) the host ingests one or more organisms capable of causing an adverse response; (iii) only a fraction of these pathogens reach a site of infection; (iv) individuals ingest a number of microorganisms representing a random sample from a Poisson

distribution; and (v) within the host, the survival of any microorganism is independent of the survival of any other microorganism (Haas *et al.* 2014).

These non-time-dependent exponential and beta-Poisson models were modified by Huang *et al.* to incorporate a time dependency of parasitic multiplication by including additional parameters of time post inoculation (TPI). This additional time parameter would quantify the time of onset of an effect presumably associated with the kinetics of *in vivo* bacterial growth. Huang developed eight separate equations based on time dependency in total (Huang *et al.* 2009; Huang & Haas 2009, 2011). These models have an empirical basis, and were developed from model parameters that were plotted against time for various survival dose-response datasets. While these TDR models take the form of the traditional dose-response equations, they do not reduce to them as time approaches infinity.

Huang's models were tested by the authors to determine if they were true cumulative distribution functions (CDFs). The criteria for determining true CDF distributions were that (i) the function at time zero is zero, and the function at time infinity is one and (ii) the functions are monotonic, which means that for all a and b, with $a \leq b$, $F(a) \leq F(b)$, and the derivative of the function is greater than or equal to zero for all of time. Of these eight models, only four were determined to be true CDFs in time, and are presented in Table 1. The models that were not true CDFs were excluded from the study.

The TDR models were fit to the dose-response datasets. Using Huang *et al.*'s methods of time-dose-response data analysis, which utilized the same statistical methods as the traditional dose-response model, TDR models were fit to

Table 1 │ True CDF time-dependent dose-response models

Model name	Parameter TPI dependency equation	Expression	Eq. number
Exponential model with exponential-reciprocal time dependency	$k = e^{-k_0/t}$	$P(d) = 1 - e^{-e^{(-k_0/t)}d}$, where $k_0 > 0$	(5)
Beta-Poisson model with exponential-reciprocal time dependency	$N_{50} = e^{(j_0/(t+j_1))}$	$P(d) = 1 - \left[1 + \frac{d}{e^{(j_0/(t+j_1))}} \times \left(2^{1/\alpha} - 1 \right) \right]^{-\alpha} where\, j_0 \gg j_1$	(6)
Beta-Poisson model with exponential time dependency	$N_{50} = e^{j_0 \times t + j_1}$	$P(d) = 1 - \left[1 + \frac{d}{e^{(j_0 \times t + j_1)}} \times \left(2^{1/\alpha} - 1 \right) \right]^{-\alpha}$	(7)
Beta-Poisson model with inverse time dependency	$N_{50} = e^{(j_0/t)+1}$	$P(d) = 1 - \left[1 + \frac{d}{((j_0/t) + 1)} \times \left(2^{1/\alpha} - 1 \right) \right]^{-\alpha}$	(8)

the datasets, parameters were optimized, and deviances were determined using maximum likelihood estimation in R.

The likelihood ratio for the incidence occurring on the jth day with i dose groups is

$$
\begin{aligned}
LR &= \prod_{j=1}^{m_{times}} \prod_{i=1}^{m_{doses}} \frac{\binom{n_{i,j}}{p_{i,j}} \pi_{i,j}^{p_{i,j}} (1-\pi_{i,j})^{n_{i,j}-p_{i,j}}}{\binom{n_{i,j}}{p_{i,j}} \left(\pi_{i,j}^{\circ}\right)^{p_{i,j}} \left(1-\pi_{i,j}^{\circ}\right)^{n_{i,j}-p_{i,j}}} \\
&= \prod_{j=1}^{m_{times}} \prod_{i=1}^{m_{doses}} \left(\frac{\hat{\pi}_{i,j}}{\pi_{i,j}^{\circ}}\right)^{p_{i,j}} \left(\frac{1-\hat{\pi}_{i,j}}{1-\pi_{i,j}^{\circ}}\right)^{n_{i,j}-p_{i,j}}
\end{aligned}
\tag{9}
$$

where m_{doses} is the total number of dose groups and m_{times} is the number of time periods during which observations were made (generally recorded in day-long segments), $p_{i,j}$ is the number of positive responses observed during time period j, and $n_{i,j}$ is number of surviving animals at the beginning of time period j. The predicted response is π_i, and the response for each set based on observations is π_i°.

The corresponding deviance is

$$
\begin{aligned}
Y &= -2 \ln LR \\
&= -2 \sum_{j=1}^{m_{times}} \sum_{i=1}^{m_{doses}} \left[p_{i,j} \ln\left(\frac{\hat{\pi}_{i,j}}{\pi_{i,j}^{\circ}}\right) + (n_{i,j}-p_{i,j}) \ln\left(\frac{1-\hat{\pi}_{i,j}}{1-\pi_{i,j}^{\circ}}\right) \right]
\end{aligned}
\tag{10}
$$

This expression is nearly the same as that used in fitting traditional dose-response data in which only the long-term endpoint is known (Haas *et al.* 1999). Differences between this expression and the expression normally used in dose-response modeling are that there is a double sum in the deviance expression and that p is the number of positive responses in a time period rather than at the end of observations.

Outbreak model

The SIR compartment model was developed by Kermack and McKendrick in their seminal papers (Kermack & McKendrick 1927), and has been extensively documented since (Anderson & May 1991). A modified SIR model was proposed by Gupta in his 1999 thesis (Gupta 1999) and utilized by the authors (Figure 2). In this model, populations are compartmentalized into groups, with individuals moving through the compartments as they become infected. The number of people in the compartments, as well as the rates of transfer are described by a system of differential equations.

In this model, populations are compartmentalized into groups, with individuals moving through the compartments as they become infected. The number of people in the compartments, as well as the rates of transfer are described by a system of differential equations. The four states proposed by Gupta are the Susceptible populations $X(t)$; Latent individuals $Y(t)$ who have been exposed to the pathogen and will ultimately become infected; those who become Symptomatically ill $Z(t)$; and those who have become infected but will remain Asymptomatic $I(t)$. The rates in these equations are given by the force of infectivity $\beta(t)$, which describes the rate of individuals going from the susceptible population to the latent population, and the rate of newly symptomatic cases is described by $Q(t)$.

Data is generally only available for the symptomatically ill populations $Z(t)$, because these patients are the ones to seek medical attention, which goes on to be reported to public health authorities. Asymptomatic data $I(t)$ is hardly ever reported, because persons who do not exhibit symptoms do not seek medical attention. Therefore, for our purposes, only the symptomatically ill
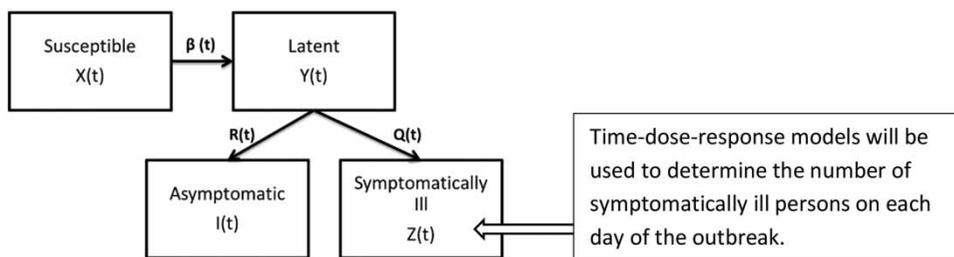

Figure 2 | Modified SIR model proposed by Gupta (Kermack & McKendrick 1927; Gupta & Haas 2004).

population will be considered, with $Z(t)$ data mined through epidemic studies.

The compartment model then reduces to the following equations:

$$\frac{dX(t)}{dt} = -\beta(t)X(t), \text{ or } \frac{1}{X}\frac{dX(t)}{dt} = -\beta(t) \tag{11}$$

$$\frac{dZ(t)}{dt} = Q(t) \tag{12}$$

The force of infection $\beta(t)$ has units of inverse time, and is a versatile parameter. In the case of a rectangular distribution for infectivity, $\beta(t)$ would be a constant for the exposure duration. However, common source outbreaks are frequently caused by time variable microbial exposures, which means that the parameter takes the form of a density distribution, generally of the following form:

$$\beta(t) = b_0 + b_1 g(t) \tag{13}$$

where $\beta(t)$ is the force of infection; b_0 is the background infectivity level, which may account for any endemic cases in the population; $g(t)$ is the ratio of susceptible persons ultimately becoming infected at time t; and b_1 is the scaling factor for increased infectivity above background. Here, $g(t)$ is modeled as a probability density function (PDF) for the exposure case. The density functions in this case that we have used are the Weibull, gamma, lognormal, and uniform, all which are standard functions which have historically been used to fit observed epidemic curves. Each of these density functions have parameter values that determine the shape and scale of the function. Integrating Equation (11) gives:

$$X(t) = X_0 exp[-B(t)], \text{ where } B(t) = \int_0^t \beta(u)du \tag{14}$$

The rate of newly ill people at any time is given by the parameter $Q(t)$. At any time t, the instantaneous rate of new illnesses can be obtained by the convolution integral:

$$\frac{dZ(t)}{dt} = Q(t) = \int_0^t \beta(\tau)X(\tau)f(t-\tau)d\tau \tag{15}$$

where $f(t-\tau)$ is the density function having incubation period of $(t-\tau)$. The function f is a PDF which has parameters that need to be estimated. This method of deriving $Q(t)$ follows the mathematics of Kermack and McKendrick, and Gupta.

Substituting Equation (14) into Equation (15) we then obtain:

$$\frac{dZ}{dt} = \int_0^t X_0\beta(\tau)\exp[-B(\tau)]f(t-\tau)d\tau \tag{16}$$

where $B(t) = b_0 t + b_1 G(t)$ and $G(t)$ is the cumulative distribution function (CDF) corresponding to the PDF $g(t)$. Since we are using the PDFs Weibull, gamma, lognormal and uniform to model $g(t)$, we necessarily must use the corresponding CDFs for $G(t)$ in each respective case.

Integrating Equation (16) directly, we can solve for the number of symptomatically ill patients per day $Z(t)$:

$$Z(t) = \int\int_0^t \beta(\tau)X_0 exp[-B(\tau)]f(\varphi-\tau)d\tau d\varphi \tag{17}$$

In Equation (17), the exposure time distribution is represented by the expression $\beta(\tau)X_0 exp[-B(\tau)]$, while the incubation distribution is represented by the expression $f(\varphi-\tau)$. The incubation distribution that will be used here will be the best-fit TDR equation. In their work, Gupta & Haas (2004) considered the Weibull, gamma, lognormal, and inverse Gaussian distributions for the exposure and incubation functions, all of which are non-time-dependent. Here, we use the Weibull, gamma, lognormal, and uniform distributions to model the exposure functions, and a time-dependent dose-response function to model the incubation function.

MATLAB model

A computer model was developed in MATLAB to best-fit the outbreak dataset. An optimization algorithm that evaluated an objective function starting from a set of assumed values for the model above-mentioned parameters was used. The objective function was the likelihood ratio

$$-2ln\left(\frac{L}{\hat{L}}\right) = 2\left[(\eta - N) - Nln\left(\frac{\eta}{N}\right)\right] \tag{18}$$

where η is the predicted number of new cases and N is the observed number of new cases. The objective function was minimized by varying a set of assumed values for the model parameters listed above in search of a global minimum value. Each unique combination of assumed values corresponds to a calculated value for the predicted number of cases per day $Z(t)$, by which a corresponding value for the daily likelihood ratio objective function can be calculated. An overall objective function is then calculated, which is the sum total of all the individual daily objective functions $-2ln(L/\hat{L})$ throughout the entire outbreak. The routine function 'fmins', which has the ability to minimize functions with several parameters and uses a Nelder-Mead simplex, was utilized to fit the dataset.

This entire process was iterative. For each exposure-incubation case, the following process was developed. A set of assumed parameter values were first input into the simulation model, and the corresponding predicted daily ill case $Z(t)$. was evaluated using the function 'quad2d'. The summation of the daily objective function $-2ln(L/\hat{L})$ over the entire duration of the outbreak was then calculated. The program varied the initial guesses, and continued to calculate the deviance objective function until convergence was assumed, which was assumed to be less than 10^{-3}.

Incorporation of the TDR into incubation distributions of the outbreak model

The TDR functions from Table 1 cannot be immediately used as incubation distributions because they are CDFs. To derive a time dose-response based incubation distribution to fit epidemiologic curves, the distributions must be PDFs. To get a PDF from a CDF, the derivative must be taken. In the case of the beta-Poisson TDR with inverse time dependency, we have the following equation:

$$R(d, t) = 1 - \left[1 + \frac{d}{(j_0/t) + 1} \times \left(2^{1/\alpha} - 1\right)\right]^{-\alpha} \quad (19)$$

$R(d, t)$ is the probability of adverse effect; α is a fitting parameter; θ is a time parameter; $(j_0/t) + 1$ is the dose that elicits a positive response in 50% of the hosts, after infinite time.

Taking the first partial derivative with respect to time, we get

$$r(d, t) = -\alpha \left[1 + \frac{d}{(j_0/t) + 1} \times \left(2^{1/\alpha} - 1\right)\right]^{-\alpha - 1} \times d\left(2^{1/\alpha} - 1\right)\frac{j_0}{t^2} \quad (20)$$

However, we want to make sure that our function is a true PDF. We want the PDF to be a true PDF with the following conditions:

Probability density functions satisfy the following conditions:

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

As time goes to infinity, we want the integral of our distribution to be exactly 1. So, we must divide by a certain value that ensures that this is true.

As $t \to \infty$, the maximum response is reached for the time dose response. Then Equation (19) no longer becomes time-dependent, and we refer to the parameter TPI dependency equation in Table 1, which assumed that $N_{50} = \frac{j_0}{t} + 1$. Plugging in N_{50} for the expression in the denominator returns us back to the regular beta-Poisson dose-response model.

$$R(d, t \to \infty) = 1 - \left[1 + \frac{d}{N_{50}} \times \left(2^{1/\alpha} - 1\right)\right]^{-\alpha} \quad (21)$$

This expression can be used as the constant by which Equation (20) can be divided. If we divide Equation (20) by Equation (21), we get the following function:

$$f(t|d) = \frac{\alpha \left[1 + \frac{d}{(j_0/t) + j_1} \times \left(2^{1/\alpha} - 1\right)\right]^{-\alpha - 1} \times d(2^{1/\alpha} - 1)j_0}{t^2 \left(-1 + \left[1 + \frac{d}{(j_0/t) + j_1} \times \left(2^{1/\alpha} - 1\right)\right]^{-\alpha}\right)} \quad (22)$$

We now have a proper PDF $f(t|d)$ that characterizes the distribution. This PDF based on the beta-Poisson TDR function distribution can now be used as a potential incubation distribution to $f(t)$ to use in the PDF used to calculate the number of ill, $Z(t)$. The functions used to model the

exposure distributions include the Weibull, gamma, lognormal and uniform distributions, and the best-fit TDR model will be used as the incubation distribution.

In this model, we have seven parameters for the number ill, $Z(t)$ in Equation (16), namely the attack parameters b_0 and b_1, two PDF parameters for the exposure distribution (Weibull, gamma, lognormal, and uniform distributions all contain 2 parameters), and the three incubation parameters (α, d, and j_0).

It is noted that the parameter values of α, d, and j values from the outbreak were used as initial guesses for those parameters of the incubation distribution in the optimization routine. In order to compare the parameters in the epidemic and dosing parameter studies to see if the differences between corresponding parameters were significant, the two-sample Kolmogorov–Smirnov (KS) test and the Anderson–Darling (AD) test were utilized. In both tests, we wanted to see if the parameters came from the same distribution. The KS test states that $D* = \dfrac{sup}{x} \mid F_1(x) - F_2(x)\mid$ (Sheskin 2003), where $F_1(x)$ and $F_2(x)$ are the empirical distribution functions of the two samples respectively, and sup is the supremum function. The null hypothesis is rejected alpha if the following equation holds true: $D* > c(\alpha)\sqrt{(n + n_1)/nn_1}$. For $\alpha = 0.05$, $c(\alpha) = 1.36$. The AD is a refinement of the KS test. The statistic that was used was of the form $A^2 = -n - \sum_{i=1}^{n} (2i - 1/n)[\ln(\Phi(Y_i)) + \ln(1 - \Phi(Y_{n+1-i}))]$ (D'Agostino & Stephens 1986). The parameter n is the number of data points in the sample, and the test statistic measures the distance between the hypothesized and empirical. The null hypothesis of normality is rejected if A^2 exceeds 0.752 for $\alpha = 0.05$.

RESULTS AND DISCUSSION

Time-dose-response

A literature review on dose-response was conducted to identify appropriate candidate models for *Giardia*. Two datasets passed the inclusion criteria for the time-dependent dose-response models. The first was a study by Erlandsen *et al.*, where beavers were exposed to different dose levels of *Giardia lamblia* cysts in order to determine an infective dose, the minimum dose required to observe cysts in the stool (Erlandsen *et al.* 1988). The cysts were obtained from

three symptomatic patients from a *Giardia* outbreak in Pittsfield, MA in December 1985. The authors reported the viability of the cysts to be 91% and administration was via the oral route. The beaver hosts were live-trapped, and were examined and treated for 30 days to ensure that they were negative for *Giardia* cysts, and were subsequently inoculated at graded doses of 48, 454, 4,460 and 5.5×10^5 cysts. None of the beavers receiving the lowest dose of cysts tested positive for the presence of fecal cysts. Two of the six beavers receiving 454 cysts, one of the three beavers receiving 4,460 cysts, and two of the three beavers inoculated with 5.5×10^5 cysts were infected, as determined by the presence of fecal cysts. The day of fecal cyst appearance post-inoculation ranged from 10 to 30.

In a second study by Rendtorff *et al.*, human volunteers were orally exposed to graded doses of *Giardia lamblia* cysts, sourced from fresh stool samples of human donors (Rendtorff 1954). The cysts were administered in either a saline or tap water solution. Infection was defined to be the appearance of cysts in stool samples. Individuals exposed to doses as low as 10 cysts were infected, and the first day of fecal cyst appearance post-inoculation ranged from 6 to 21. The adverse response for the beavers and human study is shown in Table 2.

Dose-response data for both datasets for *Giardia* inoculated orally in beavers and humans showed positive trends when subjected to the Cochran-Armitage test of trend. The Z critical value Z_{cr}, was calculated to be 2.04 for the beaver dataset, and 3.14 for the human dataset. Since these values are greater than 1.64 ($Z_{cr} > 1.64$), the null hypothesis of lack of trend is rejected and the data can be further analyzed.

Table 3 summarizes the parameters of the time-dependent dose-response data for the beaver and human model fits. The Exponential model with exponential-reciprocal time dependency did not converge for either the human or beaver datasets. This may be because this time-dependence model was not robust enough to provide adequate fitting to the datasets. In the beaver dataset, the beta-Poisson with exponential-reciprocal time dependency and beta-Poisson with exponential time dependency models provided acceptable fits, in which the minimized deviance is less than the critical chi-square value for the degree of freedom of the dose-response data at $p = 0.05$. In the beta-Poisson

Table 2 | Daily number of fecal cyst shedding of live-trapped beavers and humans exposed to graded doses of *Giardia lamblia* via the oral route

| Case | Host | No of hosts[a] | Dose[b] | Number of hosts with confirmed stool cysts on first reported day | | | | | | | | | | | | | | | | | |
				1–5	6	7	8	9	10	11	12	13	14	15	16	17–20	21	22–29	30	Total
ERL[c]	Beaver	6	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		6	454	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	2
		3	4,460	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
		3	5.5×10^5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	2
REN[d]	Human volunteers	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		2	10	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	2
		20	25	0	0	0	0	0	0	0	0	1	2	1	1	0	1	0	0	6
		2	1×10^2	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	2
		3	1×10^4	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	3
		3	1×10^5	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	3
		3	3×10^5	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	3
		2	1×10^6	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2

[a]Host refers to beavers in the Erlandsen study, and humans in the Rendtorff study.
[b]Reported in cyst numbers in beaver hosts, and colony forming units (CFUs) in human hosts.
[c]Erlandsen *et al.* (1988) data.
[d]Rendtorff *et al.* (1954) data.

Table 3 | Summary of fits of time-dependent dose-response data to beaver and human oral exposure

Case	Time dependence model	Best fit parameters	Minimized deviance	$\chi^2_{0.05,df}$	Acceptable fit?
ERL[a]	Exponential model with exponential-reciprocal time dependency	No convergence	n/a	n/a	n/a
	Beta-Poisson model with exponential-reciprocal time dependency	$\alpha = 0.11$ $j_0 = 35.7$ $j_1 = 9.35$	24.2	24.7	yes
	Beta-Poisson model with exponential time dependency[c]	$\alpha = 0.13$ $j_0 = 0.51$ $j_1 = 15.7$	22.7	24.7	yes
	Beta-Poisson model with inverse time dependency	$\alpha = 0.012$ $j_0 = 109.1$	59.9	26.1	no
REN[b]	Exponential model with exponential-reciprocal time dependency	No convergence	n/a	n/a	n/a
	Beta-Poisson model with exponential-reciprocal time dependency[d]	$\alpha = 1.11$ $j_0 = 10.4$ $j_1 = 3.12$	105.6	107.5	yes
	Beta-Poisson model with exponential time dependency	$\alpha = 5.29$ $j_0 = 92.0$ $j_1 = 3.36$	178.2	107.5	no
	Beta-Poisson model with inverse time dependency	$\alpha = 1.37$ $j_0 = 0.07$	1,493.9	108.6	no

[a]Erlandsen *et al.* beaver dataset (Erlandsen *et al.* 1988).
[b]Rendtorff *et al.* human dataset (Rendtorff 1954).
[c]Best-fit model for the Erlandsen *et al.* data (Erlandsen *et al.* 1988).
[d]Best-fit model for the Rendtorff *et al.* human dataset (Rendtorff 1954).

model with exponential-reciprocal time dependency the minimized deviance was determined to be 24.2, less than the critical chi-square value of 24.7 for the degree of freedom. In the case of the beta-Poisson model with exponential time dependency, the minimized deviance was 22.7, which was slightly lower than the previous model, and still less than the critical chi-square value of 24.7. For the beta-Poisson with inverse time dependency model, the minimized deviance was 59.9, much higher than the critical chi-square value of 26.1; this model does not

provide an acceptable fit to the beaver dataset. In the human dataset, the only model that provided an acceptable fit for TDR model was the beta-Poisson with exponential-reciprocal time dependency. The minimized deviance in this case was 105.6, less than the critical chi-square value of 107.5. The remaining models did not provide acceptable fits.

The TDR model that was ultimately chosen to be used in incubation distribution of the Pittsfield, Massachusetts *Giardia* outbreak was the best-fit model of the human dataset, the beta-Poisson with exponential-reciprocal time dependency. The advantage of using the results of the beaver dataset was that cysts used to inoculate the animals were sourced from the Pittsfield outbreak. However, the human dosing study was considered the most appropriate choice because the method of administration was via the oral route, and *Giardia* outbreaks largely are transmitted via this route.

Figures 3 and 4 present the TDR models for each of the respective studies.

Giardiasis outbreak model

While a large number of waterborne and foodborne cases of giardiasis are diagnosed each year, less than 1% of cases are associated with an outbreak (Yoder *et al.* 2012). Of these outbreaks, only a few satisfy the criteria for analysis in this study. A *Giardia* outbreak occurring between November 1985 to January 1986 in the city of Pittsfield, MA, was documented by Kent (Kent *et al.* 1988) and provided adequate data for the analysis. This outbreak was the largest ever documented for a *Giardia* pathogen. On November 5, a new auxiliary reservoir was brought online in Pittsfield to replace water from a main reservoir that was having a new filtration system installed. On November 14, the fraction of water supplied from the auxiliary reservoir increased,
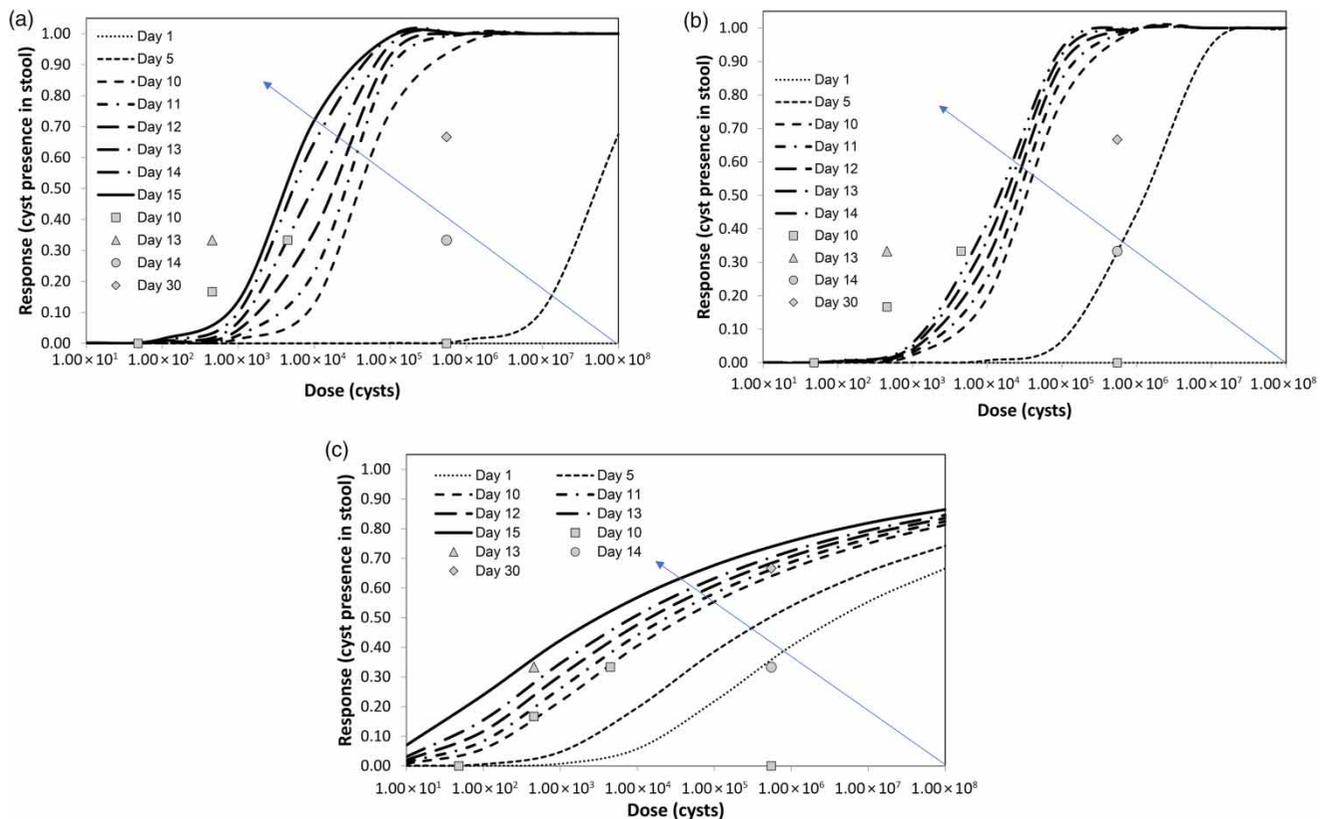

Figure 3 | Cumulative response data on each day estimated by TDR models compared with the observed adverse response of Giardia administered to beavers (Erlandsen *et al.* 1988). The arrows indicate the direction of the curves from the first day to the last. The TDR models in this figure are as follows: (a) beta-Poisson model with exponential-reciprocal time dependency; (b) beta-Poisson model with exponential time dependency; (c) beta-Poisson model with inverse time dependency.
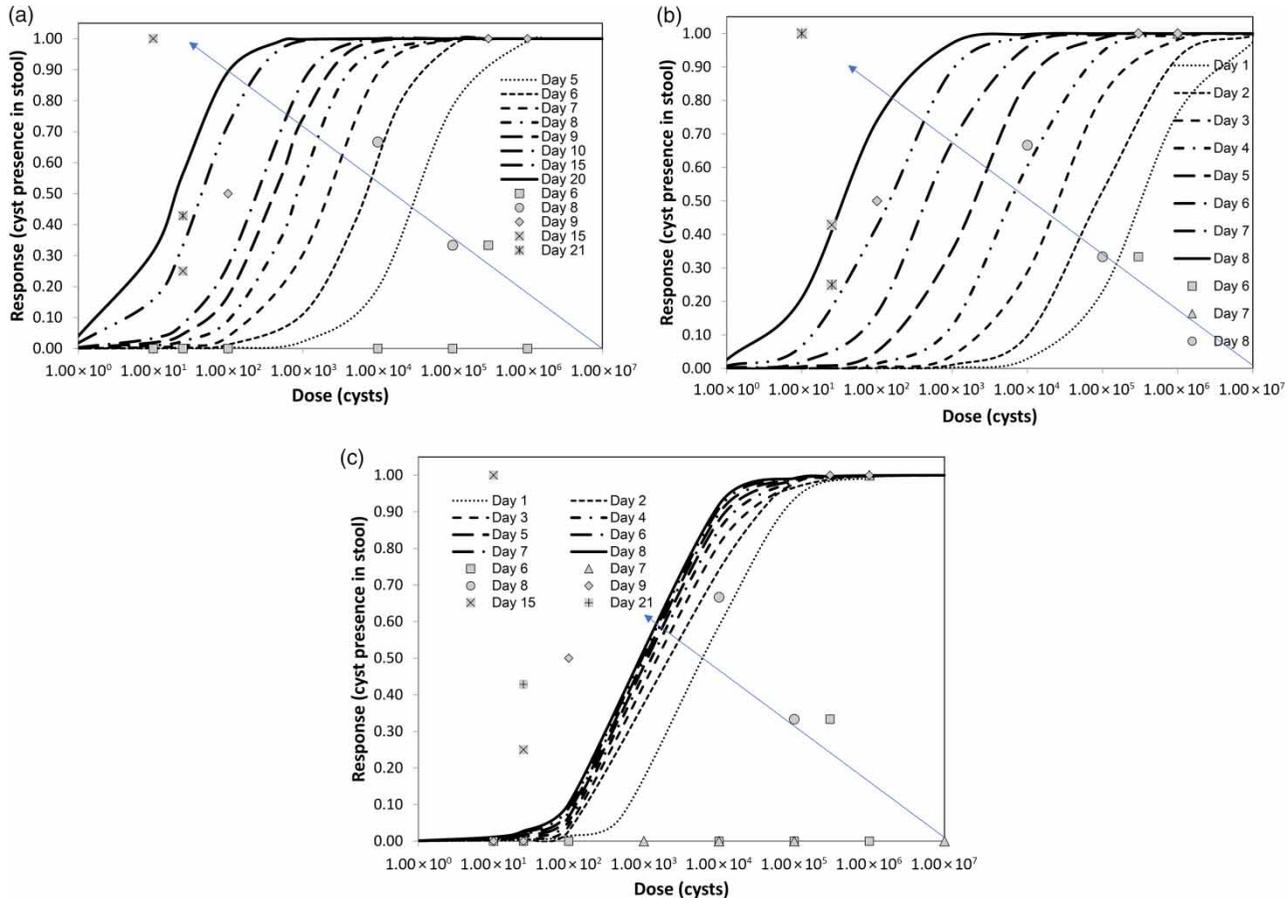
Figure 4 | Cumulative response data on each day estimated by TDR models compared with the observed adverse response of Giardia administered to human volunteers (Rendtorff 1954). The arrows indicate the direction of the curves from the first day to the last. The TDR models in this figure are as follows: (a) beta-Poisson model with exponential time dependency; (b) beta-Poisson model with exponential-reciprocal time dependency; (c) beta-Poisson model with inverse time dependency.

and residents from various parts of the city complained of cloudy water the same day. Two weeks later, health officials in Pittsfield received 70 reports of laboratory-confirmed giardiasis, once again from residents throughout the city. Health officials identified the source of the giardiasis outbreak, and the auxiliary reservoir was removed from the municipal water supply on December 21. Although the exposures appear to have spanned a period of over one week, this is still much shorter than the length of the outbreak, which spanned the course of 3 months. The exposure period continues to be modeled with the four exposure distributions: Weibull, gamma, lognormal and uniform.

Studies conducted in the area suggested that beavers and muskrats had contributed to the contamination of the reservoir, but they could have originally been infected from a human source. In a susceptible population of about

50,000, a total of 703 laboratory-confirmed cases of giardiasis were reported.

The best-fit model from the human feeding trial was the beta-Poisson with exponential-reciprocal time dependency; this TDR model was therefore chosen to be incorporated into the incubation distribution of the outbreak model. A maximum likelihood approach was used in MATLAB to predict the number ill $Z(t)$ on each respective day and each of the parameters. Table 4 summarizes the model parameters and the criterion values for the various distribution models, and Figure 5 summarizes the observed vs. expected curve fits for each PDF convolution.

The df ($m-n$) in this case was 55, the number of data points minus the 8-parameter model. The critical chi-square distribution χ^2 for 55 degrees of freedom is 73.31, so the incubation distribution convoluted with the

Table 4 │ Summary of parameter, criterion and test statistic values for the exposure and incubation periods of the model

		Model parameters								Criterion	Test statistic	
Exposure distribution	Incubation distribution	Attack param b_0	Attack param b_1	Attack dist – PDF param 1	Attack dist – PDF param 2	Incub dist – PDF param 1 alpha	Incub dist – PDF param 2 dose	Incub dist – PDF param 3 j_0	Incub dist – PDF param 4 j_1	value deviance	KS[a]	AD[b]
Gamma	Beta-Poisson with Exponential Recip Time Dependency	0.010	3.17	25.6[c]	0.397[d]	11.4	6.0	13.4	2.4	67.4	1.13	0.52
Lognormal		0.0030	3.79	3.81[e]	0.344[f]	1.39	11.0	7.47	1.68	61.2	0.15	0.27
Uniform		0.98	0.092	0.166[g]	0.647[h]	3.07	3.7	21.8	4.46	106.6	0.32	0.44
Weibull		0.032	0.244	2.04[i]	0.66[j]	2.04	0.66	13.7	18.3	72.1	1.28	0.63

[a]Kolmogorov–Smirnov test statistic $D*$.
[b]Anderson–Darling test statistic $A^{2.}$
[c]Mean.
[d]Variance.
[e]log transform of mean.
[f]log transform of variance.
[g]Lower endpoint.
[h]Upper endpoint.
[i]Scale parameter.
[j]Shape parameter.
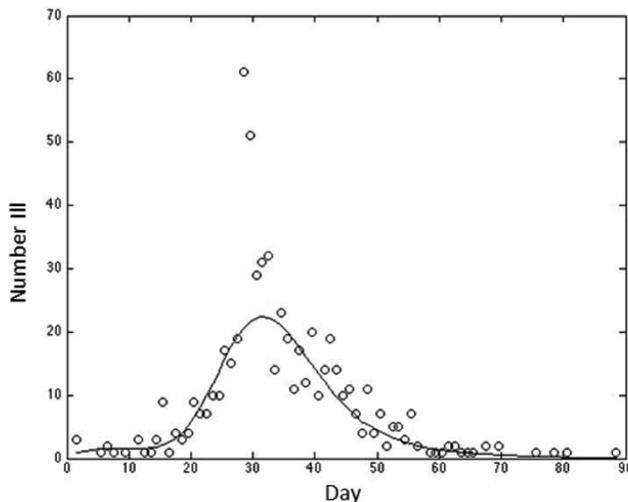
Figure 5 │ The best-fit outbreak model: the lognormal exposure distribution convoluted with the beta-Poisson with exponential-reciprocal incubation distribution.

gamma, lognormal, and Weibull distributions all provide acceptable fits (Table 4). The lognormal exposure distribution is the best fit, with the lowest deviance value at 61.15 and 75.15.

The first and second column of Table 4 are the results for the b_0 and b_1 attack rate values. These parameter values indicate the level of disease present in the population before the symptomatically ill persons were reported. The low b_0 and b_1 attack rate values indicates that there was a low baseline incidence rate of *Giardia* among this susceptible population.

The 3rd and 4th columns are referred to as 'PDF param 1' and 'PDF param 2', and they differ per the distributions used. For the gamma exposure distribution, PDF param 1 and PDF param 2 are the mean and variance, at 25.6 and 0.397, respectively. In the lognormal distribution case, PDF param 1 and PDF param 2 are the log transform of the mean and variance, at 3.81 and 0.344, respectively. In the Weibull distribution, PDF param 1 and PDF param 2 are the scale parameter a and shape parameter b, at 2.04 and 0.66, respectively. The uniform similarly follows the Weibull, and uses parameters a and b in the PDF, with a being a lower endpoint (minimum) and b being an upper endpoint (maximum) of 0.166 and 0.647, respectively. These parameters describe the exposure distributions that model the mechanisms by which the *Giardia* pathogen spreads in this particular outbreak.

Columns 5 through 8 present the $\alpha, d, j_0,$ and j_1 parameters. The α parameter ranges from 1.39 to 11.4, and the d parameter ranges from 0.66 to 11.0. The j_0 ranges between 7.47 and 21.8, while the j_1 ranges from 1.68 and 18.3. These values correspond to the incubation distribution modeled by the beta-Poisson with exponential-reciprocal functions, which are convoluted with the gamma, lognormal, Weibull and uniform functions, respectively.

The data was further analyzed by comparing the parameters of the incubation distribution of the outbreak model with the human dosing experiment to see if there was a significant difference between the parameters of the two models. These parameters that were compared were the α, d, j_0, j_1 of the TDR model with the α, d, j_0, j_1 from the outbreak fitting. The last two columns in Table 4 summarize the output of the KS and AD tests. In each case, the D^* value of the KS test is less than the critical χ^2 for $\alpha = 0.05$, which is 1.36. The AD test, which comes from the KS test, was also positive. The A^2 values for each distribution scenario were less than the critical χ^2 for $\alpha = 0.05$, which is 0.752. From these results, it was concluded that there was no significant difference in the parameters of the distributions, and they came from the same distribution.

To determine which distribution convolution fit the data best, we look to the deviance. The incubation distribution parameters are the parameters from the beta-Poisson with exponential-reciprocal time dependency model. The best-fit exposure function that was convoluted with this incubation function was the lognormal, with a deviance value of 61.2. The gamma exposure gives the next-best fit, at 67.4, followed by the Weibull at 72.1. The uniform exposure fits the data the worst, with a deviance value of 106.6. Figure 5 shows the plot for this best-fit model, the lognormal exposure function convoluted with the beta-Poisson with exponential-reciprocal dependency. All models were limited by their inability to capture the data points in which the number of ill was at its peak, around day 30 of the outbreak.

DISCUSSION

These results suggest that the incorporation of time-to-infection into the incubation function of an epidemic curve can provide acceptable fits to the outbreak data. One strength in this work is that the endpoint is the same for the dosing trials and outbreak dataset: all report the presence of laboratory-confirmed *Giardia* cysts in stool. An interesting connection between the beaver trial and the outbreak is that the beavers were inoculated with cysts from three symptomatically ill patients from the Pittsfield outbreak. The human volunteers, on the other hand, were administered cysts that were sourced from fresh stool samples of human

donors. Despite the connection with the original outbreak, the human TDR results were ultimately chosen over the beaver results to be incorporated into the outbreak model because the human *in vivo* response was thought to best model the human outbreak.

It is also acknowledged that datasets are limited for the type of modeling performed in this work. In both feeding trials, the incubation time is the time between exposure and onset of shedding, which is monitored daily for presence of *Giardia*. However, in the outbreak data, a patient must first feel symptomatically ill to go to visit the doctor, who then sends a stool sample to the laboratory for analysis. So, there is a time between exposure and onset of symptoms identified by the infected individuals themselves. Currently, there is no available data to account for this difference. However, the time difference between fecal shedding and onset of symptoms is not necessarily significant. The incubation period is estimated to be between 3–25 days, with the median number being between 7–10 days (Heyman 2004). The first appearance of cysts in beavers was most commonly at 10 days and between 13–14 days. In the human trials, there was a much larger spread – between 6 and 17 days, with a mode of 10 days. These numbers are well within the incubation period of 3–25 days, and are at the tail end of the median incubation period of 7–10 days. However, to strengthen this model in the future, the value $Z(t)$ could be multiplied by another parameter that accounts for this discrepancy.

As with any computer model, the MATLAB optimization had challenges in the fitting of the datasets. The issue of reliably fitting eight parameters with a degree of reproducibility, given an observed epidemic outbreak, was a concern. To examine this issue, the program was run several times, with varying initial guesses. At times, a global minimal for the log-likelihood was not able to be reached; instead a local minimum value was calculated. The local minimum values may suggest a substitution effect between the parameters. However, by running the simulations several times and allowing them to compute for several hours, these local minimum values were largely able to be overcome in favor of global minimums. With each run, the models were able to consistently reach the same likelihood value with parameter values that were very close to each other. Confidence in the model structure could be increased

by generating artificial data from the different models and testing the ability of the fitting algorithms to recover the parameters. This testing procedure was outside of the scope of this work, but could be done as a follow-up study. It would also be useful to see confidence interval estimates for at least the favored outbreak model. These estimates could then be compared with the estimates from those based on the human data. However, this procedure is not within the scope of this work.

CONCLUSION

TDR models were generated for *Giardia lamblia* exposure to beaver and human volunteers. The best-fit human TDR model was the beta-Poisson with exponential-reciprocal dependency, which had a minimized deviance of 105.6; this was chosen to be incorporated into the outbreak model. The convolution of this probability distribution with the lognormal distribution yielded the best-fit for the model, with a minimized deviance value of 61.2. Analysis of the parameters from the dose-response model and the epidemic study showed that the parameters were taken from the same distribution, as they passed the AD and KS tests. This work suggests that the incorporation of the *in vivo* time factor for pathogens may provide a useful alternative approach to modeling in which the time-to-infection is considered in the fitting of a disease outbreak.

REFERENCES

Anderson, R. M. & May, R. M. 1991 *Infectious Diseases of Humans*. Oxford University Press, Oxford.

D'Agostino, R. B. & Stephens, M. A. 1986 *Goodness-of-Fit Techniques, Statistics: Textbooks and Monographs*, Vol. 68, Marcel Dekker, Inc, New York.

Erlandsen, S. L. & Meyer, E. A. (eds) 2013 *Giardia and Giardiasis: Biology, Pathogenesis, and Epidemiology*. Springer Science & Business Media, New York, NY.

Erlandsen, S., Sherlock, L., Januschka, M., Schupp, D., Schaefer, F., Jakubowski, W. & Bemrick, W. 1988 Cross-species transmission of Giardia spp.: inoculation of beavers and muskrats with cysts of human, beaver, mouse, and muskrat origin. *Applied and Environmental Microbiology* **54** (11), 2777–2785.

Gupta, M. 1999 Development and Use of a Dynamic Disease Propagation Model for Assessing Risk from Common Source Epidemics. 9923549 PhD, Drexel University.

Gupta, M. & Haas, C. 2004 The Milwaukee Cryptosporidium outbreak: assessment of incubation time and daily attack rate. *Journal of Water and Health* **2** (2), 59–69.

Haas, C. N., Rose, J. B. & Gerba, C. 1999 *Quantitative Microbial Risk Assessment*. John Wiley and Sons, New York, NY.

Haas, C. N., Rose, J. B. & Gerba, C. P. 2014 *Quantitative Microbial Risk Assessment*. Wiley, New York.

Hethcote, H. 1996 *Modeling Heterogeneous Mixing in Infectious Disease Dynamics*. Cambridge University Press, Cambridge.

Heyman, D. 2004 *Control of Communicable Diseases Manual*. American Public Health Association, Washington, DC.

Huang, Y. & Haas, C. N. 2009 Time-dose-response models for microbial risk assessment. *Risk Analysis* **29** (5), 648–661.

Huang, Y. & Haas, C. N. 2011 Quantification of the relationship between bacterial kinetics and host response for monkeys exposed to aerosolized Francisella tularensis. *Applied and Environmental Microbiology* **77** (2), 485–490.

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

Karanis, P., Kourenti, C. & Smith, H. 2007 Waterborne transmission of protozoan parasites: a worldwide review of outbreaks and lessons learnt. *Journal of Water and Health* **5** (1), 1–38.

Kent, G. P., Greenspan, J. R., Herndon, J. L., Mofenson, L. M., Harris, J., Eng, T. R. & Waskin, H. A. 1988 Epidemic giardiasis caused by a contaminated public water supply. *American Journal of Public Health* **78** (2), 139–143.

Kermack, W. O. & McKendrick, A. G. 1927 Contributions to the mathematical theory of epidemics. *Proceedings of the Royal Society of London Series A* **115**, 700–721.

Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M. & Tauxe, R. V. 1999 Food-related illness and death in the United States. *Emerging Infectious Diseases* **5** (5), 607–625.

Neuhäuser, M. & Hothorn, L. A. 1999 An exact Cochran–Armitage test for trend when dose–response shapes are a priori unknown. *Computational Statistics & Data Analysis* **30** (4), 403–412.

Olson, M. E., Goh, J., Phillips, M., Guselle, N. & McAllister, T. A. 1999 Giardia cyst and Cryptosporidium oocyst survival in water, soil, and cattle feces. *Journal of Environmental Quality* **28** (6), 1991–1996.

Rendtorff, R. C. 1954 The experimental transmission of human intestinal protozoan parasites II. Giardia lamblia cysts given in capsules. *American Journal of Epidemiology* **59** (2), 209–222.

Savioli, L., Smith, H. & Thompson, A. 2006 Giardia and Cryptosporidium join the 'neglected diseases initiative'. *Trends in Parasitology* **22** (5), 203–208.

Sheskin, D. J. 2003 *Handbook of Parametric and Nonparametric Statistical Procedures*. CRC Press, Boca Raton, Florida.

Wharton, M., Chorba, T. L., Vogt, R. L., Morse, D. L. & Buehler, J. W. 1990 Case definitions for public health surveillance. *MMWR Recommendations and Reports* **39** (RR-13), 1–43.

Yoder, J. S., Harral, C. & Beach, M. J. 2010 Giardiasis surveillance–United States, 2006–2008. *MMWR Surveillance Summaries* **59** (6), 15–25.

Yoder, J. S., Gargano, J. W., Wallace, R. M. & Beach, M. J. 2012 Giardiasis surveillance–United States, 2009–2010. *MMWR Surveillance Summaries* **61** (5), 13–23.