

An evaluation of parsimony for microbial risk assessment models

Jeffrey A. Soller^{1*,†} and Joseph N. S. Eisenberg²

¹ *Soller Environmental, Berkeley, CA 94703, USA*

² *University of Michigan, Ann Arbor, MI 48104, USA*

SUMMARY

Microbial risk assessment (MRA) is a process that evaluates the likelihood of adverse human health effects following exposure to a medium in which pathogens are present. Several different classes of models are available to quantitatively characterize risks to human health from exposure to pathogens. Herein, we consider the question of parsimony for specific realizations of representative static and dynamic MRA models and identify conditions under which the more complex dynamic model provides sufficient additional insight to justify the added modeling complexity. To address this question, a standard static individual-level risk model is compared to a deterministic dynamic population-level model that explicitly includes secondary transmission and immunity processes. Exposure parameters are based on a scenario defined by human exposure to pathogens in reclaimed water. A sensitivity analysis is implemented to identify conditions under which static and dynamic models yield substantially different results. Under low risk conditions, defined by a combination of exposure levels and infectivity of the pathogen, the simpler static model provides satisfactory risk estimates. The approach presented here provides a basis for model selection for a broad range of MRA applications. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: microbial risk assessment; sensitivity analysis; mathematical models; risk assessment; waterborne pathogens; QMRA

1. INTRODUCTION

1.1. Motivation

Microbial risk assessment (MRA) is increasingly being applied for regulatory and operational purposes. Given the unique characteristics of pathogens compared with chemicals, a better understanding of appropriate model forms is needed. Fundamentally, MRA is a process that evaluates the likelihood of adverse human health effects that can occur following exposure to pathogenic microorganisms or to a medium in which pathogens occur (ILSI, 1996). Quantitative methods to characterize human health risks associated with exposure to pathogenic microorganisms were first published in the 1970s (Fuhs, 1975; Dudley *et al.*, 1976). Since that time many MRAs have used the conceptual risk assessment framework for chemicals (National Research Council, 1983) as a basis for waterborne (Haas, 1983; Regli *et al.*, 1991; Rose *et al.*, 1991; Gerba *et al.*, 1996; Crabtree *et al.*, 1997;

*Correspondence to: J. A. Soller, Soller Environmental, Berkeley, CA 94703, USA.

†E-mail: jsoller@sollerenvironmental.com

Teunis *et al.*, 1997; Mena *et al.*, 2003) and foodborne pathogen assessments (Farber *et al.*, 1996; Buchanan *et al.*, 1998; Buchanan *et al.*, 2000). Consistent with the chemical risk framework, most of these assessments assume that the number of individuals that are susceptible to infection is not time varying (static) and, thus risk is characterized at an individual level (Eisenberg *et al.*, 2002). Static models have also been used by US EPA in the development of drinking water regulations (U.S. EPA, 2002, 2006a).

As the field of MRA developed, the advantages of modeling infectious disease processes such as person-to-person transmission of infection and immunity became apparent (Eisenberg *et al.*, 1996, 1998). Addressing these issues requires dynamic methods where the number of individuals that are assumed to be susceptible to infection is time varying and risk is manifest at the population level (Hethcote, 1976; Anderson and May, 1991; Hethcote, 2000). To facilitate addressing these needs, the US EPA developed a MRA framework that highlighted the inclusion of person-to-person transmission of infection and immunity (ILSI, 1996). This framework was tested through the conduct of two case studies (Soller *et al.*, 1999; Teunis and Havelaar, 1999) and was subsequently revised (ILSI, 2000; Schaub, 2004).

Dynamic MRA models can take two basic forms: deterministic and stochastic. Deterministic dynamic models are expressed as a set of differential equations that have defined parameters and starting conditions, which determine the rate of transfer of individuals from one epidemiologic state to another. This type of model is most suitable for large populations of individuals randomly interacting with one another (Soller *et al.*, 2003; Eisenberg *et al.*, 2005). In the stochastic form, the model incorporates probabilities at an individual level and is evaluated by an iterative process. Stochastic model forms are suitable for small populations and heterogeneous mixing patterns (Koopman *et al.*, 2002.). Dynamic MRA methods have been used for numerous specific case studies in the United States (Eisenberg *et al.*, 1996; Eisenberg *et al.*, 1998; Soller *et al.*, 1999; Koopman *et al.*, 2002; Soller *et al.*, 2003; Soller *et al.*, 2006) and recently to support regulatory decisions by US EPA (U.S. EPA, 2006b, 2006c).

From a modeling perspective, biological 'realism' can be achieved, but at the cost of analytical complexity. The increase in the complexity of the model structure increases variability due to the uncertainties associated with model specification, and increases the computational demands (U.S. EPA, 2004). On the other hand, a simpler model form involves implicit or explicit assumptions that may or may not be realistic or appropriate for a particular situation. Herein, we employ representative realizations of static and deterministic dynamic MRA models to consider the question of model parsimony and to identify conditions under which the more complex dynamic models may provide sufficient additional insight that the added complexity is warranted.

1.2. Conceptual health effects models

1.2.1. Static MRA model. Assessments using a static model for evaluating microbial risk typically focus on estimating the probability of infection to an individual as a result of a single exposure event. These assessments generally assume that multiple or recurring exposures constitute independent events with identical distributions of contamination (Regli *et al.*, 1991). In most cases when this type of model has been employed, it has been implicitly assumed that secondary (person-to-person or person-to-environment-to-person) transmission and immunity are: (1) negligible, (2) of the same magnitude and effectively cancel each other out, or (3) scale linearly with the number of predicted infections. Notable exceptions to this generality include the work by Teunis *et al.* (2002) and the U.S. EPA (2006a) on *Cryptosporidium* in which immunity is addressed in the dose-response modeling.

Inclusion of secondary transmission may increase or decrease the level of infection attributable to a specific exposure to pathogens (Eisenberg *et al.*, 2004), while inclusion of immunity will decrease the level of infection in a community attributable to a specific exposure to pathogens.

Conceptually, static models employ two or three epidemiological states: a susceptible state and an infected and/or diseased state (Figure 1). Susceptible individuals, when exposed to the pathogen of interest, move into an infected state with a probability that is governed by the dose of pathogen to which they are exposed and the infectivity of the pathogen. The critical health effects information required for the static model, therefore, is summarized in the function that represents this probability of infection, $P(d)$, the pathogen-specific dose–response function. Equation 1 describes the dose–response function employed herein (Haas *et al.*, 1999). In static models, the probability of infection is often multiplied by the number of exposed individuals to estimate the expected number of infected individuals.

$$P(d) := 1 - \left(\frac{1 + \text{dose}}{\beta} \right)^{-\alpha} \quad (1)$$

1.2.2. Dynamic MRA model. In a dynamic risk assessment model, the population is classified into a group of epidemiological states (Hethcote, 1976; Anderson and May, 1991; Hethcote, 2000). Individuals move from state-to-state based characteristics such as the duration of infection, duration of immunity, etc. Thus, only a portion of the population is susceptible at any point in time, and only those individuals can become infected or diseased through exposure to microorganisms. The probability that a susceptible person moves into an exposed state is governed by the dose of pathogen to which they are exposed, the infectivity of that pathogen, and also the number of infected/diseased individuals with whom they may come into contact (Hethcote, 1976; Anderson and May, 1991; Hethcote, 2000).

For this investigation, a dynamic model that includes person-to-person transmission, immunity, an incubation period, and asymptomatic infection was used (Figure 2). This model is representative of those used previously for waterborne microbial risks investigations (Soller *et al.*, 2003; Eisenberg *et al.*, 2004; Soller *et al.*, 2006; U.S. EPA, 2006b). In this dynamic model, the population is categorized into six epidemiological states. Rate parameters specify the movement of the population between epidemiological states. The epidemiological states and rate parameters are summarized in Tables 1 and 2, respectively. Mathematically, the movement of the population between epidemiological states is

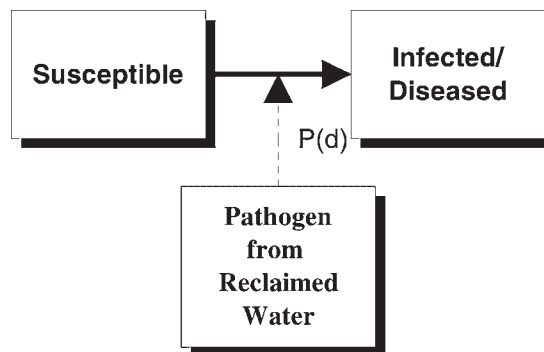


Figure 1. Static risk assessment conceptual model

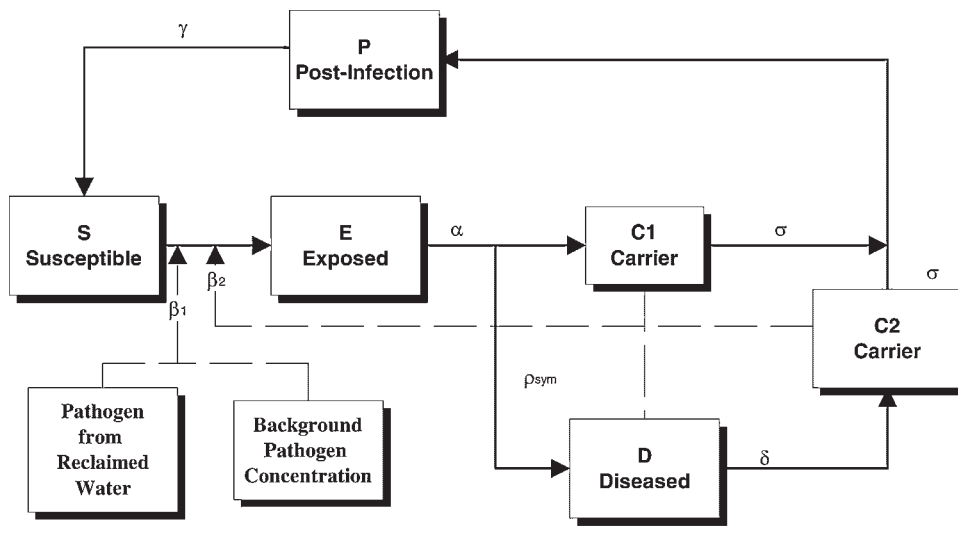


Figure 2. Dynamic risk assessment conceptual model

modeled as a distributed delay process using a series of ordinary differential equations. A full description of the mathematics for the dynamic model is presented in the Appendix.

1.3. Pathogens considered

Waterborne infectious agents of intestinal origin are discharged in human and animal feces, and the presence of microbial pathogens in wastewater has been a concern for decades. Conducting an assessment of the public health risk associated with exposure to pathogens requires the selection of a representative pathogen or set of pathogens on which to conduct the assessment.

Although a wide range of pathogens have been identified in raw wastewater, relatively few pathogens are believed to be responsible for the majority of the waterborne illnesses caused by

Table 1. Epidemiological states for dynamic model

Labels in Figure 2	Epidemiological states of population	Description
<i>S</i>	Susceptible	Individuals who are not infected and are not protected from infection
<i>E</i>	Exposed	Individuals who have been exposed to a pathogen, but are not yet infected
<i>C1</i>	Carrier 1	Individuals who are infected but do not have symptoms of disease
<i>D</i>	Diseased	Individuals who are infected and do have symptoms of disease
<i>C2</i>	Carrier 2	Individuals who were diseased and no longer exhibit symptoms of disease, but are still infected
<i>P</i>	Post-infected	Individuals who are neither infected nor symptomatic, and have resistance to infection

Table 2. Symbols used to specify rate parameters and probability of response in dynamic model

Symbol	Description
α	Rate of movement from an exposed state to a carrier (infectious and asymptomatic) state or a diseased state (infectious and symptomatic). $1/\alpha$ corresponds to the latency period prior to infection for the pathogen of interest
σ	Rate of movement from a carrier state to a post-infection state. $1/\sigma$ corresponds to the duration of infectiousness, or equivalently the duration of asymptomatic shedding of pathogen in feces
δ	Rate of movement from a diseased state (infectious and symptomatic) to an asymptomatic (carrier) state. $1/\delta$ corresponds to the duration of symptoms during infection
γ	Rate of movement from a post-infection state (not infectious, asymptomatic, and not susceptible to infection) to a susceptible state. $1/\gamma$ corresponds to the duration of immunity or protection from infection
β_1	Rate of movement from a susceptible state to an exposed state due to exposure to pathogens from an environmental source (i.e., not person-to-person transmission). It is a function of the number of pathogens to which an individual is exposed and the infectivity of the pathogen of interest. The infectivity is described quantitatively through a dose–response function which for this investigation is comprised of two parameters
β_2	Rate of movement from a susceptible state to an exposed state due to exposure to pathogens from secondary (person-to-person or person-to-environment-person) transmission
p_{sym}	Probability of a symptomatic response. Based on clinical data describing the proportion of infected individuals that develop symptoms

pathogens of wastewater origin (Mead *et al.*, 1999). The pathogens that have been reported to be responsible for the vast majority of illnesses in the United States from all sources were the focus of this investigation. From a list of pathogens of concern identified by EPA (2003) for waterborne exposures, and the Center for Disease Control estimate of disease levels in the United States (Mead *et al.*, 1999), the pathogens investigated herein include: enteroviruses, rotavirus, *Cryptosporidium parvum*, *Giardia lamblia*, *Salmonella* spp, *E. coli* O157:H7, and *Shigella* spp. Although noroviruses were also of particular interest, sufficient quantitative information was not available for inclusion in this investigation.

1.4. Human exposure

The goal of evaluating the two risk assessment model forms was to identify conditions under which the static and dynamic models predict substantially different levels of risk. To identify these conditions, we employed a case study paradigm and investigated exposure scenarios that are consistent with reclaimed water applications and with properties of the pathogens of public health concern for human exposure to reclaimed water. The use of reclaimed water is becoming increasingly common as a means to supplement municipal water supplies for nonpotable purposes, and human exposures to reclaimed water are common and vary substantially in magnitude.

2. METHODS

2.1. Simulation approach

A total of 11 parameters are required for the dynamic model, 3 of which are also used by the static model (Table 3). For each parameter, representative minimum, median, and maximum values were

Table 3. Parameters required for modeling static and dynamic disease processes

Class	Parameter	Static model	Dynamic model
Exposure-related parameters	Pathogen dose*	X	X
	Exposure intensity [†]		X
Pathogen-related parameters	Duration of incubation		X
	Probability of symptomatic response		X
	Duration of infectiousness		X
	Duration of disease		X
	Duration of protection		X
	Dose-response parameter α	X	X
	Dose-response parameter β	X	X
	Person-to-person transmission potential		X
	Effective background concentration level		X

*The product of concentration and volume yields the pathogen dose.

[†]The product of the proportion of population exposed and the frequency of exposure yields the exposure intensity.

identified from the available literature (See Subsection 2.2) resulting in 177 147 (3^{11}) possible combinations of parameters for the dynamic model, and 27 (3^3) possible combinations of parameters for the static model. The static and dynamic models were each run using the 3^{11} parameter combinations. The output from each simulation is the incidence of infection attributable to the reclaimed water exposure per 100 000 individuals. For the dynamic model, the incidence of infection is defined as the number of individuals entering either a diseased (State D) or carrier state (C_1). To determine the risk attributable to the reclaimed water exposure, the dynamic model was run twice for each parameter combination: once without the reclaimed water exposure (background), and once with the reclaimed water exposure. The number of cases of infection attributable to the reclaimed water exposure was computed as the difference between these two simulations. Thus, any infection resulting from person-to-person transmission following a primary infection from exposure to reclaimed water is included in the analysis. For the static model, the risk attributable to the reclaimed water exposure was equal to the probability of infection times the number of people exposed. Based on this methodology, a total of 3^{12} simulations were run. The attributable risk obtained from the two models was compared for each parameter combination.

2.2. Model parameterization

We conducted an extensive review of peer-reviewed literature to identify appropriate values for the pathogen-dependent variables for each of the representative pathogens of concern. Minimum, median, and/or maximum representative values were obtained to the extent feasible for each model parameter for each of the pathogens (Table 4). Blank cells in Table 4 indicate that data were not found in the literature. Based on these data, the minimum, median, and maximum values for all pathogens were noted for each pathogen-dependent parameter. These values are shown in Table 4 in the rows labeled 'Composite' and are used in the simulations. Detailed summaries for each pathogen including the literature used to support the parameter values are available elsewhere (Soller *et al.*, 2004).

Minimum, median, and maximum representative values were also developed for each of the exposure-dependent parameters (Table 5). The exposure-dependent parameter values are based on

Table 4. Pathogen-dependent parameter values

Pathogen	Duration of incubation*			Proportion of infected individuals with symptomatic response			Duration of infectiousness*		
	Min	Median	Max	Min	Median	Max	Min	Median	Max
Enteroviruses	0.29		8	0.25		0.75	0.13		4
Rotavirus	1.33		4	0.1		0.45	0.20		0.5
<i>Cryptosporidium spp</i>	0.57		4	0.2		0.7	0.15		2
<i>Giardia lamblia</i>	0.07		1.3	0.2		0.7	0.04		0.8
Salmonella	1.33		16	0.14		0.4	0.01		0.27
<i>E Coli O157:H7</i>	0.33		8	0.23		0.5	0.14		0.31
Shigella	0.57		4	0.29		0.5	0.40		0.8
Composite [¶]	0.07	1.6	16	0.1	0.4	0.75	0.01	0.3	4

Pathogen	Duration of disease*			Duration of protection*			Dose-response parameter $\alpha^{†,‡}$		
	Min	Median	Max	Min	Median	Max	Min	Median	Max
Enteroviruses	0.19		4	0.06		0.13		0.67	
Rotavirus	0.36		2	0.004		0.01	0.13		0.5
<i>Cryptosporidium spp</i>	0.15		2	0.01		0.02		0.67	
<i>Giardia lamblia</i>	0.04		0.8	0.004		0.13		0.67	
Salmonella	0.14		1.3	0.14		96	0.31		0.89
<i>E Coli O157:H7</i>	0.13		0.6	0.05		0.11		0.22	
Shigella	0.16		2	0.005		0.1		0.21	
Composite [¶]	0.04	0.7	4	0.004	0.06	96	0.13	0.5	0.89

Pathogen	Dose-response parameter $\beta^{†,‡}$			Person-to-person transmission [§]			Effective background concentration		
	Min	Median	Max	Min	Median	Max	Min	Median	Max
Enteroviruses	1.26	48	76		0.04			1.6E-04	
Rotavirus	0.21		0.84		0.06			1.4E-05	
<i>Cryptosporidium spp</i>		165			0.04			6.0E-04	
<i>Giardia lamblia</i>		34			0.01			5.0E-04	
Salmonella	2884		4.4E + 05		0.003			10	
<i>E Coli O157:H7</i>		8723			0.03			0.01	
Shigella		43			0.04			5.0E-04	
Composite [¶]	0.21	39	440000	0.003	0.04	0.06	1.4E-05	5.0E-04	10

*The parameter values shown are the 'scale' parameters, λ , of a gamma distribution. The 'shape' parameter, k , is assumed to be 4. Therefore $\lambda = 4/n$, where n is the mean value reported in literature. Unit: day⁻¹.

[†]Beta Poisson dose-response functions used for all pathogens. Refer to Soller *et al.* (2004) for details, unitless parameter.

[‡]*E. coli* dose-response data are from high dose studies and are most likely estimates, bounded by EPEC and *Shigella (flexnerii* and *dysenteriae* pooled). Enterovirus parameter estimates are for the beta Poisson dose-response function and are based on reported exponential functions for Poliovirus I, Adenovirus 4, Echovirus 12, and Coxsackie viruses, where adenovirus and coxsackievirus data are from respiratory pathway studies. *Cryptosporidium spp.* and *Giardia lamblia* beta Poisson parameters are based on reported exponential functions.

[§]Estimates based on an SIR model using data from an outbreak and from reports of the duration of the infectious period. See Soller *et al.* (2004) for details. Units are infections/person/day.

^{||}Values shown are concentrations of each pathogen required to result in a median prevalence consistent with levels reported by Mead *et al.* (1999). Units are pathogen units/liter.

[¶]Composite values are the minimum, median, and maximum values across all pathogens investigated for each parameter.

Table 5. Exposure-dependent parameter values

Component	Parameters	Min	Median	Max	Min	Median	Max
Dose from reclaimed water exposure*	Reclaimed water conc. (pathgens per liter)	0.00001	0.001	10	1.0E-08	1.0E-04	1.0
	Volume ingested (liters)	0.001	0.01	0.1			
Exposure intensity [†]	Proportion exposed (unitless)	0.001	0.01	0.1	3.33E-05	1.43E-03	0.10
	Frequency of exposure (day ⁻¹)	0.033	0.14	1			

*Dose from reclaimed water computed as (reclaimed water conc.) \times (volume ingested).

[†]Exposure intensity computed as (proportion exposed) \times (frequency of exposure).

reported data and best professional judgment to the extent necessary, and are specific to our reclaimed water scenario. Based on these data, minimum, median, and maximum values were calculated for the pathogen dose and exposure intensity model parameters. These calculated values (far right column of Table 5) were used in the simulations.

2.3. Model parsimony decision metric

The decision metric used to evaluate model parsimony was the absolute value of the difference in incidence (I_d) between the results from the static and dynamic model for each parameter combination. The simulations resulted in 3^{11} unique values for I_d .

A Classification and Regression Tree (CART) sensitivity analysis (Steinberg and Colla, 1997) was performed to determine which parameters in the model and combinations thereof impacted I_d most strongly. The CART analysis produces a tree structure that prioritizes the importance of the model parameters in determining the conditions under which I_d meet a specific criteria (Eisenberg and McKone, 1998). The CART analysis was carried out for three values of I_d : 10/100 000, 1/100 000, and 0.01/100 000. Identifying a level at which the difference in incidence between the models is considered tolerable is a risk management issue. The three threshold levels identified above were selected based on expert opinion (Soller *et al.*, 2004).

3. RESULTS

3.1. Simulation results

The I_d results from the simulations are presented in Figure 3. Substantially more positive than negative values of I_d suggests that immunity more commonly impacted the dynamic model results than person-to-person transmission relative to the static model results. A large range for both positive and negative I_d (-10^6 to $10^6/100\,000$ exposures) suggests that there are parameter combinations that strongly impact the resultant predictions of risk in the dynamic model relative to the static model. The absolute value of I_d for the simulations is presented in Figure 4.

3.2. Sensitivity analysis

CART sensitivity analysis was conducted using the absolute value of I_d to determine which of the model parameters impacted the difference in incidence between the two models. Threshold I_d values of

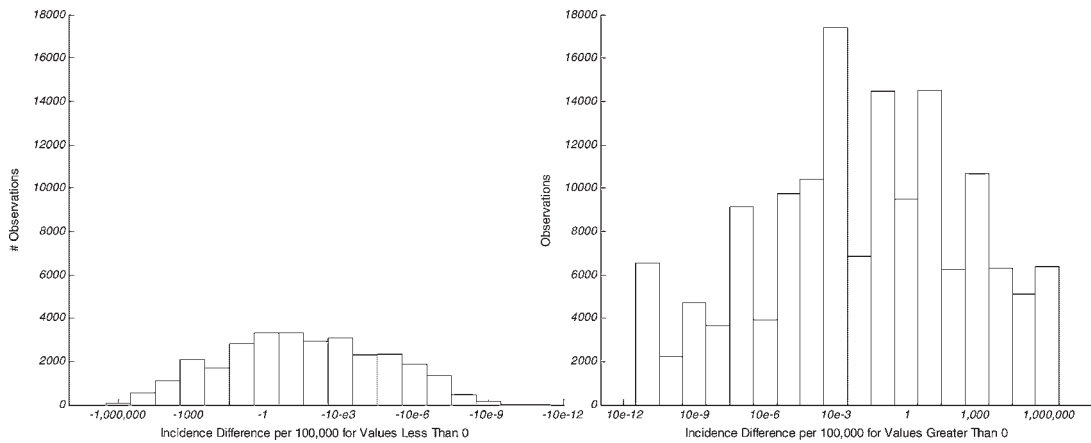


Figure 3. Incidence difference between static and dynamic simulations (per 100 000 exposures)

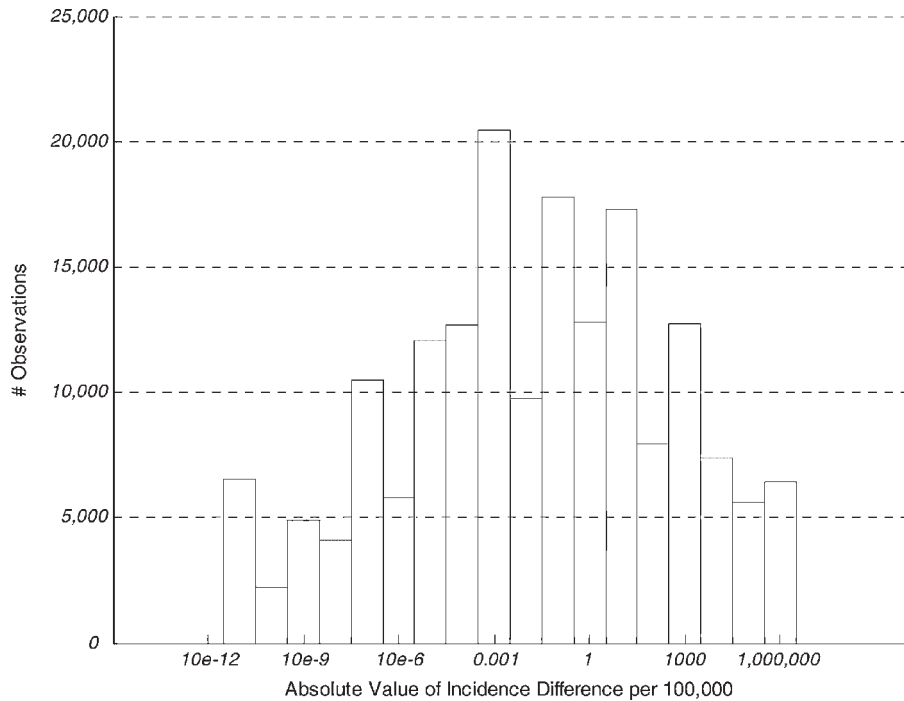


Figure 4. Absolute value incidence difference between static and dynamic simulations (per 100 000 exposures)

10/100 000, 1/100 000, and 0.01/100 000 were investigated. Results representative of these analyses are presented in Figure 5. Terminal nodes are presented as rectangles with the total number of observations in that node in bold font. Terminal nodes representing model parameter combinations with $I_d < 1/100\,000$ are white, and those with higher I_d are shaded gray.

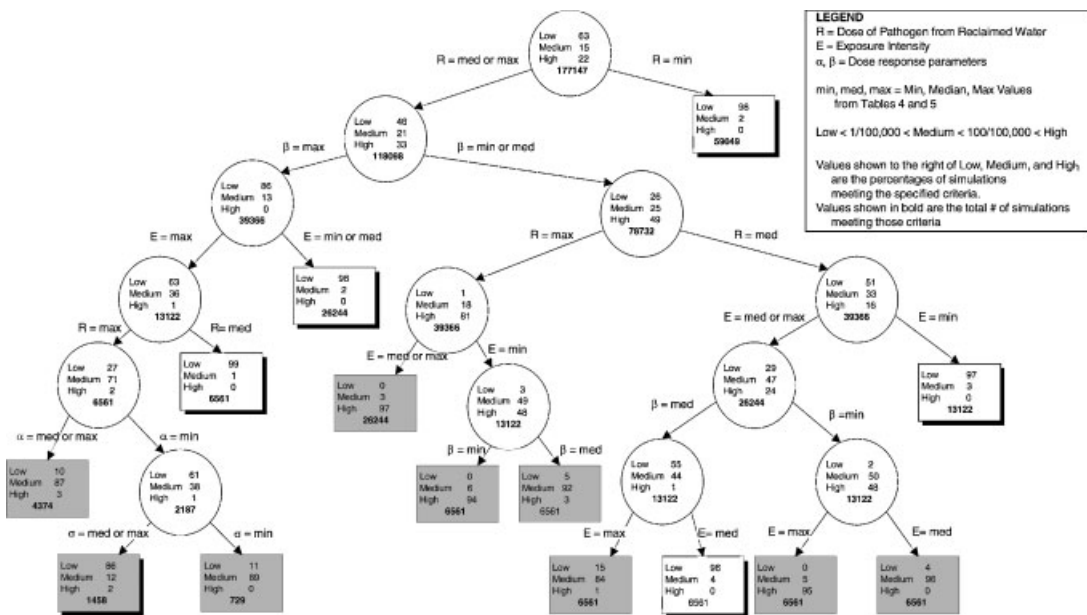


Figure 5. Representative CART output

Based on the results of the CART analyses, it may be inferred that the parameters that most strongly impact the difference in predicted incidence between the static and dynamic models are in order of decreasing importance: dose of pathogen, exposure intensity, dose–response parameter β , dose–response parameter α , and duration of infection.

3.3. Model parsimony

The combinations of model parameters that generally result in low I_d are summarized in Table 6. Inspection of Table 6 indicates that specification of three model parameters is generally sufficient to determine whether I_d will be low, depending on the threshold level selected (10/100 000, 1/100 000, or 0.01/100 000). Given that there are three levels for each of three critical parameters, there are 27 (3^3) critical parameter combinations. These critical parameter combinations are summarized graphically in Figure 6, which presents three 3×3 grids: one for each of the I_d threshold levels investigated.

In each Figure 6 grid, the levels of exposure intensity (proportion of the population exposed times the frequency of exposure) vary by row. For each threshold grid, the bottom row is low exposure intensity (3×10^{-5}), the middle row is medium (10^{-3}), and the top row is high (0.1). Similarly, in each grid the levels of average pathogen dose vary by column. The left column represents a low dose (10^{-8}), the middle column a medium dose (10^{-4}), and the right column a high dose (1.0). The relative levels of the dose–response parameter β vary by shape. In each grid, low β values (0.21) are represented by circles, medium (39) by squares, and high (440 000) by triangles. In Figure 6, gray shapes represent small differences in predicted incidence, and thus, conditions under which the static risk assessment model yields results similar to the dynamic model. White shapes indicate that the difference in predicted incidence is higher than the selected threshold, and thus are the conditions under which the dynamic risk assessment model would be necessary.

Table 6. Model parameter combinations that result in incidence differences below selected threshold

Threshold difference in predicted incidence = 10/100 000						
Parameter	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5	Condition 6
Dose of pathogen	Low	Medium/High	Medium	Medium	Medium	Medium
Exposure intensity	—	—	High	Low	Medium	Medium
Dose–response β	—	High	Medium	Low/Medium	Medium	Low
Dose–response α	—	—	—	—	—	Low
Threshold difference in predicted incidence = 1/100 000						
Parameter	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5	Condition 6*
Dose of pathogen	Low	Medium/High	Medium	Medium	Medium	High
Exposure intensity	—	Low/Medium	High	Low	Medium	High
Dose–response β	—	High	High	Low/Medium	Medium	High
Dose–response α	—	—	—	—	—	Low
Duration of infectiousness	—	—	—	—	—	Low/Medium
Threshold difference in predicted incidence = 0.01/100 000						
Parameter	Condition 1	Condition 2	Condition 3	Condition 4		
Dose of pathogen	Medium	High	Low	Low		
Exposure intensity	—	Low	Low/Medium	—		
Dose–response β	High	High	Low	Medium/High		

*Condition 6 results in low difference in 86% of cases, medium difference (1-100/100 000) 12.5% of cases, and high difference in 2% of cases.

3.4. Discussion

The concept of parsimony in model selection is well known—models should be as simple as possible, but no simpler. This has become relevant in MRA, where dynamic models are being used in conjunction with or as alternatives to static models. The sensitivity analysis employed in this study illustrates that the two classes of models do, at times produce substantially different results, and provides a quantitative means to assess whether the static model is appropriate or the more complex dynamic model is required. The results summarized in Figure 6 suggest that a static model provides satisfactory risk estimates under conditions where the risks associated with direct exposures to environmental contamination are low. These direct risks are defined in the models presented here by four factors: the percent of the population exposed, the frequency of exposure, the average dose levels to which individuals are exposed and the infectivity of the pathogen.

As the direct risks increase, the effects of secondary transmission and immunity also increase, resulting in greater differences between the static and dynamic models. These differences justify the need for a more complex dynamic model. Although the criteria for model selection provided in this analysis are specific for the exposure levels and pathogens explored in this study, analogous estimates could be obtained for alternative exposure scenarios by altering the exposure parameters in Table 5 and, if necessary the pathogen parameters in Table 4. It should however be emphasized that a relatively wide range of pathogen characteristics were evaluated in this analysis.

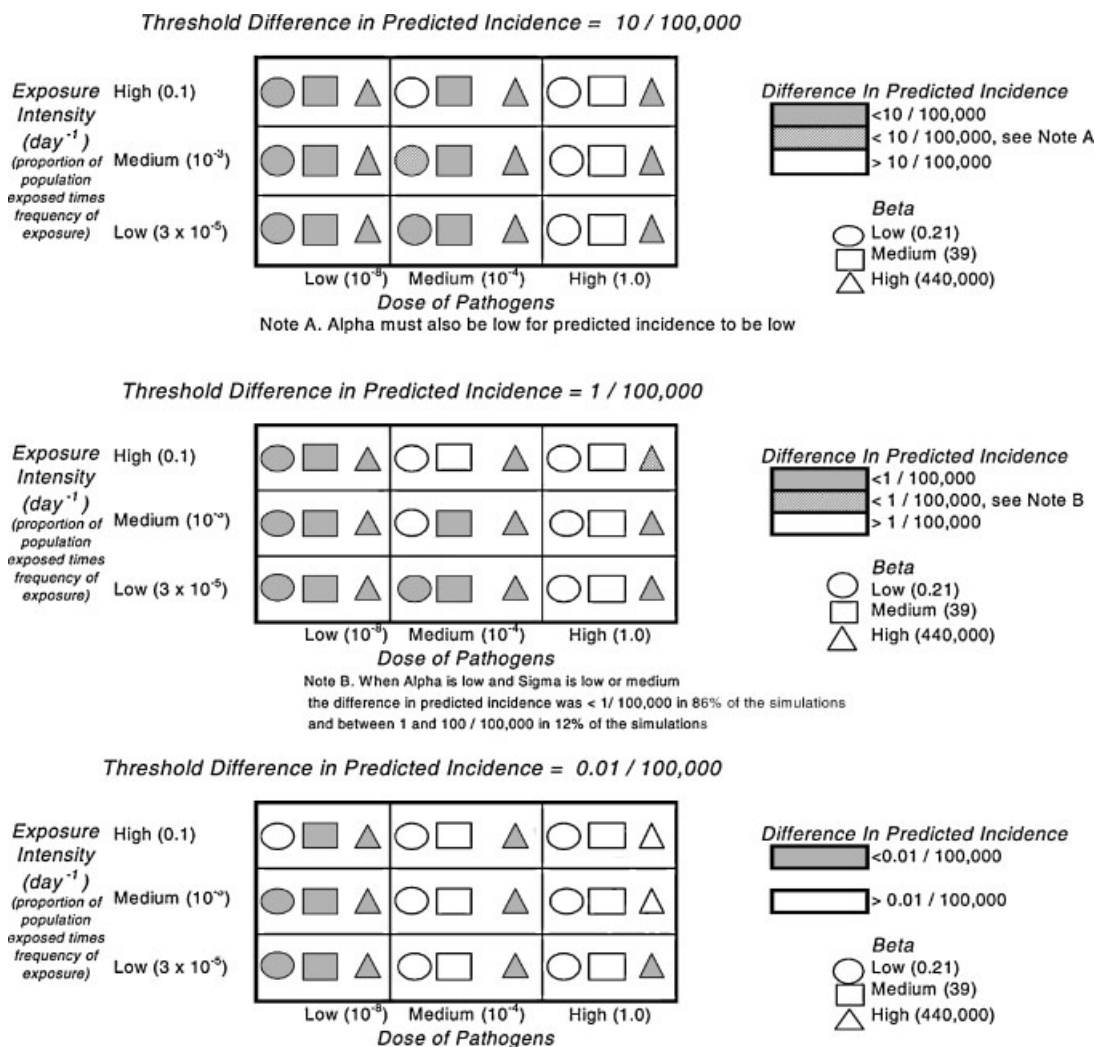


Figure 6. MRA model parsimony summary

For the specific scenario explored in this study, Figure 6 in combination with Tables 4 and 5 provides quantitative information on the definition of a low direct risk condition. For example, exposure intensity is defined as the proportion of the population that is exposed times the frequency of exposure for a given individual (exposure events per day). A low exposure intensity is defined here as 3 in 100 000, which corresponds to 1 exposure per 1000 individuals at a frequency of once per month. This level of exposure intensity could describe incidental exposures, whereas a high exposure intensity is defined here as 1 in 10. A low dose is defined as 10^{-8} (average number of pathogens ingested) compared with a high dose of 1. Finally, the infectivity ranges from highly infectious viruses such as rotavirus to less infectious bacteria such as *Salmonella*. As the tolerable level of predicted incidence difference increases, use of the static model becomes appropriate for more combinations of the critical model parameters.

Our finding that a low direct risk condition suggests that a static model is sufficient, is largely a function of the metric used for model selection, which is the absolute difference in incidence between the two models; that is, the bias of one model over another is measured by the number of cases of infection (or illness) that is either over- or under-reported by one model over the other. The reason that this metric was chosen was that incidence difference provides a public health relevant measure of bias. We use this form of model selection because traditional measures of validation that show one model form as superior over another are not appropriate. This is largely due to relatively low risk conditions (e.g., 1 in 1000 or 1 in 10 000) and the sparse nature of environmental and health data (Eisenberg *et al.*, 2002). Under these conditions, mechanistic models that incorporate what we know about the environmental and disease process become important and useful tools in risk assessment. These models provide a systematic means to integrate information on what is known about a process. Model selection techniques like the one presented here provide a framework to examine how different model forms perform under different circumstances and whether two models predict similar or different risks.

The results presented herein are only applicable within the bounds investigated and for the model forms evaluated. Thus, extrapolation of the results presented herein to routes of exposure, pathogens, and/or other model variants not investigated must be done with caution, because the parameter combinations that define when risk differences are sufficiently high to require a dynamic model are dependent both on the model structure and parameter ranges explored in the simulation analysis. Transmission rate parameters are particularly difficult to estimate and have the greatest potential for misspecification. We used a common technique to estimate the transmission rate from outbreak data where the initial and final number of susceptible individuals is known. Care must always be taken when generalizing these estimates. We use wide upper and lower bounds in the parameter estimates to help address this issue. With regards to model structure, transmission in the dynamic model presented here is characterized in a specific form, capturing one person-person pathway. Use of other model forms may result in findings different than those presented here. For example, using a model that included multiple routes of transmission including a person-environment-person route, Eisenberg *et al.* (2004) found that important factors for quantifying risk included the pathogen-shedding rate of infectious individuals, the person-person transmission potential, and the duration of immunity. Additionally, the model structure presented here assumes homogeneous mixing of individuals. Thus, further work is needed to examine the role of multiple transmission pathways and heterogeneous mixing on risk estimation.

4. CONCLUSION

We considered the question of parsimony for specific realizations of representative static and dynamic MRA models and identified conditions under which the more complex dynamic model provides sufficient additional insight to justify the added modeling complexity. The approaches presented provide a basis for model selection for a broad range of MRA applications. The results indicate that under low risk conditions, defined by a combination of exposure levels and infectivity of the pathogen, the simpler static model provides satisfactory risk estimates.

ACKNOWLEDGEMENTS

Funding for the research which provided the foundation for this manuscript was provided by the U.S. Environmental Protection Agency Cooperative Agreement No. CR-827345-01 with the Water Environment Research Foundation (Project Number 00-PUM-3). We gratefully acknowledge the contributions of Adam

Olivieri, Richard Danielson, Richard Sakaji, Vince Pon, Laura Johnson, and Carol Trese to the success of the WERF investigation. We also thank Robert Cooper, Robert Spear, George Tchobanoglous, Takashi Asano, Phil Berger, James Crook, James Geselbracht, Walt Jakubowski, and Ken Thompson for their thoughtful comments and insights during the course of the WERF investigation, and project officers Lola Olabode and Jami Montgomery. Finally, we are grateful to the anonymous peer reviewers and Alexandria Boehm for critical review of the draft manuscript.

REFERENCES

- Anderson RM, May R. 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press: New York.
- Buchanan RL, Lammerding AM, Clarke IR, van Schothorst M, Roberts TA. 1998. Potential application of risk assessment techniques to microbiological issues related to international trade in food and food products. *Journal of Food Protection* **61**: 1075–1086.
- Buchanan RL, Smith JL, Long W. 2000. Microbial risk assessment: dose-response relations and risk characterization. *International Journal of Food Microbiology* **58**: 159–172.
- Crabtree KD, Gerba CP, Rose JB, Haas CN. 1997. Waterborne adenovirus: a risk assessment. *Water Science and Technology* **35**: 1–6.
- Dudley RH, Hekimian KK, Mechalas BJ. 1976. A scientific basis for determining recreational water quality criteria. *Journal of Water Pollution Control Federation* **48**: 2661–2677.
- Eisenberg JNS, McKone TE. 1998. Decision tree method for the classification of chemical pollutants: incorporation of across-chemical variability and within-chemical uncertainty. *Environmental Science & Technology* **32**: 3396–3404.
- Eisenberg JN, Seto EYW, Olivieri AW, Spear RC. 1996. Quantifying water pathogen risk in an epidemiological framework. *Risk Analysis* **16**: 549–563.
- Eisenberg JNS, Seto EYW, Colford JM, Olivieri AW, Spear RC. 1998. An analysis of the Milwaukee cryptosporidiosis outbreak based on a dynamic model of the infection process. *Epidemiology* **9**: 255–263.
- Eisenberg JNS, Brookhart MA, Rice G, Brown M, Colford JM. 2002. Disease transmission models for public health decision making: Analysis of epidemic and endemic conditions caused by waterborne pathogens. *Environmental Health Perspectives* **110**: 783–790.
- Eisenberg JNS, Soller JA, Scott J, Eisenberg DM, Colford JM. 2004. A dynamic model to assess microbial health risks associated with beneficial uses of biosolids. *Risk Analysis* **24**: 221–236.
- Eisenberg JN, Lei X, Hubbard AH, Brookhart MA, Colford JM, Jr. 2005. The role of disease transmission and conferred immunity in outbreaks: analysis of the 1993 *Cryptosporidium* outbreak in Milwaukee, Wisconsin. *American Journal of Epidemiology* **161**: 62–72.
- Farber JM, Ross WH, Harwig J. 1996. Health risk assessment of *Listeria monocytogenes* in Canada. *International Journal of Food Microbiology* **30**: 145–156.
- Fuhs OW. 1975. A probabilistic model of bathing beach safety. *Science of the Total Environment* **4**: 165–175.
- Gerba CP, Rose JB, Haas CN, Crabtree KD. 1996. Waterborne rotavirus: a risk assessment. *Water Research* **30**: 2929–2940.
- Haas CN. 1983. Effect of effluent disinfection on risks of viral disease transmission via recreational water exposure. *Journal of the Water Pollution Control Federation* **55**: 1111–1116.
- Haas CN, Rose JB, Gerba CP. 1999. *Quantitative Microbial Risk Assessment*. J.W. Wiley, Inc: New York.
- Hethcote H. 1976. Qualitative analyses of communicable disease models. *Mathematical Biosciences* **28**: 335–356.
- Hethcote HW. 2000. The mathematics of infectious diseases. *Siam Review* **42**: 599–653.
- ILSI. 1996. A conceptual framework to assess the risks of human disease following exposure to pathogens. *Risk Analysis* **16**: 841–848.
- ILSI. 2000. *Revised framework for microbial risk assessment*. ILSI Press: Washington, DC.
- Koopman JS, Chick SE, Simon CP, Riolo CS, Jacquez G. 2002. Stochastic effects on endemic infection levels of disseminating versus local contacts. *Mathematical Biosciences* **180**: 49–71.
- Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, *et al.* 1999. Food related illness and death in the United States. *Emerging Infectious Diseases* **5**: 607–625.
- Mena K, Gerba CP, Haas CN, Rose JB. 2003. Risk assessment of waterborne coxsackievirus. *Journal American Water Works Association* **95**: 122–131.
- National Research Council. 1983. *Risk Assessment in the Federal Government, Managing the Process*. National Academy Press: Washington, D.C.
- Regli S, Rose JB, Haas CN, Gerba CP. 1991. Modeling the risk from giardia and viruses in drinking-water. *Journal American Water Works Association* **83**: 76–84.
- Rose JB, Haas CN, Regli S. 1991. Risk assessment and control of waterborne giardiasis. *American Journal of Public Health* **81**: 709–713.

- Schaub S. 2004. A risk assessment framework for waterborne pathogens and requirements for producing a complete protocol. *Human and Ecological Risk Assessment* **10**: 151–158.
- Soller JA, Eisenberg JN, Olivieri AW. 1999. Evaluation of pathogen risk assessment framework. *Prepared by EOA Inc. and U.C. Berkeley for ILSI Risk Science Institute*.
- Soller JA, Olivieri A, Crook J, Parkin R, Spear R, Tchobanoglous G, Eisenberg JNS. 2003. Risk-based approach to evaluate the public health benefit of additional wastewater treatment. *Environmental Science & Technology* **37**: 1882–1891.
- Soller JA, Olivieri AW, Eisenberg JNS, Sakaji R, Danielson R. 2004. Evaluation of microbial risk assessment techniques and applications. *Water Environment Research Foundation Report 00-PUM-3*.
- Soller JA, Eisenberg J, DeGeorge J, Cooper R, Tchobanoglous G, Olivieri A. 2006. A public health evaluation of recreational water impairment. *Journal of Water Health* **4**: 1–19.
- Steinberg D, Colla P. 1997. *CART—classification and regression trees*. Salford Systems: San Diego, CA.
- Teunis PF, Havelaar AH. 1999. Cryptosporidium in drinking water: evaluation of the ILSI/RSI quantitative risk assessment framework RIVM report no. 284 550 006. National Institute of Public Health and the Environment, Bilthoven, The Netherlands.
- Teunis PFM, Medema GJ, Kruidenier L, Havelaar AH. 1997. Assessment of the risk of infection by Cryptosporidium or Giardia in drinking water from a surface water source. *Water Research* **31**: 1333–1346.
- Teunis PF, Chappell CL, Okhuysen PC. 2002. Cryptosporidium dose-response studies: variation between hosts. *Risk Analysis* **22**: 475–485.
- U.S. EPA. 2002. National primary drinking water regulations: long term 1 enhanced surface water treatment rule. 40CFR Parts 9, 141 and 142, Volume 67, Number 9, January 14, 2002.
- U.S. EPA. 2003. Implementation and enforcement of the combined sewer overflow control policy: EPA833-R-01-003.
- U.S. EPA. 2004. Developing dynamic infection transmission models for microbial risk assessment applications, EPA-NCEA-C-1463.
- U.S. EPA. 2006a. National primary drinking water regulations: long term 2 enhanced surface water treatment rule (LT2ESWTR); final rule. 40CFR Parts 9, 141 and 142, Volume 71, Number 654, January 5, 2006.
- U.S. EPA. 2006b. National primary drinking water regulations: ground water rule. 40CFR Parts 9, 141 and 142, Volume 71, Number 216, November 8, 2006.
- U.S. EPA. 2006c. Appendices to the economic analysis for the final ground water rule, Appendix E.

APPENDIX: MATHEMATICAL EQUATIONS FOR DYNAMIC MODEL

This appendix presents the equations that correspond to the model depicted in Figure 2. The model is defined by the following six state variables:

- S Number of individuals in susceptible state.
- E Number of individuals who have been exposed but are non-infectious.
- C_1 Number of individuals in carrier state (asymptomatic and infectious).
- D Number of individuals in diseased state (symptomatic and infectious).
- C_2 Number of individuals in carrier state (asymptomatic and infectious).
- P Number of individuals in post-infection state.

and the following parameters:

- α Transition rate out of exposed state. Inverse of incubation period.
- p_{sym} Probability of symptomatic response.
- δ Transition rate out of diseased state. Inverse of disease duration.
- σ_1, σ_2 Transition rate out of carrier state. Inverse of duration of shedding.
- γ Transition rate out of post-infection state. Inverse of duration of immunity.
- β_{s1} Transmission rate (susceptible to exposed from environmental exposure).
- β_{s2} Transmission rate (susceptible to exposed from person-to-person contact).

Movement within the E , C_1 , C_2 , D , and P states is characterized by a distributed delay. This delay is realized by a series of ordinary differential equations. In such a characterization, the time spent in any state is described by a gamma distribution defined by two parameters: n , the number of first order processes; k , the rate constant governing the movement of one first-order process to the next. For example, the following set of equations depict the transition through the exposed state, where I_1 is the rate of entering and I_0 is the rate of leaving the exposed state.

$$\begin{aligned}\frac{dE_1}{dt} &= I_1 - kE_1 \\ \frac{dE_2}{dt} &= kE_1 - kE_2 \\ \frac{dE_3}{dt} &= kE_2 - kE_3 \\ \frac{dE_4}{dt} &= kE_3 - I_0\end{aligned}$$

The parameters n and k determine the mean and variance of the gamma distribution:

$$\begin{aligned}\text{Mean} &= \frac{n}{k} \\ \text{Variance} &= \frac{n}{k^2}\end{aligned}$$

For the above equations the mean transition rate out of the exposure state, α , is n/k . For computational purposes, n was set to 4 in all cases. The mean transit time for each state, therefore, is 4 divided by the corresponding transition rate parameter.

The change in the number of susceptible individuals is governed by three possible events. Two are transmission events that result in a decrease in the number of susceptible individuals through either contact with an infectious individual (a individual in either the D , C_1 , or C_2 state), or contact with pathogens from reclaimed water or from background concentrations in the environment. The third is a transition event in which a protected individual becomes susceptible again. The following equation describes this change in the number of susceptible individuals.

$$\frac{dS}{dt} = -\beta_2 S(D + C_1 + C_2) - \beta_1 S + \gamma P_4$$

For details regarding description of the parameters and their values, see the main text. The parameter β_1 is defined as follows:

$$\beta_1 := 1 - \left(1 + \frac{\text{rec_dose_exp_intensity} + \text{back_dose}}{\beta} \right)^{-\alpha}$$

where the recreational dose and exposure intensity are exposure-dependent parameters; and the background dose, β and α are pathogen-dependent parameters (effective background concentration level, dose–response parameter α , and dose–response parameter β , respectively).

To incorporate the concepts of incomplete and waning immunity, four protected states (P_1, P_2, P_3, P_4) were included in the model. In the first state (P_1), all individuals are completely immune to re-infection. In the subsequent three states (P_2, P_3, P_4) re-infection becomes possible to a varying degree; in each successive protected state the level of protection decreases until, finally, the population is completely susceptible once again. As a result, movement from three of the protected states (P_2, P_3, P_4) to the exposed state (E) is possible and is regulated by the following relationship:

$$\beta_{pji} = \frac{\beta_{st}^{(i-1)}}{n}$$

where j takes on the value of 1 or 2 corresponding to environmental or person-to-person transmission; i takes on the value of 1, 2, 3, or 4 corresponding to the particular protected state; and n corresponds to the total number of protected states (four in this case). The following equation describes the immune process.

$$\begin{aligned} \frac{dP_1}{dt} &= \sigma_1 C_{14} + k_{\sigma 2} C_{24} - \gamma P_1 \\ \frac{dP_2}{dt} &= \gamma P_1 - \gamma P_2 - \beta_2 (1/4) P_2 (D + C_1 + C_2) - (1/4) \beta_1 P_2 \\ \frac{dP_3}{dt} &= \gamma P_2 - \gamma P_3 - \beta_2 (2/4) P_3 (D + C_1 + C_2) - (2/4) \beta_1 P_3 \\ \frac{dP_4}{dt} &= \gamma P_3 - \gamma P_4 - \beta_2 (3/4) P_4 (D + C_1 + C_2) - (3/4) \beta_1 P_4 \end{aligned}$$

Individuals enter the initial protected state, P_1 , from either of the two carrier states (see model diagram in Figure 2). The transition through the protected states occurs at a rate, $k\gamma$. For the remaining protected states, P_2, P_3, P_4 , there are two transmission events possible on through contact with another infectious individual and the other through contact with a contaminated environment.

Once infected individuals are in the exposed state that represents the state of being infected but not infectious. Four transmission events enter into the initial exposed state, E_1 . This state is governed by the following set of equations.

$$\begin{aligned} \frac{dE_1}{dt} = & \beta_2 S(D + C_1 + C_2) + \beta_1 S + \beta_2(1/4)P_2(D + C_1 + C_2) + \beta_1(1/4)P_2 \\ & + \beta_2(2/4)P_3(D + C_1 + C_2) + \beta_1(2/4)P_3 + \beta_2(3/4)P_4(D + C_1 + C_2) \\ & + \beta_1(3/4)P_4 - \alpha E_1 \end{aligned}$$

$$\frac{dE_2}{dt} = \alpha E_1 - \alpha E_2$$

$$\frac{dE_3}{dt} = \alpha E_2 - \alpha E_3$$

$$\frac{dE_4}{dt} = \alpha E_3 - \alpha E_4$$

Once an individual completes the transition through the exposed state, a proportion, p_{sym} , enter the diseased state, D (infectious and symptomatic) and then transition into a carrier state, C_2 . The remaining proportion, $1 - p_{\text{sym}}$, enters a carrier state, C_1 (infectious and asymptomatic). The following equations describe the transition through the diseased and two carrier states.

$$\begin{aligned} \frac{dD_1}{dt} &= p_{\text{sym}}\alpha E_4 - \delta D_1 & \frac{dC_{2_1}}{dt} &= \delta D_4 - \sigma_2 C_{2_1} \\ \frac{dD_2}{dt} &= \delta D_1 - \delta D_2 & \frac{dC_{2_2}}{dt} &= \sigma_2 C_{2_1} - \sigma_2 C_{2_2} \\ \frac{dD_3}{dt} &= \delta D_2 - \delta D_3 & \frac{dC_{2_3}}{dt} &= \sigma_2 C_{2_2} - \sigma_2 C_{2_3} \\ \frac{dD_4}{dt} &= \delta D_3 - \delta D_4 & \frac{dC_{2_4}}{dt} &= \sigma_2 C_{2_3} - \sigma_2 C_{2_4} \end{aligned}$$

$$\frac{dC_{1_1}}{dt} = (1 - p_{\text{sym}})\delta E_4 - \sigma_1 C_{1_1}$$

$$\frac{dC_{1_2}}{dt} = \sigma_1 C_{1_1} - \sigma_1 C_{1_2}$$

$$\frac{dC_{1_3}}{dt} = \sigma_1 C_{1_2} - \sigma_1 C_{1_3}$$

$$\frac{dC_{1_4}}{dt} = \sigma_1 C_{1_3} - \sigma_1 C_{1_4}$$