

Disease Transmission Models for Public Health Decision Making: Analysis of Epidemic and Endemic Conditions Caused by Waterborne Pathogens

Joseph N. S. Eisenberg,^{1,2} M. Alan Brookhart,² Glenn Rice,³ Mary Brown,³ and John M. Colford, Jr.^{1,2}

¹Center for Occupational and Environmental Health and ²School of Public Health, University of California, Berkeley, California, USA;

³U.S. Environmental Protection Agency, Office of Research and Development/National Center for Environmental Assessment, Cincinnati, Ohio, USA

Developing effective policy for environmental health issues requires integrating large collections of information that are diverse, highly variable, and uncertain. Despite these uncertainties in the science, decisions must be made. These decisions often have been based on risk assessment. We argue that two important features of risk assessment are to identify research needs and to provide information for decision making. One type of information that a model can provide is the sensitivity of making one decision over another on factors that drive public health risk. To achieve this goal, a risk assessment framework must be based on a description of the exposure and disease processes. Regarding exposure to waterborne pathogens, the appropriate framework is one that explicitly models the disease transmission pathways of pathogens. This approach provides a crucial link between science and policy. Two studies—a *Giardia* risk assessment case study and an analysis of the 1993 Milwaukee, Wisconsin, *Cryptosporidium* outbreak—illustrate the role that models can play in policy making. **Key words:** *Cryptosporidium*, epidemic, *Giardia*, infectious disease, mathematical models, waterborne pathogens. *Environ Health Perspect* 110:783–790 (2002). [Online 17 June 2002]

<http://ehpnet1.niehs.nih.gov/docs/2002/110p783-790eisenberg/abstract.html>

Infectious diseases are a major cause of morbidity and mortality worldwide. In a recent study by the World Health Organization, ranking the global burden of diseases, five of the top seven diseases in developing countries were caused by infectious pathogens (1). Although infectious diseases are not as prevalent in developed countries, the emergence of human immunodeficiency virus, hepatitis C, Lyme disease, *Cryptosporidium*, and others has resulted in a resurgence of public health concern with infectious disease. To obtain regional estimates of disease burden, data are often collected through surveillance activities. These patterns of disease discerned through surveillance are caused by complex interactions of social, biologic, and environmental processes. Although surveillance information can be used to estimate a crude measure of disease burden, seldom can it provide information on the specific underlying causes of disease. Models of disease transmission, on the other hand, can provide a framework from which to address these questions of causality. Because an understanding of the specific causes of disease is crucial in attempts to design effective intervention and control strategies, these models can be useful in decision making.

One fundamental property of infectious diseases, including diseases caused by waterborne pathogens, is that these complex interactions always result from an infectious individual or environmental source transmitting the pathogen to a susceptible individual (2). In this article we provide a perspective suggesting that a thorough understanding of the system of interdependent transmission

pathways is crucial in formulating sound public health policy decisions. The theoretical framework proposed explicitly models the transmission pathways of waterborne infectious pathogens that cause disease. We demonstrate that this model structure offers a framework that can be applied when data are limited. With limited data, the model can be used to assess which data must be collected to improve understanding of the relevant processes, as well as to provide sensitivity information for decision making. Targeted interventions to reduce disease then may be more responsibly designed and their potential impacts more thoroughly analyzed using the model framework.

The epidemiology of diseases caused by waterborne pathogens suggests that illness arises in two broad settings: outbreaks and endemic transmission. A large number of observed cases occur from outbreaks and can be reasonably characterized by incidence data collected through outbreak investigations. By contrast, few data are available with which to evaluate endemic transmission. Therefore, the epidemic and endemic conditions necessarily require different quantitative approaches. After a brief introduction to waterborne transmission models, we present two case studies to illustrate the use of these models in both outbreak and endemic settings.

The Use of Mathematical Models as Tools for Public Health Policy

Developing effective public health policy requires integrating large collections of information that are diverse, highly variable, and

uncertain. The types of data relevant to evaluating disease attributable to waterborne pathogens range broadly from microbiologic profiles to clinical syndromes. Microbiologic data include the occurrence of pathogens in the environment and survival of the pathogen under varying climatic and environmental conditions. Clinical data include duration of illness, duration of infectivity, duration and degree of protection due to prior exposure to the pathogen, and the transmission potential in various human populations.

Transmission potential is an integrated measure of both infectivity and an individual's opportunity for contact with the environment and other infectious individuals. The microbiologic and clinical data combine to provide valuable information on the natural history of the pathogen and its ability to cause disease in human populations. Each factor involved in disease transmission is known to varying degrees of certainty. Many factors are highly uncertain because of either a limited number of studies in the literature or technologic limitations in measurement. Disease transmission also has a high degree of variability because of heterogeneity in humans, microorganisms, and environmental conditions. Heterogeneous human populations can be characterized by various factors such as age or immunocompromised status. Mutation and adaptive mechanisms result in variations in virulence factors among the genotypes of microorganisms, leading to the heterogeneity observed in these populations. Climatic variations and human interventions are major causes of the variation in environmental conditions. One important feature of a disease transmission model is the insight it provides into how each factor affects disease burden and how

Address correspondence to J. Eisenberg, School of Public Health, University of California at Berkeley, 140 Warren Hall, MC 7360, Berkeley, CA 94720-7360 USA. Telephone: (510) 643-9257. Fax: (510) 642-5815. E-mail: eisenber@socrates.berkeley.edu

We thank P. Murphy of the U.S. Environmental Protection Agency for her support for this project and her thoughtful insights into this problem area. We also thank R.M. Clark of the U.S. EPA.

This work was funded by a cooperative agreement from the U.S. EPA, Office of Research and Development/National Center for Environmental Assessment in Cincinnati (CR 827424-01-0).

Received 4 October 2001; accepted 1 February 2002.

the uncertainties and sources of variability inherent in environmental systems affect the resulting uncertainties in policy decision making.

Mathematical models that describe infectious disease transmission can be traced back to the early 1900s with Sir Ronald Ross's work on the relationship between mosquito population levels and the incidence of malaria (3), and William Hamer's work examining measles epidemic patterns (4). Both Ross and Hamer postulated that epidemic and endemic patterns depended on the rate at which susceptible individuals contacted infected individuals. This postulate has been the cornerstone of disease transmission models throughout the 20th century (5). In the case of malaria, Ross suggested that mosquitoes mediate this contact rate and used this model to justify mosquito control as a viable option to decrease disease incidence. In the case of measles, models such as that used by Hamer have been instrumental in helping to develop vaccine programs (6). In recent years, transmission models have been used extensively to study epidemic and endemic infectious disease processes for a wide array of infectious diseases such as measles, tuberculosis, and more recently human immunodeficiency virus infection (5). In general, these models follow the postulate of Ross and Hamer that the rate at which susceptible individuals within a population become infected (the transmission rate) is proportional to both the current number of infectious and current number of susceptible individuals (7). Surprisingly, despite the increase in the use of disease transmission models to study infectious disease processes and despite the known, massive public health problems associated with infectious diarrhea and gastrointestinal illness worldwide, few publications have examined disease transmission through waterborne pathogens. Two of these publications, by Eisenberg et al. (8) and Brookhart et al. (9), demonstrate the importance of both disease transmission and the immune process in understanding risk, and are discussed below.

Framework for Decision Making: Model Development

Transmission pathways: the conceptual model. Disease transmission models closely track the disease status of the population under study throughout the natural history of the disease in a population. Therefore, output of the model may consist, for example, of the number of susceptible, infectious, and protected (because of prior exposure to the pathogen) individuals as a function of time. These functions are called state variables, and the disease state at any given time

can be assessed by looking at the appropriate model output. Risk, as measured by incidence, can be estimated from the model as the per capita transmission rate of susceptible individuals into the infectious state. The specific categories of disease status used in the model are determined by the pathogen being modeled. Each category is one state variable; for example, the susceptible state variable tracks the number of susceptible individuals as a function of time. This model framework works best for pathogens that reproduce within the host, such as bacteria, viruses, and protozoa (5). An alternative framework that best addresses transmission of helminth pathogens is discussed elsewhere (5).

Once the disease states are defined, the different transmission pathways to be modeled determine the details of the model structure. The life cycle of the pathogen dictates transmission pathways. In contrast with noninfectious disease, where risk is independent of the disease status of the population, in infectious diseases the source of pathogens ultimately becomes the infected hosts. The risk of infection and illness, therefore, is related not only to the concentration of microbial pathogens in the environment but also to the number of infected hosts (2). For pathogens in which humans are the only host, the degree of environmental contamination is related to the number of infected people. For example, as more people become infected with rotavirus in a community, the likelihood that uninfected individuals will become infected rises. For many infectious diseases, the causative pathogen reproduces within the human host. In substantial contrast to noninfectious diseases (e.g., toxic exposures), the human host therefore acts as an amplifier in the disease transmission process. For a pathogen to persist in a population, it must reproduce in sufficient numbers within a given host to allow infection of additional hosts. Some infectious diseases (e.g., *Salmonella* serotypes other than *S. typhi*) have nonhuman hosts. If these diseases are maintained within an animal population and sporadically introduced to human hosts, they are referred to as enzootic. In general, humans can become infected through ingestion, inhalation, or dermal contact with pathogens.

We are concerned here with the pathways associated with waterborne pathogen transmission. One salient feature of waterborne pathogens is their ability to survive in the environment outside of a host. The duration of survival varies widely from presumably very short time periods to many years under favorable circumstances. This environmental phase largely dictates the possible transmission pathways a waterborne pathogen can exploit in completing its life

cycle. Pathways include transmission *a)* from person to person, *b)* from person to environment to person, and *c)* from environment to person. Person-person transmission is often associated with poor hygiene; such transmission can occur, for example, within households or other communal settings such as schools, day care centers, and nursing homes. Person-environment-person transmission is often associated with environmental contamination of water sources; such transmission can occur, for example, through water used for drinking or recreation. Environment-person transmission is often associated with sources of contamination external to the population under study, including animal sources and human sources from other communities.

The magnitude of these different transmission pathways often dictates the optimal intervention and control. Possible control measures include the treatment of water or other environmental media, limiting exposure to water or other environmental media, and prevention of contamination through sanitation and hygiene measures. Each of these interventions not only may reduce the disease burden associated directly with its pathway but, critically, may also reduce indirect transmission from other pathways by decreasing the amount of contamination. The degree of contamination, and therefore the degree of risk, depends on the contribution and interactions of all of the different environmental transmission pathways. This interdependency of pathways makes obvious the fact that determining the most effective control strategies requires an understanding of the complete transmission cycle.

The transmission cycle has two important features other than the interdependency of pathways. One is the potential for an individual to be infectious but not symptomatic. These asymptomatic individuals are usually mobile because of lack of illness and have a high potential to spread a pathogen widely throughout a community. The second feature is the protection conferred to a host after exposure to a pathogen, that is, acquired immunity. For some pathogens (e.g., hepatitis A), once a person has been infected, he or she will never contract the illness again; the protection is lifelong. However, for most waterborne pathogens the protection conferred to a host after exposure to the agent of disease is partial and temporary. For example, an individual with protective immunity from prior exposure may require a larger dose for infection to occur or for symptoms to develop. Such partial protection may last for months or years. This property of infectious disease has major implications for transmission both within and between populations. The greater the

number of individuals who are partially protected, the smaller the pool of susceptible individuals who are at risk of infection. This in turn implies that the pool of newly infected individuals will be smaller in the future. The decreased number of infected individuals in the future means decreasing contamination, decreasing the exposure risk. Offsetting this decrease in exposure are population dynamic processes such as birth and immigration that will increase the number of susceptible individuals.

Disease transmission models for enteric pathogens. We can conceptualize an epidemiologically based characterization of risk by dividing the population into distinct states regarding disease status. Epidemiologic states may include susceptible, diseased (infectious and symptomatic), immune (either partial or complete), and carrier (infectious but asymptomatic) populations. Over time, members of the population may move between these states. Factors affecting the rate at which members move between states include level of exposure to an environmental pathogen, intensity of exposure to individuals in the infectious or carrier state, and the temporal processes of the disease (e.g., incubation period, duration of disease, duration of protective immunity). This conceptual modeling methodology is dynamic and population based; that is, the risk of infection manifests at the population level. Specifically, in the transmission of infectious diseases (but not of diseases due to chemical exposure), the risk of disease due to pathogen exposure depends on the disease status of the population and potentially on the contact patterns within the population.

Figure 1 is a diagram of a transmission model for enteric pathogens. Each box represents one state of the system. Five of the six states represent the epidemiologic states of the population: S , susceptible; E , latent (infected but noninfectious and asymptomatic); I_S , diseased (infectious and symptomatic); I_A , carrier (infectious but asymptomatic); P , immune (either partial or complete). The sixth state, W , represents concentration of pathogens in the environment. Members of a given state may move to another state based on the causal relationships of the disease process. For example, members of the population who are in the susceptible state may move to the diseased state after exposure to a pathogenic agent.

To describe the epidemiology of enteric pathogen transmission, the conceptual model includes both state variables and rate parameters. State variables (S , E , I_A , I_S , and P) track the number of individuals in each of the states at any given point in time and are defined such that $S + E + I_A + I_S + P = N$ (i.e., the sum of the state variables equals the

total population). The rate parameters determine the movement of the population from one state to another. In general, the rate parameters are β , the rate of transmission from a noninfected state, S , to an infected state, E , due to both environmental (e.g., drinking water) and person–person exposure to a pathogen; α , the rate of movement from exposure to illness; δ and σ , the rates of recovery from an infectious state, I_S or I_A , respectively, to the postinfection state, P ; γ , the rate of movement from the postinfection state (partial immunity), P , to the susceptible state, S ; ϕ , the rate of shedding of pathogens into the environment by infectious individuals; and ξ , the per capita mortality rate of the pathogen in the environment. An additional parameter in the model, ρ , represents the proportion of asymptomatic infections. For more mathematical detail pertaining to the model, see previous publications (8,9).

The rate parameters are estimated through literature review. These parameters may be functions of other variables also determined through literature review, or may be determined through site-specific data where possible and appropriate. One technical aspect of this approach is that the distribution of time that members of the population spend in each of the states is assumed to be exponential. In some cases, this assumption is unrealistic. Relatively straightforward methods, however, modify the model structure so that the time spent in each state is other than exponential (10).

A fundamental concept in disease transmission models is the reproduction number, R_0 , defined as the number of infections that result from the introduction of one index case into a population of susceptible individuals. Therefore, R_0 is a measure of the ability of a pathogen to move through a population. $R_0 > 1$ suggests that the pathogen is multiplying within a community and that prevalence is increasing, whereas

$R_0 < 1$ suggests that the disease is dying out of the population. An average R_0 of 1 suggests that the disease is endemic in the population. Various methods estimate R_0 for different pathogens and in different environmental settings (11). Measles, for example, is a highly infectious respiratory-transmitted disease and has been estimated to have an R_0 of approximately 14. Polio, on the other hand, a waterborne pathogen, has an R_0 of approximately 6. The reproduction number for a waterborne pathogen is a function of the different pathways of transmission. For example, for the model shown in Figure 1, R_0 is composed of two terms, one for person–person transmission and the other for person–environment–person transmission:

$$\beta_2 \left(\frac{1}{\rho/\delta + (1-\rho)/\delta} \right) + \beta_1 \left(\frac{1}{\rho/\delta + (1-\rho)/\delta} \right) \left(\frac{1}{\phi + \xi} \right).$$

In this expression, each term has a β factor that represents the transmission potential of the pathogen and a factor that represents the average duration of infectiousness. The last term has a factor that represents the proportion of pathogens that survive the environmental phase of its life cycle.

Application to the Outbreak Condition

In the United States, outbreaks are responsible for a large component of the measured incidence for waterborne pathogens (12,13). Whether or not these outbreak cases represent a significant portion of the total disease burden associated with waterborne pathogens is debated because endemic cases of waterborne disease are poorly measured by existing surveillance systems (14,15). However, a thorough understanding of the causes of these outbreaks is necessary to understand

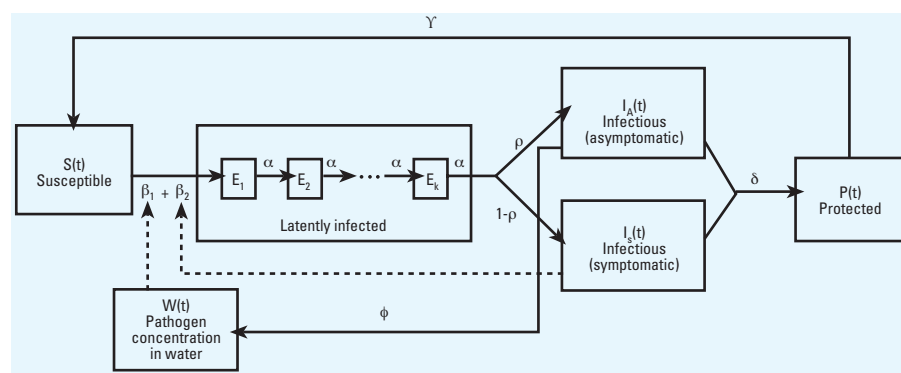


Figure 1. Schematic diagram of transmission model. t , independent variable representing time. Solid lines represent movement of individuals from one state to another. Dashed lines represent movement of pathogens either directly from infectious host to susceptible host or indirectly via the environment. State variables and parameters are defined in the text.

the full transmission process and to develop preventive measures. In this section, we describe how the information embedded in the disease transmission model structure, coupled with the incidence data from outbreaks, can inform health policy decisions. The 1993 waterborne outbreak in Milwaukee, Wisconsin, provides a good case study because extensive data were collected during the outbreak investigation.

Case study: analysis of the Milwaukee *Cryptosporidium* outbreak. In March 1993 an outbreak of cryptosporidiosis occurred in Milwaukee, caused by a contaminated water supply (16). An estimated 403,000 cases of watery diarrhea occurred in the greater Milwaukee area (Figure 2). Investigation of the water treatment facilities suggested that from 23 March to 9 April, one of the two plants serving the area failed to adequately remove *Cryptosporidium* oocysts.

We analyzed this outbreak to examine the possible combination of factors that could bring about the time series of incidence rates observed during the 1993 Milwaukee cryptosporidiosis outbreak. These factors include both clinical properties and surveillance data. Clinical properties include dose response, person–person transmission rates, incubation period, duration of immunity, duration of symptoms, percent infected that become symptomatic, and rate of pathogens shed into the environment by infectious individuals. Surveillance data include nonoutbreak incidence levels of cryptosporidiosis, levels of oocysts in surface waters before and during the outbreak, and treatment efficiency both before and during the outbreak. We used an epidemiologically based mathematical model to integrate and analyze these data in the context of the *Cryptosporidium* outbreak (10).

Traditionally, epidemic phenomena that have been analyzed through simulation studies of transmission models have provided qualitative insight into the dynamics of disease transmission. These models have proven valuable because they are process driven—they are an actual representation of how the transmission system is thought to operate. Less commonly, these models are used inferentially through statistical methods, allowing us to estimate confidence intervals on the parameters of the transmission model based on the incidence data. As described below, we used the latter, statistical approach to explore the plausibility of different parameter sets and to provide insight into the epidemic process.

Analytic approach: developing inferences from the incidence data based on model and parameter values. Fundamental to our statistical approach was the linking of a mathematical model to the data through a

statistical likelihood function. We achieved this by using a deterministic mathematical model to generate predicted incidence data. The likelihood function incorporated the predicted and observed data, returning a value that represents an empirical measure of the plausibility of the observed incidence data. Higher values of the likelihood function correspond to a closer fit between the predicted incidence from the mathematical model and the actual observed incidence.

The primary obstacle to likelihood estimation and statistical inference in mechanistic models is that they typically contain many parameters and are nonlinear or even nonidentifiable. As a consequence, the likelihood function contains many local maximum values and the solution contains complex interdependence among the parameters.

With this in mind, we considered several methods of conducting estimation and inference suggested by standard statistical theory. The approach we found most useful was based on the profile likelihood function. Maximum likelihood estimates are obtained by fixing a set of parameters of interest over a range of values and then maximizing the full likelihood over the remaining “nuisance” parameters. The likelihood ratio test provides a reference value for evaluating particular parameter sets. Those sets with profile likelihood values above the reference value are considered reasonable and are included in the confidence interval, whereas those below are rejected as implausible. This approach leads to robust confidence intervals and is computationally simpler than alternative approaches (e.g., confidence intervals based on the information matrix or Markov chain Monte Carlo methods). The computational advantages derive from the ability to fix parameters that are either interesting or troublesome, so that the remaining optimization problem is simplified. Prior information about nuisance parameters (e.g., duration of infection) can be incorporated into the estimation approach by constraining the particular parameter to a range of realistic values.

We used this approach to analyze two separate mathematical models of the Milwaukee outbreak. We constructed these models to investigate two hypotheses concerning the routes of disease transmission during the outbreak. The first model allows transmission to occur either through environmental transmission from contaminated water or through person–person transmission. This model represents the generally accepted theory that the outbreak was a direct result of a treatment failure combined with an influx of oocysts from the environment (16). In the context of Figure 1, this is realized by setting ϕ to zero, to prevent any person–environment–person transmission.

The second model we considered allows for delayed secondary transmission. This type of transmission can be expected to arise when a significant person–environment–person transmission route exists. In Milwaukee, Lake Michigan both supplies the drinking water and receives treated wastewaters. Conceivably, oocysts shed from a small number of infected people before the epidemic could have survived wastewater treatment, reentered the drinking water supply, and contributed to the outbreak.

The details of these methods and the following results are presented elsewhere (9,10). In this article, we briefly describe results in the context of health policy.

Results and conclusions. Figure 1 shows the maximum likelihood fit of the incidence data collected in a retrospective survey (16). Analysis of the first model focused on estimating the asymptomatic proportion, the rate of person–person transmission for *Cryptosporidium* infection, and the degree of water treatment failure. We found that the asymptomatic proportion in the models was critical in understanding other characteristics of the epidemic. For example, a high estimate of the asymptomatic proportion suggested that there was an exhaustion of the susceptible population: If 66% of infections were asymptomatic and 400,000 were symptomatic cases, that would suggest a total of 1.2 million infections (75% of the population). Epidemic theory suggests that the epidemic would die out because of the exhaustion of susceptible individuals. In particular, the degree of water treatment failure is closely tied to the asymptomatic proportion. Therefore, identifying the asymptomatic proportion as a key parameter was crucial to delineating a situation in which closing the treatment plant prevented a significant number of cases versus a situation in which plant closure prevented only a small fraction of cases.

Person–person transmission was not previously believed to have been a major contributor to the Milwaukee outbreak (17).

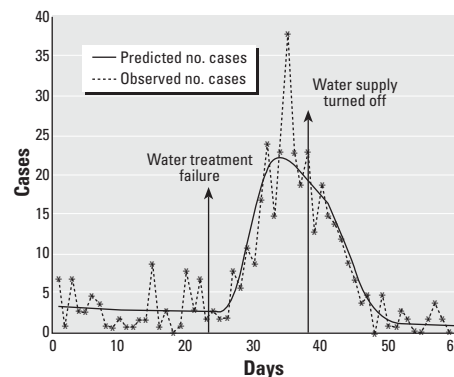


Figure 2. Maximum likelihood estimate of the outbreak.

Our analysis supported this view. In a scenario of strong person–person transmission, the number of observed cases will grow steadily over time until the susceptible population becomes exhausted. Instead, what was observed was a relatively constant rate of incidence followed by an explosive outbreak. The dynamics from a model incorporating strong person–person transmission cannot be made to exhibit this pattern of incidence.

Analysis of the second model suggests that delayed secondary transmission (person–environment–person) in combination with treatment failure could have explained the Milwaukee outbreak. Profile likelihood estimation of water treatment failure, length of the delay period, and asymptomatic rate has fielded particular values of these parameters that could have given rise to the outbreak. Although this depiction is plausible, within the context of the model and the observed data from the outbreak, more work needs to be done to understand whether these parameter sets are themselves realistic regarding other information. For example, what is the minimum time required for oocysts to move from the wastewater back to the drinking water? Or what fraction of excreted oocysts could be expected to make the journey and remain infectious? These issues can be addressed using existing knowledge about the hydrology of the water and sewer system in Milwaukee and *Cryptosporidium* epidemiology.

Application to the Endemic Condition

Although most reported cases of diseases attributed to waterborne pathogens are generally from outbreaks (because of the structure of existing surveillance systems), we have evidence that endemic transmission may account for an even larger portion of the disease burden. Estimates of the incidence of diarrheal disease come from prospective studies such as the one conducted in Tecumseh, Michigan, between 1965 and 1971 (18). More recently, Payment et al. (19) conducted a tap water intervention study that estimated a comparable incidence value. In addition, this study attributed an estimated 35% of the illnesses to the drinking water. In an analogous intervention trial, Kay et al. (20) found a significant risk associated with swimming in the ocean in four beaches in the United Kingdom. Based on these data the U.S. Environmental Protection Agency (EPA) has focused its regulatory activities on reducing risks associated with the endemic condition. As opposed to the situation with outbreaks, we have no available incidence data from which to assess risk directly. We therefore use models to provide indirect estimates of risk. The standardized modeling tool that

the U.S. EPA uses is risk assessment. The following section discusses the risk assessment methodology in the context of disease transmission models.

Risk assessment. Attempts to provide a quantitative assessment of human health risks associated with the ingestion of waterborne pathogens have generally focused on static models that calculate the probability of individual infection or disease as a result of a single exposure event (21–24). This framework is based on a model for the assessment of risk associated with chemical exposure (25) and, as such, does not address a number of properties that are unique to infectious disease transmission, including secondary (person–person or person–environment–person) disease transmission and immunity.

The limitations of treating infectious disease transmission as a static disease process, with no interaction between those infected or diseased and those at risk, have been illustrated in studies of *Giardia* (8), dengue (2), and sexually transmitted diseases (26). The U.S. EPA realized that infectious disease processes had features significantly different from disease processes caused by chemical exposure and began preliminary work by sponsoring a workshop in 1996, from which a working group developed a microbial risk framework (27). In contrast to the previously used chemical-based framework, the U.S. EPA microbial risk framework does allow incorporation of disease-specific epidemiologic data, such as incubation period, immune status, duration of disease, rate of symptomatic development, and exposure data such as those processes affecting the pathogen concentration. However, the U.S. EPA framework falls short of explicitly incorporating features of the disease transmission process that highlight the distinction between infectious disease risk and a more conventional static risk process (28).

Models using the chemical risk paradigm are static and assess risks at the individual level—the probability that a person exposed to a given concentration of pathogens will have an adverse health effect, regardless of adverse health effects on other individuals. Although this underlying assumption of independence is valid for disease associated with chemical exposure, it is not appropriate for most infectious disease processes. The risk of a person becoming infected depends not only on direct exposure to environmental pathogens via contaminated environmental media but also on exposure to other currently infected individuals through interactions within the population. One implication of this secondary infection process is that, by definition, risk manifests at a population level rather than at an individual level. Another implication is that risk calculations

are dynamic in nature; that is, the overall risk calculation is based not only on risk of current exposures (environmental or social) but also on risk of exposures from all subsequent secondary infections.

The following section describes a quantitative method of providing information for the risk assessment process in an endemic setting. Unlike for outbreak conditions, few data are available for analysis for endemic conditions, so traditional statistical techniques are not useful. We present alternative approaches that can provide information valuable for the decision making process. As we show below, an outcome of this study was the finding that, even in the presence of uncertainty and variability, a significant amount of information is obtainable in both the data collected from experiments and the mechanistic knowledge of the environmental system. Moreover, the results of these simulations inform us on the level of uncertainty in risk, as well as identifying parameters that drive uncertainty and require better definition.

Case study: exposure to *Giardia* from swimming. To illustrate the use of transmission models in risk assessment, Eisenberg et al. (8) developed an exposure scenario in which swimmers were exposed to *Giardia* from a recreational swimming impoundment filled with water reclaimed from community sewage. Because they had no incidence data from which to evaluate the model, the focus of that study was to compare two conditions: one in which the swimming impoundment water was not the exposure vehicle and one in which water was an exposure vehicle.

Analytical approach: sensitivity of parameters to outcome. In this simulation study, the model is used to assess the risk for exposure scenarios in which no incidence data are available. Because a traditional likelihood approach requires incidence data as well as a measure of uncertainty of those data, we needed an alternative approach. We based this alternative approach on the idea that rather than fitting the model to the data output as in the preceding case study, the output can be characterized as either a background endemic condition or an outbreak condition. In this manner, the output is classified into one of two categories. To obtain the output to be classified, probability distributions are assigned to each model parameter and multiple simulations are conducted. Specifically, for each simulation, a set of parameter values is obtained by randomly sampling the parameter distributions. The distributions are uniform unless the parameter range spans more than two orders of magnitude, in which case a log-uniform distribution more efficiently explores the full range of values. Assigning a bounded uniform

or log-uniform distribution to each parameter allowed us to include data from various literature sources without bias toward one value or another. We then applied a binary classification algorithm to each simulation output, in which the output either passes or fails a set of criteria. This binary classification is essentially a goodness-of-fit criterion based on whether or not the output is representative of the data. We then analyzed the multivariate parameter distribution associated with a background/above background classification using the classification and regression tree (CART) algorithm, which builds a tree based on minimizing the classification error (29). The details of this approach are given in a previous publication (8).

The model used in this study was a variant of Figure 1 in which contamination of the swimming impoundment could occur either from direct shedding of pathogens by infected swimmers, designated λ , or from community sewage that is treated before being recycled, designated T . We compared two transmission scenarios to analyze the relative risk of contracting giardiasis while swimming. The first scenario described a situation in which reclaimed water was not the exposure vehicle. We used the results to establish a baseline prevalence with which to compare the effects of the next scenario, ingesting water while swimming in an impoundment supplied with water reclaimed from the wastewater of the community. The output variable used in the analysis was the average daily prevalence of disease over the 1-year simulation period. The outcome was binary, either background prevalence levels consistent with nonoutbreak conditions and suggesting insignificant risk levels, or above background levels suggesting that a significant risk is associated with being exposed to the swimming impoundment.

Results and conclusions. Analysis of the model suggested that the output classification was most sensitive to the value of the shedding parameter, λ , the rate at which infectious swimmers shed pathogens into the swimming impoundment. We identified two regions: region 1, defined by $\lambda \leq 2 \times 10^4$ pathogens shed/swimming event, and region 2, defined by $\lambda > 2 \times 10^4$. Figure 3 illustrates these two regions. We initially classified 19% of the 3,022 simulations as an outbreak (see first node of tree). Of those 3,022 simulations, CART identified 2,266 simulations in which $\lambda \leq 2 \times 10^4$ and classified only 9% of these as an outbreak. We labeled these simulations region 1. We labeled the remaining 756 simulations region 2, in which $\lambda > 2 \times 10^4$, of which 48% were classified as an outbreak.

Within region 1, whether a simulation was classified as a background, low-risk condition was most sensitive to the value of

the water treatment parameter, T . The subregion within region 1 composed of those simulations in which $T > 2 \times 10^3$ were classified as a background region. Within region 2, on the other hand, whether a simulation was classified as a background, low-risk condition was most sensitive to the duration of a swimming event, ϵ_F , and the frequency of swimming events, ϵ_T . These two exposure parameters are incorporated into the transmission parameter, β_2 , shown in Figure 1. Figure 3 illustrates the two subregions within region 2 categorized as low exposure: $\epsilon_T \leq 1.6$ and $\epsilon_F \leq 23.3$, and $\epsilon_T > 1.6$ and $\epsilon_F \leq 9.6$.

The magnitude of the shedding parameter, λ , determined the optimal control strategy. If the magnitude was low and within region 1, then centralized control realized through water treatment (option 1) would be the optimal control option. However, if the magnitude was high and within region 2, then localized control realized through limiting exposure (option 2) would be the optimal control option. If a risk manager, therefore, must decide between option 1, ensuring that centralized treatment is above 2.6 log removal ($T > 2 \times 10^3$), and option 2, limiting swimming in the impoundment, our risk assessment suggests that λ is the crucial parameter to refine. Furthermore, if λ is already estimated, Figure 3 can help decide the best control option and help estimate the degree of confidence that should be placed on that decision. For example, if λ is estimated to be at point *A* on Figure 3, the model indicates that shedding would have to be $(2 \times 10^4 - A)/A$ greater before option 1

would cease to be the optimal control. Likewise, if λ is estimated to be at point *B*, then shedding would have to be $(B - 2 \times 10^4)/B$ less before option 2 would cease to be the optimal control.

Discussion

In this article we have demonstrated that using disease transmission models to identify data gaps and to aid in decision making provides a crucial link between science and policy while simultaneously coping in a responsible fashion with existing data gaps. In the context of issues related to infectious disease, the appropriate model describes the disease process at a system level; that is, it describes the transmission pathways. A system-level model has the potential to provide insight into which factors play a role in the transmission of pathogens to humans and therefore influence risk estimates. In the initial model development phase, features of a conceptual model are rigorously defined by a set of equations. At this level, data gaps are identified as those data required to define key factors of the disease process that are not available in the literature. In the analysis phase of the modeling process, sensitivity studies can provide information on the significance of these data gaps regarding a particular question of interest. In this analysis phase, the identification of data gaps can be translated into research needs. Although model analysis can identify research activities that would improve uncertain risk estimates, decisions often cannot wait for this research to be completed. Models also can play a role

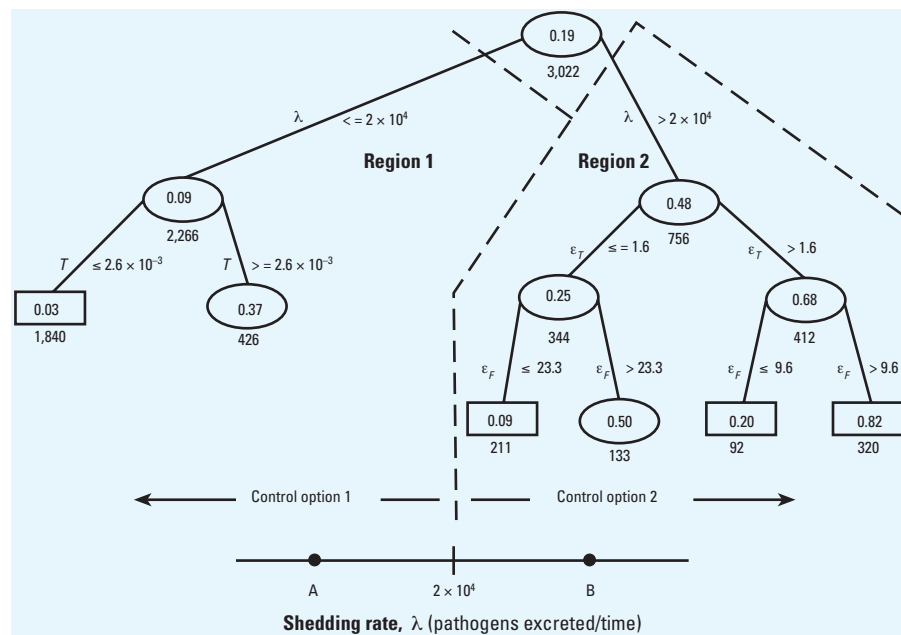


Figure 3. Classification diagram. The numbers in the ovals are the percentages of simulations that are classified as an outbreak condition, and below each oval is the number of simulations used to calculate this value. Parameters used in the classification are the shedding rate, λ ; the water treatment efficiency, T ; the time spent swimming in each event, ϵ_T , and the number of swimming events per year, ϵ_F .

in this decision-making process by defining the level of certainty associated with a given decision.

These two features of the modeling process, identifying data gaps and aiding in decision making, are illustrated by the two case studies described in this article. The *Giardia* risk assessment case study shows a clear gap in shedding rate data—the rate at which infected swimmers contaminate a swimming pool. The analysis phase was further able to quantify the sensitivity of the rate of shedding to the risk estimate. The sensitivity analysis not only identified this parameter as central to decreasing the uncertainty of the risk estimate but also defined the necessary degree of resolution. In the *Giardia* example, we were interested in the sensitivity of parameters to control decisions: option 1, to improve the level of water treatment, and control option 2, limiting exposure. The analysis showed that, in general, if the decision is to adopt control option 1, the model provides the degree to which a parameter value estimate, such as the rate of shedding, would have to vary for control option 1 to cease to be the optimal option. This analysis also places a limit on the degree of resolution needed; we needed only enough data on shedding to be confident that the rate of shedding estimate is within region 1 rather than region 2.

In the analysis of the *Cryptosporidium* outbreak, we found that when we constrained the proportion asymptomatic to the lower part of its distribution, [0, 0.6], its parameter value was insensitive regarding the incidence time series estimate of the outbreak. However, when constrained to the upper part, [0.6, 1.0], its parameter value was quite sensitive. For example, an asymptomatic rate of 0.7 would suggest that we had exhausted the pool of susceptible individuals and that the outbreak would have ended independent of whether or not the drinking water plant was closed on 9 April. Recent serologic data from the Milwaukee outbreak suggest that the asymptomatic proportion of individuals was between 0.5 and 0.7 (30). As with the *Giardia* risk assessment example, the data gap was evident at the model development phase; however, not until we concluded the analysis phase was it evident how and why the asymptomatic rate was a sensitive parameter requiring better identification.

Risk managers accept the responsibility of using risk analyses to choose the best course of action to reduce risk of disease. The best course of action is determined after considering a series of feasible intervention options. In a decision process, risk managers identify specific objectives that may be attained through implementing different

interventions. When faced with choices among different interventions targeting infectious diseases, risk managers can use disease transmission models to project the reduction in infectious diseases attained by each intervention. When making a decision, risk managers would consider the reduction in the predicted number of infectious disease cases associated with each intervention along with other factors such as economic and social impacts.

The systemic perspective of disease transmission models provides an opportunity to quantify the impacts of interventions at different points in the pathogen transmission cycle. For example, interventions in the treatment and delivery of drinking water can occur at 4 levels: *a*) The risk manager could improve the watershed quality of the source water through reduction of combined sewer overflows or improvements in private septic systems; *b*) drinking water treatment systems can be improved by increasing the number of barriers in the treatment process or by increasing the efficacy of an individual barrier; *c*) delivery systems can be improved through replacement of older pipes to reduce the risk of incursions or the addition of booster systems to maintain a high residual; *d*) finally, individuals may purchase a filter or a disinfection device for water in their homes.

This last intervention affects only the residents of the home. Watershed and treatment system improvements potentially reduce the entire population's exposure to pathogens from drinking water. Interventions at the other two levels reduce drinking water pathogen exposures to a fraction of the population, but this fraction may be more susceptible to the pathogen. For example, the interventions may target a retirement community. Each of these interventions can be accounted for in the disease transmission model, in the context and presence of different modes of disease transmission, person–person, person–environment–person, and person–environment.

Identifying data gaps and obtaining information on parameter sensitivity that would help in the decision making require a modeling approach to risk assessment with two essential features. The first important feature is the use of a mechanistic model in which the parameters all have biologic meaning. The second essential feature is the use of analytic tools that can assess uncertainty and sensitivity associated with complex models. In statistics, models are used to analyze data and therefore the data drive the structure of these models. A goal of statistical analysis is to assess how well the model can predict the data. Statistical techniques provide approaches for making inferences about the parameter values and for assessing statistical

significance. The parameters of these statistical models have no biologic meaning. In contrast, the model structure introduced in this article describes the transmission process of an infectious disease, so the parameters of the model have biologic meaning. Rather than being driven by the data, the model structure is driven by the process. In this way, the model structure is a summary of the relevant information and is independent of the data. In addition, these model parameters also can be constrained from independent information. Therefore, incidence data can be interpreted in the context of what is already known about the disease process. Although this is clearly a powerful approach, the downside of this approach is that these models are generally more complex than a traditional statistical model and therefore are more difficult to analyze.

Besides the fact that disease transmission models are complex, another technical challenge is the variable availability and quality of the data used in the analysis. The two case studies presented here illustrate the range in data quality. The *Cryptosporidium* outbreak study had data that represented accurately the incidence during the outbreak period, so a more traditional likelihood approach was promising. Given the complexity of our model, however, we considered a profile likelihood or Bayesian approach more tractable because they allow analysis of models with large numbers of parameters. Because we were interested in obtaining confidence intervals, we chose to use the profile likelihood approach. An example of using the Bayesian approach in public health decision making is given elsewhere (31).

For the *Giardia* study, we had no available data from which to analyze the model. For this analysis, an interest in those parameter combinations that produce a high-risk condition suggested a binary classification approach. A likelihood approach simplifies to a binary classification when the uncertainty of the data is uniform. In general, when one is not confident enough in the data to support a point-by-point goodness of fit, and information exists on general features of the data (such as timing and number of modes), a criterion-based approach can be developed to implement the binary classification.

Our perspective is that the power of this modeling approach lies in its ability to provide sensitivity information for the decision-making process. Decision makers need to know how sensitive a given decision is to the uncertainties associated with the disease process. The models presented here help quantify this sensitivity by allowing us to estimate the level of confidence that can be attributed to a given decision. Additionally, these analyses provide information on the

types of research needed to increase that confidence. In this way policy, can be informed by the current state of scientific knowledge, based on both what is known and what is unknown.

REFERENCES AND NOTES

- Murray CJL, Lopez AD. The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020. Cambridge, MA:Harvard School of Public Health for the World Health Organization and the World Bank, 1996.
- Koopman JS, Longini IM. The ecological effects of individual exposures and nonlinear disease dynamics in populations. *Am J Public Health* 84:836–842 (1994).
- Ross R. Report on the Prevention of Malaria in Mauritius. London:Waterlow, 1908.
- Hamer WH. Epidemic disease in England. *Lancet* 733–739 (1906).
- Anderson RM, May R. Infectious Diseases of Humans: Dynamics and Control. New York:Oxford University Press, 1991.
- Fine PEF. Herd immunity: history, theory, practice. *Epidemiol Rev* 15:265–302 (1993).
- Ross R. Some *a priori* pathometric equations. *Br Med J* 2818:546–547 (1915).
- Eisenberg JN, Seto EYW, Olivieri AW, Spear RC. Quantifying water pathogen risk in an epidemiological framework. *Risk Anal* 16:549–563 (1996).
- Brookhart MA, Hubbard AH, Vanderlaan MJ, Colford JM, Eisenberg JNS. Statistical estimation of parameters in a disease transmission model: analysis of a *Cryptosporidium* outbreak. *Stat Med* (in press).
- Eisenberg JNS, Seto EYW, Colford J, Olivieri AW, Spear RC. An analysis of the Milwaukee *Cryptosporidium* outbreak based on a dynamic model of disease transmission. *Epidemiology* 9:255–263 (1998).
- Dietz K. The estimation of the basic reproduction number for infectious diseases. *Stat Methods Med Res* 2:23–41 (1993).
- Levy DA, Bens MS, Craun GF, Calderon RL, Herwaldt BL. Surveillance for waterborne-disease outbreaks—United States, 1995–1996. *Morb Mortal Wkly Rep* 47:1–34 (1998).
- Barwick RS, Levy DA, Craun GF, Beach MJ, Calderon RL. Surveillance for waterborne-disease outbreaks—United States, 1997–1998. *Morb Mortal Wkly Rep* 49:1–21 (2000).
- Craun GF. Waterborne Diseases in the United States. Boca Raton, FL:CRC Press, 1986.
- Craun GF. Waterborne disease outbreaks in the United States of America: causes and prevention. *World Health Stat Q* 45:192–199 (1992).
- MacKenzie WR, Hoxie NJ, Proctor ME, Gradus MS, Blair KA, Peterson DE, Kazmierczak JJ, Addis DG, Fox KR, Rose JB, et al. A massive outbreak in Milwaukee of *Cryptosporidium* infection transmitted through the public water supply. *N Engl J Med* 331:161–167 (1994).
- MacKenzie WR, Schell WL, Blair KA, Addiss DG, Peterson DE, Hoxie NJ, Kazmierczak JJ, Davis JP. Massive outbreak of waterborne *Cryptosporidium* infection in Milwaukee, Wisconsin: recurrence of illness and risk of secondary transmission. *Clin Infect Dis* 21:57–62 (1995).
- Monto AS, Koopman JS. The Tecumseh Study. XI. Occurrence of acute enteric illness in the community. *Am J Epidemiol* 112:323–333 (1980).
- Payment P, Richardson L, Siemiatycki J, Dewar R, Edwards M, Franco E. A randomized trial to evaluate the risk of gastrointestinal disease due to consumption of drinking water meeting current microbiological standards. *Am J Public Health* 81:703–708 (1991).
- Kay D, Fleisher JM, Salmon RL, Jones F, Wyer MD, Godfree AF, Zelenauch-Jacquette Z, Shore R. Predicting likelihood of gastroenteritis from sea bathing: results from randomised exposure. *Lancet* 344:905–909 (1994).
- Haas CN. Estimation of risk due to low doses of microorganisms: a comparison of alternative methodologies. *Am J Epidemiol* 55:573–582 (1983).
- Anderson MA, Stewart MH, Yates MV, Gerba CP. Modeling the impact of body contact recreation on pathogen concentrations in a source water drinking water reservoir. *Water Res* 32:3293–3306 (1998).
- Fuhs GW. A probabilistic model of bathing beach safety. *Sci Tot Environ* 4:165–175 (1975).
- Regli S, Rose JB, Haas CN, Gerba CP. Modeling the risk from *Giardia* and viruses in drinking water. *J Am Water Works Assoc* 83:76–84 (1991).
- NRC. Risk Assessment in the Federal Government: Managing the Process. Washington, DC:National Research Council, National Academy Press, 1983.
- Koopman JS, Longini IM, Jacquez JA, Simon CP, Ostrow DG, Martin WR, Woodcock DM. Assessing risk factors for transmission of infection. *Am J Epidemiol* 133:1199–1209 (1991).
- ILSI Risk Science Institute Pathogen Risk Assessment Working Group. A conceptual framework to assess the risks of human disease following exposure to pathogens. *Risk Anal* 16:841–848 (1996).
- Soller JA, Eisenberg JN, Olivieri AW. Evaluation of Pathogen Risk Assessment Framework. Washington, DC: International Life Sciences Institute, Risk Science Institute, 1999.
- Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees. Pacific Grove, CA:Wadsworth, 1984.
- McDonald AC, MacKenzie WR, Addiss DG, Gradus MS, Linke G, Zembrowski E, Hurd MR, Arrowood MJ, Lammie PJ, Priest JW. *Cryptosporidium parvum*-specific antibody responses among children residing in Milwaukee during the 1993 waterborne outbreak. *J Infect Dis* 183:1373–1379 (2001).
- Spear RC, Long S, Hubbard AH. Disease transmission models for public health decision making: Designing intervention strategies for *Schistosomiasis japonicum*. *Environ Health Perspect* (in press).