The Role of Disease Transmission and Conferred Immunity in Outbreaks: Analysis of the 1993 *Cryptosporidium* Outbreak in Milwaukee, Wisconsin

Joseph N. S. Eisenberg, Xiudong Lei, Alan H. Hubbard, M. Alan Brookhart, and John M. Colford, Jr.

From the School of Public Health, University of California, Berkeley, Berkeley, CA.

Received for publication May 2, 2004; accepted for publication June 29, 2004.

The authors examined two competing hypotheses regarding the cause of the 1993 *Cryptosporidium* outbreak in Milwaukee, Wisconsin. The first was that oocyst contamination of the drinking-water influent, coupled with a treatment plant failure, resulted in a point-source outbreak. The second was that the outbreak was the result of transmission processes that amplified the oocyst concentration in the drinking-water effluent. Analysis of the model suggested that 1) transmission directly from person to person contributed 10% (95% confidence interval: 6%, 21%) of the total cases; 2) closing the drinking-water plant prevented 19% (95% confidence interval: 17%, 21%) of the additional cases of disease that occurred compared with the scenario in which the plant had not been closed, a result primarily driven by conferred immunity that resulted in depletion of the susceptible population; and 3) the outbreak was caused by a transmission cycle due to infectious persons shedding pathogens into the sewage, environmental transport of these pathogens via Lake Michigan to the drinking-water plant, and infection of susceptible persons via exposure to drinking water. The incidence data were consistent with this hypothesis. Further simulations suggested that increasing the distance between the wastewater effluent and the drinking-water influent may have prevented the outbreak.

*Cryptosporidium*; disease outbreaks; disease transmission; drinking; models, theoretical; water

Abbreviations: CI, confidence interval; MH, Metropolis-Hastings; MLE, maximum likelihood estimate.

Waterborne disease outbreaks are generally thought to occur in a similar fashion. That is, the outbreak is caused by direct exposure to a contaminated water source and is terminated when the contamination is removed. This event is often called a point-source outbreak (1–3). Another potential cause of an outbreak is a setting in which an infection is propagated through secondary transmission; that is, a given infectious person on average transmits the pathogen to more than one susceptible person. This transmission of disease can be through a direct person-person pathway or an indirect person-environment-person pathway. Outbreaks propagated in this way have been studied for a wide variety of pathogens by using disease transmission models (4–6). Because waterborne pathogens can exploit multiple transmission pathways (e.g., person-person and environment-person-environment), outbreaks due to waterborne pathogens may be the result of a combination of point-source and propagated processes. Eisenberg et al. (7) examined the environment-person pathway as the primary causal pathway of the outbreak. Brookhart et al. (8) additionally examined person-person transmission.

In this paper, we examine the contributions of three transmission pathways—environment-person, person-person, and person-environment-person—in producing cases associated with the outbreak. We used a disease transmission model to examine the role of these transmission processes as well as conferred immunity in the 1993 *Cryptosporidium* outbreak in Milwaukee, Wisconsin. Examination of these transmission pathways requires the ability to estimate parameters of interest. A secondary goal, therefore, was to quantify the benefits of improved data collection on parameter estimation. Two types of data are generally collected when waterborne pathogens are studied: water quality data and incidence data. In endemic conditions, water quality data are often collected; in outbreak settings, incidence data are...
often collected. Seldom are both types of data collected at the same time.

In the spring of 1993, a massive outbreak of Cryptosporidium occurred in Milwaukee, Wisconsin, in which an estimated 400,000 people became ill (9). This incidence estimate, developed from a retrospective survey, has been corroborated in a more recent serologic analysis of serum collected from children during the outbreak period (10). The cause of the outbreak was traced to the water supply, but the origin of the oocysts remains unknown. Several theories have been proposed to explain the origin of these oocysts. A widely accepted hypothesis was based on the concurrence of two events. First, increased water flows into Lake Michigan because of a late spring thaw, coupled with greater-than-normal winds, resulted in increased flows into the drinking-water inlet. Second, a treatment plant failure on March 23 resulted in diminished efficiency (11).

However, recent evidence has cast doubt on this hypothesis. In an analysis of isolates from cryptosporidiosis patients, a recent study (12) found that human rather than bovine strains were the cause of the outbreak, suggesting that the oocyst source was of human origin. One potential oocyst source was the city’s wastewater effluent, which was in close proximity to the drinking-water influent. We examined these two competing hypotheses in the context of the measured incidence data and what is known about transmission of Cryptosporidium.

**MATERIALS AND METHODS**

**Model description**

A disease transmission model was used to generate the expected incidence data for the likelihood analysis (figure 1). The model used in this study tracks the number of persons in different disease states. Specifically, our model was constructed with five disease states, defined as follows:

1. \( S \): those persons who are susceptible to disease,
2. \( E \): those persons who have been exposed to pathogens but are asymptomatic and noninfectious,
3. \( I_s \): those persons who are infectious and asymptomatic,
4. \( I_c \): those persons who are infectious and symptomatic, and
5. \( P \): those persons who are noninfectious and are protected from becoming reinfected with Cryptosporidium.

In addition, one state, \( W \), accounts for the concentration of oocysts in the water. The rates of movement of persons from one disease state to another are parameters represented by Greek letters in figure 1. Four of these parameters represent processes that occur postinfection: 1) \( \pi \), the rate of movement from the exposed state to an infectious state; 2) \( \delta \), the rate of movement from the symptomatic infectious state to the protected state; 3) \( \rho \), the proportion of symptomatic infections; and 4) \( \phi \), the rate that infectious persons shed pathogens into the environment. The parameter \( \pi \) represents the number of stages that one must pass through to become infectious, assuming that the duration of latency is distributed as a gamma random variable with rate parameter \( \pi \) and shape parameter \( \pi_n \) (8).

Two transmission-rate parameters, \( \beta_p \) and \( \beta_s \), account for infection events and represent two distinct transmission routes (refer to the dotted lines in figure 1). The subscript \( p \) represents a secondary infection event in which pathogens move directly from infectious person to susceptible person (labeled person-person in this figure), whereas the subscript \( p \) represents a primary infection event in which pathogens move from infectious person to the environment and subsequently from the environment to a susceptible person (labeled person-environment-person). Infectious persons can also shed pathogens into the sewage system that, at rate \( \phi \), can potentially make their way into the wastewater and be transported to the drinking-water system. During this transport, oocysts are depleted through dilution, natural die-off, and water treatment. This depletion is represented by the parameter \( \mu \), whereas transport time is represented by the parameter \( d \).
The outbreak was modeled by varying the transmission-rate parameter, \( \beta_p \), using the following equation:

\[
\beta_p(t) = I(t < 23) \beta_{p0} + I(23 < t < 38) \beta_{p1} + I(t > 38) \beta_{p2},
\]

where \( I \) is the indicator function defined as 1 if the conditions within the parentheses are met and 0 otherwise. In this realization, treatment failure occurred on day 23 and the treatment plant closed on day 38. The parameters \( \beta_{p0}, \beta_{p1}, \) and \( \beta_{p2} \) represent the transmission potential due to exposure from drinking water before the treatment failure, after the treatment failure, and after plant closure, respectively (8).

The model assumes that during the 1–2-month epidemic period, infected persons experienced complete conferred immunity; that is, the rate of movement from the removed state to the susceptible state, \( \gamma \), was equal to zero. This assumption is consistent with a laboratory controlled dosing trial, where persons previously exposed to *Cryptosporidium* were rechallenged 1 year later (13). The extrapolated infection rate was zero for exposure levels of 100 oocysts, approximately the assumed levels during the outbreak.

**Analysis**

The model described above was used for two distinct purposes. First, we examined the hypothesis that the outbreak was due to external contamination of the drinking water, assuming that the person-environment-person transmission pathway was negligible. We then used the incidence data to assess the fraction of outbreak cases attributable to person-person transmission and the preventable fraction of outbreak cases associated with closing the treatment plant. Second, we examined the hypothesis that person-environment-person transmission significantly contributed to the outbreak cases. In this application, we used parameter values obtained from a profile likelihood estimation procedure to predict the attack rate under the hypothetical conditions that the drinking-water outlet was moved progressively further from the wastewater outlet, thereby increasing the transport time of the pathogens in the water.

**Likelihood function.** March 1–April 29 incidence data for watery diarrhea were obtained from a random digit dialing telephone survey (\( N = 1,663 \)) of the greater Milwaukee area conducted at the conclusion of the outbreak (9). Baseline incidence of watery diarrhea was estimated in a subsequent survey 2 months after the outbreak ended. An estimate of the incidence of cryptosporidiosis was obtained by subtracting the baseline incidence rate (0.5 percent per year) from that obtained by the outbreak investigation.

Each day, persons in the cohort were labeled either a case or not. After becoming a case, a person was removed from the cohort of potential cases. For days 1–60, therefore, there were estimates of the number of new cases of diarrhea, represented by \( Y_1, Y_2, \ldots, Y_{60} \). The distributions of the number of cases for any given day were assumed to be binomial with parameter \( u_i \), the conditional probability that a randomly selected person becomes ill on day \( i \) given that the person was not observed to have become ill before day \( i \), and \( N_i \), the sample size for day \( i \). On the basis of these definitions, the likelihood of the full data can be written as the product of binomial likelihoods for each observed day (8):

\[
L \propto \prod_{i=1}^{60} u_i^{Y_i} (1 - u_i)^{N_i - Y_i}.
\]

**Model 1: role of person-person transmission.** In this first model, we modified figure 1 by setting the shedding rate, \( \phi \), to zero, eliminating the person-environment-person transmission pathway and thus yielding the model described by Brookhart et al. (8). The model was used to estimate the fraction of outbreak cases associated with person-person transmission and the preventable fraction of outbreak cases associated with the closing of the drinking-water treatment plant.

Table 1 lists the parameters for model 1. There are three classes of parameters: 1) the fixed parameters, where \( \pi_t \) was set to 7 and \( \gamma \) was set to zero (8); 2) the profiled parameters, where the duration of infectiousness, \( 1/\delta \), took on one of the values 1, 3, 7, or 10 days (14) and the fraction infected who were asymptomatic took on one of the values 0.1, 0.3, 0.65, or 0.7 (10); and 3) the estimated parameters, which were left unconstrained and were estimated in the analysis. For every combination of the profiled parameters, \( \rho \) and \( \delta \), a posterior distribution for the six estimated parameters was estimated by using a Monte Carlo Markov chain technique. To implement this technique, we used a novel two-step approach to the Metropolis-Hasting (MH) procedure (15) to handle the potential colinearities (nonidentifiabilities) of parameter estimates in these data (refer to the Appendix for details).

Once a posterior distribution was estimated, it was used to calculate two cumulative incidence distributions: 1) \( I_1 \), produced by a random draw from the posterior distribution; and 2) \( I_0 \), obtained by setting the secondary transmission parameter, \( b_2 \), to zero while keeping all other parameters the same as those used to obtain \( I_1 \). Finally, the attributable risk associated with secondary transmission was estimated by calculating the difference between the two incidence estimates, \( I_1 - I_0 \). The attributable risk distribution was estimated by repeating this procedure 12,000 times.

To estimate the preventable fraction associated with closing the drinking-water treatment plant, an analogous procedure was followed. The only difference was that, to estimate \( I_o \), the parameter \( \beta_2 \) was set equal to \( \beta_1 \).

The 95 percent confidence intervals were estimated by using the 2.5th and 97.5th percentiles of the empirical distribution of the 12,000 attributable risk estimates.

**Model 2: role of person-environment-person transmission.**

In this second model, we modified figure 1 by setting the secondary transmission rate, \( \beta_p \), to zero, eliminating the person-person transmission pathway. To examine the role of person-environment-person transmission in the outbreak, we used a profile likelihood approach (8). Table 2 summarizes the parameters used in model 2. For the fixed parameters, \( \pi_t = 7 \) and \( \gamma = 0 \) (as in model 1); \( \beta_0 = 0 \) (by definition); \( \rho = 0.65 \) (the maximum likelihood estimate (MLE) from model 1 (8)); \( b_1 = 10b_2 \), which modeled 1 log decrease in treatment efficiency due to the treatment plant failure (based on the turbidity data from the outbreak (9)); and \( d_2 = 25 \), the shape parameter for the gamma-distributed oocyst environmental
transport time. This shape parameter value provided a reasonable amount of variation for mean delay times of 1–40 days. The three profiled parameters were the transport delay, \(d\); the oocyst die-off rate in the water, \(\mu\); and the rate at which persons left the infectious state, \(\delta\). The delay ranged from 1 to 40 days. The die-off rates for Cryptosporidium in fresh water will vary with temperature and pH (16). To account for this variability, the die-off rate parameter, \(\mu\), took on a value of either 0.1 or 0.01 corresponding to a half-life of 10 or 100 days, respectively. Lastly, \(\delta\) took on the same values as in model 1.

The likelihood was calculated for all combinations of \(d\) and \(\delta\). The largest likelihood from these combinations was determined to be the MLE. The 95 percent confidence intervals were estimated by using a \(\chi^2\) threshold in which points above the threshold line represented parameter values consistent with the data (8).

To estimate the number of cases averted as the distance from the wastewater outlet and drinking-water inlet increased, we took the MLE as the “truth” and computed the incidence due to the outbreak as the delay, \(d\), increased. A preventable fraction for a given delay value, \(d\), was calculated by using the following equation: preventable fraction = 1 – \(AR_{\text{MLE}}/AR_d\), where \(AR_{\text{MLE}}\) is the attack rate for the MLE of the model and \(AR_d\) is the attack rate for a given value of \(d\), keeping all other parameter values at their MLE.

### TABLE 1. Parameters for model 1 used to assess the cause of the 1993 Cryptosporidium outbreak in Milwaukee, Wisconsin

<table>
<thead>
<tr>
<th>Parameter type</th>
<th>Description</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\gamma)</td>
<td>Rate at which persons leave the protected state and become susceptible (1/day)</td>
<td>0</td>
</tr>
<tr>
<td>(\pi_n)</td>
<td>Shape parameter for the gamma distribution describing the incubation period</td>
<td>7</td>
</tr>
<tr>
<td>Profiled</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\rho)</td>
<td>Fraction of those persons infected who remain asymptomatic</td>
<td>(0.1, 0.3, 0.65, 0.7)</td>
</tr>
<tr>
<td>(\delta)</td>
<td>Rate at which persons leave the infectious state (1/day)</td>
<td>(1, 1/3, 1/7, 1/10)</td>
</tr>
<tr>
<td>Estimated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\pi)</td>
<td>Rate at which persons leave the latent state (1/day)</td>
<td></td>
</tr>
<tr>
<td>(b_0)</td>
<td>Rate of primary transmission prior to treatment failure (1/day)</td>
<td></td>
</tr>
<tr>
<td>(b_1)</td>
<td>Rate of primary transmission during treatment failure (1/day)</td>
<td></td>
</tr>
<tr>
<td>(b_2)</td>
<td>Rate of primary transmission after treatment failure (1/day)</td>
<td></td>
</tr>
<tr>
<td>(\beta_s)</td>
<td>Rate of secondary transmission (((\text{probability of infection}) \times (\text{no. of contacts/day})))</td>
<td></td>
</tr>
<tr>
<td>(E_0)</td>
<td>Initial fraction of those persons latently infected</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 2. Parameters for model 2 used to assess the cause of the 1993 Cryptosporidium outbreak in Milwaukee, Wisconsin

<table>
<thead>
<tr>
<th>Parameter type</th>
<th>Description</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\gamma)</td>
<td>Rate at which persons leave the protected state and become susceptible (1/day)</td>
<td>0</td>
</tr>
<tr>
<td>(\beta_s)</td>
<td>Rate of secondary transmission (((\text{probability of infection}) \times (\text{no. of contacts/day})))</td>
<td>0</td>
</tr>
<tr>
<td>(\pi_n)</td>
<td>Shape parameter for the gamma distribution describing the incubation period</td>
<td>7</td>
</tr>
<tr>
<td>(\rho)</td>
<td>Fraction of those persons infected who remain asymptomatic</td>
<td>0.65</td>
</tr>
<tr>
<td>(b_1)</td>
<td>Rate of primary transmission during treatment failure (1/day)</td>
<td>10(b_2)</td>
</tr>
<tr>
<td>(\mu)</td>
<td>Die-off rate of oocysts in the water (1/day)</td>
<td>0.1, 0.01</td>
</tr>
<tr>
<td>(E_0)</td>
<td>Initial fraction of those persons latently infected</td>
<td>0.05</td>
</tr>
<tr>
<td>(d_n)</td>
<td>Shape parameter for the gamma distribution describing (d), the time from wastewater to drinking-water taps</td>
<td>25</td>
</tr>
<tr>
<td>Profiled</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(d)</td>
<td>Time for oocysts to transport from wastewater in the sewage to the drinking-water taps of homes (days)</td>
<td>(1, 40)</td>
</tr>
<tr>
<td>(\delta)</td>
<td>Rate at which persons leave the infectious state (1/day)</td>
<td>(1, 1/3, 1/7, 1/10)</td>
</tr>
<tr>
<td>Estimated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\pi)</td>
<td>Rate at which persons leave the latent state (1/day)</td>
<td></td>
</tr>
<tr>
<td>(b_0)</td>
<td>Rate of primary transmission prior to treatment failure (1/day)</td>
<td></td>
</tr>
<tr>
<td>(b_2)</td>
<td>Rate of primary transmission after treatment failure (1/day)</td>
<td></td>
</tr>
</tbody>
</table>
To estimate the increased power achieved if both the incidence and the concentration of pathogens were measured during the outbreak, we created data for the concentration of pathogens in the water by using the simulated values from the MLE and adding random noise.

RESULTS

Model 1: role of secondary transmission and conferred immunity

Figure 2A contains the posterior distribution for $\beta$, that is, the probability distribution of $\beta$, given the outbreak incidence data and the prior distribution of all model parameters. The median value of $\beta$ was 0.06 (95 percent confidence interval (CI): 0.02, 0.36). The estimate of the attributable fraction due to person-person transmission was 0.1 (95 percent CI: 0.06, 0.21) (figure 2B). An attributable fraction of 0.1 suggests that 10 percent of the outbreak cases were due to person-person transmission. The preventable fraction associated with closing the drinking-water treatment plant strongly depended on the estimate of the fraction of infections that were asymptomatic, $\rho$ (figure 3). For example, the median preventable fraction for $\rho = 0.1$ was 0.36 (95 percent CI: 0.35, 0.38), for $\rho = 0.3$ was 0.34 (95 percent CI: 0.33, 0.35), for $\rho = 0.65$ (the MLE) was 0.19 (95 percent CI: 0.17, 0.21), and for $\rho = 0.7$ was 0.12 (95 percent CI: 0.09, 0.14).

Model 2: role of environmental transmission

The ability to define the role of environmental transmission and to estimate the delay parameter strongly depended on $\mu$, the pathogen die-off rate in the water. For $\mu = 0.1 \, d^{-1}$ (half-life of 10 days), the MLE for the delay parameter was 11 days (95 percent CI: 8.3, 19) (figure 4A). For $\mu = 0.01 \, d^{-1}$ (half-life of 100 days), the MLE for the delay parameter was 10.5 days and the lower confidence bound was still 8.7. However, the upper confidence bound was not bounded for delay values as large as 40 days (figure 4B). When we incorporated (pseudo) information on $W$, the concentration of oocysts in the water, the estimation error of the delay parameter, $d$, for $\mu = 0.1$ decreased approximately threefold, from a confidence interval of 8.3, 18 to 9.6, 12.6 days (figure 5). For $\mu = 0.01$, adding (pseudo) water quality information reduced the upper bound estimate from $\infty$ (no information) to 13 days.

The prediction for the preventable fraction due to increasing time from contamination site to drinking-water exposure site is shown in figure 6. At the MLE of $d = 11$, the preventable fraction was defined as 0.0. As the transport time increased from 11 days to 40 days, the preventable fraction increased to 0.8.

DISCUSSION

Analysis of epidemics can provide valuable information on disease processes. Incidence data from outbreak investigations form the basis for this analysis and traditionally rely on a parametric statistical model that is determined by the data. For example, a normal distribution is chosen if the data are symmetric, a lognormal distribution if the data are skewed. The parameters of these statistical models generally do not have any biologic interpretation. In contrast, a disease transmission model is mechanistically driven; its structure is chosen on the basis of what is known about a specific pathogen. Inference regarding these model parameters can provide insight into the epidemic process. For example, a disease transmission model enables the identification of both the parametric values of important and measurable factors, such as the transport delay of oocysts in the environment, and the factors that strongly affect these estimates. Thus, a transmission model can provide explanatory information about the process. The analysis herein focused on the role of three specific attributes of this epidemic: person-person transmission, environmentally mediated transmission, and conferred immunity.

Person-person transmission has been documented for Cryptosporidium (17–23). For the Milwaukee outbreak, it was estimated that 5 percent of the outbreak cases were due to within-household transmission (24). Since that study considered household contacts only, this estimate provided a lower-bound estimate for person-person transmission. Consistent with these data, our analysis suggested that 6–21 percent of the cases of disease were due to person-person transmission. A biologic interpretation is that secondary spread could not have been the principal component of the outbreak because the rate of increase in the incidence in Milwaukee occurred too quickly. A propagated outbreak driven by person-person transmission increases geometrically; that is, an index case transmits to secondary cases, and the secondary cases in turn transmit to tertiary cases. This geometric process takes time to build up. Even a pathogen that is exclusively driven by secondary spread, such as measles, takes about 1 month to unfold from the index case to peak incidence. Given what is known about the natural history of Cryptosporidium, it would be impossible for an outbreak driven exclusively by secondary spread to generate the observed epidemic curve. Even though person-person was probably not the primary component of the outbreak, the analysis suggests that secondary spread could have accounted for 21 percent of the cases.

Likewise, conferred immunity has been clearly documented for those persons previously exposed to Cryptosporidium (13, 25, 26). Seroprevalence data derived from children in the Milwaukee area during the 1993 outbreak suggested that a significant number of children were infected both prior to (15 percent) and during (80 percent) the outbreak (10). Compared with the attack rate estimates of 19 percent in children (24), these data suggest that approximately 60 percent of the children were asymptomatic. This estimate is probably higher in adults, who tend to have higher asymptomatic rates. Consistent with these data was our MLE for the proportion who were asymptomatic, $\rho = 0.65$. Our analysis suggested that the fraction of cases prevented by closing the drinking-water treatment plant was highly sensitive to this estimate of $\rho$; that is, the higher the estimate of $\rho$, the greater the estimate for the number of susceptible persons exposed and infected during the outbreak. As the number of susceptible persons declines, so does the incidence.

Few empirical data examine the person-environment-person transmission pathway for waterborne pathogens. However, data do exist for certain pathogens on certain components of this pathway. For example, a number of studies have demonstrated a correlation between cases of poliovirus caused by specific serotypes and those same serotypes isolated in sewage (27–31). More recent studies have looked at other viruses such as enteroviruses (32) and hepatitis E (33). Data from these studies tend not to distinguish between the two possible causal pathways: humans contaminating the sewage system and causing increased levels of pathogens, or pathogens originating from the sewage causing human infection. A recent study by Yamamoto et al. (34) suggests that a 1996 Cryptosporidium drinking-water outbreak was potentially the result of a person-environment-person transmission pathway in which the drinking-water facility was located downstream from the wastewater treatment plant. As mentioned at the beginning of this paper, both the proximity of the wastewater outlet into Lake Michigan and the drinking-water inlet from Lake Michigan, along with

FIGURE 2. A: Posterior distribution of $\beta_s$, the secondary Cryptosporidium transmission-rate parameter for model 1; B: distribution of attributable risk associated with secondary transmission for model 1.
the data suggesting that the outbreak was caused by the human genotype, indicate that a person-environment-person pathway had the potential to cause the outbreak in Milwaukee.

Providing this explanatory insight into the disease process is a clear benefit of disease transmission models. One limitation associated with these models is that they inherently contain a large number of parameters that are often highly collinear. To address these complexities, we explored two analytical techniques: 1) the profile likelihood (an approach to deal with large numbers of parameters) and 2) the Bayesian estimation technique (an approach to deal with collinearity). Coupled with the issues of using models that contain relatively large numbers of parameters is the fact that data are often limited or sparse. One approach that we explored to address this last issue was simultaneous measurement of two state variables. In the context of a waterborne pathogen, it is conceivable that the concentration of oocysts in the source water can be measured, either by directly measuring the pathogen or by measuring a surrogate indicator, simultaneously with the incidence of infection or disease. In our analysis, we examined the additional statistical power that could be achieved if incidence and water quality were measured simultaneously. These additional data provided increased resolution of parameter estimates. This finding has strong implications for future research involving statistical analysis of nonidentified/highly parameterized disease transmission models. Our experience suggests that statistical analysis of these models will benefit more from improved data collection that can help measure previously unobserved state variables than from more sophisticated statistical approaches.

FIGURE 3. Distribution of the preventable fraction of Cryptosporidium associated with closing the drinking-water treatment plant for model 1 using four values for the proportion of asymptomatic persons: A: $\rho = 0.1$; B: $\rho = 0.3$; C: $\rho = 0.65$; and D: $\rho = 0.7$. 
Another potential limitation of this analysis is that our model does not account for heterogeneous factors in the population, such as age. Two reasons for using a model that was homogeneous with respect to age were that 1) expanding the model to include age-specific parameters increased the number of parameters, and 2) stratifying the data across age decreased the effective sample size. Under these conditions, the statistical analysis became intractable. Another reason for using a model homogeneous for age was that the incidence data were consistent with a homogeneous

FIGURE 4. Log-likelihood for model 2 as a function of the delay parameter, \( d \), for two values of the oocyst die-off rate: A: \( \mu = 0.1 \); B: \( \mu = 0.01 \). The points at which the graph crosses the horizontal line represent the lower and upper 95% confidence limits.
model; that is, for most age groups, the attack rate was at least 20 percent (except for those aged >70 years, for whom the attack rate was 14 percent), and no age group had an attack rate greater than 34 percent (9). Furthermore, we found no evidence that the changes in the instantaneous rate of disease onset differed by age group; a graph of the instantaneous rate of change versus time was parallel for the age groups 0–5, 6–65, and >65 years (unpublished age-stratified incidence data from the Milwaukee outbreak, J. P. Davis, Wisconsin Department of Health, 2004). However, many of

![Log-likelihood graph for model 2](image1)

**FIGURE 5.** Log-likelihood for model 2 as a function of the delay parameter, $d$, when both incidence and water quality data are available ($\mu$, the oocyst die-off rate = 0.1). The points at which the graph crosses the horizontal line represent the lower and upper 95% confidence limits.

![Preventable fraction graph](image2)

**FIGURE 6.** Preventable fraction for model 2 as the delay parameter, $d$, increases from 1 to 40 days. The maximum likelihood estimate, $d = 11$, is associated with a preventable fraction of 0.
the model parameters are known to vary by age; for example, compared with the remaining population, children and the elderly are more susceptible to severe disease outcomes, tend to have a longer duration of disease, and potentially have greater rates of secondary transmission.

To examine whether these age-dependent factors might have changed any of the conclusions of our study, we developed an age-stratified model consisting of three age groups: 0–5, 6–65, and >65 years. Simulations were conducted based on the assumption that 83 percent of the population was between 6 and 65 years of age and the remaining 17 percent were equally divided into the other two age groups. Parameterization of the age-dependent factors depended partially on existing knowledge of how the young and elderly differ from the remaining population and partially on the age-stratified incidence data from the outbreak. On the basis of these data, we assumed that factors such as susceptibility to infection and disease, and disease duration in the young and the elderly, were either decreased or increased by as much as 50–100 percent of the values assigned to the majority age group (6–65 years). Simulations with varying degrees of age dependency produced incidence curves nearly identical to those in the homogeneous model. Since our estimates were based on fitting the model to the total number of cases, we would expect that the age-averaged results would not be affected by age stratification and would therefore provide no new insights.

When it is assumed that outbreaks are of a point-source nature, the transmission processes for waterborne pathogens are generally ignored. These analyses suggest specific roles of transmission and conferred immunity in the 1993 Milwaukee Cryptosporidium outbreak, and they further demonstrate the importance of understanding the natural history of the Cryptosporidium disease process.

ACKNOWLEDGMENTS

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

REFERENCES


**APPENDIX**

**Two-step MH Procedure**

When all of the parameters were allowed to vary freely within their respective limits, the MLE could not be found (refer to Brookhart et al. (8)) because of the nonidentifiability or near nonidentifiability of the underlying model. Specifically, the combination of parameter values giving the best (or very near best) fit was not unique. However, when certain parameters were treated as fixed, the MLE was identifiable; thus, we profiled over these parameter sets to define the global MLE.

If the information matrix were invertible at the global MLE and we assumed asymptotic multivariate normality, then the posterior distribution of the parameter estimates could be directly estimated from the MLE and estimated variance-covariance matrix. Because of the colinearities, we could not analytically estimate the posterior distribution and thus had to rely on an MH algorithm to generate a random draw from this distribution. However, the colinearities that plague finding the MLE can also affect performance of the MH procedure, specifically when plausible sets of parameters occur in narrow regions of the parameter space. Thus, we used a two-step procedure to generate samples from the posterior parameter distribution. First, we created a dense grid of the parameter subspace defined by the parameters, over which we profiled. For each point in this multidimensional grid, we ran the MH algorithm over the remaining parameters to generate a posterior distribution, assuming that the parameters defining the grid were fixed. When this procedure was finished, the result was a sample of all parameters and the associated likelihood. Finally, we used a random MH algorithm that uses as trial parameter sets rows from this data set. After this second step was performed, the result was a sample of the multivariate posterior distribution of all of the parameters.