18 MICROBIAL RISK ASSESSMENT FOR DRINKING WATER

Stephen E. Chick[1]
Sada Soorapanth[2]
James S. Koopman[3,4]

[1]Technology Management Area, INSEAD, Fontainbleau, France

[2]Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI

[3]Department of Epidemiology, University of Michigan, Ann Arbor, MI

[4]Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI

Abstract: Infectious microbes can be transmitted through the drinking water supply. Recent research indicates that infection transmission dynamics influence the public health benefit of water treatment interventions, although some risk assessments currently in use do not fully account for those dynamics. This chapter models the public health benefit of two interventions: improvements to centralized water treatment facilities, and localized point-of-use treatments in the homes of particularly susceptible individuals. A sensitivity analysis indicates that the best option is not as obvious as that suggested by an analysis that ignores infection dynamics suggests. Deterministic and stochastic dynamic systems models prove to be useful tools for assessing the dynamics of risk exposure.

Key words: Microbial risk, Epidemic model, Water treatment, Stochastic infection model, Ornstein-Uhlenbeck process

18.1 INTRODUCTION

A cryptosporidiosis outbreak linked to *Cryptosporidium* oocysts in Milwaukee's drinking water caused over 400,000 cases of diarrhea and 1,000 hospitalizations in 1993. The outbreak played a role in the death of more than 50 individuals, primarily individuals with AIDS [1, 2]. The World Health Organization indicates that disease caused by these and other waterborne microbes is involved in the death of millions of people every year [3], and the illness of many more. One culprit is the lack of a safe water supply and basic sanitation [4]. Endemic infection is a significant concern, not just outbreaks.

This chapter reviews recent progress in merging infection transmission models with microbial risk assessments. The goal of that work is to better represent the dynamics of infection in such risk assessments. This chapter also presents sensitivity analyses that provide policy regions that indicate when it is better to use centralized water treatment alternatives versus local water treatment measures, as a function of infection transmission parameters. We also discuss how different model structures, including ordinary differential equations (ODEs), stochastic Markov chains of individual infection and recovery events, and Ornstein-Uhlenbeck (OU) diffusion approximations may be useful for policy region assessment and inference for parameters whose values are poorly understood.

Although the focus of this chapter is risk assessment for water treatment interventions and their public health consequences, the idea of modeling risk exposure as a dynamic function of a system's state is rather general. Other microbial applications include the protection of the food supply chain, and biological warfare preparedness. Specific issues that have attracted public interest recently include so-called "mad cow disease" (bovine spongiform encephalopathy) and the threat of anthrax and smallpox attacks. *E. coli* and Norwalk-like viruses can be found in both the water system and the food chain [5]. The importance of dynamics for risk exposure assessments is not exclusive to infectious diseases. For example, weather dynamics can influence risk exposure to radiation in the aftermath of nuclear accidents [6]. The need for dynamic systems models of risk exposure, then, has a much wider application than the scope presented here, and the tools available to approximate those exposures continue to be developed.

Drinking water can be protected from microbes with a series of barriers starting with source water protection, centralized municipal water treatment, filters or other local point-of-use treatments, and wastewater treatment. Centralized drinking water treatments improve water quality for the entire community. Options include filtration, chlorination, and ozone pretreatment.

Ozone pretreatment may reduce *Cryptosporidium* oocysts in water by 40-60%, but may be quite costly. A facility for a particular California reservoir is estimated to cost $154-190 million initially, and $3.8-5.2 million per year thereafter [7]. Local treatment can also be used for population subgroups that require particularly effective pathogen removal. Options include copper-silver ionization and chlorine dioxide generation in hospitals and nursing homes [8], and reverse osmosis filters in the homes of immunocompromised individuals. Such filters may costs hundreds of dollars per home, and require regular maintenance. These costs justify a formal assessment of the public health benefit of each treatment option.

A standard approach to risk assessment for chemicals and microbes is to identify hazards, quantify occurrence and exposure, assess the dose-response relationship, and identify human health consequences. Exposure is generally taken to be from drinking water in this context. The probability of infection is assessed with a dose-response curve, where dose is a function of microbes in consumed water. The health effects of any resulting disease are then quantified. But microbes present additional risk exposures that chemicals do not usually exhibit. Microbes can circulate through two secondary transmission routes: interpersonal human contact, and a water loop where infected individuals recontaminate water through recreational use or waste [9, 10].

Some analyses (e.g., [7]) account for secondary transmission in the water loop by using the prevalence of infection in the population to assess the amount of microbes shed into recreational water, then estimating increased contamination in drinking water. That approach is consistent with risk calculations used by the Environmental Protection Agency (EPA) [11]. Such an approach does not fully model the fact that effective water treatment changes the prevalence of infection, which is an input to the assumed risk exposure model. Although this indirect effect of treatment on risk exposure due to secondary transmission is not modeled by that approach, there is a recognized need to do so to inform water treatment policy [12].

Other analyses [13, 14] use dynamic systems models to represent the dynamics of risk exposure. The models are based on deterministic ordinary differential equations (ODEs). Such models find that the public health benefit of water treatment interventions depends strongly on how infection is circulated. For example, Milwaukee residents with AIDS suffered particularly extreme consequences from cryptosporidiosis during the 1993 epidemic [15]. Some have proposed that highly effective filters that eliminate *Cryptosporidium* oocysts would effectively protect individuals with AIDS from similar risks in the future. This would be the case if there were no additional exposure from secondary transmission to that subgroup

from human contact. However, it is likely that some secondary transmission occurred [1, 13]. Depending on the average number of secondary transmissions, and the relative probability of infection given exposure for those with AIDS, improving a standard municipal facility by adding ozone pretreatment may be more effective than filters [14]. If secondary transmission is sufficiently high, ozone can reduce secondary exposure in the AIDS subgroup by reducing cryptosporidiosis prevalence in the general population – and that reduction can outweigh the benefits of completely effective filters on the water taps of individuals with AIDS.

Section 18.2 extends previous work [14] by presenting a sensitivity analysis for those policy regions (ozone pretreatment versus local filters) with respect to several infection parameters, and by using a more refined model of the natural history of infection of cryptosporidiosis. The policy region is quite sensitive to the efficacy of ozone for inactivating oocysts, but is not very sensitive to the size of the sensitive population subgroup, as long as it is not too large, nor to the rate of exogenous introduction of oocysts into the water supply. Ozone becomes less effective, relative to filters in the susceptible subpopulation, as the water loop becomes relatively more important for secondary transmission than human contact.

Deterministic infection models ignore variability that arises in real infection transmission systems. Further, standard ODE parameter fitting tools make normal distribution assumptions that may not be satisfied in practice [16]. Stochastic infection models explicitly account for this variability, and may provide a mechanism to further incorporate infection dynamics into the parameter inference process. Parameter inference is important because the secondary transmission parameters for a number of microbes of interest to the EPA are poorly understood at present. Several researchers have examined mechanisms to infer parameters of various infection models given outbreak or intervention trial data [13, 17-19], or using endemic data [16]. Those works attempt to incorporate the dynamics of infection into the likelihood model using a variety of approximations (e.g., binomial distributions for discrete-time models, normal approximations for larger populations using moment methods).

Section 18.3 extends that work by suggesting that diffusion process approximations be used to model the stochastic infection dynamics. The idea is to apply stochastic process results [20-23] to approximate the underlying discrete-state Markov chain model of infection and microbe contamination with a continuous-state Ornstein-Uhlenbeck (OU) process. We present diffusion approximation formulas for the stationary mean and covariance of the underlying infection model. Section 18.4 describes potential areas for

further research for water treatment policy, risk analysis, and epidemic modeling research.

18.2 ODE MODELS TO EVALUATE POLICY REGIONS

Our goal is to develop a mathematical model that captures the dynamics of three modes of infection transmission: infection from microbes in the drinking water that come from exogenous sources, secondary infection from microbes in drinking water that result from contamination of source water from modeled individuals, and secondary transmission from human-to-human contact. The model must account for multiple subgroups with different infection susceptibility and outcome parameters, and further allow for the assessment of public health benefits of both local and municipal level interventions. We first describe an ODE infection model. Many parameters are not well understood for most microbes on the EPA's Candidate Contaminant List. We therefore present a sensitivity analysis that could be applied for those agents. The analysis here is consistent with current knowledge about cryptosporidiosis.

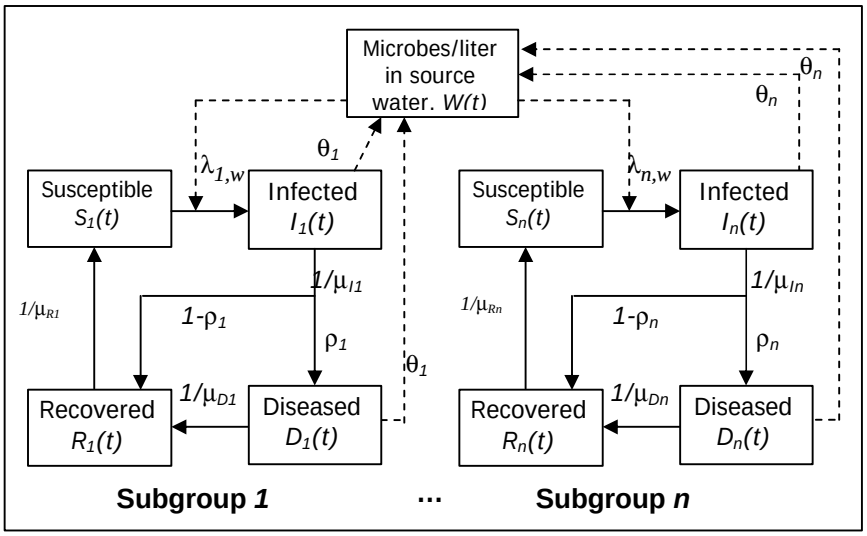
18.2.1 Deterministic infection transmission system model

Figure 18.1. illustrates that humans are assumed to change health status from susceptible (S), infected (I), diseased (D), and recovered (R) as a result of microbial infection. Microbes can be shed by infected and diseased individuals into the water supply, which in turn can reinfect susceptible individuals. We further assume that there are n different subgroups that interact according to a proportional mixing pattern [24]. Individuals in different subgroups may have different mixing and infection parameters. Here we are particularly interested in the case of two subgroups: immunocompetent and immunosuppressed individuals. A more detailed study might also model special characteristics of the young and the aged.

The N_i individuals in subgroup i are counted as to whether they are infected $I_i(t)$ (infectious, but asymptomatic), diseased $D_i(t)$ (infectious and symptomatic), recovered $R_i(t)$ (temporarily immune to reinfection), or susceptible $S_i(t)$. These values vary through time as the system evolves. For simplicity, the argument t is dropped below except when we wish to emphasize dependence of these values on time.

Microbe concentration in the water supply, $W(t)$, shown in the upper portion of the figure, is influenced by the rate γ of exogenous introduction of microbes, the rate α that microbes leave the system from water flow or

Figure 18.1 An SIDRS/W infection model with water loop and proportional mixing



inactivation, and the rate θ_i that infected individuals contaminate the water supply. This leads to the microbe concentration dynamic in equation 1.

$$\frac{dW}{dt} = \gamma - \alpha W + \sum_{i=1}^{n} \theta_i (I_i + D_i) . \qquad (1)$$

Each susceptible individual in subgroup *i* has the potential of becoming infected after being exposed. The rate of exposure for each susceptible individual depends on two main sources. Exposure from water consumption is determined by the number of microbes per unit volume in the source water, *W*, the fraction of microbes that remain after treatment τ_i, the volume of drinking water consumed per day ϕ_i, and the probability of infection per ingested microbe, r_i. Exposure from secondary transmission depends on the number of individuals in each subgroup, N_j, the number of contacts per day, c_j, and the probability β_i that a potentially infectious contact will infect an individual in subgroup *i*. This results in an overall exposure rate $\lambda_{i,W}$ for subgroup *i*, when the microbe concentration is *W*.

$$\lambda_{i,W} = r_i \tau_i \phi_i W S_i + \sum_{j=1}^{n} c_j (I_j + D_j) \frac{c_i N_i}{\sum_{k=1}^{n} c_k N_k} \frac{S_i}{N_i} \beta_i \qquad (2)$$

The first term models exposure from drinking water. The second term sums the exposures from each subgroup to susceptibles in *i*: there are $c_j(I_j+D_j)$ potentially infectious contacts, of which a fraction $c_iN_i / \Sigma c_kN_k$ are with members of group *i*. The probability that a member of subgroup *i* is susceptible is S_i/N_i, and the probability of infection given the contact is β_i.

After becoming infected, only a fraction ρ_i become diseased; the rest recover and become immune for some duration of time, μ_{Ri}. The mean duration of infection is μ_{Ii}, and the mean duration of disease is μ_{Di}. Since the dynamics of microbial infection are on a much faster time scale than the lifetimes of humans, we assume a closed population.

$$\begin{aligned} dS_i/dt &= -\lambda_{i,w} + R_i/\mu_{Ri} \\ dI_i/dt &= \lambda_{i,w} - I_i/\mu_{Ii} \\ dD_i/dt &= \rho_i I_i/\mu_{Ii} - D_i/\mu_{Di} \\ R_i &= N_i - S_i - I_i - D_i \end{aligned} \qquad (3)$$

In summary, the infection transmission model is specified by equation (1) and equation (3). We refer to this as an SIDRS/W model. The parameters in the above equations, as well as values that are consistent with *Cryptosporidium*, are presented in Table 18.1. Parameters without base values are functions of other parameters, or are unknown or varied in the sensitivity analysis to follow. The term in brackets is the unit of measure for the parameter values in the table.

18.2.2 Policy regions for water treatment decisions

This section presents a sensitivity analysis for water treatment policy regions for centralized versus local treatment interventions. We consider *n=2* population subgroups, (1) immunocompetent and (2) immunocompromised individuals, and their exposure to *Cryptosporidium*. The centralized water treatment considered here is ozone pretreatment, which can remove 40-60% of *Cryptosporidium* oocysts from water. This has the effect of reducing τ_i by an appropriate percentage for the entire population. The local treatment considered here is a filter that essentially removes exposure from drinking water (as an extreme case) for the immunocompromised subgroup. This sets $\tau_2 = 0$ for the immunocompromised subgroup, but leaves τ_1 unchanged for the immunocompetent subgroup.

Table 18.1 Summary of notation for SIDRS/W model, ranges for Cryptosporidium, and values used in a base analysis of the deterministic ODE model

Symbol	Meaning	Range	Base value
N	Number of subgroups in human population	1, 2, …	2
N	Total # individuals in human population	> 0	1.6×10^6
N_i	Total # individuals in subgroup i	> 0	
γ	Rate of exogenous introduction of microbes [microbes/liter/day]	10^{-6}-10^{2}	10^{-6}
α	Rate microbes become inactivated [1/day]	.05	.05
θ_i	Rate an infectious individual sheds microbes [microbes /liter/day]	≥ 0	0
r_i	Probability an ingested microbe causes infection	.0021-.0076	.00428
ϕ_i	Water consumption [liters/day]	.017-2	1
τ_i	Fraction of microbes surviving water treatment	10^{-6}-1	10^{-3}
ρ_i	Probability that infection progresses to disease	.38-.81	.61
μ_{Ii}	Mean incubation period [days]	1-12	7
μ_{Di}	Mean duration of disease stage [days]	1-55	9
μ_{Ri}	Mean duration of recovered/immune stage [days]	60-120	90
	Fraction of oocysts viable after ozone pre-treatment	.2-.8	.4
c_i	Human contact rate for members of subgroup i [contacts/day]	≥ 0	
β_i	Probability a susceptible member of subgroup i becomes infected from a potentially infectious human contact	0.0-1.0	
$\lambda_{i,W}$	Force of infection to subgroup i, given microbe concentration W.	See equation (2)	

We define the 'better' treatment in this chapter as that which leads to the lowest endemic prevalence of cryptosporidiosis in the immunocompromised subgroup. This objective is motivated by the extreme effects of

cryptosporidiosis in that subgroup during the 1993 Milwaukee outbreak. A similar analysis can be run for other outcome measures of merit, including quality-adjusted and disability-adjusted life years, and cost effectiveness ratios, but we do not do so here.

A risk assessment that ignores the dynamics of secondary transmission would conclude that the filter is more successful that ozone pretreatment for the immunocompromised subgroup. If secondary transmission is significant, however, secondary transmission from the immunocompetent subgroup can result in significant infection in the immunocompromised subgroup. In fact, if human-to-human secondary transmission is high enough, then removing all microbes from the water will still not prevent endemic transmission. In that case, water treatment makes almost no impact on the prevalence of infection.

Before presenting policy regions, we introduce notation to describe secondary transmission. Let $R_{0h,ij}$ be the mean number of secondary transmissions from human contact by an infected individual of subgroup *j* to individuals in subgroup *i*, assuming that all individuals in subgroup *i* are susceptible (c_j contacts per unit time, a fraction $c_i N_i / \Sigma\, c_k N_k$ of them with subgroup *i*, of which β_i are infective, for a mean duration of $\mu_{Ij} + \rho_j \mu_{Dj}$).

$$R_{0h,ij} = c_j \frac{c_i N_i}{\sum c_k N_k} \beta_i \left(\mu_{Ij} + \rho_j \mu_{Dj}\right). \qquad (4)$$

Let $R_{0w,ij}$ be the analogous number of secondary transmissions through the water loop from an infective in subgroup *j* to individuals in subgroup *i*, assuming that all individuals in subgroup *i* are susceptible (see Appendix).

$$R_{0w,ij} = \frac{N_i r_i \tau_i \phi_i \theta_j}{\alpha} \left(\mu_{Ij} + \rho_j \mu_{Dj}\right). \qquad (5)$$

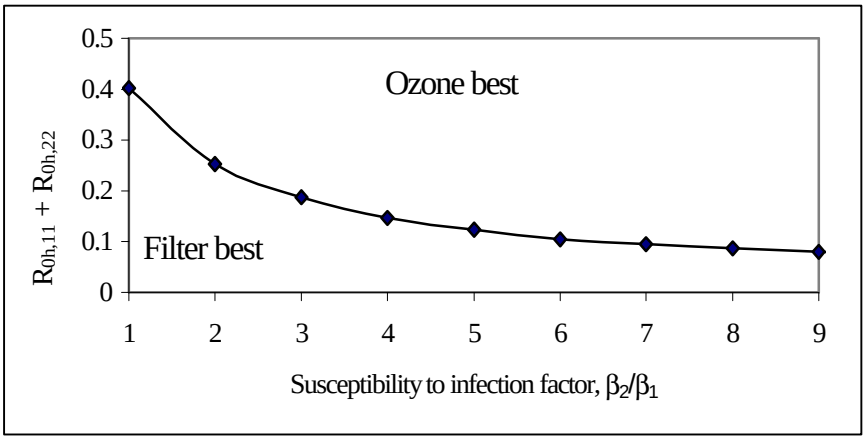
The Appendix proposes two different arguments to show that the basic reproduction number, R_0, is key to determining the infection dynamics. It can be related to the expected number of secondary infections needed to sustain endemic infection.

$$\begin{aligned} R_0 = {} & R_{0h,11} + R_{0w,11} + R_{0h,22} + R_{0w,22} - R_{0h,11}R_{0w,22} - \\ & R_{0h,22}R_{0w,11} + R_{0h,12}R_{0w,21} + R_{0h,21}R_{0w,12}. \end{aligned} \qquad (6)$$

If $R_0 > 1$, then infection remains endemic even if no exogenous introduction of microbes occurs ($\gamma = 0$).

Somewhat surprisingly, it is still possible for a municipal improvement like ozone pretreatment to outperform filters on the taps of immunocompromised individuals, even if endemic infection is not sustainable through secondary transmission. The reason is that cryptosporidiosis prevalence in the immunocompetent subgroup can be significantly reduced with ozone pretreatment. This in turn reduces secondary exposure of cryptosporidiosis to the immunocompromised subgroup. Figure 18.2 illustrates that ozone pretreatment is more successful at reducing endemic cryptosporidiosis infection in the immunocompromised subgroup if the secondary transmission rate from human contact is high enough. This graph assumes that all secondary transmission occurs from human contact ($R_0 = R_{0h,11}+R_{0h,22}$, because $\theta_1 = \theta_2 = 0$), and that other parameters take on the base values for *Cryptosporidium* given in Table 18.1. If immunocompromised individuals are much more susceptible to cryptosporidiosis than immunocompetent individuals (larger β_2/β_1), then ozone pretreatment is attractive at even lower levels of secondary transmission. The values for $c_i\beta_i$ are chosen to give rise to the corresponding value of R_0 on the y-axis.

Figure 18.2 Ozone pre-treatment is better for larger values of secondary transmission or the relative susceptibility of immunocompromised individuals

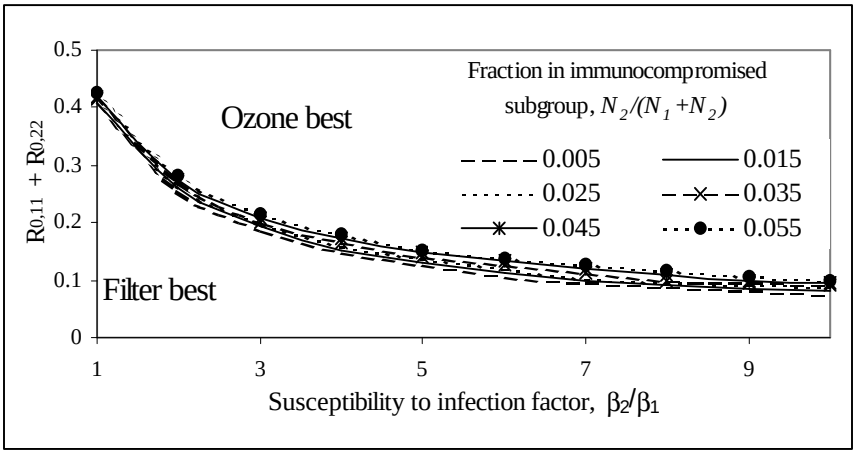


These observations are qualitatively similar to results in our previous work [14]. The policy region boundary is somewhat lower here than in [14] for several reasons: the natural history of infection is more realistic here (including two infectious periods, the infected/asymptomatic and diseased/symptomatic states), a more effective ozone pretreatment process is assumed (60% of oocysts are removed rather than 50%), and a few other

parameters are changed. The qualitative shape of the policy region, however, is the same. We now extend the results by assessing the sensitivity of the policy region to several parameters that may affect transmission dynamics.

Figure 18.3 shows that the policy region is relatively insensitive to the fraction $N_2/(N_1 + N_2)$ of individuals in immunocompromised subgroup, at least when base case parameter values are used, and the fraction of immunocompromised individuals is relatively small (under 5% or so). If that fraction increases, the policy region boundary would rise, as direct exposure would become relatively more important than secondary transmission from the smaller immunocompetent subgroup. The policy region is similarly insensitive [25] to the rate γ of exogenous introduction of microbes, except if rates would lead to oocyst concentrations found during outbreaks with plant failures.

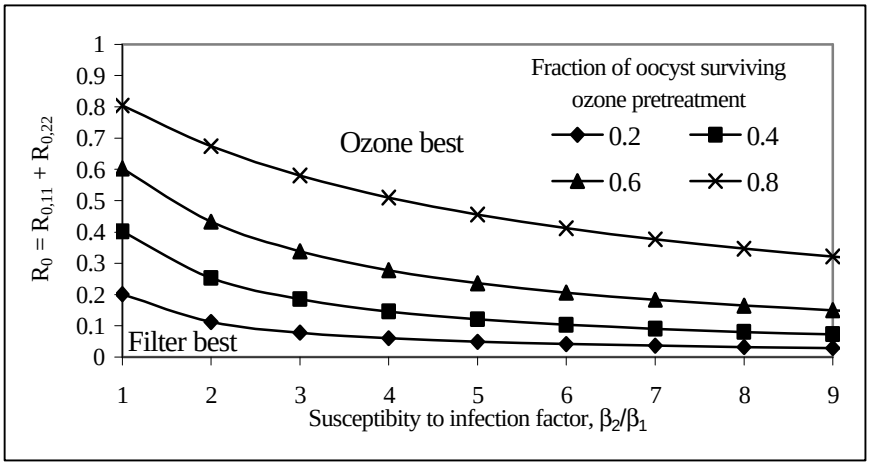
Figure 18.3 The policy region is relatively stable over a range of values for the fraction of population that is immunocompromised



The 1993 Milwaukee outbreak data has been used to estimate [13] the secondary transmission rate as R_0=0.15. The secondary transmission rate during endemic situations is unknown, but individuals may be more conscientious about secondary transmission during an outbreak than when infection is transmitted silently in the background. It seems reasonable to assume that immunocompromised individuals may be somewhat more susceptible to cryptosporidiosis infection due to human transmission ($\beta_2/\beta_1 > 1$), but there is inconclusive data one way or the other [26].

Figure 18.4 illustrates the sensitivity of the policy region to ozone pretreatment efficiency. Ozone pretreatment outperforms filters in this analysis even at relatively low values of secondary transmission, assuming that 80% of oocysts can be inactivated during the pretreatment. Although the values of secondary transmission parameters are not completely understood, this would put the treatment policy boundary near educated approximations for the parameter estimates. On the other hand, a risk assessment that assumes that there is no secondary transmission from interpersonal contact would indicate that filters are much more effective at reducing the endemic prevalence of cryptosporidiosis in the immunocompromised subgroup.

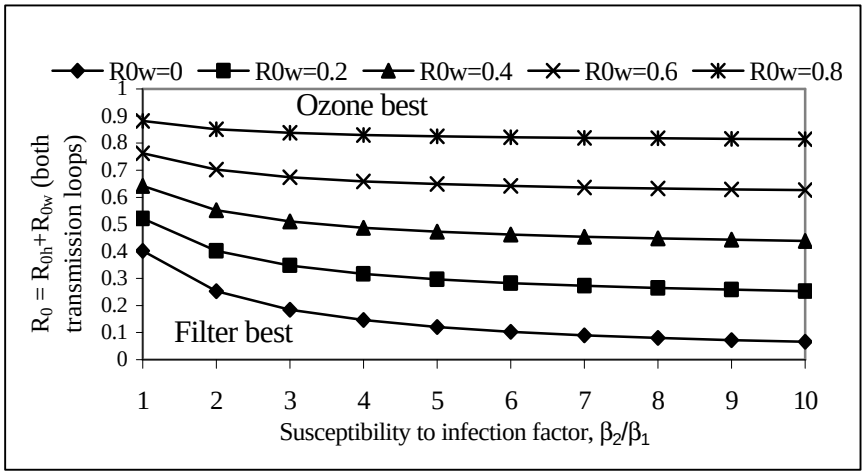
Figure 18.4 The policy region is highly dependent upon the effectiveness of ozone pre-treatment for removing oocysts



The graphs above assume that human contact is the sole exposure for secondary transmission, with no active water loop. This may be appropriate where there is no potential for recreational activities to contaminate source water. In some regions, however, recreational use can pose a distinct risk for water loop transmission [7]. Figure 18.5 shows that as the water loop increases in importance for transmission (increasing $R_{0w} = R_{0w,11} + R_{0w,22}$), the policy region boundary rises. Conceptually, this matches the notion that if *all* secondary transmission occurs through the water loop, with no human contact, then filters for the immunocompromised subgroup are more effective than ozone pretreatment in reducing cryptosporidiosis prevalence in that subgroup (filters are assumed here to be 100% effective, but ozone is only partially effective at removing oocysts). This means that filters are always more effective, relative to this objective, when there is no human-to-

human transmission. Filters may still be an effective intervention if secondary transmission occurs primarily through the water loop.

Figure 18.5 Filters are much more effective if the water loop increases in importance relative human-to-human secondary transmission



18.3 VARIATION IN INFECTION OUTCOMES

Infection and recovery times are stochastic, not deterministic; this is one source of variation in prevalence and microbial contamination data. How much variation in infection outcomes should one expect, even if all infection transmission parameters are known precisely? Another important related question is how to estimate unknown infection parameters, given field data. While the policy regions like those in Section 18.2 are useful for qualitative insights into the effects of treatment given transmission parameter assumptions, the precise values of parameters are still poorly understood for several microbes transmitted through the water system. A model of the random variation in infection prevalence and microbe concentration can be used as a likelihood function to help infer the unknown parameters. Ideally, such a model would be easy to simulate quickly.

18.3.1 Stochastic model background

Several authors have incorporated stochastic system dynamics to infer the parameters of infection models. Deterministic ODE infection models may have stochastic analogs that are derivable as large population limits [20-23, 27]. A continuous time stochastic analog of the deterministic SIS/W model

with n closed subgroups, the model in Section 18.2 without the extra disease states, has a state $(S_1, \ldots, S_n, Y)$, where $Y=WN\Delta$ is the total oocyst count in the drinking water supply[1]. The state space is a lattice, $\{\prod_{i=1}^{n}\{0, 1, \ldots, N_i\}\}x\{0, 1, \ldots\}$. The state does not include I_i since $I_i = N_i - S_i$ by assumption here. State transition rates are determined by the associated rate in the ODE. For example, the transition rate from $(S_1,\ldots,S_i,\ldots, S_n,Y)$ to $(S_1,\ldots,S_i+1\ldots, S_n, Y)$ is I_i / μ_{Ii}, based on the recovery rate $1/\mu_{Ii}$ of each individual. Infection transitions to $(S_1, \ldots, S_i-1 \ldots, S_n, Y)$ occur with rate

$$r_i\phi_i\tau_i WS_i + \sum_{j=1}^{n} c_j I_j \frac{c_i N_i}{\sum_{k=1}^{n} c_k N_k}\frac{S_i}{N_i}\beta_i . \qquad (7)$$

The analogy of these rates with Equations (2) and (3) should be clear. Rate terms in the ODE dynamics correspond to infinitesimal state transition rates in the Markov chain model, and transition rates for microbe immigration and inactivation occur similarly. Figure 18.6 shows a sample path for the number infected in a 3-subgroup model as it varies about the trajectory of the analogous ODE model. An alternate approach is to use a closely related discrete-time Reed-Frost epidemic model [18], or to also incorporate social network information into the state with a stochastic graph [19].

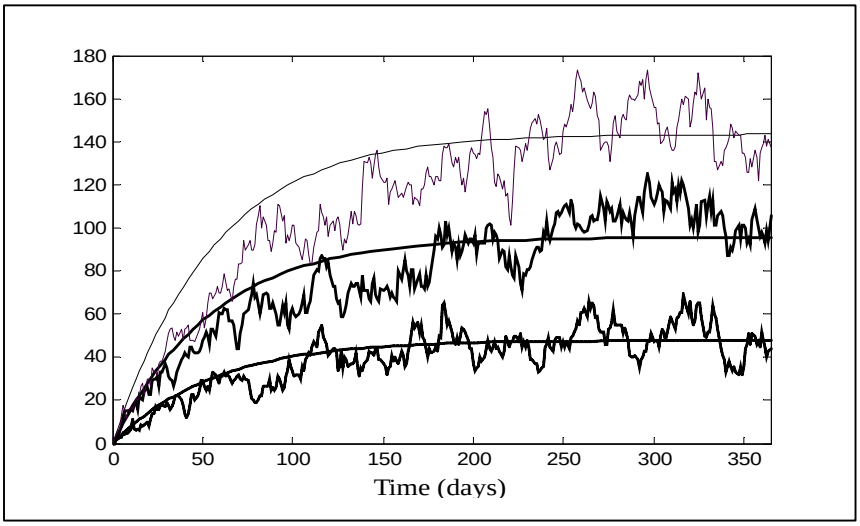
Several researchers (e.g., see [19] and references therein) have developed likelihood models for Bayesian inference that incorporate infection dynamics into the likelihood function for parameters as a function of data that might be obtained from tracing an outbreak, or closely monitoring an intervention trial. Interesting properties of quasistationary distributions [28], the long run distribution assuming that infection remains endemic, of infection models have been studied as well.

A recent proposal to infer infection parameters with endemic data provides a statistical tool that provide an alternative to waiting for, identifying, and measuring an outbreak [16]. The work uses stationary distributions of a combined stochastic-deterministic SIS/W infection model in a homogenous population to model endemic data. Infection and recovery events were assumed to be stochastic, but water contamination was assumed to be deterministic, given the number infected. Because a closed form for the stationary distribution is not known and there are situations when normal approximations used by standard ODE least square estimators are not fully justified, the authors develop two likelihood approximations.

---

[1] The total number of microbes, not microbes per volume, in a volume $N\Delta$ of water that scales with N is needed to obtain diffusion approximation results. See [20-23], Appendix 18.A.4.

Figure 18.6 A sample path for the number infected in three subgroups for a stochastic model varies about the trajectory of the analogous ODE



The first uses the stationary distribution of a closely related lattice Markov chain whose state is the number infected. That likelihood approximation has good bias and root mean square error (RMSE) properties, but may be computationally intensive when extended to populations with multiple subgroups, or if the natural history of infection is more complex. The second likelihood approximation uses a normal distribution approximation that takes advantage of relationships between low order moments that are determined by the Kolmogorov forward equations, but is somewhat more biased or may give confidence regions that are too small, particularly near $R_0 = 1$ or when populations are small, where the normal approximation may be suspect. Further, the continuous dynamics for the water, combined with the moment relationships, may or may not give a full specification of the system with more complicated natural histories of infection, or with multiple subpopulations (e.g., higher dimensions).

Here we take an alternate approach to approximating the stationary distribution of the number of infections: a diffusion approximation [20-23]. While statistical bias issues may remain to be resolved if the populations are small or if R_0 is near 1, the approach appears to be more generalizable to higher dimensions. While the mixed stochastic/deterministic model in [16] cannot directly use diffusion approximation results, a slight change to use the stochastic model on the lattice introduced at the beginning of Section

18.3.1 makes those results applicable. In particular [20-23] illustrate that density-dependent processes, which include many epidemic models like the lattice-state model above, can be approximated (in law) by an Ornstein-Uhlenbeck (OU) process near an endemic equilibrium point, as N grows. The stationary mean is approximated by the ODE's asymptotically stable endemic infection level (which is positive if there is exogenous contamination, $\gamma > 0$, and other parameters are not 0), and the stationary covariance matrix Σ can be approximated by appropriately rescaling the solution to a Lyapunov equation, as overviewed in Appendix 18.A.4.

18.3.2 Preliminary results for diffusion approximation

This chapter presents only preliminary results for the OU approximation. We simulated the continuous time SIS/W stochastic process with proportional mixing and compared sample statistics for the stationary mean and variance with the endemic ODE mean and OU approximation to the variance.

Simulated population sizes were 60, 600 and 6000 individuals in $n=3$ subgroups, with 1/6 of the individuals in subgroup 1, 1/3 in subgroup 2, and 1/2 in subgroup 3. Parameters were chosen so that a fair amount of secondary transmission would be observed. Parameters were chosen to be the same for each subpopulation, with $c_i=c$, $\beta_i=\beta$, $\mu_{Ii}=\mu_I=7$ days, etc. so that $R_{0h} = c\beta\mu_I = 0.875$, $R_{0w} = 0.05$. Table 18.2 provides some summary sample statistics for the stationary mean $\bar{i}_j$ and standard deviation σ_j of the number infected in subgroup j. The statistics were based on 150 years of simulated infection and water contamination. The means are time averages, and the standard deviations are based upon sampling the number infected once per month. The OU approximation for the mean equals the ODE endemic equilibrium, and the standard deviations are computed as described in the previous section and Appendix 18.A.4. As observed elsewhere [29, 30], the mean number infected estimated by the simulations is lower than predicted by the deterministic model for smaller populations. Correlation is strong between subgroups and water contamination levels in simulations with significant secondary transmission, matching simulations with a single subgroup in [16].

The OU approximation for the mean and variance of the number infected provides yet another likelihood approximation for inferring infection parameters from endemic data, to complement the two approximations in

Table 18.2 Comparison of some estimates from long simulation runs versus the Ornstein-Uhlenbeck (OU) approximation for the stationary mean and covariance

Population Size, N	Approx-imation	Statistic			
		$\bar{i}_2$	$\bar{i}_3$	σ_2	σ_3
60	Sim.	0.72	1.10	0.927	1.18
	OU	0.95	1.44	1.10	1.44
600	Sim.	9.28	13.9	3.72	5.32
	OU	9.56	14.4	3.81	5.25
6000	Sim.	95.7	143	13.5	18.82
	OU	95.5	143	14.3	20.97

[16] (the two ideas there were to compute the stationary distribution, and to use the Kolmogorov forward equations to establish relationships between moments). How strongly the bias in the estimates of the means and variances might influence the bias and RMSE of parameter estimators based upon an OU likelihood approximation is an area for further study.

18.4 CONCLUSIONS AND FUTURE RESEARCH

18.4.1 Water treatment policy

Infection transmission dynamics can strongly influence the public health benefit of water treatment interventions. Ignoring secondary transmission in a risk assessment, or examining only first order effects, can suggest misleading conclusions. System dynamics models can help quantify the complex infection dynamics that some microbes transmitted through the drinking water system may have. Policy decisions regarding the recreational use of public waterways that are source water directly influence the potential for secondary transmission, too.

While infection transmission parameters may be important determinants of the health benefit of interventions, their values are not well understood for a number of microbes. The use of stationary distributions as likelihood functions for unknown parameters allows endemic data to be used in the inference process. This complements tools by others to infer parameters in an intervention trial or with outbreak data [17-19].

Here we considered only one microbial agent. In reality, there are many strains of many microbes. A comprehensive risk management program must consider multiple microbes and multiple intervention options. Further, some

coordination may be required between different governmental agencies. The Centers for Disease Control and Prevention are historically responsible for outbreak and infection data, whereas the EPA is historically responsible for water quality data.

18.4.2 Infection modeling

One advantage of the OU approximation, at least for large populations with nontrivial endemic levels, is that the mean and covariance matrix are readily computed. Furthermore, transient probability distributions can be estimated with this OU approximation under certain conditions [23]. In principle, this would allow for data from outbreaks, intervention trials and/or endemic data to be used to infer transmission parameters.

The OU approximation has a statistical bias when the population size is small, or there are small numbers of individuals per subgroup, such as occurs when family units or small work sites form the subgroups [30]. The stationary and quasistationary mean prevalence of the lattice-based Markov chain infection model may be lower than the scaled endemic equilibrium infection level. The difference goes to zero in the large population limit, but may be nontrivial for small populations. A rigorous exploration of this bias is an area for further research.

Such bias holds implications not only for parameter inference, but also for speeding up simulations of infection processes. The OU process might ignore every infection and recovery event in a large process, but may require small time steps to insure that bias is avoided. An interesting simulation question is to evaluate effective ways to simulate the approximating OU process in a way that faithfully represents important low order statistical properties of the original Markov process. Approximations that work well when almost everybody or almost nobody in a subgroup is infected are an open area of research, and have implications for simulating small populations and subgroups, such as family units or daycare centers that are participating in water treatment intervention or vaccine trials.

18.4.3 Other modeling applications

The food supply chain is complex and presents another potential route for the transmission of microbes. Some infection models have examined the dynamics of growth of microbes as food passes from the farm to the fork [31] in one context. Others have examined the dynamics of infection in herds [32, 22] and the ensuing impact on the livestock industry. One area for further development is the integration of infection dynamics models in

animals, microbes in the food supply chain, and primary and secondary exposure in human populations.

Acknowledgments

We appreciate discussions with Brenda Boutin, PhD, and the financial support of the US EPA National Center for Environmental Assessment (CR-827427-01-0).

References

[1] Hoxie, N.J., J.P Davis, J.M Vergeront, R.Nashold and K. Blair (1997). Cryptosporidiosis-associated mortality following a massive waterborne outbreak in Milwaukee, Wisconsin. *American Journal of Public Health*, 87, 2032-2035.

[2] MacKenzie, W.R., et al. (1994). A massive outbreak in Milwaukee of Cryptosporidium infection transmitted through the public water supply. *New England Journal of Medicine*, 331, 161-167.

[3] Iley, K. (2002). Aid groups urge action on water-borne diseases. Reuters News Service at www.planetark.org (March 25, 2002).

[4] Cowdy, H. (2002). Millions at risk from contaminated water. Reuters News Service at www.planetark.org (March 25, 2002).

[5] Centers for Disease Control and Prevention (2001). Norwalk-like viruses: Public health consequences and outbreak management. *Morbidity and Mortality Weekly Report*, 50(RR-9), 1-17.

[6] Cooke, R. and B. Kraan (2000). Processing expert judgements in accident consequence modelling. *Radiation Protection Dosimetry*, 90, 311-315.

[7] Stewart, M.H., M.V. Yates, M.A. Anderson, C.P. Gerba, J.B. Rose, R.DeLeon and R.L. Wolfe (2002). Predicted public health consequences of body-contact recreation on a potable water reservoir. *Journal of the American Water Works Association*, 94, 84-97.

[8] Stout, J.E., Y-S E. Lin, A.M. Goetz and R.R. Muder (1998). Controlling legionella in hospital water systems: Experience with the superheat-and-flush method and copper-silver ionization. *Infection Control and Hospital Epidemiology*, 19, 663-674.

[9] Current, W. (1994). Cryptosporidium parvum: Household transmission. *Annals of Internal Medicine*, 120, 518-519.

[10] Medema, G. and J. Schijven (2001). Modeling the sewage discharge and dispersion of cryptosporidium and giardia in surface water. *Water Research*, 35, 4370-4316.

[11] Regli, S., J.B. Rose, C.N. Haas and C.P. Gerba (1991). Modeling the risk from giardia and viruses in drinking water. *Journal of the American Water Works Association*, 83, 76-84.

[12] International Life Sciences Institute (2000). *Revised Framework for Microbial Risk Assessment*. ILSI Risk Science Institute workshop report, International Life Sciences Institute, Washington, DC.

[13] Eisenberg, J.E., E.Y.W. Seto, J.M. Colford Jr, A. Olivieri and R.C. Spear (1998). An analysis of the Milwaukee Cryptosporidiosis outbreak based on a dynamic model of the infection process. *Epidemiology*, 9, 255-263.

[14] Chick, S.E., J.S. Koopman, S. Soorapanth, M. E. Brown (2001). Infection transmission system models for microbial risk assessment. *Science of the Total Environment*, 274, 197-207.

[15] Frisby, H.R., D.G. Addiss, Reiser, et al. (1997). Clinical and epidemiologic features of a massive waterborne outbreak of WJ Cryptosporidiosis in persons with HIV infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 16, 367-373.

[16] Chick, S.E., J.S. Koopman, S. Soorapanth, B.K. Boutin (In press). Inferring infection transmission parameters that influence water treatment decisions. *Management Science*.

[17] Brookhart, M.A., A.E. Hubbard, M.E. van der Laan, J.M. Colford, J.N.S. Eisenberg (In press). Statistical estimation of parameters in a disease transmission model. *Statistics in Medicine*.

[18] O'Neill, P.D. (In press). Perfect simulation for Reed-Frost epidemic models. *Statistics and Computing*.

[19] Britton, T. and P.D. O'Neill (2002). Statistical inference for stochastic epidemics in populations with random social structure. *Scandinavian Journal of Statistics*, 29, 375-390.

[20] Kurtz, T.G. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, 7, 49-58.

[21] Kurtz, T.G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8, 344-356.

[22] Clancy, D. and N.P. French (2001). A stochastic model for disease transmission in a managed herd, motivated by Neospora caninum amongst dairy cattle. *Mathematical Biosciences*, 170, 113-132.

[23] Pollett, P.K. (1990). On a model for interference between searching insect parasites. *Journal of the Australian Mathematical Society, Series B*, 31, 133-150.

[24] Nold, A. (1980). Heterogeneity in disease-transmission modeling. *Mathematical Biosciences*, 52, 227-240.

[25] Soorapanth, S. (2002). *Microbial Risk Models Designed to Inform Water Treatment Policy Decisions.* PhD Thesis, University of Michigan, Ann Arbor, MI.

[26] Osewe, P., D.G. Addiss, K.A. Blair, A. Hightower, M.L. Kamb, J.P. Davis (1996). Cryptosporidiosis in Wisconsin: A case-control study of post-outbreak transmission. *Epidemiology and Infection*, 117, 297-304.

[27] Altmann, M. (1998). The deterministic limit of infectious disease models with dynamic partners. *Mathematical Biosciences*, 150, 153-175.

[28] Nåsell (1996). The quasi-stationary distribution of the closed endemic SIS Model. *Advances in Applied Probability*, 28, 895-932.

[29] Chick, S.E. (2002). Approximations of stochastic epidemic models for parameter inference. Working Paper, INSEAD, Fontainbleau, France.

[30] Koopman, J.S., S.E. Chick, C.S. Riolo, C.P. Simon, G. Jacquez (In press). Stochastic effects on endemic infection levels of disseminating versus local contacts. *Mathematical Biosciences*.

[31] McNab, B.W. (1998). A general framework illustrating an approach to quantitative microbial food safety risk assessment. *Journal of Food Protection*, 61, 1216-1228.

[32] Ferguson, N.M., C.A. Donnelly, and R.M. Anderson (2001). The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. *Science*, 292, 1155-1160.

[33] Simon, C.P. and J. A. Jacquez (1992). Reproduction numbers and the stability of equilibria of SI models for heterogeneous populations. *SIAM Journal of Applied Mathematics*, 52, 541-576.

18.A Appendix

18.A.1 Algebraic stability conditions for the SIRS/W model

Consider first the SIRS/W infection transmission model in a homogenously mixing population (a special case of the general model, with *n=1* subgroup, $\rho=0$, so D=0 and we drop subscripts in this section). The expected number R_{0h} of secondary transmissions, due to human contact, caused by one infective in an otherwise susceptible population, is the contact rate *c*, times the infection probability per contact β, times the duration of infection μ_I.

$$R_{0h} = c\beta\mu_I . \qquad \text{(A.1)}$$

The analogous number of secondary transmissions through the water loop is qualitatively derived by noting that an infected individual raises the concentration of microbes by θ microbes per day for μ_I days, the microbes remain viable for $1/\alpha$ days, and each of *N* susceptibles consumes a fraction $\tau\phi$ of available microbes, each of which causes infection with probability *r*.

$$R_{0w} = \frac{Nr\tau\phi\theta}{\alpha}\mu_I . \qquad \text{(A.2)}$$

Then $R_0=R_{0h}+R_{0w}$ is the total number of secondary transmissions, on average.

Theorem 1: If there is no exogenous source of microbes ($\gamma=0$), and there is homogeneous mixing (*n=1*), then

- The disease free equilibrium ($S^* = N$, $I^* = R^* = W^* = 0$) is locally asymptotically stable if $R_0 < 1$, and unstable if $R_0 > 1$.
- The endemic equilibrium ($S^*=N/R_0$, $I^* = N(1-1/R_0)\mu_I/(\mu_I + \mu_R)$, $W^*= \theta I^*/\alpha$) is locally asymptotically stable if $R_0 > 1$, and is not realizable if $R_0 < 1$.

Proof: The equilibrium values are determined by setting derivatives to 0. The stability result is proven by linearization in [25].

18.A.2 Algebraic stability conditions for the ODE in Section 18.2, (n=2)

By analogy with A.1, let $R_{0h,ij}$ and $R_{0h,ij}$ be as in Equations (4) and (5).

Theorem 2: Suppose there is no exogenous source of microbes ($\gamma = 0$), and there are $n=2$ subgroups with proportional mixing, as in Section 18.1. Consider the following two conditions.

(i) $$R_{0h,11} + R_{0w,11} + R_{0h,22} + R_{0w,22} - R_{0h,11}R_{0w,22} - R_{0h,22}R_{0w,11} + R_{0h,12}R_{0w,21} + R_{0h,21}R_{0w,12} < 1$$

(ii) $$R_{0h,11} + R_{0w,11} + R_{0h,22} + R_{0w,22} < 1$$

Then:

- Conditions (i) and (ii) are sufficient for the disease free equilibrium to be asymptotically stable ($S_1^* = N_1$, $I_1^* = R_1^* = W^* = 0$).
- If the inequality in condition (i) is reversed, then the zero equilibrium is not stable, resulting in positive endemic infection.

Proof: The stability result is proven by linearization in [25]. The two conditions are equivalent when $c_1\theta_2 = c_2\theta_1$. If $c_1\theta_2 \neq c_2\theta_1$, then the linearization leads to a quintic equation after some factorization, which is not solvable in closed form. The two conditions together are sufficient to insure that the dominant eigenvalue falls in the left hand complex plane.

We have not yet developed characterizations for $n>2$ subgroups when the water loop is active. [33] use Lyapunov functions to characterize stability for $n\geq 1$ subgroups with proportional and other mixing patterns for human-to-human transmission, but do not account for the water loop.

18.A.3 Alternate stability conditions for the ODE in Section 18.2

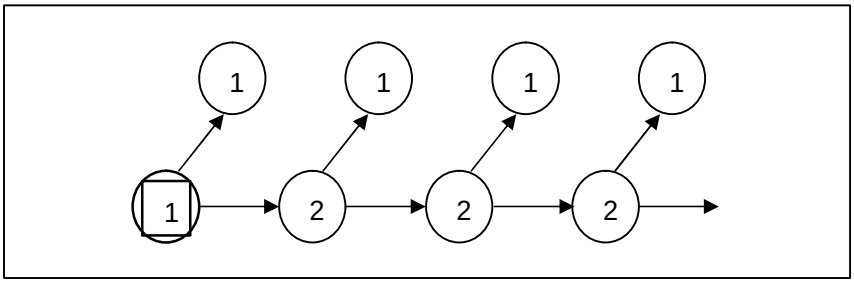
Sections 18.A.1 and 18.A.2 provide *population thresholds* to characterize stability based on an algebraic analysis. An alternate heuristic to assess whether endemic infection is sustainable even if no exogenous introduction of microbes occurs ($\gamma = 0$) is to assess an *individual level* endemic threshold using probabilistic arguments. This section overviews such an argument for the $n = 2$ subgroup model. Let $R_{0h,ij}$ and $R_{0w,ij}$ be as in 18.A.2, and denote the total mean number of secondary transmissions to a completely susceptible subgroup i from an index case in subgroup j by

$$R_{0,ij} = R_{0h,ij} + R_{0w,ij}. \qquad \text{(A.3)}$$

The individual level threshold is established by assessing whether the mean number of new infections in subgroup i caused by an initial index case in i is at least 1, when the whole chain of infection is considered. For example, an individual in subgroup 1 can infect someone in subgroup 2, who then infects

another person, who then eventually infects someone in subgroup 1. The expected number of infections (directly or indirectly caused) in the chain (see Figure 18.7) should be at least 1 for at least one subgroup.

Figure 18.7 The chain of infection from an index case in subgroup 1 can result in infections in subgroup 1 directly, or indirectly through subgroup 2



If $R_{0,11} > 1$ or $R_{0,22} > 1$, then a given subgroup can sustain infection within itself, and therefore infection remains endemic. Consider the case where $R_{0,11}$ and $R_{0,22}$ are both at least 0 but neither exceeds 1. If a single individual in subgroup 1 is infected, and the population is otherwise susceptible, then the expected number R of additional cases through the whole chain of transmission that eventually reach subgroup 1 is

$$\begin{aligned} R &= R_{0,11} + R_{0,21}\left(R_{0,12} + R_{0,22}\left(R_{0,12} + R_{0,22}(\cdots)\right)\right) \\ &= R_{0,11} + R_{0,21}R_{0,12} + R_{0,21}R_{0,12}R_{0,22} + R_{0,21}R_{0,12}R_{0,22}^2 + \cdots \quad \text{(A.4)} \\ &= R_{0,11} + \frac{R_{0,21}R_{0,12}}{1 - R_{0,22}}. \end{aligned}$$

The last equation holds because $R_{0,22} \in [0,1)$. An individual level threshold says that endemics cannot be sustained without exogenous sources of infection if $R<1$, or $R_{0,11} + R_{0,22} - R_{0,11}R_{0,22} + R_{0,12}R_{0,21} < 1$. Substituting the definition of $R_{0,ij}$ in Equation (A.3) gives an individual level threshold that is equivalent to the population threshold in condition (i) of Theorem 2 above.

18.A.4 Ornstein-Uhlenbeck approximation to SIS/W process

The OU approximation to the stochastic SIS/W model with proportional mixing can be derived by examining the ODE analog of that model along with the transmission rates of the stochastic model $(S_1,\ldots, S_n, W)$ summarized in Section 18.3.1. The idea (e.g., [22, 23]) is to first find a representation so

that the state scales up with N, the total population size, then to look at a rescaled version of that process. The S_1 already scale directly with N. To get the microbe contamination to scale with N, we model the total oocyst count Y in the drinking supply, rather than microbe concentration, and suppose that the drinking supply scales with the population size (e.g., contains a total of $N\Delta$ liters, and water drunk by individuals is replaced with fresh water so that the total volume remains constant). This means that $Y = N\Delta W$, γ is the oocyst contamination rate per unit time per Δ liters of water, and the rescaled process of interest is $x=(S_1,\ldots, S_n, Y)/N$.

Let $dx/dt = f(x_0)$ be a vector valued function that describes the dynamics of the scaled ODE model, and let x_0 be an asymptotically stable equilibrium in the interior of the scaled state space, with $f(x_0) = 0$. Let $A = \nabla f(x)\big|_{x=x_0}$ be the matrix containing the gradient of the dynamics $f(x)$ of the scaled process $x=(S_1,\ldots,S_n, Y)/N$, evaluated at x_0, and let $\zeta_i = N_i/N$. Let the matrix G be the local covariance of a scaled version of the state over a short time δt, given that the state is currently x_0. For the SIS/W model with proportional mixing, G is a diagonal matrix, and is determined by evaluating the following at x_0.

$$G_{ii} = \frac{(\zeta_i - x_i)}{\mu_{Ii}} + r_i\phi_i\tau_i w x_i + x_i \frac{c_i N_i \beta_i}{N_i \sum_{k=1}^{n} c_k N_k} \sum_{j=1}^{n} c_j(\zeta_j - x_j), \text{ for } i = 1,\ldots,n$$

$$G_{n+1,n+1} = \gamma + \sum_{j=1}^{n} \theta_j (N_j - S_j) + \alpha w \qquad \text{(A.5)}$$

The matrix G will have nonzero off-diagonal elements for the SIDRS/W model, since an increase in S_i means a decrease in I_i.

The stationary distribution for the OU process can be approximated with a normal distribution with mean Nx_0 and covariance matrix $N\Sigma$, where Σ solves Equation (A.6) (e.g., see [22, 23] for similar models without water transmission).

$$A\Sigma + \Sigma A^T + G = 0 \qquad \text{(A.6)}$$