A Simulation Model of Waterborne Gastro-Intestinal Disease Outbreaks: Description and Initial Evaluation

Anya Okhmatovskaia[1], Aman D. Verma[1], Benoit Barbeau[2], Annie Carriere[2], Romain Pasquet[1], David L. Buckeridge[1]
[1]McGill University, Montreal, QC; [2]École Polytechnique de Montréal, Montreal, QC

Abstract

We present an agent-based simulation model for generating realistic multivariable outbreak signals. The model defines a synthetic population and simulates the dissemination of pathogenic organisms through a municipal water distribution system, the mobility of individuals between geographic locations, their exposure to pathogens through water consumption, and disease progression in infected individuals. We present the results of an initial evaluation of the model – a simulation study replicating the historical outbreak of cryptosporidiosis in Milwaukee in 1993.

Introduction

Large waterborne disease outbreaks occur regularly in developed countries, resulting in considerable morbidity and cost [1-3]. Water treatment, outbreak detection, and control measures are public health strategies aimed at preventing future morbidity and mortality from GI outbreaks.

Proposed public health strategies to detect and control GI outbreaks should be evaluated. Data collected from historical events, such as the outbreak of cryptosporidiosis in Milwaukee in 1993 [2], can be helpful in testing potential strategies. There are, however, a limited number of historical outbreaks. Moreover, each outbreak is only one possible instance of what may have occurred given the wide variation in factors that influence epidemic dynamics.

Simulation offers considerable potential for the evaluation of public health policies. It allows one to generate multiple outbreaks under the same set of conditions and evaluate surveillance across the full distribution of outbreaks that may occur with a given mechanism of exposure, magnitude of exposure, and so on. Simulation also allows one to vary the outbreak conditions systematically and determine how detection and control strategies vary across different types of outbreaks that may occur in the future.

Objective

We aimed to develop a simulation model that generates epidemiological data sufficiently complex and realistic to evaluate public health strategies for detection and control of waterborne gastrointestinal disease outbreaks.

Methods

The proposed model uses an agent-based approach with an hourly temporal resolution. There are four main processes simulated by the model corresponding to four event types: a) contamination of the water by pathogenic organisms, b) population mobility (commute), c) water consumption by individuals, and d) disease progression in infected individuals. When multiple events are scheduled to occur at the same time, they are processed in the above order.

We use part of the Island of Montreal as the setting for our simulation, and we represent individuals and locations (census tracts and water sources). The model includes components for defining the synthetic population and simulating population mobility and water consumption patterns, and parameterized submodels for simulating pathogen dispersion, exposure, infection, and disease progression. In the following sections we explain the model and its components in more detail, and describe our approach to software implementation and validation of the model.

Model Description

Assumptions

We assume that person-to-person transmission is not an important factor in the progression and detection of waterborne disease outbreaks. This assumption is critical to simplify modeling and to reduce the computational requirements for software implementation of the model. Multiple studies of historical waterborne disease outbreaks suggest that cases attributable to secondary transmission constitute a minority and have a late onset of symptoms, well after the peak of the epidemic curve [1].

Age progression of individuals is not represented in our model, because the typical duration of a waterborne disease outbreak does not exceed several weeks. For the same reason we do not model seasonal variation in water consumption patterns and concentration of infectious organisms due to their changing viability.

Input and Parameters

The input to our model is a definition of the contamination scenario for the Pathogen Dispersion Component. This definition includes concentration of

C.parvum oocysts in the source water (St. Lawrence River) and the dynamics of the effectiveness of water treatment. In the simplest case, the treatment plant failure can be modeled as a step decrease in oocyst removal rate, which effectively increases the concentration of cryptosporidium in the treated water. The simulation control parameters include population sample size, the overall duration of the simulation in hours, and the initial seed for a random number generator. The parameters used by individual components of the model are summarized in Table 1. All parameters are interpreted as probabilities or are drawn from empirical probability distributions.

Components

Synthetic Population. Approximately 1 million people reside in the center-east of the Island of Montreal, represented in our model. The population component draws on Canadian census data and the literature to create a synthetic population of agents representing the true population of Montreal. We begin with census data, generating simulated individuals within each census tract (CT) so that, on average, the agents have the same demographic properties (age, gender) as the true population. Each simulated individual is then assigned a health status of immune competent or compromised based on known prevalence rates [4].

Population Mobility. The people in our model can change their location several times during the day, and we assign each individual a fixed daily commute schedule drawn from the Enquête origine-destination (EOD). The EOD is a survey of 58,000 Montreal residents to determine their commuting patterns. The commute schedule lists trips with their times and destinations. To generate such schedule for an individual agent in our model, we randomly select a respondent from the EOD with matching demographic characteristics and residence CT, and assign this respondent's itinerary to the agent. For each CT in the itinerary, we select a water source randomly from among nodes of the water system within a 2 km radius from the center of CT – this node will be used by the agent. The probability of selection is inversely related to a node's distance from the CT center. The resulting commute schedule thus lists water sources for all destinations.

Water Consumption. This component relies on the results from cross-sectional studies of water consumption [5-7]. These data provide the average daily volume of ingested water stratified by age and sex, and the distribution of the number of water servings per day. Depending on the outbreak scenario, the timing of exposure to contaminated sources of drinking water can critically affect the adverse health outcomes of the outbreak. Davis & Janke [8] have developed a probabilistic timing model for the ingestion of tap water, which assumes that most of the daily water volume is consumed around meal times. We use this model to generate a series of ingestion times that constitute the agent's water consumption schedule. To determine the amount of water consumed as one drink for each individual, we equally divide the total daily volume by the number of water servings per day. In our model, we also account for the proportion of consumed water that is treated (boiled, bottled), since 22% of the general population consume treated water (bottled, boiled) for over three-quarters of their daily intake [6], and this proportion is even higher in immunocompromised populations [9]. We use the share of untreated water in the overall daily ingested volume as the probability of drinking from the tap (for each ingestion).

Pathogen Dispersion. This component of the model simulates the dissemination of infectious organisms through the municipal water system following the introduction of organisms into the water system. A calibrated EPANET hydraulic model was used for this purpose. Part of the distribution system (2500 km, Figure 1) supplied by the two treatment plants (Atwater and Charles J.-Des Baillets) is modeled hourly, taking into account the intra-daily consumption pattern. We assume that *Cryptosporidum* is not

Table 1. Model parameters

Population	Proportion of population immunocompromised	0.0076
Water Consumption	Total daily volume of ingested water, L (by age & sex)	0.238 – 1.045
	Proportions of treated water in daily volume, proportion of population 0%, under 25%, 25-75%, over 75%	0.674, 0.014, 0.084, 0.228
	Number of servings of water per day, proportion of population < 4, 4-8, > 8	0.303, 0.481, 0.217
	Proportion of immunocompromised treating all ingested water	0.48
Disease Progression	Symptomatic rate of C.parvum infection (healthy, immunocompromised)	0.611, 0.767
	Infectivity of single organism (r parameter)	0.09
	C.parvum incubation period, days	Log-normal (1.8, 0.4)
	Duration of symptoms, days	Log-normal (2.4, 0.5)
	Duration of symptoms in immunocompromised, multiplier	5
	Mortality rate with prolonged illness (over 60 days, immunocompromised)	0.438

inactivated during water transit. The output from this model is the number of oocysts of Cryptosporidium per liter of water over time for each node of the water system (C_{n_t}).

Exposure Component. This component determines if and to what extent an individual becomes exposed to contaminated water. The cumulative exposure is calculated for each person over the period of 24 hours, summing exposure associated with each single drink of tap water:

$$D = \sum C_{n\_t} \times V_t ,$$

where V_t is the size of the drink taken at time t, and C_{n_t} is the concentration of C.parvum at that time at the individual's current location.

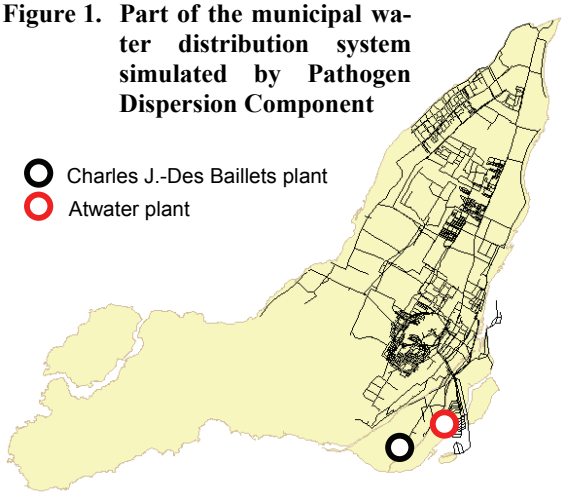
Infection Component. Teunis et al. [10] suggest an exponential function of the number of ingested C.parvum oocysts to determine the dose-dependent probability of infection:

$$P_{inf} = 1 - e^{-rD},$$

where r is the infectivity parameter specifying the probability of a single organism surviving in the host body and initiating an infection. The value of r was estimated to be 0.09 in the recent analysis performed by the US EPA for the Long Term 2 Enhanced Surface Water Treatment Rule [11]. There is no evidence in the literature suggesting higher infection rates among immunocompromised individuals; therefore we use the same infection model for all simulated agents.

Disease Component. Infected agents progress through a period of latent infection or incubation period, may progress to a symptomatic phase, and then on to either recovery or death. The Disease Component calculates the probability and timing of these transitions. According to the analysis of human infections with Cryptosporidium following waterborne exposure, 62% of infected patients develop watery diarrhea and other symptoms following an incubation period of about 7 days [2]. The reported median duration of symptoms for otherwise healthy adults ranges from 4 to 12 days[2, 12]. Given that the incubation period for most infectious disease is modeled well by a log-normal distribution, we model the duration of the latent infected and the symptomatic disease phases as log-normal, parameterized to follow the expected distribution of times for incubation and symptoms. Although cryptosporidiosis is self-limiting in most infected individuals, in people with compromised immune system, the disease may become chronic, and even fatal. To reflect longer illness in immunodeficient individuals, we scale the duration of illness drawn from a log-normal distribution for healthy adults by a factor of 5, expecting the median duration for immunodeficient to be around 60 days. When the calculated length of the symptomatic phase exceeds 60 days, a person with a compromised immune status in our model has a 50% chance of dying, which complies with the findings reported in the literature [13].

Figure 1. Part of the municipal water distribution system simulated by Pathogen Dispersion Component



Implementation

We have implemented our model in Java using MASON and SSJ libraries. MASON is a fast discrete-event multi-agent simulation library core that contains both a low-level model library and an optional suite of visualization tools. SSJ is a Java library for stochastic simulation that provides facilities for generating uniform and non-uniform random numbers.

A single simulation of a large outbreak (25% attack rate) with a population of 1 million and a duration of 60 days currently takes approximately 6 minutes on a single core of a 2.8GHz Nehalem EP (Xeon X5560) processor with 3GB of RAM. Execution time increases linearly with population size and outbreak duration, and negligibly with the number of infected individuals.

Validation

Because there are no historical data describing a waterborne outbreak of acute GI illness on the Island of Montreal, true validation of the entire model against external data is not yet possible. To evaluate individual components of the model and verify the correctness of their encoding in the software, we first examined the Java code, performed a systematic step through the program using a small sample from the population, and tested the components one at a time through a series of runs to compare intermediate model outputs against expected distributions. We also

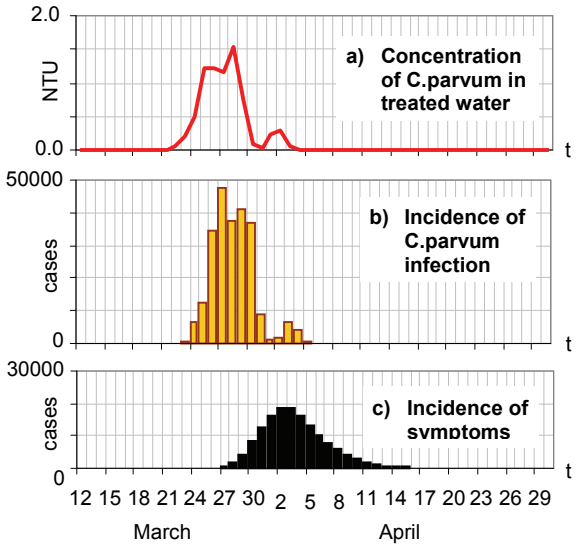


Figure 2. Simulation results for Milwaukee scenario, March 12 - April 30, 1993

used animation to validate Pathogen Dispersion, Mobility and Disease Components by monitoring visually the spread of contamination, movement of individual agents, and spatio-temporal distribution of disease cases.

To verify whether the presented simulation model as a whole produces credible output given a real outbreak scenario as an input, we have attempted to replicate in our model the largest documented waterborne epidemic of Cryptosporidiosis, the Milwaukee outbreak of 1993. During this outbreak, over 400,000 people were estimated to have watery diarrhea attributable to acute Cryptosporidium infection. The investigation of the Milwaukee Water Works (MWW) treatment facilities revealed an unprecedented increase in the turbidity of treated water from the southern plant during the period from March 21 through April 5 [2].

To reproduce the Milwaukee outbreak scenario as it would unfold within the setting of the Island of Montreal, we simulated the failure of the Atwater treatment plant (decline in the oocyst removal rate) while the Charles-J. DesBaillets plant would still work properly. First, mean Cryptosporidium concentration in treated water was calculated. Then, the source water concentration was back-calculated based on a linear relationship between Cryptosporidium log removal and turbidity log removal [14] using published turbidity values recorded during the Milwaukee outbreak [15]. This calculation yielded an average of 18 oocysts per 1 L in source water, a value assumed to be stable during the time of treatment failure (turbidity above 0.5 NTU, 13 days). Finally, *Cryptosporidium* concentration in treated water was calculated using this constant source water concentration, the turbidity profile during the outbreak and the Nieminski relationship [14]. The resulting concentration of oocysts in the treated water of the Atwater plant over time is plotted in Fig. 2a.

We ran the entire model for the period of 60 simulated days with an onset of the Atwater plant failure on day 10 and recorded the daily numbers of exposed, infected and symptomatic individuals, and the number of deaths (while most deaths fall outside the scope of the 60-day simulation, we count them toward outbreak-related mortality). We gathered the outbreak statistics for each census tract in the model, as well as cumulative counts. The simulation was repeated 1000 times using different seeds for the random number generator.

Results

Figure 2 displays the levels of contamination and the incidence of infections and symptoms over time during the outbreak period, from one simulation run. The incidence of infection appears to be strongly correlated with the concentration of Cryptosporidium in drinking water. The distribution of symptoms incidence is log-normally shaped and follows the infection curve with a lag of about 5-6 days, as expected given incubation times for C.parvum.

Table 2 lists the aggregate rates of exposure, infection, symptoms (attack rate) and mortality, along with the corresponding numbers observed or estimated during the Milwaukee outbreak. While the attack rate is similar to reported numbers, the mortality rate is almost 5 times higher than what was documented in Milwaukee.

Figure 3 compares the shape of the simulated epidemic curve (scaled down to account for the differences in sample size) to previously published plots of historical data [2]. The time series display visible similarity and have similar skewness and kurtosis. The simulated curve is slightly more positively skewed and peaks one to two days earlier than the historical ones.

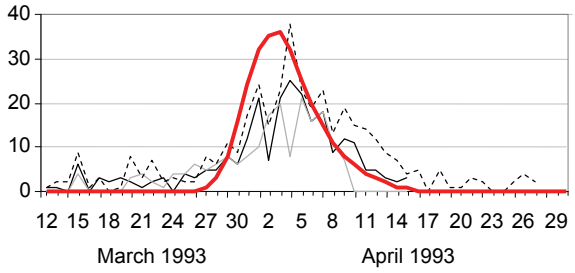
Discussion

We have demonstrated that the proposed simulation model can generate a realistic time series of gastroin-

Table 2. Aggregate simulation results

	Mean (min, max)	Historical
Exposure rate	0.588 (0.587,0.589)	-
Infection rate†	0.422 (0.420,0.424)	-
Attack rate‡	0.258 (0.256,0.260)	0.260
Mortality rate	0.00016 (0.00013,0.00019)	0.000034

† *N infected / N exposed* ‡ *N symptomatic / N exposed*



Data source	Peak	Skewness	Kurtosis
Milwaukee lab-confirmed cases	Apr 4	1.51	4.24
Clinical infection in MWW area	Apr 5	1.55	4.40
Phone survey of larger Milwaukee area	Apr 4	1.41	4.78
Simulation (scaled)	Apr 3	1.74	4.66

Figure 3. Onset of illness in historical [2] vs. simulated data

testinal cases resulting from a waterborne disease outbreak. The simulated outbreak signal is consistent with the observations during the 1993 cryptosporidiosis outbreak in Milwaukee.

In our study, only 58% of the simulated population became exposed to pathogenic organisms through drinking water. This number relies on the rates of bottled water usage and patterns of pathogen dispersion specific to the water distribution system. Unfortunately, true exposure rates are not available for historical waterborne outbreaks, thus making it problematic to validate the simulated exposure rates. The significantly elevated mortality rate obtained in our study was unexpected. We speculate that it is due to an overestimated proportion of immunocompromised individuals in the synthetic population. There is also considerable uncertainty in the literature about the distribution of symptoms duration and associated mortality rates in immunocompromised, which necessitates more careful validation of the corresponding parameters.

Although our model is specific to the geographical setting of the Island of Montreal, the comparison to the Milwaukee outbreak suggests that the model produces data similar to what one would observe in another large urban area. Moreover, due to the modular nature of our model, it would be straightforward to use alternative Dispersion and Mobility components to model other urban areas.

We are currently developing an extension to the presented model, which takes into account health-care system utilization by the population and under-reporting in public health. The model can also be extended to simulate waterborne outbreaks caused by other pathogenic organisms, such as Campylobacter jejuni or Giardia liamblia: this would require modifications of the Disease Progression Component.

References

1. Liang JL, Dziuban EJ, Craun GF, et al. Surveillance for waterborne disease and outbreaks associated with drinking water and water not intended for drinking--United States, 2003-2004. MMWR Surveill Summ. 2006 Dec 22;55(12):31-65.
2. Mac Kenzie WR, Hoxie NJ, Proctor ME, et al. A massive outbreak in Milwaukee of cryptosporidium infection transmitted through the public water supply. N Engl J Med. 1994 Jul 21;331(3):161-7.
3. Schuster CJ, Ellis AG, Robertson WJ, et al. Infectious disease outbreaks related to drinking water in Canada, 1974-2001. Can J Publ Health. 2005 Jul-Aug;96(4):254-8.
4. Casman EA, Fischhoff B, Palmgren C, Small MJ, Wu F. An integrated risk model of a drinking-water-borne cryptosporidiosis outbreak. Risk Anal. 2000;20:495-511.
5. Payment P, Berte A, Prevost M, Menard B, Barbeau B. Occurrence of pathogenic microorganisms in the Saint Lawrence River (Canada) and comparison of health risks for populations using it as their source of drinking water. Can J Microbiol. 2000 Jun;46(6):565-76.
6. Jones AQ, Majowicz SE, Edge VL, et al. Drinking water consumption patterns in British Columbia: an investigation of associations with demographic factors and acute gastrointestinal illness. Sci Total Environ. 2007 Dec 15;388(1-3):54-65.
7. Garriguet D. Beverage consumption of Canadian adults. Health Rep. 2008 Dec;19(4):23-9.
8. Davis JP, Janke R. A Probabilistic Timing Model for the Ingestion of Tap Water. Journal of Water Resources Planning and Management. 2009;135(5):397-405.
9. Eisenberg JN, Wade TJ, Hubbard A, et al. Associations between water-treatment methods and diarrhoea in HIV-positive individuals. Epidemiol Infect. 2002 Oct;129(2):315-23.
10. Teunis PF, Chappell CL, Okhuysen PC. Cryptosporidium dose-response studies: variation between hosts. Risk Anal. 2002 Jun;22(3):475-85.
11. US EPA. National Primary Drinking Water Regulations: Long Term 2 Enhanced Surface Water Treatment Rule. Federal Register. 2003;68(154):47640-88.
12. Jokipii L, Jokipii AM. Timing of symptoms and oocyst excretion in human cryptosporidiosis. N Engl J Med. 1986 Dec 25;315(26):1643-7.
13. Pozio E, Rezza G, Boschini A, et al. Clinical cryptosporidiosis and human immunodeficiency virus (HIV)-induced immunosuppression: findings from a longitudinal study of HIV-positive and HIV-negative former injection drug users. J Infect Dis. 1997 Oct;176(4):969-75.
14. Nieminski EC. Giardia and Cryptosporidium cysts removal through direct filtration and conventional treatment. Americal Water Works Association Annual Conference; 1994; New York City, NY; 1994.
15. Fox KR, Lythe DA. Milwaukee's crypto outbreak: Investigation and recommendations. American Water Works Association Journal. 1996;88(9).