# The potential implications of person-to-person transmission of viral infection for US EPA's Groundwater Rule

Jeffrey A. Soller

## ABSTRACT

The risk characterization method employed by US EPA to quantitatively characterize the benefits of the Groundwater Rule (GWR) for drinking water computes person-to-person transmission intensity as the product of the number of primary illnesses and a static secondary morbidity factor. A population level infectious disease health effects model is used here to evaluate the implications of secondary transmission on exposures to viruses that are relevant to the GWR. These implications are evaluated via a hypothetical case study in which it is assumed that a tour group from a large population centre visits an outlying area that is served by a non-community water system with untreated or inadequately treated groundwater that is contaminated with a highly infectious virus. It is assumed that some of the exposed individuals become infected and then return home. Numerical simulations are used to estimate the subsequent number of additional infections and illnesses due to secondary transmission within the large community. The results indicate that secondary transmission could substantially impact the predicted benefits of the GWR depending on the suite of population dynamic elements and assumptions employed.

**Key words** | drinking water regulation, microbial risk assessment, risk assessment, waterborne pathogens

**Jeffrey A. Soller** (corresponding author)
Soller Environmental,
3022 King Street,
Berkeley, CA 94703,
USA
E-mail: *jsoller@sollerenvironmental.com*

## INTRODUCTION

In the United States, drinking water is regulated by the Safe Drinking Water Act (SDWA). The SDWA requires many actions to protect drinking water for public water systems (PWS) and its sources, which include rivers, lakes, reservoirs, springs and groundwater wells. In 2006, the US Environmental Protection Agency (EPA) published the Groundwater Rule (GWR) (US EPA 2006*c*) which applies to all community and non-community PWS that use groundwater as a water source. An estimated 147,330 PWS in the United States, serving over 114 million people, use groundwater as their primary water source (US EPA 2006*c*). Unlike surface waters, prior regulations in the US did not require filtration or disinfection of groundwater sources. If pathogenic protozoa or other parasitic organisms such as helminths are found in groundwater PWS supplies, those PWS are regulated as groundwater under the direct influence of surface water under the Long Term 2 Enhanced Surface Water Treatment Rule (LT2ESWTR) (US EPA 2006*a*). Thus, the GWR addresses the human risks of illness due to non-parasitic faecal contamination of groundwater.

doi: 10.2166/wh.2009.018

### Microbial risk assessment

Microbial risk assessment (MRA) is a process that is used to evaluate the likelihood of adverse human health effects that can occur following exposure to pathogenic microorganisms or to a medium in which pathogens are present (ILSI 1996). The risk characterization method employed by the base analysis in the GWR is based on an individual-level (also known as a static) model (US EPA 2006*b*) (Figure 1). As shown in Figure 1, the predicted intensity of person-to-person (secondary) transmission is computed as the product of the number of primary illnesses and a static secondary morbidity factor.

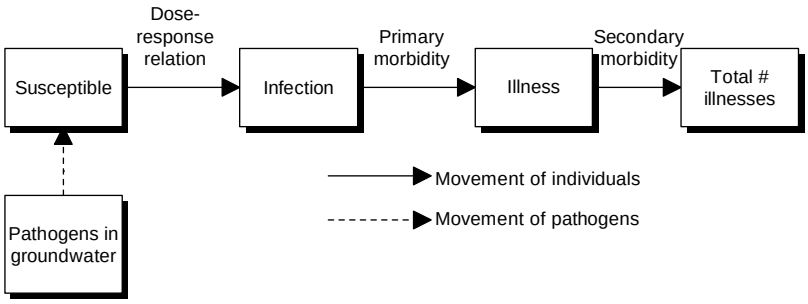


**Figure 1** | Conceptual model for the GWR base analysis.

## Motivation

The limitations of treating infectious disease transmission as a static disease process with no interaction between those infected or diseased and those at risk has been illustrated for various infectious diseases including giardiasis (Eisenberg *et al.* 1996), dengue fever (Koopman *et al.* 1991*b*) and sexually transmitted diseases (Koopman *et al.* 1991*a*). Further, a variety of model forms can be employed to characterize infectious disease transmission. In this context, a model form is a mathematical representation of the epidemiological status of the population together with rules that define the movement of individuals and pathogens among sub-populations with defined characteristics (e.g. infected individuals, ill individuals, immune individuals). Particular characteristics of each model form capture different aspects of the disease transmission system. However, it is unrealistic to presume that one model form is most appropriate for all waterborne microbial risk assessments (Soller & Eisenberg 2008).

This manuscript summarizes how a dynamic infectious disease model was used (Soller 2006; US EPA 2006*d*) to explore the potential implications of secondary transmission of infection and subsequent illness in the GWR relative to the static method employed by the base analysis (US EPA 2006*b*). For this work, secondary transmission includes infections due to both person-to-person contacts and person-to-environment-to-person contacts.

## METHODS

Potential implications associated with secondary transmission of infection and subsequent illness in the GWR are evaluated via a hypothetical case study. In this case study, it is assumed that a large tour group from a total population of 100,000 visit an outlying area that is served by a non-community water system (NCWS) with untreated or inadequately treated groundwater that is contaminated with a highly infectious virus that causes adverse health effects of low severity (type A virus). It is assumed that 100 individuals become infected and then return home to the larger community that is assumed to be served by a community water system. The population is then assumed to mix in a spatially and temporally homogeneous manner.

In the GWR base analysis, rotavirus is employed as the prototype type A virus (US EPA 2006*b*). Noroviruses are also described as epidemiologically important type A viruses (US EPA 2006*a*,*c*); however, illnesses from norovirus infection are not quantified in the GWR. Two types of analysis were conducted in this investigation, a median value analysis and a sensitivity/uncertainty analysis. In the median value analysis, the model type A virus employed is assumed to have the clinical properties and infectivity of rotavirus. However, it is assumed that age is not an important factor relative to infection and that individuals of all ages are equally likely to become infected and subsequently propagate infection via person-to-person transmission. The analysis is conducted in this manner to explore potentially important characteristics of type A viruses as a class, as opposed to rotavirus for which age structure is relevant. The sensitivity/uncertainty analysis (referred to as uncertainty analysis hereafter) explores how uncertainty in model parameters and the model form influence the potential insights provided by the median value analysis and considers how those insights may change for other type A viruses, such as noroviruses,

with somewhat different clinical or environmental properties.

Numerical simulations are used to estimate the number of additional infections and illnesses due to person-to-person transmission within the large community after the group of infected individuals returns home. Cumulatively, the results of the median value and uncertainty analyses are used to identify conditions under which the predicted magnitude of person-to-person transmission of illness in the community is substantially different from that predicted under the GWR base analysis.

## Overview of scenario evaluated

The case study scenario comprises the following assumptions. A hypothetical community with a population of 100,000 individuals is considered. From that population, a tour group is assumed to visit an area served by a NCWS that is contaminated with a type A virus. One hundred members of the tour group become infected and then return home. The type A virus has the infectivity of rotavirus among adults in the US population, the background incidence levels of rotavirus in the US population and the morbidity of rotavirus among adults and children in the US population based on weighted population averages. Secondary transmission is homogeneous among individuals of all ages within the larger population. The median value analysis is based on community level data describing the proportion of household level infections in the community. Incubation and clinical severity are based on unweighted data on rotavirus from children and adults. Immunity data are based primarily on rotavirus data from children but informed by a qualitative assessment of adults. None of the infected individuals would be considered 'super-spreaders'. Immunocompromised individuals are not considered separately from the rest of the population.

## Risk assessment methodology

In the base analysis for the GWR, the magnitude of secondary transmission is estimated by multiplying the number of children younger than 3 years who are ill by a constant secondary morbidity factor (US EPA 2006*b*). In this investigation, secondary transmission is estimated using a dynamic infectious disease methodology. The numerical modelling approach derives from the use of dynamic population models in the study of epidemics (Hethcote 1976, 2000; Anderson & May 1991) and environmental processes and disease (Koopman *et al.* 1991*a*, 2001, 2002). Furthermore, the microbial risk assessment approach builds on previous studies addressing human health risks from exposure to waterborne pathogens (Eisenberg *et al.* 1996, 1998; Soller *et al.* 1999, 2003, 2004, 2006) and is consistent with the EPA/ILSI (International Life Sciences Institute) framework for microbial risk assessment (ILSI 1996). Two routes of transmission are accounted for in this investigation: primary transmission by environmental exposure and secondary transmission via person-to-person transmission. The conceptual model for health effects associated with exposure to the model type A virus is presented in Figure 2.

The model is composed of five state variables that are used to track the number of individuals in each epidemiological state over time: S–individuals susceptible to infection (Susceptible); E–individuals who have been exposed but have yet to become infectious (Exposed); C–infectious but asymptomatic individuals (Carrier); D–infectious and symptomatic individuals (Diseased); and P–post-symptomatic and non-infectious individuals (with limited term protection from infection) (Protected). Eleven model parameters are used to define the model (Tables 1 and 2).

Assuming that the population is large and that primary and secondary transmission processes are independent (Hethcote 1976; Anderson & May 1991), the change in the fraction of the population in any state from one time period

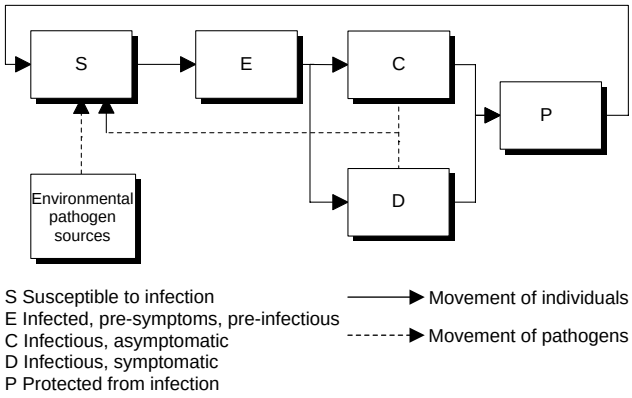


**Figure 2** | Conceptual health effects model.

**Table 1** | Model parameters derived directly from scientific literature

| Model parameter | Description | Comments |
|---|---|---|
| Background incidence | Incidence of illness in US | |
| $\alpha$ | Beta Poisson does response parameter | Maximum liklihood estimate fit to beta-Poisson |
| $\beta_{dr}$ | Beta Poisson does response parameter | Maximum liklihood estimate fit to beta-Poisson |
| $\zeta$ | Inverse of duration of incubation (days) | Inverse represent rate of movement out of state E |
| $P_{sym}$ | Probability of symptomatic response | |
| $\sigma$ | Inverse of duration of asymptomatic infection (days) | Inverse represent rate of movement out of state C |
| $\delta$ | Inverse of duration of symptomatic infection (days) | Inverse represent rate of movement out of state D |
| $\gamma$ | Inverse of duration of protection from infection (days) | Inverse represent rate of movement out of state P |

to the next may be modelled as a series of first order differential equations (Table 3). The model is implemented via numerical simulation and integration using MathCad version 13 (Mathsoft, Inc.). The model parameter values used in the analyses were determined through literature review and are summarized in Table 4. A description of that review is available separately (Soller 2006; US EPA 2006*d*) and a brief summary of the parameter value selection basis is provided in the Appendix.

## Numerical simulation methodology

Numerical simulation was used to determine the number of individuals in each epidemiological state over time. The first step is a calibration that is used to identify the proportion of individuals in each state for endemic conditions and to find appropriate values for the model parameters. At the inception of each calibration simulation, it is assumed that 95% of the population is in state S (susceptible) and 5% is in state E (exposed) (Soller *et al.* 1999, 2006). The parameter values shown in Table 4 are used to conduct the median value analysis calibration and subsequently the median value analysis. Using those values with $\beta$pp set to zero, an 'endemic dose' (and resulting value for $\beta$) is found that results in a proportion of the population in state D (diseased) that is consistent with the background level of illness incidence in the United States and the assumption that 20% of cases are due to person-to-person transmission (Koopman *et al.* 1989). Using those values for $\beta$ and 'endemic dose', a value of $\beta$pp is then identified so that the incidence of illness in the population was consistent with the background level shown in Table 4.

After the calibration is complete the median value analysis is conducted. The median value analysis is used to assess the magnitude of illness attributable to the 100 individuals returning with type A virus infections after visiting a contaminated NCWS, assuming that all parameter values are set equal to the median values of those reported in the literature. Mathematically, the simulation is conducted in a similar manner to that described above for

**Table 2** | Model parameters that are computed based on other parameters

| Mode parameter | Dependent variables | Description | Comments |
|---|---|---|---|
| End_dose | All other model parameters | Background level of pathogen that results in a disease prevalence consistent with estimated levels in US | End_dose is derived for each simulation. It is the dose results in an average prevalence consistent with a specified level of endemic disease |
| $\beta$ | End_dose, $\alpha$, $\beta_{dr}$ | Probability of Infection from environmental exposure | $1 - \exp\left(-\text{end\_dose}\cdot\frac{\alpha}{\alpha+\beta_{dr}}\right)$ |
| $\beta$_pp | All other model parameters | Probability of infective contact due to person-to-person exposure | $\beta$_pp is derived for each simulation. It is a rate that results in a specified level of endemic disease due to secondary transmission |

**Table 3** | First order differential equations used for MRA modelling

$$\frac{d}{dt}S(t) = -\beta \cdot S(t) - \beta\_pp \cdot S(t) \cdot (C(t) + D(t)) + \gamma \cdot P(t)$$
$$\frac{d}{dt}E(t) = \beta \cdot S(t) + \beta\_pp \cdot S(t) \cdot (C(t) + D(t)) - \xi \cdot E(t)$$
$$\frac{d}{dt}C(t) = \zeta \cdot (1 - P_{sym}) \cdot E(t) - \sigma \cdot C(t)$$
$$\frac{d}{dt}D(t) = \zeta \cdot P_{sym} \cdot E(t) - \delta \cdot D(t)$$
$$\frac{d}{dt}P(t) = \delta \cdot D(t) + \sigma \cdot C(t) - \gamma \cdot P(t)$$

the median value calibration except that 1) the initial conditions are set equal to the steady state conditions identified in the calibration including values for $\beta$ and $\beta$pp and then 2) 100 people are removed from state S (susceptible) and moved into state E (exposed). (Simulation files are available from the author on request.)

Similar simulations are employed to conduct the uncertainty analysis, except different parameter values or model forms are used. For each simulation conducted for the uncertainty analysis, a calibration simulation is performed as described above, followed by a simulation representing the case study.

## Uncertainty analysis

The goal of the uncertainty analysis is to consider how uncertainty in the model form and parameter values could impact the model output and the subsequent interpretation of the results. The uncertainty analysis plan encompasses four broad categories: model parameter uncertainty issues; transmission parameter uncertainty issues; model form uncertainty issues; and population uncertainty issues.

### Model parameter uncertainty

The first model parameter to be considered is the background level of illness. This parameter is important in determining the status of the population under endemic conditions. Specifically, the calibration step is used to determine an 'endemic dose' and resulting values for $\beta$ and $\beta$pp that result in a proportion of the population in state D (diseased) that is consistent with the background level of illness incidence. The background level of illness may be substantially different from the value estimated using the median value for a number of reasons, including: 1) the background level of rotavirus illness in the US is uncertain; 2) the values presented are the best available estimates; however, it is possible that the true incidence could be higher or lower than that estimated; 3) the background level of illness varies between communities; 4) other type A viruses may have substantially different illness incidence levels from that investigated in the median value analysis; and 5) the general classification of type A viruses includes noroviruses. Noroviruses are estimated to cause 23,000,000 illnesses annually in the United States (Mead *et al.* 1999). However, noroviruses as a class are a genetically and antigenically diverse group of viruses (Ando *et al.* 2000).

**Table 4** | Parameter values used in simulations

| Primary variables | Description | Median value or duration | Minimum and maximum reasonable range | Median rate* |
|---|---|---|---|---|
| Background incidence | Incidence of illness in US | 3,900,000 | 500,000–23,000,000 | |
| $\alpha$ | Beta Poisson dose response parameter | 0.26 | 0.126–0.52 | |
| $\beta_{dr}$ | Beta Poisson dose response parameter | 0.42 | 0.21–0.84 | |
| $\zeta$ | Inverse of duration of incubation (1/days) | 2.5 | 1–5 days | 0.4 |
| $P_{sym}$ | Probability of symptomatic response | 0.45 | 0.1–0.6 | |
| $\sigma$ | Inverse of duration of asymptomatic infection (1/days) | 5 | 2–8 days | 0.2 |
| $\delta$ | Inverse of duration of symptomatic infection (1/days) | 6 | 2–8 days | 0.17 |
| $\gamma$ | Inverse of duration of protection from infection (1/days) | 548 | 7–30 months | 0.0018 |

*The rates shown are the inverse of the median values and correspond to the mean amount of time spent in the corresponding states.

Furthermore, the potential for norovirus cross-strain immunity is not well understood.

To explore how the background level of illness could have an impact on the insights provided by this analysis, simulations were conducted in which the endemic level of illness in the US population varied between 50,000 and 23,000,000. In each of these simulations, it is assumed that $\beta$pp (the probability of an infective contact due to person-to-person exposure) is not affected by changing the background level of disease in the community.

There are also seven other model parameters that affect the rate at which individuals move between epidemiological states. Those parameters are: $\zeta$, $\sigma$, $\delta$, $\gamma$, $\alpha$, $\beta\_{dr}$ and $P_{sym}$. Based on the peer-reviewed literature, there is substantial uncertainty in each of these values. To determine how the uncertainty in these parameter values affects the potential propagation of secondary transmission of illness, the minimum and maximum reasonable values are explored for each variable. Each unique combination of these minimum and maximum values was simulated (resulting in 128 unique simulations). Minimum and maximum reasonable values for $\alpha$ and $\beta\_{dr}$ are the lower and upper 95% confidence values of the maximum likelihood fits for the beta-Poisson dose response function and the minimum and maximum reasonable values for $P_{sym}$, $\zeta$, $\sigma$, $\delta$ and $\gamma$ were based on data from the literature review. In these analyses, the background incidence of illnesses is assumed to be the same as in the median value analysis.

### Transmission parameter uncertainty

The two transmission parameters in the analysis are $\beta$ (also referred to as $\beta$env). To be consistent with the base analysis in the GWR, $\beta_{env}$ was computed using the following formula which is an approximation to the beta-Poisson model (US EPA 2006*b*):

$$\beta_{env} := 1 - \exp\left(\text{dose}\cdot\frac{\alpha}{\alpha+\beta_{dr}}\right) \quad (1)$$

Approximations to the beta-Poisson model are commonly employed in microbial risk assessment investigations because the exact form is complex and computationally intensive (Teunis *et al.* 1996). There are, however, other approximations to the beta-Poisson function that are also reported and used in the literature, the most common of which are the following:

$$\beta_{env} := 1 - \left(1 - \frac{\text{dose}}{\beta_{dr}}\right)^{-\alpha} \quad (2)$$

$$\beta_{env} := \text{dose}\cdot\left[\frac{\alpha}{(\alpha+\beta_{dr}}\right] \quad (3)$$

To determine if the functional form of the dose response model has a substantial impact on the interpretation of this analysis, additional analyses are conducted using the dose response model forms described above.

The approach employed for estimating the proportion of person-to-person transmission in a community is to solve the system of differential equations (Table 3) for $\beta_{pp}$ so that the results are consistent with those reported in the Tecumseh study (Koopman *et al.* 1989). To determine how the uncertainty in $\beta_{pp}$ impacts the propagation of secondary transmission of illness for the case study scenario, person-to-person transmission intensity within the minimum (10%) and maximum (60%) reasonable values for $\beta_{pp}$ are evaluated in the uncertainty analysis.

### Model form uncertainty issues

The conceptual model used in the median value analysis is similar to conceptual models that have been used in other peer-reviewed studies to investigate health effects associated with waterborne pathogens (Eisenberg *et al.* 1996, 1998, 2004; Soller *et al.* 2003, 2004, 2006). However, a variety of model forms can be employed to characterize infectious disease transmission and to evaluate the potential for effective interventions. The selection of a model involves trade-offs. Biological or demographic 'realism' can be achieved, but often comes at the cost of analytical complexity and/or unrealistic data requirements (US EPA 2004). To gain insight into how specific model forms impact the predicted magnitude of secondary transmission of illness, two alternative model forms were evaluated through simulation using a median value analysis. Relative to the model shown in Figure 2, the first model adds a carrier (infectious and asymptomatic) state after the diseased (infectious and symptomatic) state and prior to the state

which provides protection from infection. This modification accounts for the shedding of pathogens after symptoms subside. This characteristic has been reported in outbreak literature (Pickering *et al.* 1988; Katz *et al.* 2006) and clearly has the potential to affect the predicted propagation of person-to-person transmission of illness in a community.

In the second model, movement within states E, C, D and P is characterized by a distributed delay process (Eisenberg *et al.* 2004) and, thus, the time that individuals spend in each these states is described by a gamma distribution. In the model presented in Figure 2, the time that individuals spend in each of these states is characterized by an exponential distribution. The two parameters describing the gamma distribution are the number of sub-states per state (4) and the rate constant governing the movement between sub-states. This model form was selected for evaluation based on models in peer-reviewed literature (Eisenberg *et al.* 1996) and the facts that incubation period is reasonably well characterized by a gamma distribution (Brookhart *et al.* 2002; Eisenberg *et al.* 2005) and that zero latency which is likely in the exponential model is not very plausible.

### Population uncertainty issues

One important assumption employed in this analysis is that the type A virus under investigation does not cause infection in any particular portion of the population differently from any other (for example the model does not differentiate between children and adults). The purpose of this component of the uncertainty analysis is to determine how sensitive the predicted number of secondary cases in the community is to this assumption.

The type A virus described in the GWR is meant to represent highly infectious viruses with low severity of illness. In considering rotavirus and noroviruses, both of which are considered type A viruses and account for a large percentage of illnesses from known pathogens (Mead *et al.* 1999), some portion of the population is likely to be either semi-permanently immune or genetically immune (Lindesmith *et al.* 2003; Anderson & Weber 2004).

To evaluate the uncertainty associated with portions of the population never moving into a 'susceptible' state, a series of simulations are conducted in which 10, 25, 50 and 75% of the population never move out of state P (protected).

## RESULTS

The results of the median value analysis and the corresponding uncertainty analyses summarize the output from over 300 individual simulations including the required calibration simulations.

### Median value analysis

The median value calibration (not shown) indicated that a steady-state condition is reached relatively quickly in the simulation (about 60 days) and that the vast majority of the population is in state S (susceptible) (95.2%) under endemic conditions. For a population size of 100,000 individuals, the calibration also indicates that, under endemic conditions, approximately 22–24 individuals would be in states E, C and D and 4,728 individuals would be in state P (protected) (Soller 2006).

The dynamic nature of the model under the case study scenario (for median parameter values) is illustrated in Figure 3 for state D (diseased) and compared with endemic conditions. As illustrated, endemic steady state conditions are re-established approximately 50 days after the infected individuals return home. The number of illnesses attributable to the case study is obtained as the difference of the areas under the curves shown in Figure 3 divided by the average duration of symptomatic illness. Mathematically, the number of illnesses attributable to the case study is computed as follows:

$$\text{Avg\_Attrib\_Inc} := \left[\frac{\int_0^T (\text{D\_gwc}(t) - \text{D\_end}(t))\,dt}{\text{delt\_avg}}\right] \quad (4)$$

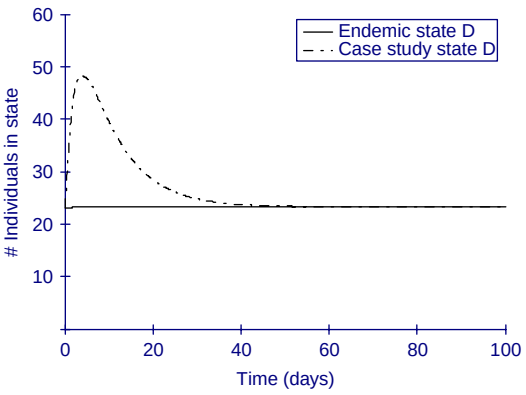


**Figure 3** | Median value analysis compared with endemic conditions.

where:

D_gwc is the number of individuals in state D under the case study scenario; D_end is the number of individuals in state D under the endemic conditions; and delt_avg is the average duration of symptomatic illness.

The results of the median value analysis are summarized in Table 5 and indicate that approximately 53 illnesses were attributable to the case study: 45 of those illnesses were from primary exposure (the NCWS) and 8 from subsequent person-to-person transmission. Based on these results there were, on average, an additional 0.18 illnesses attributable to person-to-person transmission for each ill individual returning to the community. Alternatively, the number of additional illnesses can be computed based on the number of infected individuals returning to the community, in which case, there were on average an additional 0.08 illnesses attributable to person-to-person transmission (for each infected individual returning to the community).

## Uncertainty analysis

### Background incidence

The results of the uncertainty analysis for background level of illness are presented in Table 6. Referring to Table 6, the magnitude of person-to-person transmission attributable to the case study is inversely related to the background level of disease in the community. For example, at an average background incidence level of 50,000 illnesses annually in a population of 275,000,000, ~57 illnesses were attributable to the case study, 12 of which were due to person-to-person transmission (representing a 50% increase over the median value analysis). These results indicate an average additional 0.27 illnesses attributable to person-to-person transmission for each ill individual returning to the community, compared with 0.18 additional illnesses for each ill individual returning to the community in the median value analysis. Similarly, it can be seen that, as the background incidence level increases, the number of illnesses attributable to the case study decreases.

From the data presented in Table 6, it is apparent that the proportion of individuals in state S (susceptible) decreases as the endemic level of disease increases. Thus, it is likely that the inverse relation between attributable illness levels and background levels of disease is due to decreasing numbers of successful contacts with individuals in state S as the endemic level of disease increases.

**Table 5** | Median value analysis results

| Endemic level of disease in US | 3,900,000 |
|---|---|
| Incidence per 100,000/year | 1,418 |
| Proportion in state S | 0.95204 |
| Proportion in state E | 0.00022 |
| Proportion in state C | 0.00024 |
| Proportion in state D | 0.00023 |
| Proportion in state P | 0.04728 |
| Attributable number of illnesses | 53.2 |
| # Illnesses per primary illness | 1.18 |
| # Illnesses per primary infection | 0.53 |

### Other literature-derived model parameters

The results of the uncertainty analysis for other literature-derived model parameters indicated that the attributable number of illnesses varied between approximately 7 and 75 and the number of illnesses per primary illness varied between approximately 0.7 and 1.25 for these simulations. Recall that the results of the median value analysis were an attributable number of illnesses of 53 and a corresponding 1.18 illnesses per primary illness.

For the attributable number of illnesses, the results are explained by the probability of symptomatic response (the probability of symptomatic response was 0.1 in 50% of simulations and 0.6 in the remaining 50%). That is, in the simulations in which the probability of symptomatic response was low, the attributable number of illnesses was also low and vice versa. The interpretation of the number of illnesses per primary illness is more complex. However, the combination of the probability of symptomatic response and the duration of immunity seem to interact to yield the lowest and highest values observed. Specifically, when the probability of symptomatic response was high and the duration of immunity was short, the highest values of illnesses per primary illness were observed. Similarly, when the probability of symptomatic response was low and the

**Table 6** | Uncertainty analysis for background level of illness

| | Endemic level of disease in US | | | | | |
|---|---|---|---|---|---|---|
| | 50,000 | 500,000 | 1,500,000 | 3,900,000 | 10,000,000 | 23,000,000 |
| Incidence per 100,000/year | 18 | 182 | 545 | 1,418 | 3,636 | 8,364 |
| Steady state conditions | | | | | | |
| Proportion in state S | 0.99930 | 0.99378 | 0.9815 | 0.95203 | 0.8771 | 0.7172 |
| Proportion in state E | 0.0000028 | 0.000028 | 0.000083 | 0.00022 | 0.00056 | 0.00127 |
| Proportion in state C | 0.0000030 | 0.000030 | 0.000091 | 0.00024 | 0.00061 | 0.0014 |
| Proportion in state D | 0.0000030 | 0.000030 | 0.000090 | 0.00023 | 0.00060 | 0.00138 |
| Proportion in state P | 0.0006875 | 0.00613 | 0.01823 | 0.04728 | 0.12119 | 0.27880 |
| Results of perturbed steady state | | | | | | |
| Attributable number of illnesses | 57.3 | 56.79 | 55.74 | 53.2 | 47.4 | 36.27 |
| # Illnesses per incident illness | 1.27 | 1.26 | 1.24 | 1.18 | 1.05 | 0.81 |
| # Illnesses per incident infection | 0.57 | 0.57 | 0.56 | 0.53 | 0.47 | 0.36 |

Notes: Proportion ill based on duration of symptoms of 6 days (average symptomatic duration) In these results the probability of symptomatic response is 0.45.

duration of immunity was long, the lowest values of illnesses per primary illness were observed.

### Transmission parameter uncertainty

Two 'variants' were run for each dose response model form investigated. In variant 1, $\beta$pp was kept constant relative to the median value analysis and the 'endemic dose' was varied to ensure that the incidence was consistent with the median value analysis. In variant 2, $\beta$pp and 'endemic dose' were kept constant relative to the median value analysis and the incidence was allowed to vary. Based on the simulation results, the magnitude of person-to-person illness propagation in the case study is not related in a substantial way to the form of the dose response function, for the functional forms evaluated.

The results of the uncertainty analysis for the proportion of person-to-person transmission in a community are presented in Table 7. As shown, the proportion of person-to-person transmission does have an impact on the magnitude of the incidence attributable to the case study. For example, if person-to-person transmission accounted for 10% of the illness rather than 20%, the estimated number of illnesses due to person-to-person transmission decreases from 8 to 2 and, thus, the number of illnesses per primary illness decreases from 1.18 to 1.05. Similarly, if person-to-person transmission accounts for 60% of the illness in the community, the estimated number of illnesses from person-to-person transmission increases from 8 to 57 and, thus, the number of illnesses per primary illness increases to 2.27. In interpreting the results presented in this section it should be noted that a static model would also predict increased attributable illnesses as the proportion of infections due to person-to-person transmission increases. Thus, the results observed in this portion of the uncertainty analysis are not necessarily due to the fact that a dynamic model was employed, but rather may be due to the particular scenario investigated.

### Model form uncertainty

The results of the simulations with the distributed delay model were similar to those for the median value analysis, and the attributable number of illnesses for the model employing a post-disease carrier state is slightly higher than those for the median value analysis for both variants investigated. The observed increased attributable number of illnesses for the model with a post-disease carrier state is probably due to the fact that there are individuals that are shedding pathogens for a longer period of time and thus the potential for person-to-person transmission is increased.

**Table 7** | Uncertainty analysis for person-to-person transmission intensity

| | Median value analysis | | | |
|---|---|---|---|---|
| **P-P transmission %** | **10%** | **20%** | **40%** | **60%** |
| Steady state conditions | | | | |
| Proportion in state S | 0.95205 | 0.95203 | 0.95201 | 0.95200 |
| Proportion in state E | 0.00022 | 0.00022 | 0.00022 | 0.00022 |
| Proportion in state C | 0.00024 | 0.00024 | 0.00024 | 0.00024 |
| Proportion in state D | 0.00023 | 0.00023 | 0.00023 | 0.00023 |
| Proportion in state P | 0.04727 | 0.04728 | 0.04731 | 0.04732 |
| Results of perturbed steady state | | | | |
| Attributable number of illnesses | 47.4 | 53.2 | 70.4 | 102.4 |
| # Illnesses per incident illness | 1.05 | 1.18 | 1.56 | 2.27 |

### Population issues uncertainty

The results of the uncertainty analysis for population uncertainty issues are presented in Table 8. As shown, two 'variants' were run for each proportion of the population that was permanently removed from state S (susceptible). In variant 1, the 'endemic dose' was kept constant relative to the median value analysis and $\beta$pp was varied to ensure that the incidence was consistent with the specified level (same as the median value analysis). In variant 2, $\beta$pp was kept constant relative to the median value analysis and the 'endemic dose' and was varied to ensure that the incidence was consistent with the specified level. In both variants, the total incidence rate is the same as in the median value analysis; therefore, the endemic incidence in the *genetically susceptible* part of the population needs to be progressively greater as this portion of the population becomes smaller.

Variant 1 effectively keeps constant the number of pathogens to which individuals are exposed from the environment. Thus, as the number of individuals in state S decreases, the intensity of person-to-person transmission must increase for a given level of incidence. Similarly for variant 2, with a constant probability of infective contact ($\beta$pp), the endemic dose must increase as the number of individuals in state S decreases for a given level of incidence. It is not known which of these two processes is more likely to represent a real world situation (or whether it is a combination thereof).

Based on the results presented in Table 8, the number of illnesses attributable to the case study increases under variant 1 assumptions rather dramatically. For example, if 25% of the population is permanently removed from state S (susceptible), the estimated number of illnesses due to person-to-person transmission increases from 8 to 24 and the number of illnesses per primary illness increases from 1.18 to 1.53. This type of analysis is intended to simulate the population dynamics that may be occurring in the general population for noroviruses (Lindesmith *et al.* 2003) or in the adult population for rotavirus (Anderson & Weber 2004). Similarly, if 50% of the population were permanently removed from state S, the estimated number of illnesses due to person-to-person transmission increases from 8 to 48 and the number of illnesses per primary illness increases from 1.18 to 2.06 relative to the median value analysis.

One subtle but important point associated with this portion of the uncertainty analysis is that these simulations involved a constant number of infected individuals regardless of the proportion of susceptible individuals in the population. As the genetically susceptible portion of the population becomes smaller, this scenario requires more individuals to be exposed. Thus, caution is needed in interpreting these results.

## DISCUSSION

### Interpretation and insights gained from analysis

An infectious disease paradigm was employed to evaluate the potential implications of secondary transmission of infection and illness in the GWR relative to the intensity of secondary transmission estimated using the GWR base analysis. The

**Table 8** | Uncertainty analysis for population uncertainty

| | Median value analysis | Variant 1 | Variant 2 | Variant 1 | Variant 2 | Variant 1 | Variant 2 | Variant 1 | Variant 2 |
|---|---|---|---|---|---|---|---|---|---|
| **Percent of population always in state p** | **0%** | **10%** | **10%** | **25%** | **25%** | **50%** | **50%** | **75%** | **75%** |
| Steady state conditions | | | | | | | | | |
| Proportion in state S | 0.95203 | 0.85202 | 0.85204 | 0.70202 | 0.70205 | 0.45204 | 0.45204 | 0.20205 | 0.20204 |
| Proportion in state E | 0.00022 | 0.00022 | 0.00022 | 0.00022 | 0.00022 | 0.00022 | 0.00022 | 0.00022 | 0.00022 |
| Proportion in state C | 0.00024 | 0.00024 | 0.00024 | 0.00024 | 0.00024 | 0.00024 | 0.00024 | 0.00024 | 0.00024 |
| Proportion in state D | 0.00023 | 0.00023 | 0.00023 | 0.00023 | 0.00023 | 0.00023 | 0.00023 | 0.00023 | 0.00023 |
| Proportion in state P | 0.04728 | 0.14730 | 0.14727 | 0.29730 | 0.29727 | 0.54728 | 0.54728 | 0.79726 | 0.79727 |
| Results of perturbed steady state | | | | | | | | | |
| Attributable number of illnesses | 53.2 | 58.7 | 51.6 | 68.9 | 49.1 | 92.9 | 44.6 | 106.7 | 37.6 |
| # Illnesses per incident illness | 1.18 | 1.30 | 1.15 | 1.53 | 1.09 | 2.06 | 0.99 | 2.37 | 0.84 |
| # Illnesses per incident infection | 0.53 | 0.59 | 0.52 | 0.69 | 0.49 | 0.93 | 0.45 | 1.07 | 0.38 |

evaluation was conducted using a hypothetical case study scenario of a type A virus in a community with a population size of 100,000.

The predicted magnitude of person-to-person transmission of illness in the community was reported and discussed in terms of the number of illnesses per primary illness based on a number of individuals from that community visiting a NCWS contaminated with a type A virus. The number of illnesses per primary illness is a normalized metric to characterize person-to-person transmission intensity. It is robust to other formulations of the case study provided that the ratio of infected individuals to the total population is small. A series of sensitivity analyses (not shown) were conducted to verify that the normalized results are robust within the range of 0.001% to 0.1% of the population assumed to be infected.

Because the number of illnesses per primary illness is robust to other case study scenarios and is, thus, scalable, it is a useful metric for providing insight into the expected implications for other situations in which either more or less individuals are exposed to or infected with type A viruses from untreated or inadequately treated groundwater. For example, another scenario relevant to the GWR may be one in which low numbers of individuals return to a large community infected with a type A virus after visiting an area served by a NCWS with untreated or inadequately treated groundwater. Based on these results, the average number of illnesses per primary illness can be used to predict the expected magnitude of person-to-person transmission from one or more event of this type.

From the results of the uncertainty analysis it was found that several parameters and assumptions have the potential to substantially influence the magnitude of disease propagation predicted by the median value analysis for the case study investigated. For example, the probability of symptomatic response and duration of immunity were both found to be important factors in predicting the propagation of illness via person-to-person transmission. Similarly, transmission parameter and population issue uncertainties were found to strongly affect the predicted magnitude of disease propagation.

One of the most important differences between the assumptions employed in this investigation and those employed in the GWR base analysis was the relaxation of the assumption that only children are sources of secondary

infections. In the GWR base analysis, primary waterborne illnesses in young children are multiplied by a secondary transmission factor to estimate the number of secondary cases in individuals of all ages. In the analysis described herein, it is assumed that individuals of all ages are equally likely to become infected and subsequently propagate infection via person-to-person transmission. By relaxing the GWR assumption and deriving a different multiplicative factor for secondary spread based on the epidemiological data reported by Koopman *et al.* (1989), it is possible to illustrate the importance of this assumption. Those calculations (not shown) produce similar, albeit slightly higher estimates of person-to-person transmission than those derived with the dynamic model. The assumption that individuals of all ages are equally likely to become infected and subsequently propagate infection via person-to-person transmission, coupled with a simple epidemiologically based multiplicative factor tends to systematically overestimate the total number of illnesses attributable to NCWS outbreaks relative to the dynamic model. In this case, the dynamic model accounts for temporary removal of the primary infections (and all secondary infection cases that they generated) from the pool of susceptible individuals. By way of contrast, both the dynamic model and the simple epidemiologically based multiplicative factor model with the relaxed assumption about sources of secondary infection, both predict substantially higher person-to-person transmission than the GWR base analysis, with specific results depending on the parameter values employed.

This analysis was intended to complement the base analysis in the GWR. In that context, this analysis provides a perspective on the relative magnitude of uncertainty that may be associated with the benefits of the GWR due to secondary transmission of infection. Although the GWR base analysis and this work both focus on characterizing the risk associated with illness from groundwater sources, each analysis is built on a different set of assumptions. To the extent feasible, similar parameter values and assumptions were applied in this analysis compared with those used in the base analysis. However, because the two analyses were intended to address slightly different questions, the results are not necessarily directly comparable.

This analysis employed a model virus that was intended to be representative of highly infectious viruses that have low severity of clinical illness (type A viruses). The model virus in the median value analysis was assumed to have the clinical properties and endemic level of illness of rotavirus. As discussed above, it was further assumed that no specific portion of the population was any more or less likely to be infected upon exposure to the model virus. Uncertainty analysis was used to explore the potential implications of type A virus characteristics different from those of the model virus employed. Of particular note and concern are noroviruses, which are estimated to cause approximately 23,000,000 illnesses in the United States annually (Mead *et al.* 1999) and are associated with up to 90% of the epidemic nonbacterial gastroenteritis worldwide (Lindesmith *et al.* 2003). Although noroviruses are highly infectious, volunteer studies have shown that some subjects remain uninfected even after challenges with high doses (Johnson *et al.* 1990; Matsui & Greenberg 2000). Recent research indicates that resistance to norovirus infection is multifactorial and that a substantial portion of the population (approximately 20%) may not be susceptible to infection at any point in time (Lindesmith *et al.* 2003).

Rigorous modelling of norovirus transmission is extremely difficult at the present time for a number of reasons including: 1) little is known about the potential for cross-strain immunity; and 2) the person-to-person transmission potential appears to be substantial based on outbreak data. Approximately 300 norovirus outbreaks were documented in the US between 1993 and 1999 and the genetic diversity of the noroviruses responsible for those outbreaks encompassed approximately 68 strains (Ando *et al.* 2000). Clearly, the transmission of norovirus infection in a community is much more complicated than that presented herein. Nevertheless, several salient properties of noroviruses were explored in the uncertainty analysis in an attempt to provide some perspective on the potential importance of secondary transmission for noroviruses with respect to the GWR.

## Limitations

Given the complexity of characterizing disease propagation from a case study such as that presented herein, a number of methodological assumptions were required. The limitations of the analyses presented generally stem from those assumptions.

One fundamental limitation of any analysis such as this one is that the results are applicable only for parameter

values within the range of those investigated. Although a detailed literature review was conducted and the data from the literature were evaluated carefully to identify appropriate parameter values for the investigation, it is possible that parameter values could be refined or changed based on future research.

A related assumption was that the epidemiological status of the population could be approximated reasonably with the relatively simple structure of the disease transmission model (Figure 2). Although several alternative model forms were also investigated, it is possible that other types of model could yield additional and/or alternative insights. For example: 1) 'super-spread events' may be important in characterizing disease transmission magnitude (Riley *et al.* 2003); 2) there may be different transmission rates of viruses within and outside of households (but within the same community); and 3) viral infection rates may vary by season. Any or all of these factors are worthy of consideration for future extensions of this or related work.

Other simplifying assumptions were made to streamline both the analysis and the interpretation of the results presented herein. The most important of those assumptions were that insights could be drawn from the one hypothetical case study scenario that was investigated and that the model type A virus, as constructed in the median value analysis and the uncertainty analyses, is representative of the type A viruses of concern for groundwater contamination scenarios.

## CONCLUSIONS

Secondary transmission of infection and subsequent illness has the potential to have a substantial impact on the predicted benefits of the GWR depending on the suite of population dynamic elements and assumptions employed. Potentially important factors include clinical disease character, aetiologic agent infectivity, population immunity and person-to-person transmission characteristics of the aetiologic agents of interest. Specific results indicate that: 1) the uncertainty bounding the magnitude of person-to-person transmission for type A virus illnesses is substantial; and 2) depending on the assumptions employed, the predicted number of additional illnesses due to secondary transmission could be greater than that predicted by the GWR base analysis by approximately an order of magnitude or could be as low as effectively zero. A similar analysis could also be conducted for other drinking water regulations and may or may not yield similar insights.

## ACKNOWLEDGEMENTS

Philip Berger, Stig Regli and Michael Messner provided insightful comments on the design of the investigation and interpretation of the results. Mary Rothermich and Andrey Egorov supplied a critical review of the draft manuscript and George Tchobanoglous contributed constructive suggestions and thought-provoking recommendations. This investigation was funded by the US EPA Office of Groundwater and Drinking Water under subcontract to the Cadmus Group, Inc. The views expressed in this article are those of the author and do not necessarily reflect the views or policies of the US Environmental Protection Agency.

## REFERENCES

Anderson, E. J. & Weber, S. G. 2004 Rotavirus infection in adults. *Lancet Infect. Dis.* **4**(2), 91–99.

Anderson, R. M. & May, R. 1991 *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, New York.

Ando, T., Noel, J. S. & Fankhauser, R. L. 2000 Genetic classification of 'Norwalk-like viruses'. *J. Infect. Dis.* **181**, S336–S348.

Barron-Romero, B. L., Barreda-Gonzalez, J., Doval-Ugalde, R., Zermeno-Eguia Liz, J. & Huerta-Pena, M. 1985 Asymptomatic rotavirus infections in day care centers. *J. Clin. Microbiol.* **22**(1), 116–118.

Bernstein, D. I., Sander, D. S., Smith, V. E., Schiff, G. M. & Ward, R. L. 1991 Protection from rotavirus reinfection: 2-year prospective study. *J. Infect. Dis.* **164**(N2), 277–283.

Brookhart, M. A., Hubbard, A. E., van der Laan, M. J., Colford, J. M. & Eisenberg, J. N. S. 2002 Statistical estimation of parameters in a disease transmission model: analysis of a *Cryptosporidium* outbreak. *Stat. Med.* **21**(23), 3627–3638.

Champsaur, H., Questiaux, E., Prevot, J., Henry-Amar, M., Goldszmidt, D., Bourjouane, M. & Bach, C. 1984 Rotavirus carriage, asymptomatic infection and disease in the first two years of life. I. virus shedding. *J. Infect. Dis.* **149**(5), 667–674.

Chiba, S., Nakata, S., Ukae, S. & Adachi, N. 1993 Virological and serological aspects of immune resistance to rotavirus gastroenteritis. *Clin. Infect. Dis.* **Suppl2**, S117–S121.

Eisenberg, J. N., Seto, E. Y. W., Olivieri, A. W. & Spear, R. C. 1996 Quantifying water pathogen risk in an epidemiological framework. *Risk Anal.* **16**(4), 549–563.

Eisenberg, J. N. S., Seto, E. Y. W., Colford, J. M., Olivieri, A. W. & Spear, R. C. 1998 An analysis of the Milwaukee cryptosporidiosis outbreak based on a dynamic model of the infection process. *Epidemiology* **9**(3), 255–263.

Eisenberg, J. N. S., Soller, J. A., Scott, J., Eisenberg, D. M. & Colford, J. M. 2004 A dynamic model to assess microbial health risks associated with beneficial uses of biosolids. *Risk Anal.* **24**(1), 221–236.

Eisenberg, J. N., Lei, X., Hubbard, A. H., Brookhart, M. A. & Colford, J. M., Jr 2005 The role of disease transmission and conferred immunity in outbreaks: analysis of the 1993 *Cryptosporidium* outbreak in Milwaukee, Wisconsin. *Am. J. Epidemiol.* **161**(1), 62–72.

Ferson, M., Streingfellow, S., McPhie, K., McIver, C. & Simos, A. 1997 Longitudunal study of rotavirus infection in child-care centres. *Paediatr. Child Health* **33**(2), 157–160.

Flewett, T. H. & Woode, G. N. 1978 The rotaviruses: brief review. *Arch. Virol.* **57**, 1–23.

Flewett, T. H., Bryden, A. S. & Davies, H. 1975 Epidemic viral enteritis in a long-stay children's ward. *Lancet* **1**(7897), 4–5.

Gomez-Barreto, J., Plamer, E. L., Nahmias, A. J. & Hatch, M. H. 1976 Acute enteritis associated with Reovirus-like agents. *J. Am. Med. Assoc.* **235**, 1857–1860.

Gurwith, M., Wenman, W., Hinde, D., Heltham, S. & Greenberg, H. 1981 A prospective study of rotavirus infection in infants and young children. *J. Infect. Dis.* **144**(3).

Haas, C. N., Rose, J. B. & Gerba, C. P. 1999 *Quantitative Microbial Risk Assessment*. John Wiley & Sons, New York.

Hethcote, H. 1976 Qualitative analyses of communicable disease models. *Math. Biosci.* **28**, 335–356.

Hethcote, H. W. 2000 The mathematics of infectious diseases. *Siam Rev.* **42**(4), 599–653.

ILSI 1996 A conceptual framework to assess the risks of human disease following exposure to pathogens. *Risk Anal.* **16**(6), 841–848.

Johnson, P. C., Mathewson, J. J., DuPont, H. L. & Greenberg, H. B. 1990 Multiple-challenge study of host susceptibility to Norwalk gastroenteritis in US adults. *J. Infect. Dis.* **161**(1), 18–21.

Katz, D. E., Heisey-Grove, D., Beach, M., Dicker, R. C. & Matyas, B. T. 2006 Prolonged outbreak of giardiasis with two modes of transmission. *Epidemiol. Infect.* **134**(5), 935–941.

Kim, H. W., Brandt, C. D., Kapikian, A. Z., Wyatt, R. G., Arrobio, J. O., Rodriguez, W. J., Chanock, R. M. & Parrott, R. H. 1977 Human reovirus-like agent infection: occurrence in adult contact of pediatric patients with gastreoenteritis. *J. Am. Med. Assoc.* **238**(5), 404–407.

Koopman, J. S. & Monto, A. S. 1989 The Tecumseh Study XV: Rotavirus infection and pathogenicity. *Am. J. Epidemiol.* **130**(4), 750–759.

Koopman, J. S., Monto, A. S. & Longini, I. M. 1989 The Tecumseh study XVI: family and community sources of rotavirus infection. *Am. J. Epidemiol.* **130**(4), 760–768.

Koopman, J. S., Longini, I. M., Jacquez, J. A. & Simon, C. P. 1991*a* Assessing risk factors for transmission of infection. *Am. J. Epidemiol.* **133**(12), 1199–1209.

Koopman, J. S., Prevots, D. R., Vaca Marin, M. A., Gomez Dantes, H., Zarate Aquino, M. L., Longini, I. M. Jr & Sepulveda Amor, J. 1991*b* Determinants and predictors of dengue infection in Mexico. *Am. J. Epidemiol.* **133**(11), 1168–1178.

Koopman, J. S., Jacquez, G. & Chick, S. E. 2001 New data and tools for integrating discrete and continuous population modeling strategies. *Ann. NY Acad. Sci.* **954**, 268–294.

Koopman, J. S., Chick, S. E., Simon, C. P., Riolo, C. S. & Jacquez, G. 2002 Stochastic effects on endemic infection levels of disseminating versus local contacts. *Math. Biosci.* **180**(Sp. Iss. SI), 49–71.

Lindesmith, L., Moe, C., Marionneau, S., Ruvoen, N., Jiang, X., Lindbland, L., Stewart, P., LePendu, J. & Baric, R. 2003 Human susceptibility and resistance to Norwalk virus infection. *Nat. Med.* **9**(5), 548–553.

Lycke, E., Blmber, J., Berg, G., Eriksson, A. & Madsen, L. 1978 Epidemic acute diarrhea in adults associated with infantile gastroenteritis virus. *Lancet* **2**(8098), 1056–1057.

Matsui, S. M. & Greenberg, H. B. 2000 Immunity to calicivirus infection. *J. Infect. Dis.* **181**(Suppl 2), S331–S335.

Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M. & Tauxe, R. V. 1999 Food related illness and death in the United States. *Emerg. Infect. Dis.* **5**(5), 607–625.

Pickering, L., Bartlett, A., Reves, R. & Morrow, A. 1988 Asymptomatic excretion of rotavirus before and after rotavirus diarrhea in children in day care centers. *J. Pediatr.* **112**(3), 361–365.

Regli, S., Rose, J. B., Haas, C. N. & Gerba, C. P. 1991 Modeling the risk from giardia and viruses in drinking-water. *J. Am. Water Works Assoc.* **83**(11), 76–84.

Riley, S., Fraser, C., Donnelly, C. A., Ghani, A. C., Abu-Raddad, L. J., Hedley, A. J., Leung, G. M., Ho, L. M., Lam, T. H., Thach, T. Q., Chau, P., Chan, K. P., Lo, S. V., Leung, P. Y., Tsang, T., Ho, W., Lee, K. H., Lau, E. M., Ferguson, N. M. & Anderson, R. M. 2003 Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science* **300**(5627), 1961–1966.

Rodriguez, W. J., Kim, H. W., Brandt, C. D., Yolken, R. H., Richard, M., Arrobio, J. O., Schwartz, R. H., Kapikiam, A. Z., Chanock, R. M. & Parrott, R. H. 1979 Common exposure outbreak of gastroenteritis due to type 2 rotavirus with high secondary attack rate within families. *J. Infect. Dis.* **140**(3), 353–357.

Rodriguez, W. J., Kim, H. W. & Brandt, C. D. 1987 Longitudinal study of rotavirus infection and gastroenteritis in families served by a pediatric medical practice: clinical and epidemiologic observations. *Pediatr. Infect. Dis. J.* **6**(2), 170–176.

Shepherd, R. W., Truslow, S., Walkersmith, J. A., Bird, R., Cutting, W., Darnell, R. & Barker, C. M. 1975 Infantile gastroenteritis: clinical-study of reovirus-like agent infection. *Lancet* **2**(7944), 1082–1084.

Soller, J. A. 2006 *Potential Implications of Population Dynamics and Secondary Transmission of Infection on the Benefits of the Groundwater Rule*. Prepared by Soller Environmental for US EPA Office of Ground Water and Drinking Water.

Soller, J. A. & Eisenberg, J. N. S. 2008 An evaluation of parsimony for microbial risk assessment models. *Environmetrics* **19**(1), 61–78.

Soller, J. A., Eisenberg, J. N. & Olivieri, A. W. 1999 *Evaluation of Pathogen Risk Assessment Framework*. Prepared by EOA Inc. and UC Berkeley for ILSI Risk Science Institute.

Soller, J. A., Olivieri, A., Crook, J., Parkin, R., Spear, R., Tchobanoglous, G. & Eisenberg, J. N. S. 2003 Risk-based approach to evaluate the public health benefit of additional wastewater treatment. *Environ. Sci. Technol.* **37**(9), 1882–1891.

Soller, J. A., Olivieri, A. W., Eisenberg, J. N. S., Sakaji, R. & Danielson, R. 2004 *Evaluation of Microbial Risk Assessment Techniques and Applications*. Water Environment Research Foundation Report 00-PUM-3.

Soller, J. A., Eisenberg, J., DeGeorge, J., Cooper, R., Tchobanoglous, G. & Olivieri, A. 2006 A public health evaluation of recreational water impairment. *J. Water Health* **4**(1), 1–19.

Teunis, P. F., van der Heijden, O. G., van der Giessen, J. W. B., Havelaar, A. H. 1996 *The Dose-Response Relation in Human Volunteers for Gastro-intestinal Pathogens*. RIVM Report. 284550002.

Tucker, A. W., Haddix, A. C., Bresee, J. S., Holman, R. C., Parashar, U. D. & Glass, R. I. 1998 Cost-effectiveness analysis of a rotavirus immunization program for the United States. *J. Am. Med. Assoc.* **279**(17), 1371–1376.

Tufvesson, B., Johnsson, T. & Persson, B. 1977 Family infections by reo-like virus: comparison of antibody titres by complement fixation and immunoelectroosmophoresis. *Scand. J. Infect. Dis.* **9**(4), 257–261.

US Census Bureau 2000 Population Estimates, http://www.census.gov/popest/estimates.php [accessed May 1, 2006].

US EPA 2004 *Developing Dynamic Infection Transmission Models for Microbial Risk Assessment Applications*. EPA-NCEA-C-1463.

US EPA 2006*a National Primary Drinking Water Regulations*: Long term 2 enhanced surface water treatment rule (LT2ESWTR); final rule. 40CFR Parts 9, 141 and 142, volume 71, Number 654, 5 January 2006.

US EPA 2006*b Economic Analysis for the Ground Water Rule*. EPA 815-R-06-014.

US EPA 2006*c National Primary Drinking Water Regulations*: Ground Water Rule. 40CFR Parts 9, 141 and 142, Volume 71, Number 216, 8 November 2006.

US EPA 2006*d Appendices to the Economic Analysis for the Final Ground Water Rule* Appendix E. EPA 815-R-06-014.

Velazquez, F. R., Calva, J. J., Guerrero, M. L., Mass, D., Glass, R. I., Pickering, L. K. & Ruiz-Palacios, G. M. 1993 Cohort study of rotavirus serotype patterns in symptomatic and asymptomatic infections in Mexican children. *Pediatr. Infect. Dis. J.* **12**(1), 54–61.

Ward, R. L. & Bernstein, D. I. 1994 Protection against rotavirus disease after natural rotavirus infection. *J. Infect. Dis.* **169**, 900–904.

Ward, R. L., Bernstein, D. I., Young, E. C., Sherwood, J. R., Knowlton, D. R. & Schiff, G. M. 1986 Human rotavirus studies in volunteers: determination of infectious dose and serological response to infection. *J. Infect. Dis.* **154**(5), 871–879.

Wenman, W. M., Hinde, D., Feltham, S. & Gurwith, M. 1979 Rotavirus infection in adults: results of a prospective family study. *N. Engl. J. Med.* **301**(6), 303–306.

First received 27 February 2007; accepted in revised form 20 April 2008. Available online February 2009

## APPENDIX: MODEL PARAMETER VALUES

From the literature review, median values were determined for each of the model parameters (Table 4). Minimum and maximum reasonable values were also identified for use in the uncertainty analysis.

### Background incidence

The median estimate of background incidence of rotavirus illness in the US is 3,900,000 cases per year (Tucker *et al.* 1998; Mead *et al.* 1999) based on available primary data (Gurwith *et al.* 1981; Rodriguez *et al.* 1987). For the purposes of the uncertainty analysis, the minimum and maximum reasonable values used are 50,000 and 23,000,000 cases per year, respectively (Mead *et al.* 1999). The minimum and maximum values are selected with the understanding that noroviruses are also type A viruses and this wide parameter range was investigated to account for the estimated incidence of norovirus illness in the US, their genetic diversity and the uncertainty associated with norovirus cross-genotype immunity.

### Dose response parameters

The best fitting beta-Poisson dose response model has point estimates of $\alpha = 0.26$ and $\beta_{dr} = 0.42$ (Ward *et al.* 1986; Regli *et al.* 1991; Haas *et al.* 1999). The corresponding 95% confidence intervals range from 0.126 to 0.52 for $\alpha$ and from 0.21 to 0.84 for $\beta_{dr}$. The best fitting point estimates are employed in the median value analyses and the lower and upper 95% confidence values are used as the minimum and maximum reasonable values, respectively, in the uncertainty

analysis. Unfortunately, the shape of the 95% confidence region of the rotavirus dose response maximum likelihood estimate (MLE) is not defined by a simple formula (Regli *et al.* 1991). The minimum and maximum values employed herein represent a reasonable approximation to the 95% confidence region.

### Duration of incubation

The median estimate for the duration of incubation of 2.5 days and the minimum and maximum reasonable values are 1 and 5 days, respectively (Flewett *et al.* 1975; Shepherd *et al.* 1975; Flewett & Woode 1978; Rodriguez *et al.* 1979; Ward *et al.* 1986).

### Probability of symptomatic response

The median probability of symptomatic response for adults is 0.43 (Kim *et al.* 1977; Tufvesson *et al.* 1977; Wenman *et al.* 1979; Ward *et al.* 1986). The minimum and maximum reasonable values for adults are 0.1 and 0.57, respectively.

For children, the median probability of symptomatic response is 0.67. The minimum and maximum reasonable values are 0.17 and 0.82, respectively (Wenman *et al.* 1979; Gurwith *et al.* 1981; Champsaur *et al.* 1984; Barron-Romero *et al.* 1985; Bernstein *et al.* 1991; Velazquez *et al.* 1993; Ferson *et al.* 1997).

Values used in the analysis are weighted averages for children and adults, based on census data from 2000 (US Census Bureau 2000), assuming that children are represented by those under age 5 years. Thus, a probability of symptomatic response of 0.45 is used in the median value analyses and 0.1 and 0.6 are used as the minimum and maximum reasonable values, respectively.

### Duration of symptomatic infection

A duration of symptomatic infection of 6 days is used in the median value analyses and 2 and 8 days are used as the minimum and maximum reasonable values, respectively (Flewett *et al.* 1975; Shepherd *et al.* 1975; Gomez-Barreto *et al.* 1976; Flewett & Woode 1978; Lycke *et al.* 1978; Gurwith *et al.* 1981; Ward *et al.* 1986).

### Duration of asymptomatic infection

The average duration of asymptomatic infections is 5 days with durations of asymptomatic infections ranging from 1 to 12 days (Ward *et al.* 1986). Based on these data, a duration of asymptomatic infection of 5 days is used in the median value analyses. Given the sparse data available for this parameter, the minimum and maximum reasonable values are the same as those used for the duration of symptomatic infections (2 and 8 days, respectively).

### Duration of protection from infection

Based on the available data (Koopman & Monto 1989; Bernstein *et al.* 1991; Chiba *et al.* 1993; Ward & Bernstein 1994), it is inferred that the duration of protection from infection is highly uncertain and not well understood. From studies on children it has been found that infection confers immunity (typically from illness but not necessarily from infection) for a duration of not less than 1 year. Based on data from a lifetime analysis it appears that the duration of protection may be longer for adults than for children. Thus, a duration of 18 months is used in the median value analyses and durations of 12 and 60 months are used as the minimum and maximum reasonable values, respectively.

### Person-to-person transmission

During the Tecumseh, Michigan, study, Koopman *et al.* estimated the proportion of rotavirus infections acquired in the household (17–20%) compared with those acquired outside the household (Koopman *et al.* 1989). For the purposes of this investigation, it is assumed that infections acquired in the household are through a person-to-person route of transmission and infections acquired in the community are acquired either through person-to-person transmission or from environmental sources. Based on those data, 20% is used in the median value analyses as the proportion of rotavirus infections acquired via person-to-person transmission and a range of 10 and 60% is used to encapsulate the minimum and maximum reasonable values, respectively. This relatively wide range employed in the uncertainty analysis is intended to be representative of the uncertainty associated with: 1) extrapolating the results from Tecumseh to the US population; and 2) the assumption that infections acquired in the community are due to a combination of environmental sources and person-to-person contacts outside of the household.