Risk Analysis, Vol. 13, No. 5, 1993

Risk Assessment of Virus in Drinking Water

Charles N. Haas,[1] Joan B. Rose,[2] Charles Gerba,[3] and Stig Regli[4]

Received December 29, 1992; revised May 7, 1993

The reevaluation of drinking water treatment practices in a desire to minimize the formation of disinfection byproducts while assuring minimum levels of public health protection against infectious organisms has caused it to become necessary to consider the problem of estimation of risks posed from exposure to low levels of microorganisms, such as virus or protozoans, found in treated drinking water. This paper outlines a methodology based on risk assessment principles to approach the problem. The methodology is validated by comparison with results obtained in a prospective epidemiological study. It is feasible to produce both point and interval estimates of infection, illness and perhaps mortality by this methodology. Areas of uncertainty which require future data are indicated.

KEY WORDS: Virus; water supply; infectious disease.

1. INTRODUCTION

Recent changes in regulation of drinking water[1], most particularly in the area of increasing stringency of disinfection (surface water treatment rule, groundwater treatment rule) and minimizing formation of disinfection byproducts warrant a detailed examination of risks associated with ingestion of drinking water containing infectious microorganisms at current, and possible future regulatory levels. It has always been recognized that disinfection of a drinking water does not mean sterilization, nor can any water supply be treated so that there is absolute certainty that zero pathogens are present. Hence, there will always remain some level of residual microbial risk from that treated supply, although the goal is to reach a *de minimus* level. The first purpose of this paper is to describe methodology for computing such risks, using viral agents as a specific example.

A large number of viruses have long been known to be present in sources of drinking water.[2] In more recent years, additional viral agents of disease have been identified.[3] There also remain a substantial number of unidentified agents of waterborne gastroenteritis, many of which may be viral. In prior work, equations for dose–response relationships for virus and other organisms have been tested, and the beta-poisson model has been identified as one which fits most data and which provides a conservative method for low dose extrapolation.[4,5] There has also been a prospective intervention epidemiological study in Montreal examining the proportion of gastroenteritis that may be attributable to waterborne agents.[6] However, to date, there has not been a comparison between estimates based on a risk assessment methodology and empirical data on waterborne infectious disease prevalence. It is the second purpose of this paper to attempt such a comparison.

2. METHODOLOGY

A formal risk assessment of microbial exposure may be conducted under the standard framework used for

[1] Department of Civil and Architectural Engineering and Environmental Studies Institute, Drexel University, Philadelphia, PA 19104.
[2] School of Public Health, University of South Florida, Tampa, Florida.
[3] Department of Soil and Water Science, University of Arizona, Tucson, Arizona.
[4] Office of Ground Water and Drinking Water, U.S. Environmental Protection Agency, Washington, D.C.

0272-4332/93/1000-0545$07.00/1

chemical risk assessment. This consists of three steps[7]—dose–response assessment, exposure assessment, and risk characterization. The microbial risk quantification may be performed using point estimation methods, such as maximum-likelihood estimates of dose–response and maximum exposed individual, or estimates for the entire population based on the frequency distribution of the exposure. Both of these approaches will be discussed.

2.1. Dose–Response Assessment

The process of contracting an undesirable consequence (illness, fatal disease) as a result of exposure to a waterborne virus may be described by a series of sequential events. This is depicted schematically in Fig. 1.

There are three sequential events useful to characterize the course of an illness. A susceptible individual ingesting water containing infectious microorganisms may contract an infection, providing that the dose ingested is sufficient to allow organism survival and multiplication to overcome internal defense mechanisms. The probability of this occurring on the basis of a single exposure is P_I. Once infected, there is a probability that a single infection will result in illness denoted by the conditional expression $P_{D:I}$. An ill individual stands a risk of death (mortality), with a further conditional probability of $P_{M:D}$.

The process of risk assessment therefore may be described as the task of defining the relationship between the three probabilities and the level of exposure. The level of exposure is a function of water ingestion and the quality (specific pathogen amount) of the water ingested.

2.2. Assessment of Infection Probability (P_I)

If a dose of microorganisms is ingested (in the amount N), then the beta-poisson model[4] gives the following expected relationship:

$$P_I = 1 - \left[1 + \frac{N}{N_{50}}(2^{1/\alpha} - 1)\right]^{-\alpha} \quad (1)$$

In this relationship, N_{50} is the microbial dose eliciting 50% infections in the exposed population, and α is a slope parameter. This model can be derived from certain assumptions about the nature of the infection process, including a measure of host heterogeneity.[4,8,9] Both of these parameters must be estimated from dose–response experiments. As the α parameter approaches infinity, the relationship in Eq. (1) approaches an exponential relationship. This behavior is illustrated in Fig. 2.

The effect of decreasing values of α is to reduce the slope of the dose–response relationship, reflecting an increasing degree of heterogeneity in the microorganism-host interaction leading to infection. At very low pathogen concentrations, the relationship between risk of infection and microbial dose is approximately linear, according to this model.

Given a dose–response experiment in which infection in humans has been assessed as a function of ingested pathogen number, the parameters in the dose–response relationship can be assessed. Formally, if k groups of persons (with number T_1, $T_2 \ldots T_k$ in each group)

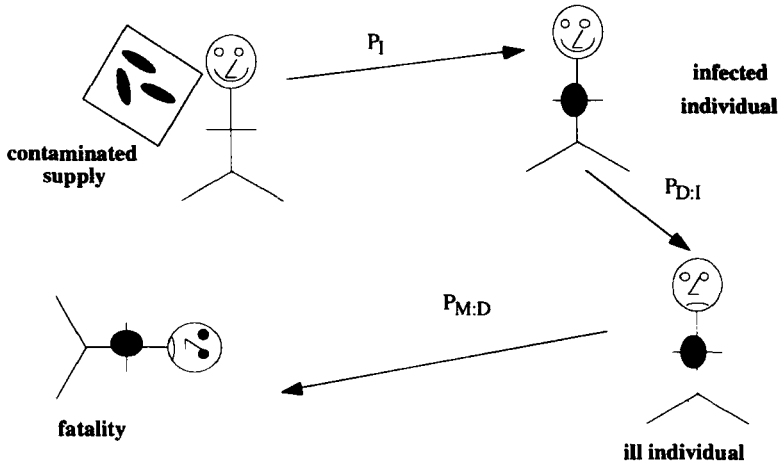

Fig. 1. Conceptual paradigm of the consequences of exposure to water containing microorganisms.

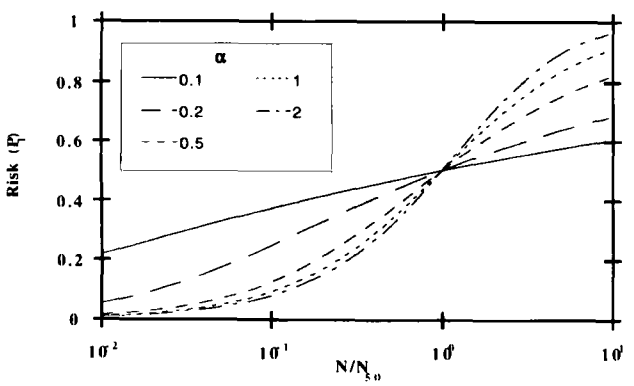

Fig. 2. Effect of α on the shape of the microbial dose–response relationship for the Beta-Poisson model.

are exposed to doses of N_1, N_2 ... N_k pathogens (where this is the estimated mean number) and I_1, I_2 ... I_k persons are found to have become infected, then the ln-likelihood function can be written as (assuming the beta-poisson model):

$$Y = 2\sum_{i=1}^{k}\left\{I_i \ln\left(\frac{I_i}{T_i\pi_i}\right) + (T_i - I_i)\ln\left[\frac{T_i - I_i}{T_i - T_i\pi_i}\right]\right\}$$

$$\pi_i = 1 - \left\{1 + \frac{N_i}{N_{50}}(2^{1/\alpha} - 1)\right\}^{-\alpha} \qquad (2)$$

The values for the dose–response parameters (α, N_{50}) which maximize the log-likelihood function are accepted as the maximum-likelihood estimators of these parameters. Goodness-of-fit can be assessed by comparison of the minimum Y with the chi-square distribution at k-2 degrees of freedom—if Y is less than the critical value, the model should not be rejected.[10,11]

Earlier reviews of the available data for various viruses have shown that rotavirus is the most infectious waterborne virus for which dose–response information is available.[5] If this is assumed to represent the most infective virus likely to be present in a drinking water, then a plausible upper-limit risk assessment can be based on the dose–response properties of this organism.

Using the maximum-likelihood approach, the best fit parameters for rotavirus using the experimental data of Ward[12] were found to be $N_{50}=5.60$ and $\alpha=0.265$. The low value of the latter parameter shows that the dose–response relationship was significantly shallower than exponential (see Fig. 2). This was confirmed by a likelihood ratio test of the difference of α from ∞. The qualitative nature of the fit was found to be good, as shown by the plot in Fig. 3, which also includes the likelihood-ratio based binomial confidence limits for the infectivity probabilities at different dosages. This dose–response relationship is subsequently used to assess the daily risk of infection from waterborne viruses (P_I).

2.3. Probability of Morbidity

Infection is not the primary consequence of interest; however, it is a prerequisite to frank disease. Therefore, the process of development of infection to symptomatic disease may be regarded as a conditional probability that, once having been infected, a particular individual contracts a disease.

There is far less quantitative data available to analyze this process. In particular, the relationship (if any) between the level of exposure (ingested dose) and the chance of contracting disease (once having been infected) is unknown. The simplest assumption to make is that this conditional probability, designated $P_{D:I}$, is independent of exposure level. Since the level of exposure in treated drinking water may rarely exceed 1 organism per unit consumption, this assumption may not be that critical. Hence, the probability of contracting disease from a single exposure can be written as:

$$P_D = P_{D:I} \times P_I \qquad (3)$$

The morbidity data for a variety of waterborne infectious agents have been reviewed elsewhere.[5,13] For enteric viruses, morbidity rates between 1% and 97% have been observed, depending upon the virus and the age of the subject. Until further data is available, it appears reasonable to use a midpoint value of 50% for the morbidity rates for the waterborne viruses. Ideally, a probability distribution, reflecting age, underlying health status (including immunological competence), and other variables influencing sensitivity to morbidity, should be used for a given virus.

2.4. Probability of Mortality

It is known that, at least for some agents, disease may result in death of a small proportion of those who contract an illness. While there is a paucity of data, the available information suggests that mortality rates of 0.12–0.94% can occur for coxsackie and echo viruses.[13] It is perhaps too early to compute estimated fatalities resulting from waterborne exposure to viruses; however, the mortality is clearly not negligible.

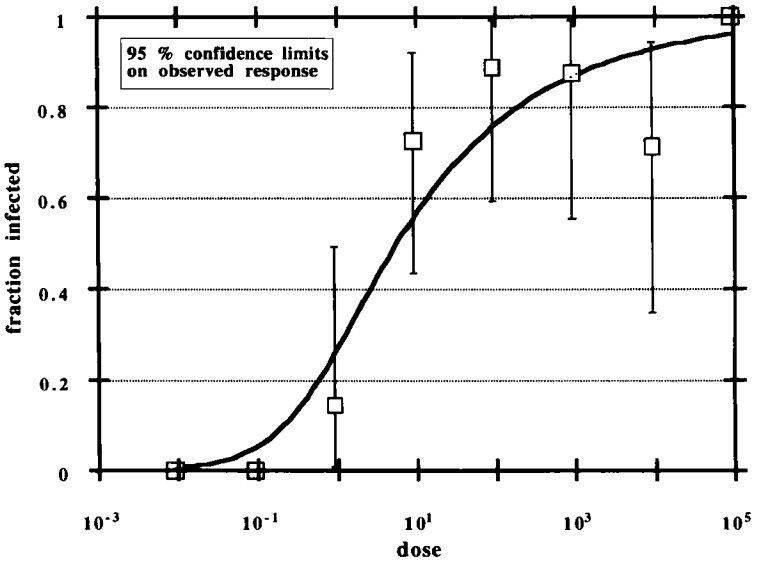

Fig. 3. Comparison of beta-poisson model to rotavirus infectivity data of Ward *et al.* Error bars show 95% confidence limits on binomial response (likelihood based).

2.5. Consequences of Multiple Exposures

In the particular case of drinking water ingestion, exposure to a microbiological contaminant may occur rarely and sporadically, or may occur relatively frequently. If each exposure is regarded as statistically independent [i.e., the chance of developing an infection (or illness or death) from one exposure is not related to any prior exposures and effects], then simple probability relationships may be invoked. For example, regarding a given day as a discrete exposure, it $P^{(j)}_{*,d}$ is the chance of consequence "*" (I = infection; D = disease) resulting from the exposure on a single day, then the chance $P_{*,y}$ of an individual having at least one occurrence of that consequence resulting from an annual exposure, providing that risks are statistically independent, can be written as:

$$P_{*,y} = 1 - \prod_{j=1}^{365} (1 - P^{(j)}_{*,d}) \qquad (4)$$

which can be simplified to (if the daily risks are furthermore assumed to be identically distributed):

$$P_{*,y} = 1 - (1 - P_{*,d})^{365} \qquad (5)$$

The major process that can cause lack of independence of multiple exposures is temporary or permanent immunity. Much more information is needed, particularly on strains of pathogens of most significance for waterborne infection. It is clear that different organisms show different properties with respect to multiple exposure. Rotavirus infections appear to confer highly strain-specific immunity for periods of perhaps years; however, Norwalk virus infection results only in short-term, but not long-term, immunity.[14] It is conceivable that initial exposures to an agent may result in development of a hypersensitivity. Due to the paucity of data, independence of exposure will be assumed.

2.6. Uncertainties in Administered Microorganisms and Host Susceptibility

A major uncertainty in application of available dose–response information is that human experiments have been performed using healthy adult volunteers. In the general population exposed to an infectious agent, the overall health status is presumed to be somewhat poorer, and hence the susceptibility to adverse effects assumed to be greater. However, the quantitative effect of this difference is unknown.

A related issue is that the dose–response studies have generally been performed using well-characterized laboratory strains of pathogens. The intrinsic infectivity (and potential to cause morbidity) may differ between laboratory maintained cultures and indigenous viruses; however, the magnitude of these differences is currently not clear.

2.7. Exposure Estimation

To determine the risk from microbial exposure, an exposure assessment must be performed. There are two essential components to the problem at hand. First, the consumption of water must be determined. Second, the concentration of pathogenic viruses in water must also be assessed. Data on both facets are available, although clearly the latter component is highly site-specific, being a function of both the raw water quality and the degree of treatment and intermediate virus inactivation prior to ingestion.

The U.S. EPA generally employs 2 L/person-day for risk estimation from drinking water. This number will be used for a point risk estimate. This number appears to emanate from an *ad hoc* review of early data, with rounding upward.[15]

More recent analysis has provided the detailed frequency distribution of water consumption, which has been found to be log-normal with a geometric mean of 1.7 L/cap-day.[16] The information from this latter study will be used in interval risk assessment.

Payment *et al.*[17] found an average of 0.0006 viruses per L in finished drinking water in the Montreal area. The overall occurrence distribution was found to be log-normal. For a point risk assessment, to approximate a most exposed individual (MEI), twice this value will be used, (i.e., 0.0012/L).

2.8. Point Risk Assessment

Using the information above, a point risk estimate of exposure to virus may be made. This can be compared to the observed risk of disease found in a prospective intervention study conducted in the Montreal area.[6] In that study, an annual risk of illness of 0.24 cases/person-year was found. The area in which the epidemiological study was conducted corresponded to that in which prior virus measurements were taken,[17] and are believed to characterize the distribution to which individuals in the epidemiological study were exposed (Payment, personal communication). By application of Eq. (5), this is equivalent to a mean daily risk of 0.00082.

2.9. Uncertainty Analysis of Dose–Response Parameters

To perform a full uncertainty analysis, it is necessary to estimate the uncertainty distribution of the dose–response parameters. Computationally, the simplest accurate estimator of this distribution appears to be the simple bootstrap estimation procedure.[18,19] The process may be described by the following steps:

1. Construct a hypothetical data set of experimental data with $I^*_1 \ldots I^*_k$ infected individuals randomly selected as binomial variates from a distribution with parameters $(I_1\ T_1,T_1)\ \ldots\ (I_kT_k,T_k)$ at the respective doses $N_1 \ldots N_k$
2. Fit the hypothetical data set to the beta-poisson model, determining the MLE estimates (α^*,N_{50}^*).
3. Repeat the process a large number of times.
4. The bivariate distribution of values (α^*,N_{50}^*) represents the uncertainty in the true dose–response parameters.

Computationally, the above process was performed using a custom program written in TURBO Pascal and employing the Nelder-Mead polytope algorithm to optimize the likelihood function.[20]

Figure 4 presents results of this analysis. One thousand bootstrap replications were used in construction of this figure. The bivariate sampling distribution for the dose–response parameters (uncertainty distribution) is approximated by a discrete distribution with unit mass on each of the bootstrap points. For comparison, the likelihood based 95% confidence region is shown; this latter region was previously reported.[5] Note that there is good agreement between the two methods. The bootstrap based approach is used in the interval estimation procedure, since it is computationally easier.

2.10. Characterization of Uncertainty (Interval Estimation)

It is desirable to present some measure of uncertainty in the estimation of risk. The arguments for conveying uncertainty have been presented,[21] along with the framework for a methodology for estimating such uncertainties. In the context of the present problem, it is necessary to first determine uncertainties in the underlying exposure estimates and dose–response relationships and then to compound these into an overall uncertainty estimate. With currently available tools, it is relatively straightforward to compute the uncertainty in the daily or annual risk using Monte-Carlo analysis. The formal computations were conducted using @Risk.[22] Except for the correlations between the dose–response parameters noted above, all distribution functions were assumed to be statistically independent. The empirical bootstrap distribution of dose–response parameters was used as a random bivariate input to the computation.

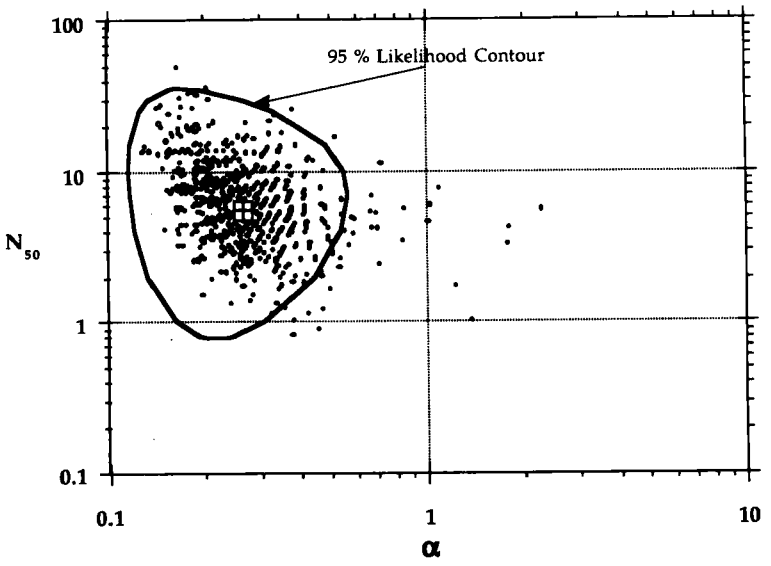

Fig. 4. Bootstrap sampling distribution for rotavirus dose–response parameters (1000 replicates). The box indicates the maximum-likelihood estimate. Contour is the 95% likelihood based confidence boundary.

To compare the uncertainty which may be attributed to the actual variability in exposure (water consumption rates, virus densities) to that which may be due to uncertainty in virus infectivity, morbidity, and mortality ratios, the Monte-Carlo analysis was also conducted using the point estimates of infectivity (N_{50}, α) and morbidity parameters, in conjunction with the exposure and consumption distributions. This latter analysis is termed the "partial Monte Carlo" analysis, in contrast to the full Monte Carlo analysis described in the previous paragraph.

3. RESULTS AND DISCUSSION

Using the above assumptions, the point estimate of risk was computed at a plausibly most exposed individual. By employing Monte Carlo simulation, the uncertainty distribution of risk was also estimated. In the Monte Carlo computation, 10,000 iterations were employed. This large number of iterations was used to produce a reliable estimate of the 95% confidence limit (which excluded the upper and lower 250 values).

3.1. Point Risk Estimate

The point risk estimate of the daily probability of disease was found to be 0.000717. This is equivalent to an annual risk of disease of 0.23. This compares favorably with the daily and annual risks in the Payment study[6] of 0.00082 and 0.24.

3.2. Uncertainty Estimate

Figure 5 presents the full uncertainty distribution of the daily risk of disease. The mean and median daily risk of disease were found to be 0.000443 and 0.000276, respectively. The 95% confidence interval of the estimated daily disease risk was found to be 0.0000317 - 0.00188. As expected, the confidence band for the partial Monte Carlo analysis is somewhat narrower (since it includes fewer sources of variation). However, it is particularly interesting that the upper tail (0.975 cumulative point) of the distribution, which would represent an upper confidence limit to the risk estimate, is less than a factor of 2 greater for the full Monte Carlo analysis than the partial Monte Carlo analysis, indicating that there is not an extreme degree of uncertainty in the risk estimate provided by the estimation of dose–response parameters (as opposed to the intrinsic variability in the exposure itself).

4. DISCUSSION

The proposed methodology for estimation of risk associated with waterborne exposure to infectious mi-

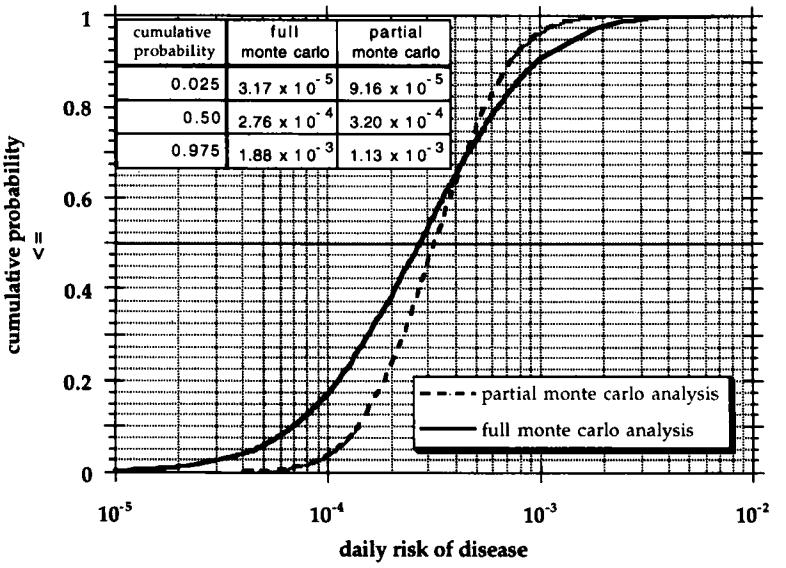

Fig. 5. Cumulative probability distribution of estimated daily risk of morbidity. Comparison of full and partial (excluding dose–response and infectivity uncertainty) Monte Carlo analysis.

croorganisms appears to produce estimates which are consistent with independent estimates obtained using a prospective epidemiological approach. The confidence limits to the risk estimate appear to be reasonably tight, especially when compared to risk estimates in the field of chemical risk assessment.

The proposed methodology may also be adopted to problems other than drinking water risk assessment. For example, the risks due to recreational exposure to waterborne virus[23] or protozoan cyst,[24] ingestion of contaminated food, or exposure to shellfish grown in contaminated waters may all be approached using the proposed methodology, providing applicable occurrence and exposure data are available.

It is instructive to consider the potential for mortality from exposure to infectious agents. Using the daily morbidity risk interval estimate of 0.0000317 - 0.00188, and a mortality ratio of 0.001, the daily risk of death is computed to be 3.2×10^{-8} - 1.9×10^{-6}. Based on a 70-year exposure, and independent identically distributed daily risks, the lifetime risk of death is then determined as 0.0008 - 0.047. In other words, there may be a lifetime risk of death as high as 1 in 20 from exposure to waterborne virus based on the above assumptions. This is clearly not negligible (particularly in viw of the 10^{-4} to 10^{-6} lifetime risk that is often regarded as an action level for carcinogens), and must be considered in the risk management decision-making process with respect to drinking water treatment.

CONCLUSIONS

It is possible to employ a risk assessment paradigm to the estimation of likelihood of various consequences resulting from environmental exposure to infectious microorganisms. In the particular case of drinking water, the results of a sample risk analysis appear to be in reasonable agreement with estimates of infectious disease risk obtained from an epidemiological study. The risk from mortality in such cases cannot be regarded as being negligible. Further research is needed to clarify and confirm some details needed to a full uncertainty analysis of this problem; however, use of the methodology in its current state of development would appear to be possible with respect to decision-making.

ACKNOWLEDGMENTS

This work has not been subject to review by the U.S. EPA, and thus should not be considered to represent agency policy. The authors acknowledge useful consultation with Dr. Bruce Macler, and useful comments from the referees.

REFERENCES

1. J. A. Cotruvo and C. D. Vogt, *Rationale for Water Quality Standards and Goals, in* F. W. Pontius (ed), *Water Quality and Treatment* (McGraw-Hill, New York, 1990), pp. 1–62.

2. J. L. Melnick, *et al.*, "Viruses in Water," *Bulletin of the World Health Organization* **56,** 499–508 (1978).
3. C. P. Gerba and J. B. Rose, *Viruses in Source and Drinking Water, in* G. A. MeFetes (ed.), *Drinking Water Microbiology*. (Springer-Verlag, New York, 1990).
4. C. N. Haas, "Estimation of Risk Due to Low Doses of Microorganisms: A Comparison of Alternative Methodologies," *American Journal of Epidemiology* **118,** 573–582 (1983).
5. S. Regli, *et al.*, "Modeling Risk for Pathogens in Drinking Water," *Journal of the American Water Works Association* **83,** 76–84 (1991).
6. P. Payment, *et al.*, "A Randomized Trial to Evaluate the Risk of Gastrointestinal Disease Due to Consumption of Drinking Water Meeting Current Microbiological Standards," *American Journal Public Health* **81,** 703 (1991).
7. National Academy of Sciences, *Risk Assessment in the Federal Government: Managing the Process* (Washington, D.C., National Academy Press, 1983).
8. W. A. Furumoto and R. Mickey, "A Mathematical Model for the Infectivity-Dilution Curve of Tobacco Mosaic Virus: Theoretical Considerations, *Virology* **32,** 216 (1967).
9. W. A. Furumoto and R. Mickey, "A Mathematical Model for the Infectivity-Dilution Curve to Tobacco Mosaic Virus: Experimental Tests," *Virology* **32,** 224 (1967).
10. R. H. Norden, "A Survey of Maximum Likelihood Estimation," *International Statistical Review* **40,** 329 (1972).
11. R. H. Norden, "A Survey of Maximum Likelihood Estimation: Part II," *International Statistical Review* **41,** 39 (1973).
12. R. L. Ward, *et al.*, "Human Rotavirus Studies in Volunteers: Determination of Infectious Dose and Serological Response to Infection, *"Journal of Infectious Disease* **154,** 871 (1986).
13. C. P. Gerba and C. N. Haas, "Assessment of Risks Associated with Enteric Viruses in Contaminated Drinking Water," *ASTM Special Technical Publication* **976,** 489–494 (1988).
14. G. Cukor and N. R. Blacklow, "Human Viral Gastroenteritis," *Microbiological Reviews* **48,** 157 (1984).
15. National Academy of Sciences, *Drinking Water and Health* (Washington, D.C. National Academy of Sciences, 1977).
16. A. M. Rosebury and D. E. Burmaster, "Log-Normal Distributions for Water Intake by Children and Adults," *Risk Analysis* **12,** 99 (1992).
17. P. Payment, *et al.*, "Elimination of Viruses and Indicator Bacteria at Each Step of Treatment During Preparation of Drinking Water at Seven Water Treatment Plants," *Applied Environmental Microbiology* **49,** 1418 (1985).
18. K. S. Crump and R. B. Howe, *A Review of Methods for Calculating Statistical Confidence Limits in Low Dose Extrapolation, in* D. B. Clayson *et al.* (eds.), *Toxicological Risk Assessment* (CRC Press, Boca Raton, Florida, 1985), pp. 187–203.
19. D. V. Hinckley, "Bootstrap Methods," *Journal of the Royal Statistical Society B* **50,** 321–337 (1988).
20. J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *Computer Journal* **7,** 298–313 (1964).
21. A. M. Finkel, *Confronting Uncertainty in Risk Management* (Washington, DC., Resources for the Future, Center for Risk Management, 1990).
22. Palisade Corporation, *@RISK, Version 2.0 for Lotus* (1991).
23. C. N. Haas, "Effect of Effluent Disinfection on Risks of Viral Disease Transmission via Recreational Exposure," *Journal of the Water Pollution Control Federation* **55,** 1111–1116 (1983).
24. J. B. Rose, *et al.*, "Risk Assessment and the Control of Waterborne Giardiasis," *American Journal of Public Health* **81,** 709–713 (1991).