Technical Note

Cryptosporidium and *Giardia* Distribution in Water: Re-Evaluation

Jerry E. Ongerth, Ph.D., F.ASCE[1]

Abstract: The data and the analysis applied to it in a 1996 paper on the statistical distribution of *Cryptosporidium* in a reservoir are re-examined with the objective of clarifying general understanding of the way in which *Cryptosporidium* oocysts and *Giardia* cysts are distributed in water. General objectives of monitoring for these protozoan parasites are summarized along with essential assumptions used in statistical analysis. The 1996 analysis is reviewed and an alternative analysis is proposed. The key distinction is that the Poisson model assumption of a homogeneous population was not appropriate as applied to the data set consisting of 52 consecutive weekly samples, leading to likely misinterpretation of the data. The critical issue is interpretation of negative (zero) results, whether as absence and hence intermittent presence or as continuous presence but below the limit of detection. The alternate analysis shows that the typically skewed annual data set can be effectively described as lognormal. The lognormal distribution of the 1996 data is compared to previously published data on *Cryptosporidium* and *Giardia* from reservoirs elsewhere with apparent similarity. The application of the Poisson model to understanding the relation between ambient concentrations of *Cryptosporidium* and *Giardia*, sample volumes, and recovery efficiency is described with its importance to effective planning of monitoring and data analysis. **DOI: 10.1061/(ASCE)EE.1943-7870.0001193.** *© 2017 American Society of Civil Engineers.*

Introduction

The purpose of this paper is to clarify understanding regarding the distribution of *Cryptosporidium* oocysts and *Giardia* cysts in a body of water at any given time. Key issues are (1) the appropriate selection of a statistically valid "population"; (2) the interrelation between true ambient concentration in the population, the sample volume, and the appropriate statistical model; and (3) the selection of an appropriate sample volume to characterize organism occurrence at any sampling location. These concepts are fundamental to a clear understanding of the occurrence of *Cryptosporidium* and *Giardia* in water and convey the ability to plan for effective monitoring and to interpret resulting data without misunderstanding and misconception.

Valid application of virtually any statistical analysis of water quality data such as the occurrence of *Cryptosporidium* oocysts and *Giardia* cysts in a water body (e.g., river, lake, or reservoir) depends on critical assumptions. The most important of these is that the measurements are from a single population and that the sample is representative of the population—hence, the term representative sampling. For example, in sampling from a flowing stream the key issues are that water at the sampling location is well mixed and that flow conditions are stable. In a lake or reservoir, similarly for samples to accurately represent the entire system the sampling location must be as representative of the total volume as possible, likely below a stratified surface layer and at a middepth. Often sampling of such systems is conducted at an outflow or intake location because that water is directly relevant to the point of use (e.g., a water treatment plant).

An effective sampling plan for *Cryptosporidium* and *Giardia* monitoring requires several key elements: (1) the objective of the monitoring and intended use of resulting data; (2) the analytical procedure; and (3) the sample volume to be collected (Ongerth 2013a). A common objective of monitoring for these organisms is simply to satisfy regulatory requirements—for example, the USEPA's Long Term 2 Enhanced Surface Water Treatment Rule (LT2ESWTR or simply LT2) (USEPA 2006). A utilitarian objective would be to identify their ambient concentrations and their characteristic variations over a typical annual cycle. To satisfy this objective would require that the sample volumes collected be sufficiently large to permit finding the target organism(s) in a majority of the samples given the analytical procedure to be used. The most common analytical procedure currently is USEPA Method 1622/23 (USEPA 2005). No analytical method can find all of the target organisms present in a given sample. Typically, with Method 1622/23 the recovery efficiency for *Cryptosporidium* ranges between 15 and 40% and for *Giardia* between 25 and 60% (Messner 2011). To satisfy the objective of positive results in a majority of samples, the volume to be analyzed must take recovery efficiency into account. If no prior data on *Cryptosporidium* and *Giardia* concentrations are available from the intended sampling location, a trial and error process is needed to determine the required sample volume.

Analysis

In a paper demonstrating the applicability of the Poisson statistical model to the distribution of *Cryptosporidium* oocysts in surface water, Haas and Rose (1996) presented data on the "density" of *Cryptosporidium* oocysts in water samples identified only as "obtained from the water supply reservoir of a utility in the Northwest of the United States" (Table 1). In a paper providing laboratory data on the distribution of *Cryptosporidium* and *Giardia* at concentrations above and below the normal limit of detection for USEPA

[1]Principal Honorary Fellow, Dept. of Civil, Mining, and Environmental Engineering, Univ. of Wollongong, CME Bldg., 4 EEC, Wollongong, NSW 2522, Australia. ORCID: http://orcid.org/0000-0003-0339-8942. E-mail: jongerth@uow.edu.au

Note. This manuscript was submitted on August 28, 2016; approved on October 14, 2016; published online on February 11, 2017. Discussion period open until July 11, 2017; separate discussions must be submitted for individual papers. This technical note is part of the ***Journal of Environmental Engineering***, © ASCE, ISSN 0733-9372.

Method 1623, Ongerth and Saaed (2013) demonstrated that both organisms measured in a homogeneous system at concentrations near the limit of detection are distributed according to the Poisson statistical model.

The analysis of Haas and Rose can easily be misconstrued with respect to sampling for *Cryptosporidium* and *Giardia* in water and to analyzing the resulting data. The Poisson statistical model specifically applies to rare objects randomly distributed in a homogeneous system; that is, it applies strictly to a defined population. As Haas and Rose stated (p. 2252), "if a single mean density characterizes the water from which all samples are collected," essentially the Poisson model applies. However, the data used by Haas and Rose are from 52 different populations present in the reservoir from which the samples were collected on 52 consecutive weekly occasions. In this context, the inference suggested by the preponderance of zero results (33 of 52, or 63%) is that in those samples the organisms were not present at the sampling times.

A reasonable alternative interpretation of the data presented is that they in fact describe different "mean densities" corresponding to individual weekly sample results. Furthermore, the samples in which no organisms were found were simply too small so that the concentrations at the sampling times were below the limit of detection (Ongerth 2013b). Ample data from elsewhere demonstrate that populations characterized by water quality parameters, including *Cryptosporidium* oocysts and *Giardia* cysts and similar discrete physical entities, are not constant from week to week (Ongerth 1989; Hansen and Ongerth 1991). Rather, the *Cryptosporidium* and *Giardia* populations present at any given week are more accurately characterized as unique populations.

Examining their *Cryptosporidium* data, Haas and Rose (1996) could readily see that they were not normally distributed. In that time period (roughly 1985–1995) whether *Cryptosporidium* and *Giardia* occurrence in surface water was continuous or intermittent was a subject of controversy. A project designed to test this question resulted in data from the Northwest United States (rivers near Seattle and Tacoma, Washington) strongly suggesting that *Cryptosporidium* are present continuously (Hansen and Ongerth 1991). Previous data suggested that this was the case for *Giardia* also (Ongerth 1989). Despite this evidence, others viewing the relatively infrequent finding of *Cryptosporidium* (or *Giardia*) resulting from surface water sampling persisted in the interpretation of intermittency. It might follow logically that when generalized conceptually to the distribution of such discrete and virtually inert (nonmultiplying) organisms, the idea that the distribution throughout the entire volume of a reservoir might follow the Poisson model.

The crux of the issue rests ultimately on the true concentration in the system being characterized. If the concentration is small with respect to the sample volume, the Poisson distribution of organisms is relevant (Ongerth and Saaed 2013). For example, if the analytical method is USEPA Method 1622/23 using a volume of 10 L and the concentration is only 1 in 10 L (0.1/L), a common level reported in U.S. surface water, then the probability of finding none is important—in fact, for this example it is approximately 90% (Ongerth and Saaed 2013). For a given concentration using the example of 0.1/L, as the test volume is increased, for example to 100 or 1,000 L, the number of organisms in other samples of that volume from the test system begins to approach a normal distribution (Ongerth and Saaed 2013). Application of the Poisson distribution to the distribution of *Cryptosporidium* throughout the entire Northwest U.S. reservoir, at weekly intervals, is thus conceptually misleading. Analysis of a "representative" volume taken anywhere in the reservoir, at the weekly sampling intervals, would have resulted in weekly estimates of *Cryptosporidium* occurrence. The actual data (Table 1) approximate this with the disadvantage of having a high proportion of negative (zero) results.

The Haas and Rose data (Table 1) consisted of *Cryptosporidium* oocyst numbers identified in sample volumes that ranged approximately 20–225 L, as analyzed essentially by what was at the time known as the "ICR Method" (USEPA 1995) and specifically not adjusted for recovery efficiency. Previous reports of *Cryptosporidium* (and *Giardia*) concentration measurement in comparable Pacific Northwest surface waters have demonstrated that—as is the case commonly for many environmental measurements, including water quality parameters—their distribution over time (e.g., a typical annual cycle) is skewed and reasonably represented as approximately lognormal (Chou 1954). For this paper, the Haas and Rose *Cryptosporidium* data were used to calculate occurrence over the 52-week sampling period in terms of raw numbers of oocysts/L from the tabulated sample volumes and then examined in terms of a cumulative probability distribution. The resulting distribution was approximately lognormal (Fig. 1). The distribution was truncated by the limit of detection but its essential characteristics, median and standard deviation, were readily discernible (Ongerth 2013b).

The Pacific Northwest reservoir data in this form were then added to a similar presentation of data from other reservoirs elsewhere in the region (Fig. 2). Previously published data representing the other distributions of *Cryptosporidium* and *Giardia* in Fig. 2 are specifically concentration data calculated from observed numbers/L and the recovery efficiency measured with each observation. Accordingly, the Haas and Rose data are likely a factor of 4 to 5 or more lower simply because of the typical recovery of the ICR method of approximately 5% compared with the 20–30% measured for the published data.

Comparison with other data from reservoirs in the U.S. Northeast provides an interesting example (Fig. 3). The comparison in this case is for *Giardia*, using data from the New York City Department of Environmental Protection (NYC DEP) website (http://www.nyc.gov/html/dep/html/drinking_water/pathogen.shtml). The year of observations selected for the comparison was 2002, which was the closest period to that of the Haas and Rose data following the NYC beginning analysis by Method 1623. The NYC DEP

Table 1. Cryptosporidium Oocysts Found in 52 Weekly Samples from a Pacific Northwest Reservoir

Number	Sample volume[a] (L)	Oocysts[a]	Oocysts (number/L)
1	100	3	0.0300
2	98.4	3	0.0305
3	89.8	3	0.0334
4	227.1	2	0.0088
5	223.7	2	0.0089
6	223.7	2	0.0089
7	95.8	2	0.0209
8	193	1	0.0052
9	183.5	1	0.0054
10	101.3	1	0.0099
11	101.1	1	0.0099
12	100	1	0.0100
13	100	1	0.0100
14	100	1	0.0100
15	100	1	0.0100
16	100	1	0.0100
17	99.9	1	0.0100
18	74.1	1	0.0135
19	18.4	1	0.0543
20–52	+33 samples having volumes 48–191.4 L	0	+33 at 0.0

[a]Data from Haas and Rose (1996).

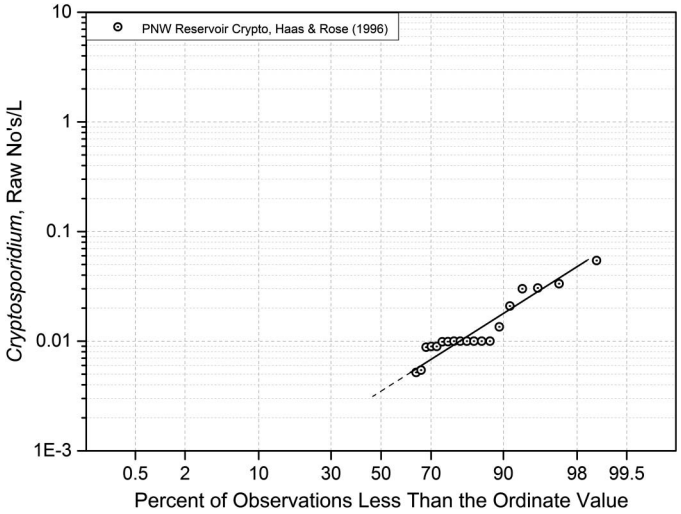

Fig. 1. Cumulative frequency plot of *Cryptosporidium* raw numbers/L in 52 weekly samples from a water supply reservoir in the U.S. Pacific Northwest (data from Haas and Rose 1996)

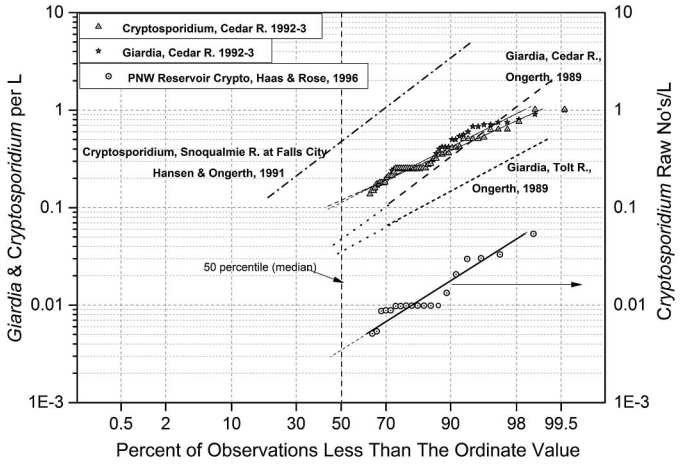

Fig. 2. Cumulative frequency plot of Pacific Northwest *Cryptosporidium* raw numbers/L with similar plots of previously published *Cryptosporidium* and *Giardia* concentrations

website also includes data on *Cryptosporidium* occurrence that are nearly an order of magnitude lower than for *Giardia*. That places them in the same range as the Pacific Northwest reservoir data, although the variability of the latter (Fig. 2) appears to be characteristically higher than observed at the NYC reservoir sampling sites.

The 1996 analysis of Haas and Rose interpreting the distribution of *Cryptosporidium* occurrence as conforming to the Poisson model was supported by statistical evaluation using the Fisher chi-square test. Thus, the researchers concluded (p. 2253) that the methods of analysis in use for *Cryptosporidium* did not convey "excessive variation relative to intrinsic variability in environmental samples." Haas and Rose asserted that their analysis in part validated "the use of Poisson-based assumptions" in assessing risk and treatment needs for similar waters. Although this may indeed be true, it is important that a distinction be drawn between the typically skewed, often lognormally distributed, occurrence of *Cryptosporidium* and *Giardia* in annual data sets and the classically Poisson-distributed oocysts and cysts in water samples from a homogeneous water system at any time at which the ambient concentration is near the limit of detection with the analytical method.

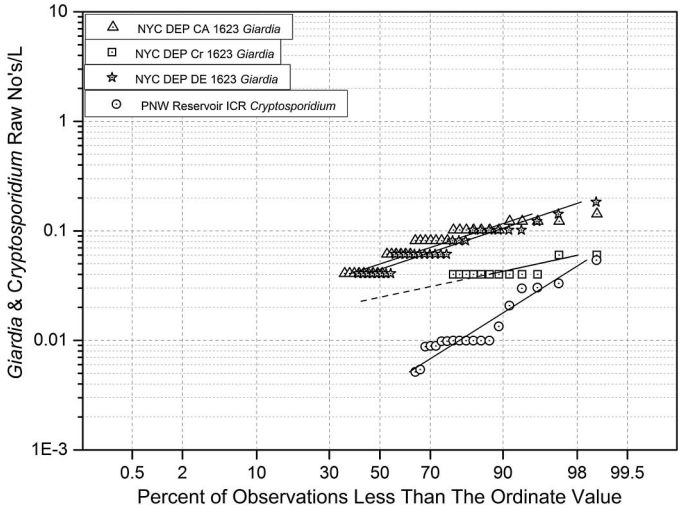

Fig. 3. Cumulative frequency plot of Pacific Northwest *Cryptosporidium* raw numbers/L with similar plots of *Giardia* raw numbers/L from three sampling sites at discharge locations from NYC DEP water supply reservoirs for the year 2002

Summary and Conclusions

The analytical method for *Cryptosporidium* that produced the 1996 data has indeed been significantly improved by general adoption of USEPA Method 1622/23. Improvements in analysis methods for *Cryptosporidium* and *Giardia* have undoubtedly increased the accuracy with which the presence of these organisms can be measured. Nevertheless, it is important to recognize the difference between the Poisson distribution of oocysts and/or cysts in a water sample of any defined volume taken from a surface water at a specific time and place and the skewed distribution of *Cryptosporidium* and *Giardia* occurrence (better yet, concentrations) over a typical annual data set, conforming to any statistical model. The critical importance lies in planning for sampling and interpretation of analytical results. Understanding the relation between sample volume and concentration in light of Poisson distribution principles makes the difference between successful and unsuccessful results of any monitoring program. Analysis of sample volumes that are small relative to the occurrence of the target organism inevitably result in a high proportion of negative (zero) results (Ongerth and Saaed 2013). The problem of determining how large a sample volume must be to be representative is specific to individual sampling locations and their specific water quality and variations over time. Because, at the outset of monitoring, the concentration of the target organism is not known, adequate sample volumes must be determined on a trial-and-error basis, adjusting sample volume to produce required data in light of prevailing water quality conditions. Variation in sample volumes during a monitoring program to accommodate changes in water quality conditions may be required.

References

Chow, V. T. (1954). "The log-probability law and its engineering applications." *Proc. Am. Soc. Civ. Eng.*, 80(11), 1–25.

Haas, C. N., and Rose, J. B. (1996). "Distribution of *Cryptosporidium* oocysts in a water supply." *Water Res.*, 30(10), 2251–2254.

Hansen, J., and Ongerth, J. E. (1991). "Effects of time and watershed characteristics on the concentration of *Cryptosporidium*

oocysts in river water." *Appl. Environ. Microbiol.*, 57(10), 2790–2795.
Messner, M (2011). "LT2 round 1 *Cryptosporidium* matrix spike recovery." ⟨http://water.epa.gov/lawsregs/rulesregs/sdwa/lt2/upload/lt2round1crypto.pdf⟩ (Dec. 7, 2011).
Ongerth, J. E. (1989). "*Giardia* cyst concentrations in river water." *J. Am. Water Wks. Assoc.*, 81(9), 81–86.
Ongerth, J. E. (2013a). "LT2 *Cryptosporidium* data: What do they tell us about *Cryptosporidium* in surface water in the United States?" *Environ. Sci. Technol.*, 47(9), 4029-4038.
Ongerth, J. E. (2013b). "The concentration of *Cryptosporidium* and *Giardia* in water: The role and importance of recovery efficiency." *Water Res.*, 47(7), 2479–2488.
Ongerth, J. E., and Saaed, F. M. A. (2013). "Distribution of *Cryptosporidium* oocysts in water above and below the normal limit of detection." *Parasitol. Res.*, 112(2), 467–471.
USEPA. (1995). "ICR protozoan method for detecting *Giardia* cysts and Cryptosporidium oocysts in water by a fluorescent antibody procedure." *EPA/814-B-95-003*, Washington, DC.
USEPA. (2005). "Long term 2 enhanced surface water treatment rule." ⟨https://nepis.epa.gov/Exe/ZyPDF.cgi/20017JBU.PDF?Dockey=20017JBU.PDF⟩ (Sep. 12, 2016).
USEPA. (2006). "Method 1623: *Cryptosporidium* and *Giardia* in water by filtration/IMS/FA." ⟨https://www.epa.gov/sites/production/files/201507/documents/epa-1623.pdf⟩ (Sep. 10, 2016).