# LT2 *Cryptosporidium* Data: What Do They Tell Us about *Cryptosporidium* in Surface Water in the United States?

Jerry E. Ongerth*

Civil, Mining, and Environmental Engineering, CME Bldg 4 EEC, University of Wollongong, Wollongong NSW 2522, Australia

ⓢ *Supporting Information*

ABSTRACT: Beginning in 2006 a United States Federal regulation required public water suppliers using surface water serving more than 10,000 population to analyze for *Cryptosporidium* in at least 24 consecutive monthly samples from each surface water source. In July 2012, the U.S. EPA released the resulting data consisting of ca. 45,000 records. No *Cryptosporidium* were found in 93% of samples and no *Cryptosporidium* were found in any samples analyzed from over half of 1670 locations sampled. Nevertheless, at 250 locations representing every region of the U.S., *Cryptosporidium* were found in sufficient numbers of samples to provide a picture of their occurrence nationwide. Data from about 100 sites reporting the highest numbers were examined in detail. Although analysis of matrix spikes was required for quality control, the results do not permit estimating organism concentrations. The data reported at each of the individual sample locations were analyzed in the form of cumulative probability distributions to describe key risk-related features of median level and variability. Taken as a whole, the data describe a spectrum of median *Cryptosporidium* occurrence in surface waters of the U.S. ranging from ca. 0.005 to ca. 0.5 oocysts/L. The variability at individual sites ranged from ca. 1 to 15 r.s.d. Based on the LT2 positive data, comparison to measurements of other water quality parameters, and independent means of estimating organism production from watersheds reported in the literature, the hypothesis is offered that *Cryptosporidium* may be found in surface water anywhere worldwide continuously and within the spectrum defined above.
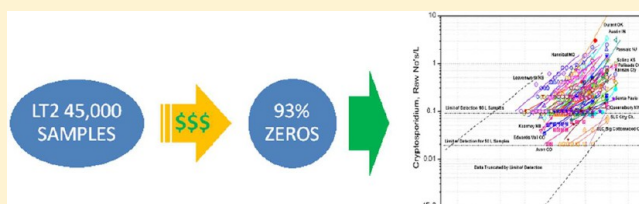
1. INTRODUCTION

Following waterborne outbreaks of cryptosporidiosis in Swindon, England,[1] Carrolton, GA,[2] and Milwaukee WI[3] and on the heels of the initial phase of the AIDS epidemic in which cryptosporidiosis played an unfortunate role (accounting for ca. 5% of AIDS mortality until the advent of modern chemotherapy),[4] the U.S. EPA, working with the U.S. public water supply industry, developed and implemented a regulation to control *Cryptosporidium* in public water supplies. The regulation is the Long-term Stage 2 Enhanced Surface Water Treatment Rule (LT2ESWTR) known simply as LT2.[5] A major part of LT2 required all public water suppliers using surface water sources and serving populations >10,000 to monitor their sources for *Cryptosporidium* by analyzing at least 24 consecutive monthly samples. Resulting data from 1670 sampling sites were released by EPA in July 2012.[6] This paper will focus on site-by-site analysis of these unique U.S.-wide data on *Cryptosporidium* occurrence following a brief summary of factors unique to *Cryptosporidium* as a human pathogen and its relatively recent recognition and addition to the list of contaminants regulated in drinking water. The LT2 data as analyzed below provide a new and thought provoking perspective for further regulatory action related to *Cryptosporidium*.

1.1. Human Cryptosporidiosis. The entire history of cryptosporidiosis affecting the human population as recorded in the scientific literature is remarkably brief. The first case of human cryptosporidiosis was reported in 1976[7] although

subsequent evidence indicates that it undoubtedly evolved in forms infectious to humans along with human evolution.[8] Its lack of earlier recognition may be attributed to (1) relatively innocuous aspects of the illness similar to viral diarrheas acquired typically in early childhood, and (2) the lack of effective analytical and diagnostic tools. In the period from 1995 to 2008 cryptosporidiosis occurred in the U.S. population at a rate of ca. 1−3/100,000 per year. [9] In the same period giardiasis in the U.S. population averaged ca. 10/100,000 per year. [10] Relevant to the significance of *Cryptosporidium* (and *Giardia*) in drinking water, a recent review has summarized over 500 waterborne outbreaks of protozoan illness worldwide, with the vast majority attributed to *Cryptosporidium* (60%) and *Giardia* (35%).[11]

1.2. Analysis of *Cryptosporidium* and *Giardia* in Water. The first effective clinical assay for *Cryptosporidium*, a modified acid-fast stain, was reported by Bronsdon in 1984.[12] Shortly thereafter an IFA for *Cryptosporidium* was developed[13] and applied to finding *Cryptosporidium* in surface water.[14] Over the next 15 years surveys revealed the presence of *Cryptosporidium* (and *Giardia*) in surface water throughout the U.S. and elsewhere.[15−17] Analytical procedures evolved through this

period resulting in the current EPA Method 1623, now widely accepted as the best available for analysis of *Cryptosporidium* and *Giardia* in water.[18,19]

Some critical features of analyzing water samples important to understanding and interpreting data describing *Cryptosporidium* and *Giardia* occurrence include the following: (1) the recovery efficiency is finite, typically ranging from ca. 5% to 60% averaging ca. 30% for *Cryptosporidium* and ca. 45% for *Giardia*;[18] (2) recovery efficiency measured in samples from an individual sampling location varies significantly over annual cycles;[21] (3) recovery efficiency variations are unique to individual sampling locations even on the same surface water source;[22] (4) application of an analytical method, e.g. Method 1623, to a water sample has a limit of detection dependent only on the sample volume processed to completion and recovery efficiency, i.e.

$$\text{L.D.} = \frac{1 \text{ organism}}{\text{sample volume (L)} \times \text{recovery fraction}} \qquad \text{(Eq. 1)}$$

and (5) ambient concentrations have been found most often below the limit of detection (L.D.) using the most common 10 L minimum sample volume required by Method 1623.[23]

1.3. The LT2 Regulation. A principal objective of LT2 was to determine for individual surface water source locations throughout the U.S., the degree of risk due to the level of *Cryptosporidium* to be established through the required monitoring program. Based on data produced in previously required monitoring,[18] four risk categories were established, termed BINs.[24] Source water in the lowest BIN requires no additional treatment. Sources in successively higher BINS require successively increased treatment for risk control. BIN determination for any sampling point is based on the highest 12-month running average of *Cryptosporidium* numbers found per L of sample, not adjusted for recovery efficiency. Although LT2 required monitoring by water systems serving populations <10,000, only the large system data were released in 2012. These data are the exclusive subject of analysis herein. The LT2 regulation requires a second round of monitoring to commence in 2015−2016 and the regulation is currently in the formal 6-year regulatory review process, lending significance to understanding the first round monitoring data.

2.0. METHODS AND PROCEDURES

Methods and procedures for this work included elements of the LT2 regulation that specified what data were required to be collected and analytical procedures affecting the resulting data. Also described are procedures used in analyzing the LT2 data.

2.1. EPA LT2 Data Collection. The LT2 regulation includes detailed requirements for sampling, analysis, and reporting.[5] Briefly, for each individual sampling site monthly samples of at least 10 L were collected according to a preapproved schedule. Samples were analyzed for *Cryptosporidium* oocysts by a EPA-approved laboratory according to Method 1622. Method 1622 requires "matrix spike" (MS) recovery efficiency measurement accompanying one out of every 20 samples processed by each approved lab. The EPA expected that in the course of analyzing the required samples one MS sample would result for each sampling site per year providing some information on matrix effects at each site. However, MS recovery efficiencies were not used to calculate oocyst concentrations. In fact previous examination of the LT2 data found that among all 45,000 records, only 319 MS

measurements corresponded to a positive field sample (FS) result.[21]

2.2. LT2 *Cryptosporidium* Analysis. Method 1622 begins with sample collection by pumping the desired sample volume, most often ∼10 L (> 97% of all LT2 samples) through a pleated cartridge filter in the field. Filters are shipped on ice to the lab where they are eluted and centrifuged to pellet. Pellets are resuspended and oocysts are selectively concentrated using immuno-magnetic separation (IMS). Oocysts are released from the IMS beads in minimal volume, stained with fluorescent-labeled antibody, counter-stained to reveal oocyst internal contents, and deposited on glass slides. Finally, slides are examined by UV epi-ilumination microscopy to enumerate oocysts. Data were reported to EPA electronically using a Data Collection and Tracking system (DCTS).[24]

Data on findings of *Cryptosporidium* oocysts in the field samples (FS) were reported exclusively as raw numbers, termed occurrence in this paper. The fact that the data are not concentration, lacking incorporation of recovery efficiency, is critically important to understanding the ultimate meaning and interpretation of the data and any analysis of it including quantitative comparison among sites.[21] The compiled data were subjected to quality control procedures and the resulting data file was released on the EPA LT2 Web site on July 29, 2012. The file is in Microsoft Excel ".csv" format and includes about 50 columns of information on a total of about 45,000 field and matrix spike analyses from a total of 1670 sampling sites in 1375 systems in all 50 states and Puerto Rico. The Puerto Rico data (3378 records from 133 sites) were excluded from analysis in this paper.

2.3. Analysis of the EPA LT2 Data File. Analysis of the EPA LT2 data for this paper was performed using Microsoft Excel functions to sort the file to generate subfiles containing the data for individual sampling points at which multiple field samples contained *Cryptosporidium* oocysts. The entire LT2 file (primary file) as downloaded from the EPA Web site, http://water.epa.gov/lawsregs/rulesregs/sdwa/lt2/upload/cryptodatareported.csv, includes 45,033 file records. Sort criteria were established to permit identifying *Cryptosporidium* oocyst findings at individual sampling sites serving individual surface water source use points, and groupings by state representing limited geographical association. The primary file was sorted in the following order: (1) field sample (FS) and matrix spike (MS); (2) state; (3) sampling site code; (4) *Cryptosporidium* found, low to high. The resulting table of FS data was then manually scanned to identify sites having sufficient numbers of nonzero (positive) *Cryptosporidium* findings to permit analysis by forming log-probability distributions.

Previously published descriptions of *Cryptosporidium* and *Giardia* concentrations at individual sampling locations have identified two key features that are meaningful in terms of relative risk and permit quantitative comparison both in time (year to year) and from site to site.[25,26] Those features are the median concentration of an annual data set and the degree of variation represented by the slope of a log-probability plot from which the standard deviation can be deduced. Statistical distributions of oocyst occurrence for individual sampling sites were formed to identify key descriptive parameters, median level, and standard deviation, using OriginPro 8.6 (OriginLab, Northhampton, MA). Log-probability distributions were formed initially using the Excel table of FS data for an individual site. The numbers of oocysts reported were

normalized by dividing by the sample volume and ranked high to low. For each individual site the table of ranked occurrence/ L values including zeros was then exported to an OriginPro 8.6 worksheet. To construct an accurate distribution of the positive results for each site, the zero values, in reality below the limit of detection (LD)[23] were then replaced by dummy values at ca. 1-log below the lowest occurrence/L value for the site, e.g. 0.05, 0.01, 0.005. A log probability plot was then produced using the "Probability Plot" function under Plot → Statistics. Using the "Mask Data Points" function (Data → Mask Data Points) the dummy values were masked and then the line of best fit was established for the unmasked data (all points above the LD) using Analysis → Fitting → Linear Fit. Detailed statistical description of the resulting least-squares line of best fit was produced with the graph permitting objective comparison between limited sets of data from individual sites. As described previously, the Kruskal−Wallace (K−W) statistic was applied for pairwise comparison of oocyst observations between sampling sites.[25] The Kolmogorov−Smirnov test was also applied to test the difference between pairs of the distributions of oocyst observations at any selected pair of sample sites. All tests were made for a critical p-value of 0.05. Statistical calculations were made from the previously ranked site by site occurrence/L data using XLSTAT (Addinsoft, Paris FR) added on to Microsoft Excel.

3.0. DATA AND ANALYSIS

The LT2 data have been summarized by EPA. Subsequent to the LT2 data release by EPA, the data for individual sampling sites were examined to provide a perspective relevant to individual water systems as reported here. The EPA summary and individual site analyses are included below.

3.1. EPA LT2 Data Summary. In December 2012 the EPA presented a summary of the LT2 data submitted by water systems serving >10,000 population.[27] The summary was based on data compiled through November 2011 and did not included data submitted as "grandfathered". With only a few corrections the data summarized were the same as those released in July 2012 and analyzed herein. Data were submitted by a total of 1376 systems for 1670 sampling sites. The data comprised nearly 43,000 records consisting of analysis results from 39,676 field samples and 3,234 matrix spikes. More than half (51%) of facilities reported no Cryptosporidium-positive findings. Overall, no Cryptosporidium were found in nearly 93% (38,869 of 39,676) of the field samples analyzed. The EPA summary, Figure 1, did not include any analysis of data from individual sampling sites.

3.2. LT2 Individual Sampling Site Data. Minor unreconciled differences exist between summary totals derived from the LT2 data presented above and those developed for this work and cited below. Examination of the LT2 data file sorted as described above revealed a total of 1332 sampling sites for which 918 (69%) reported finding no Cryptosporidium. Of the remaining 414 sites reporting at least one Cryptosporidium positive result, 246 (18%) had at least 3 positive samples totalling 5 or more Cryptosporidium, judged to be the minimum for even the crudest estimation of slope from a log-probability plot. The data from these sites were used in further analysis.

Examining the distribution of sites at which Cryptosporidium were found in a significant portion of samples provides the first suggestion of anomalies in the LT2 data The numbers of sampling sites reported by State ranged from 2 (MS) to 172 (PA), with 7 states reporting less than 5 sites (AK, HI, ID, MS,
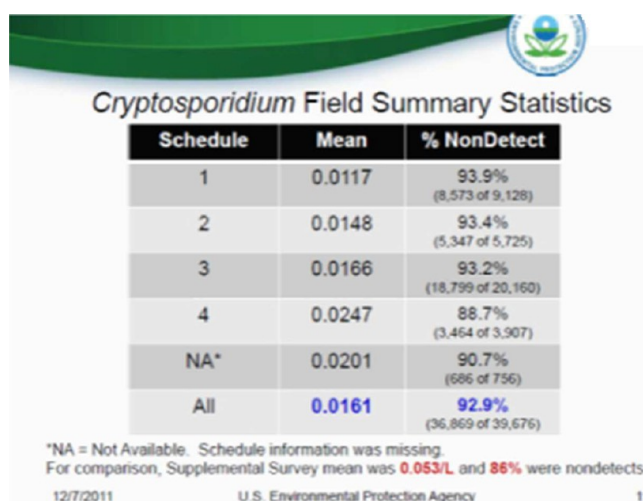

Figure 1. EPA summary of all U.S. LT2 Cryptosporidium occurrences. Reprinted from ref 27.

NE, RI, and WY) and 12 states reporting 50 or more sites (AL, CA, GA, KY, MA, NC, NY, OH, PA, TN, TX, and VA). No sites reporting positive findings above the minimum for analysis occurred for 7 of the 50 states: GA, HI, MS, NH, NM, RI, and VT. Eleven states (AL, AZ, FL, GA, ME, NH, NM, NV, NY, RI, and VT) reported >70% of sites that found no Cryptosporidium. In 6 states (DE, IN, KS, MO, NE, and UT) > 40% of sites reported numbers of Cryptosporidium-positive samples sufficient for further analysis. The geographic distribution of sites reporting multiple positive findings is provided by a map of the continental U.S. including the highest 160 sampling sites, Figure 2.

A summary of Cryptosporidium occurrence reported among all nonzero FS analyses provides a context for examining individual site data. In the entire 40,000 records of LT2 data a total of about 2900 positive findings were reported by about 400 of the more than 1600 LT2 sampling sites. About 150 of the 400 reported only between 1 and 4 positive samples and were not considered further. On a volumetric occurrence/L basis, i.e. Cryptosporidium numbers per liter of sample processed (Col AF/Col AD of EPA file), the vast majority of findings were of a single oocyst in sample volumes between 9 and 12 L, i.e. ca. 10 L. About 2000 of the 2900 C numbers/L observations were less than 0.12/L. In 450 additional samples two oocysts were found in volumes of close to 2 L or about 0.2/L. Numbers of oocysts greater than two declined steeply. A total of 57 samples from the entire LT2 data set had occurrence greater than 1/L reflecting 10 or more oocysts found in the most common volume of about 10 L. Only 17 of the samples were reported to have greater than 2/L with the maximum of 16/L. The distribution of oocyst findings independent of sample volume was: 1720 1's; 547 2's; 453 3−4; 117 5−9; and 58 of 10 or more.

The distributions of oocyst/L findings for individual sampling sites were pursued selectively. Initially, distributions were formed for sampling sites with the greatest proportion of nonzero FS results. The first group is presented to illustrate the features described above (Figure 3). It is apparent that all the distributions are truncated by the limit of detection imposed by individual sample volumes. The distribution for Hannibal, MO on the Mississippi River appears at the top of the spectrum of reported LT2 oocyst occurrence [O], numbers/L, having a
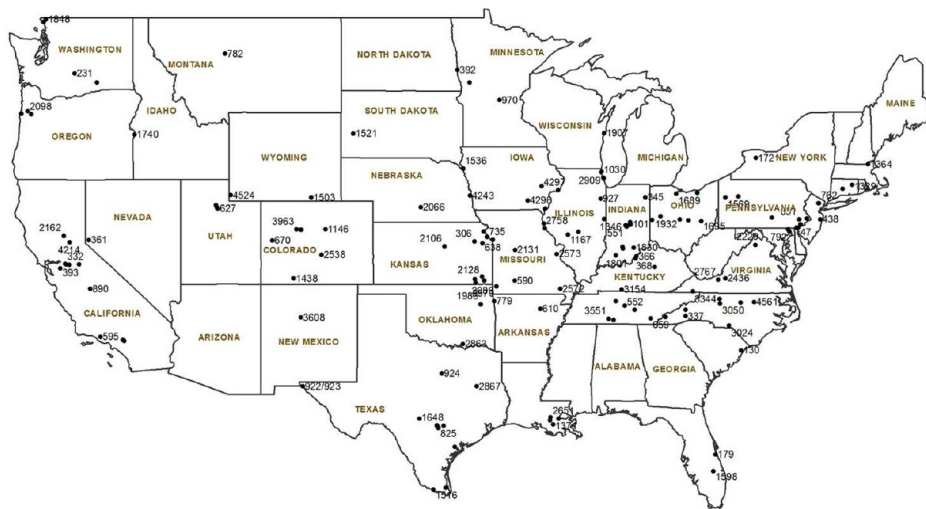
Figure 2. Locations of 160 LT2 sampling sites reporting highest numbers of nonzero *Cryptosporidium* oocyst FS findings.
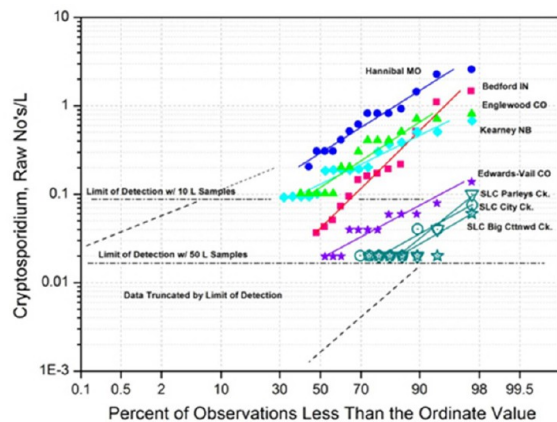

Figure 3. Log probability distributions of LT2 *Cryptosporidium* occurrence data for eight sampling sites in the central U.S. illustrating key features.

median level of ca. 0.25/L and a relative standard deviation RSD, $([0_{84.16\%ile}] - [0_{50\%ile}])/(0_{50\%ile})$, of ca. 4.3. The distributions for the three Salt Lake City (SLC) sites appear at the lower end of the spectrum. The sample volume used in all SLC samples was 50 L permitting measurement of occurrence at a limit of detection 1/5 of that used at Hannibal MO. The occurrence levels above the LD for the SLC distributions were thus measurable at lower levels although the distributions were still truncated requiring extrapolation to estimate a median value. Median SLC levels were ca. 0.003–0.006/L with variability indicated by RSD's little different from that of Hannibal MO. The Edwards-Vail, CO data were also derived from 50 L samples with an estimated median level ca. 0.015/L and an RSD similar to those of Hannibal and SLC sites. The Bedford, IN data were derived from 30 L samples having a LD between that of Hannibal and most other sites and that of the SLC and Eagle-Vail samples. The variability observed at Bedford, IN was clearly greater than the other sites in this example with an RSD ca. 9. The other sites included in Figure 5 are Englewood, CO on the South Platte and

Table 1. Significant Differences in *Cryptosporidium* Occurrence/L between Sampling Sites by K–W Statistic, $p \leq 0.05$

	Hannibal MO	Kearney NE	Bedford IN	Englewood CO	Eagle-Vail CO	SLC City Creek	SLC Parleys	SLC Big Cttnwd Ck
Hannibal MO		yes	no	no	yes	yes	yes	yes
Kearney NE	yes		no	no	yes	yes	yes	yes
Bedford IN	no	no		no	no	yes	no	yes
Englewood CO	no	no	no		yes	yes	yes	yes
Eagle-Vail CO	yes	yes	no	yes		no	no	yes
SLC City Ck	yes	yes	yes	yes	no		no	no
SLC Parleys Ck	yes	yes	no	yes	no	no		no
SLC Big Cttnwd	yes	yes	yes	yes	yes	no	no	
p-values of K–W statistic comparisons for *Cryptosporidium* occurrence/L between pairs of sampling sites, $p \leq 0.05$								
	Hannibal MO	Kearney NE	Bedford IN	Englewood CO	Eagle-Vail CO	SLC City Creek	SLC Parleys	SLC Big Cttnwd Ck
Hannibal MO	1	0.008	0.065	0.143	0.000	0.003	0.011	0.005
Kearney NE	0.008	1	0.991	0.880	0.002	0.002	0.024	0.003
Bedford IN	0.065	0.991	1	0.909	0.274	0.014	0.071	0.015
Englewood CO	0.143	0.880	0.909	1	0.003	0.002	0.017	0.004
Eagle-Vail CO	0.000	0.002	0.274	0.003	1	0.060	0.234	0.019
SLC City Ck	0.003	0.002	0.014	0.002	0.060	1	1.000	0.987
SLC Parleys Ck	0.011	0.024	0.071	0.017	0.234	1.000	1	0.994
SLC Big Cttnwd	0.005	0.003	0.015	0.004	0.019	0.987	0.994	1

D

dx.doi.org/10.1021/es4006509 | *Environ. Sci. Technol.* XXXX, XXX, XXX–XXX

Kearney, NE downstream on the Platte River below the confluence of the North and South forks. Direct comparisons between sites are simply not valid due to the lack of recovery efficiency measurement and its use to calculate concentration. Their median occurrence differed little although the distribution at Englewood, CO reflected somewhat more variation.

As an example, statistical comparisons were made for this group of example sites. Results indicated that the observations between most sites were distinguishable at the level of $p = 0.05$ (Table 1). The exceptions were between observations at Hannibal and both Englewood and Bedford though not Kearney, and between Eagle-Vail and both SLC City Ck and SLC Parleys Ck though not SLC Big Cottonwood Ck. The three SLC sites were not significantly different from each other.

As preparation of the distributions proceeded they were grouped into geographic regions including single river groups for sites on the Mississippi and the Missouri River (Figures 4
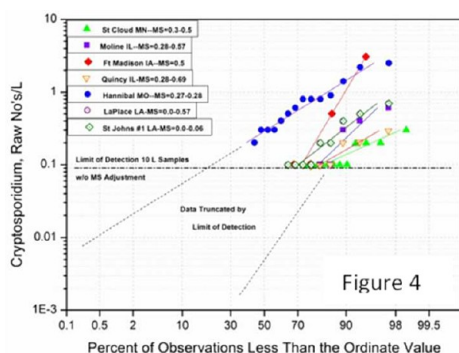

Figure 4. Log probability LT2 *Cryptosporidium* oocyst distributions at seven sampling sites on the Mississippi River and six sampling sites on the Missouri River.

Figure 5. Log probability LT2 *Cryptosporidium* oocyst distributions at seven sampling sites on the Mississippi River and at six sampling sites on the Missouri R.

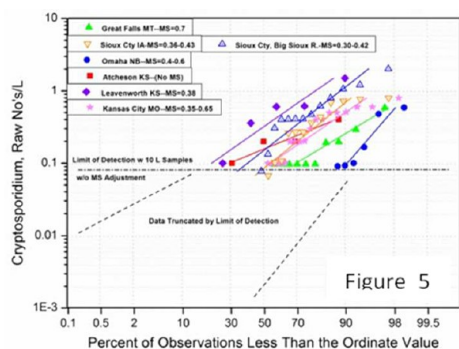and 5). *Cryptosporidium* occurrence at sites on the Mississippi River from St Cloud, MN downstream to just above New Orleans, LA follows a generally increasing trend with highest oocyst occurrence at Hannibal, MO but with particularly low occurrence at the LA sites. This inconsistency illustrates the inability to make direct quantitative comparison between data not expressed as concentration taking recovery efficiency into account. *Cryptosporidium* occurrence data at sites on the Missouri R, from Great Falls, MT downstream to Kansas City, MO ranged from a median of ca. 0.003 raw numbers/L at

Omaha, NE to 0.03 raw numbers/L at Leavenworth, KS. As with all the LT2 data, quantitative comparisons cannot be made and apparent relations between occurrence from site to site can only be made on a tentative basis, taking the few MS recovery observations into account in a general way.

Cryptosporidium occurrence distributions for more than 100 additional sites in 28 states across the U.S. are shown in Figure 6a−f and Figure 7a−e). Features of importance contained in the figures include the following: (1) The range (spectrum) of median occurrence levels spanned about 2 logs from ca. 0.002 to 0.2 raw numbers/L. (2) The variability ranged from sites having little variation and an RSD ca. 1.0, e.g. Carson City, NV, to sites like Durant. OK having an RSD ca. 12. (3) Data at all sites were truncated by the limit of detection dependent only on the sample volume since recovery efficiencies were not measured or taken into account. (4) Occurrence levels at sites for which sample volumes >10 L were analyzed enabled detection of oocyst occurrence at lower levels approaching 0.01 raw numbers/L for 50 L samples. (5) The range in occurrence and the degree of variability appeared to have some regional similarity, although, due to the lack of meaningful recovery efficiency measurement and application to calculate concentrations, the distributions must only be considered as indicative of occurrence at individual sites. The typically two reported MS values per site were included with the site location on each of the Figures 4 through 8, which summarize the oocyst occurrence distributions. The MS values can be taken into account in a general way in an effort to make the most appropriate comparison between sites.

A compilation of occurrence distributions from 50 representative sites across the U.S., Figure 8, clearly suggests a spectrum of occurrence and range of variability for *Cryptosporidium* occurrence virtually anywhere in the country. The highest levels of occurrence were reported at sites on the Mississippi River at Hannibal, MO and on the Missouri River at Lawrence, KS. Lowest *Cryptosporidium* oocyst occurrence levels were reported at sites in the mountain west. Sites for which larger volumes were analyzed provided lower limits of detection, e.g. SLC City Ck, Parleys Ck, and SLC Big Cottonwood Ck, and revealed occurrence at lower levels. The compilation also indicates that the other major risk-related parameter, slope or variability also had a limited range from as low as ca. 1.0 to as much as 12−15 at sites such as Durant OK, Austin IN, Passaic NJ, and Jefferson Co, MO.

4.0. DISCUSSION

The EPA LT2 data are a unique and valuable resource describing *Cryptosporidium* occurrence at a wide range of surface water locations distributed throughout the U.S. The sampling approach, including timing, and the analytical requirements, including laboratory certification and quality control, produced data that provide a view of *Cryptosporidium* occurrence not available anywhere else and not likely to be duplicated except for the prospect of a second sampling under the LT2 regulation.

Taken as a whole subdivided only by population-served categories, surface water type, and treatment applied, the data appear disappointing having a predominant proportion of negative results, i.e. 93% zeros overall and no organisms found at over half of sites sampled. Characterizing *Cryptosporidium* occurrence, as in the EPA summary, statistically representing features of the entire or subdivided by population is therefore dominated by the negative results. This approach on the one
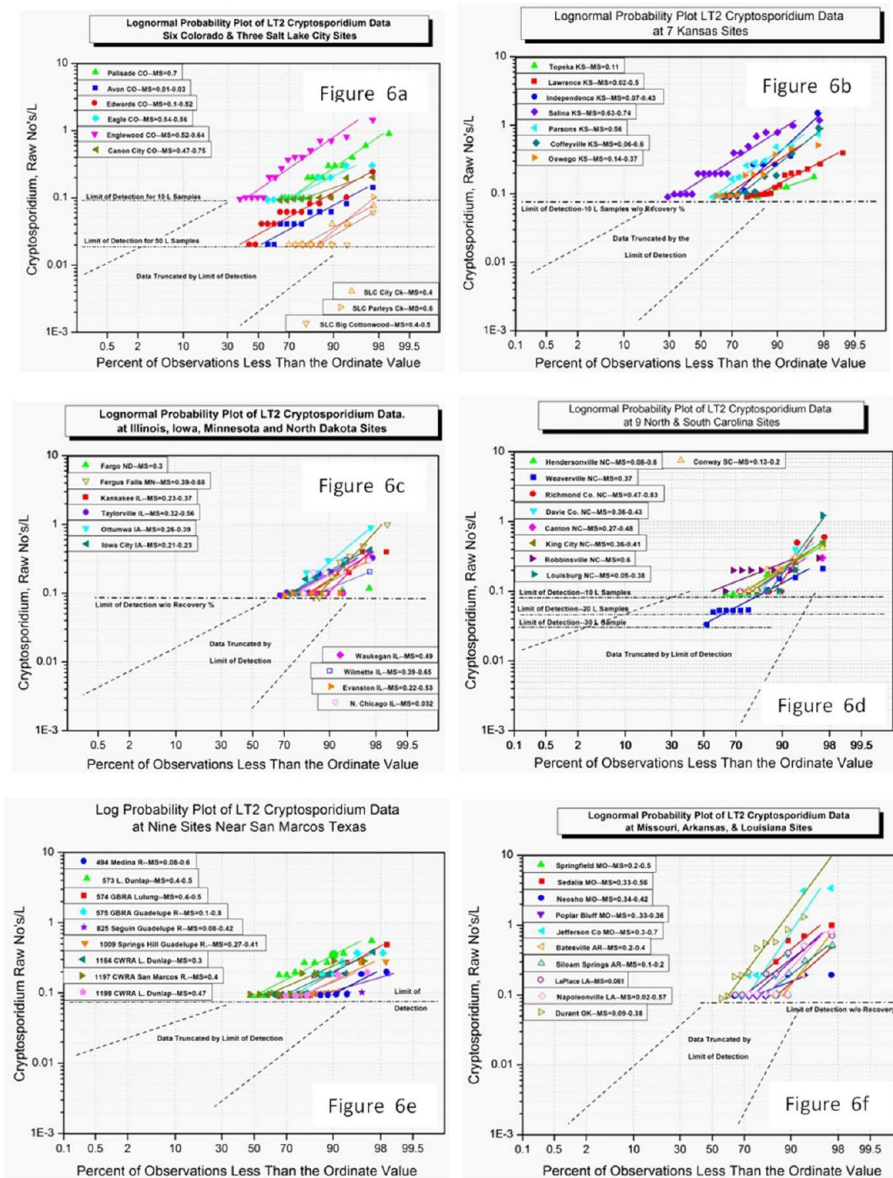
E

dx.doi.org/10.1021/es4006509 | *Environ. Sci. Technol.* XXXX, XXX, XXX–XXX

Figure 6. Log probability plots of *Cryptosporidium* oocyst occurrence, raw numbers/L, including site locations and reported MS recovery fractions, at six regional groupings of LT2 sampling sites: (a) mountain west, (b) Kansas, (c) upper midwest, (d) Carolinas, (e) Texas, and (f) lower Mississippi.

hand cannot avoid minimizing apparent occurrence and on the other hand has little meaning or utility for individual water providers using water sources having definable individual characteristics.

Examination of the LT2 data site-by-site indicates that *Cryptosporidium* occurrence has site-specific characteristics. Even in geographically related systems, e.g. Mississippi River, Missouri River, and Ohio River, *Cryptosporidium* occurrence has features of individual populations. Combining data from disparate sites implies homogeneity in the population inconsistent with individual watershed factors including *Cryptosporidium* oocyst sources from animals both wild and domestic, human sources including sewage discharges, and geo-climatic factors that control the transport and distribution of this organism in surface water.[21]

The LT2 data from individual sampling sites at which multiple positive results were reported describe distributions of *Cryptosporidium* oocyst occurrence in annual data sets from single locations similar to both *Cryptosporidium* and *Giardia*

concentration data reported previously.[25,26] The log-normal occurrence distributions are also analogous to the distribution of other water quality parameters widely reported and familiar to environmental scientists and engineers, e.g. total coliforms.[26] Indeed, it would be both remarkable and difficult to understand were *Cryptosporidium* oocyst distributions from individual sampling locations not similar to those of other similar water quality parameters.

Perhaps the most interesting and important feature of the LT2 data compiled from the individual multiple-positive sites is the clear description of a spectrum of *Cryptosporidium* occurrence having both upper and lower limits. Based on even the most intuitive understanding of watershed and water quality features associated with the individual sites, the appearance of highest *Cryptosporidium* oocyst occurrence at a downstream site on the intensively used Mississippi River below the confluence of both the Missouri River and the Ohio River is not surprising. The tributary drainage area is both very large (ca. 175,000 mi^2) and includes extensive sources from
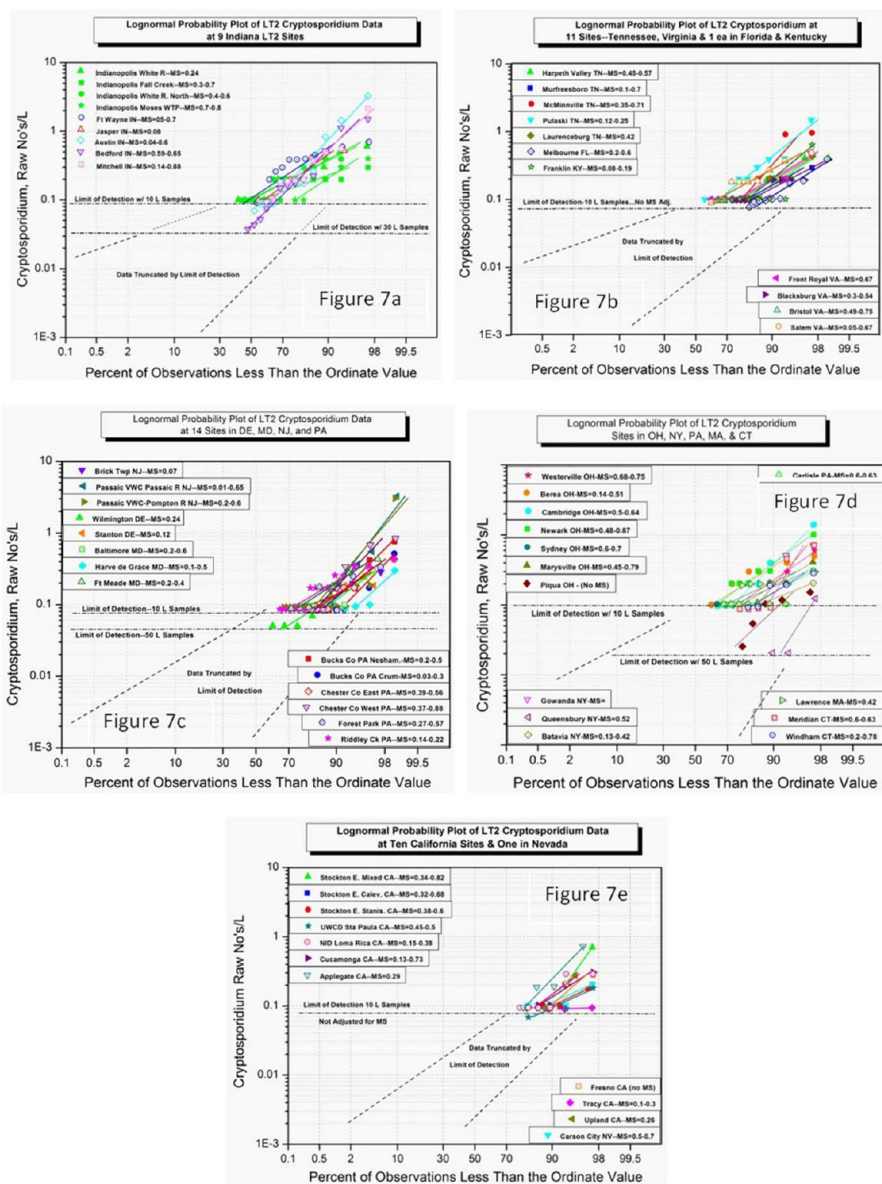
Figure 7. Log probability plots of *Cryptosporidium* oocyst occurrence, raw numbers/L, including site locations and reported MS recovery fractions, at six regional groupings of LT2 sampling sites: (a) Indiana, (b) Appalachia, (c) northeast, (d) mid Atlantic, and (e) west.
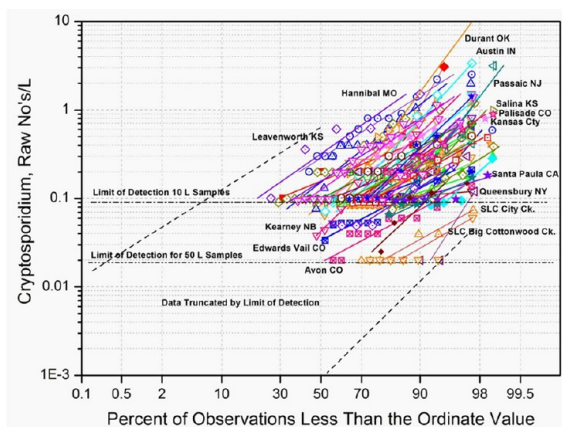

Figure 8. Compilation of log-normal probability distributions of LT2 *Cryptosporidium* occurrence data for 50 representative U.S. sampling sites.

natural, agricultural, and municipal activities. Similarly, it should not be surprising to find the appearance of lowest *Cryptosporidium* oocyst occurrence at sites on streams from mountain watersheds having only sparse human activity, e.g. Salt Lake City's City, Parleys, and Big Cottonwood Creeks.

Comparison of the 2- to 3-log range in *Cryptosporidium* occurrence, Figure 8, can be made to other parameters describing the quality of surface water.[28] For example, the entire range of TDS in rivers of the U.S. is from <10 mg/L to rarely >1000 mg/L. [29] Total suspended solids concentrations for public water supply sources should be below ca. 100 mg/L and only occasionally higher to permit effective conventional water treatment. The range in occurrence variability also appears to follow intuitive expectations. Highest variability is apparent in the distributions at sites on relatively small watersheds. Overall, the range in variability is remarkably low with similarities between sites more evident than differences. Caution should be used in effort to make more quantitative comparisons because of the lack of recovery efficiencies as was

mentioned above. Yet, the information on *Cryptosporidium* occurrence from any of the sampling sites included in the multiple positive LT2 data should be of interest and value to any water utility using the water as it describes oocyst occurrence at each site relative to the others on reasonably equivalent terms. Those terms are the ones imposed by equivalence of sampling plans and analytical requirements.

The value and utility of statistical comparisons of LT2 data from different sites is an important issue, and understanding the limitations of both the data and of statistical comparisons is equally important. The occurrence data at individual sites describe distributions consistent with previously published observations of both *Cryptosporidium* and *Giardia* concentrations.[25,26] Whether or not the distribution of observations at one specific site can be distinguished statistically from the distribution at any other site is largely beside the point in the context of the nation-wide LT2 data. If indeed the true concentration distributions, represented in the LT2 data by occurrence, of *Cryptosporidium* oocysts at all watersheds throughout the U.S.—or any other large geographic area—describe a spectrum, i.e. a continuous range of distributions, the distributions within a sequence of slices of the spectrum from high to low at some minimal width will not be statistically distinguishable, while distributions in individual non-adjacent slices will be. The existence of a spectrum is clearly illustrated in the LT2 data (Figure 8) and the range from high to low occurrence is consistent with general characteristics of the associated watersheds from heavily used downstream locations associated with high occurrence to upstream and pristine watersheds associated with low occurrence. More detailed statistical comparisons will be of more value and importance with the meaningful measurement of recovery efficiency and development of concentration data. An important goal in planning and implementation of *Cryptosporidium* and *Giardia* monitoring for water supply agencies (and regulatory agencies) should be to provide concentration data for individual sampling locations with a minimum of negative results that will provide statistical power for discriminating between sampling locations and time periods. An important regulatory question for which statistical comparison is critical is the assignment of sampling results to BIN categories. The LT2 data suggest that assignment based on the existing data would be difficult to support.

The feature of the LT2 data most difficult to understand and to explain is the appearance of so many sites at which no positive findings were reported, particularly when sites reporting multiple positives were located in the same region and often on the same or at least similar streams in the same region. Several types of information can be used to shed light on this issue. Previously published data have indicated widespread distribution of *Cryptosporidium* in U.S. surface waters.[15,16,20] Considering the distribution of sites represented in the LT2 data (Figure 2), examination of watershed characteristics is relevant. Previous reports from sampling on watersheds including a range of human and animal *Cryptosporidium* (and *Giardia*) source activity have shown a relation between those watershed characteristics and measured oocyst (and cyst) concentrations.[25,26] These early reports included calculations of *Cryptosporidium* oocyst and *Giardia* cyst production rates for the watersheds represented providing a basis for comparison to watersheds elsewhere. *Cryptosporidium* oocyst production rates ranged from 2×10^6 oocysts per mi^2/ day from a mountain watershed having little human activity to

2×10^8 oocysts per mi^2/day at a downstream location 25 miles below a community and dairy farming areas. *Cryptosporidium* production rates can be estimated from LT2 data on median occurrence numbers/L. Production rates were calculated for sites on of the Mississippi River from St. Cloud, MN to Hannibal, MO, based on USGS gauged flow rates (Table 2).

Table 2. Estimated Mean Annual *Cryptosporidium* Oocyst Production Rates for Watershed Areas Tributary to Three LT2 Sampling Sites on the Mississippi River Based on LT2 Data and on USGS Gauging Data and Watershed Areas

sample code-location	*Cryptosporidium* median numbers/L	watershed area, mi^2	avg. annual USGS[a] flow rate, ft^3/sec	*Cryptosporidium* rate, number/mi^2/day
970-St Cloud MN	0.05	13,320	6,515	0.151×10^6
2713-Moline IL	0.02	85,600	51,204	0.072×10^6
1977-Hannibal MO	0.25	171,300	126,700	1.809×10^6

[a]USGS Surface-Water Annual Statistics for the Nation, http://waterdata.usgs.gov /nwis/annual.

For example, using the closest USGS gauge data, i.e. for Grafton, IL, watershed area = 171,000 mi^2, and assuming a typical *Cryptosporidium* oocyst MS recovery rate for Hannibal of 25%, would give an estimated mean annual production rate of ca. 1.8×10^6 oocysts per mi^2/day. Pursuing this approach should permit estimation of approximate *Cryptosporidium* and *Giardia* concentrations from basic watershed characteristics for any surface water intake location at which the watershed area and mean flow rate were known for corresponding *Cryptosporidium* concentration measurements. This would be an approach useful to water suppliers using water from a location from which no *Cryptosporidium* or *Giardia* data are available. Information on the watershed is available from water source evaluation reports required as part of the Surface Water Treatment Rule.

It might be suggested that *Cryptosporidium* occurrence at the all-zero sites was simply an extension of the spectrum describing distributions at progressively lower levels of occurrence than illustrated for example in Figure 8. Information on the limited concentration spectrum of other water quality parameters and the similarity of watershed characteristics and particularly the extent and distribution of *Cryptosporidium*-generating sources in similar watersheds represented among the multiple-positive LT2 sampling sites argue against occurrence of *Cryptosporidium* distributions at levels significantly lower than those described by observations from LT2 sites at the lower end of the occurrence spectrum (Figure 8).

Using the 11 geographically related distribution data (Figure 6a−f and Figure 7a−e) individual sampling site comparisons could be made between sampling sites upstream and downstream of the multiple-positive sites. Where this has been done, (in Colorado; along the Mississippi, Missouri, and Ohio Rivers; in Ohio, Indiana, and Michigan; between South Carolina, Louisiana, Mississippi, and Alabama), no features of watershed characteristics suggesting the absence of *Cryptosporidium* sources to surface water can be found. Other related questions include: Why only six sites with positives on the Missouri River and only five on the Mississippi River to Hannibal, MO? Why

were positives found only at Louisville on the Ohio River and why not at Cincinnati or other upstream and downstream locations? Why did occurrence increase down the Eagle River to Edwards but none were found further downstream at Livingston or Grand Junction in Colorado? Why were so many positives reported in North Carolina, Ohio, Indiana, and Illinois but so few in Wisconsin, Michigan, and New York?

A small list of factors potentially accounting for the all-zero sampling sites includes the following. First, as previously observed, sampling and analysis for *Cryptosporidium* oocysts in water is complex and challenging. On the one hand, Method 1622/1623 is an effective and reliable method and is supported by QA/QC, laboratory certification, and oversight by the responsible regulatory agencies. Prior to development of LT2, or even priot to the Information Collection Rule (ICR) applied to the earlier stages of the Surface Water Treatment Rule preceding LT2, only a modest number of qualified laboratories experienced in analysis of *Cryptosporidium* and *Giardia* existed, barely more than a dozen, divided among University research, large water utility, and commercial laboratories. Experience among this group was developed largely through examining individual water sources with which each lab developed familiarity and a basis for understanding water quality at specific sites and what to expect in terms of *Cryptosporidium* and *Giardia* occurrence. The basic validity of individual analytical results could be assessed based on recent experience and knowledge of current water quality conditions at the site.

The leap from such experience to production line analysis of nearly 45,000 samples by nearly 100 laboratories in a period of essentially 5 years is huge. Analytical results of any individual sample became anonymous without relation to any site or water quality base. Responsibility for the entire process, from sampling to reporting, was distributed with little if any opportunity for feedback based on an assessment of the data being generated. In addition, LT2 required that the data being generated be reported to the DCTS system as-produced, minimizing opportunity for evaluating whether or not results were consistent with expectations. In addition, because of the risk mitigation requirements of LT2 based on BIN level, at least a minor incentive in favor of negative findings was present. These observations are not intended as "Monday morning quarterbacking" and indeed the potential for explaining an effect on analytical results is only speculative. It would have been difficult if not impossible to foresee if indeed such factors might actually account for some or even many of the all-zero sampling site results.

The two features of the process leading to the LT2 data as they exist that have objectively contributed to mostly zero findings clearly include the following: (1) analysis of predominantly minimum volume (10 L) samples where the ambient concentration was evidently close to the limit of detection; and (2) the lack of recovery efficiency measurement as an important adjunct to interpretation of each analytical result, precluding calculation of concentration and preventing meaningful quantitative comparisons from the data as collected.

A final observation on the LT2 data collection process concerns the omission of *Giardia* from analysis requirements. The extent and consequences of giardiasis are greater than those of cryptosporidiosis.[9,10] Most analysts agree that *Giardia* is generally easier to find in water samples than *Cryptosporidium*, as evidenced by the typically higher recovery efficiency for *Giardia* (ca. 40−60%) compared to *Cryptosporidium* (ca. 20−40%).[30] Typically, *Giardia* are reported more frequently, at

higher concentration, and are also frequently associated with waterborne illness.[11] Experience of this analyst in all phases of the Method 1622/1623 analysis process indicates that examination of samples for both organisms assists in the finding of both.

Information on *Cryptosporidium* in surface water provided in the data collected under the EPA drinking water regulation LT2 is extensive and informative. Despite being dominated by 93% negative results overall and with no positive findings at over half of the 1670 sampling sites, the information provided in positive *Cryptosporidium* oocyst findings at nearly 250 of the sampling sites provides a unique picture of critical occurrence parameters, i.e. level and variability, broadly distributed across the U.S.

Based on *Cryptosporidium* occurrence data and their distributions at over 150 sampling sites in surface water sources broadly representative of watersheds and watershed characteristics throughout the U.S., the spectrum of median occurrence of raw oocyst numbers/L in annual data sets lies between ca. 0.005/L and ca. 0.5/L. Taking general MS recovery efficiencies into account, highest true *Cryptosporidium* oocyst concentrations of 10 oocysts/L may be expected with concentrations ca. 1/L occurring at half of sites >25% of the time.

The information provided from the multiple positive LT2 sites combined with other published information suggests statement of a general hypothesis that *Cryptosporidium* oocysts occur routinely at levels detectable at all surface water locations represented in the LT2 sampling sites and in surface water elsewhere throughout the world. An approach essential to demonstrating this must include analysis of sufficient volume collected as a grab sample accompanied by recovery efficiency measurement to result in positive finding. The approach can be applied most efficiently by trial and error to establish the volume necessary to provide positive results in light of recovery efficiency measured with each sample. Future requirements for sampling and analysis for protozoan pathogens under LT2ESWTR should include analysis of representative sample volumes and analysis for both *Cryptosporidium* and *Giardia*.

Implications of this analysis of the LT2 site by site data implementation of the second round of sampling required by LT2 must be considered seriously by both the water supply industry and by responsible regulatory agencies.

◼ ASSOCIATED CONTENT

ⓢ Supporting Information
Full size copies of Figures 2, 3, 4, 5, 6a−f, 7a−e and 8, not practical to include in the body of the article, are provided. This information is available free of charge via the Internet at http://pubs.acs.org.

◼ AUTHOR INFORMATION

Corresponding Author
*Tel: ++61 0408 548 427; fax: ++61 2 4221 3238; e-mail: jongerth@uow.edu.au.

Notes
The authors declare no competing financial interest.

◼ REFERENCES

(1) Richardson, A. J.; et al. An outbreak of waterborne cryptosporidiosis in Swindon and Oxfordshire. *Epidemiol. Infect.* **1991**, *107* (3), 485.
(2) Hayes, E. B.; Matte, T. D.; O'Brien, R. T.; McKinley, T. W.; Logsdon, G. S.; Rose, J. B.; Ungar, B. L. P.; Word, D. M.; Pinsky, P. F.;

I

dx.doi.org/10.1021/es4006509 | *Environ. Sci. Technol.* XXXX, XXX, XXX−XXX

Cummings, M. L.; Wilson, M. A.; Long, E. G.; Hurwitz, E. S.; Juranek, D. D. Large community outbreak of cryptosporidiosis due to contamination of a public water supply. *N. Engl. J. Med.* **1989**, *320* (21), 1372.

(3) MacKenzie, W. R.; et al. A massive outbreak in Milwaukee of *Cryptosporidium* infection transmitted through the public water supply. *N. Engl. J. Med.* **1994**, *331* (3), 161-.

(4) Current, W. L. *Cryptosporidium* and cryptosporidiosis. In Gottlieb, M. S., Groopman, J. E., Eds.; *Acquired Immune Deficiency Syndrome*; Allen R. Liss, Inc.: New York, 1984; pp 355−373.

(5) U.S. EPA. *NPDWR Long Term 2 Surface Water Treatment Rule*; 2006. https://www.federalregister.gov/articles/2006/01/05/06-4/national-primary-drinking-water-regulations-lo.

(6) U.S. EPA. *Cryptosporidium* "cleaned up" data from monitoring under the LT2 Rule (CSV) (7MB); 2012. http://water.epa.gov/lawsregs/rulesregs/sdwa/lt2/upload/cryptodatacleaned1.csvng-term-2-enhanced-surface-water-treatment-rule.

(7) Meisel, J. L.; Perera, D. R.; Meligro, C.; Rubin, C. E. Overwhelming watery diarrhoea associated with a *Cryptosporidium* in an immunocompromised patient. *Gastroenterology* **1976**, *70* (6), 1156−1160.

(8) Barta, J. R.; Thompson, R. C. A. What is *Cryptosporidium*? Reappraising its biology and phylogenetic affinities. *Trends Parasitol.* **2006**, *22* (10), 463−468.

(9) Yoder, J. S.; Harral, C.; Beach, M. J. *Cryptosporidiosis Surveillance --- United States, 2006−2008 Morbidity & Mortality Weekly Report*, June 11, 2010; Vol. *59* (SS06), pp 1−14.

(10) Yoder, J. S.; Harral, C.; Beach, M. J. *Giardiasis Surveillance --- United States, 2006−2008 Morbidity & Mortality Weekly Report*, June 11, 2010; Vol. *59* (SS06), pp 15−25.

(11) Baldursson, S.; Karanis, P. Waterborne transmission of protozoan parasites: Review of worldwide outbreaks − An update 2004−2010. *Water Res.* **2011**, *45* (20), 6303−6314.

(12) Bronsdon, M. A. Rapid dimethyl sulfoxide-modified acid-fast stain of *Cryptosporidium* oocysts in stool specimens. *J. Clin. Microbiol.* **1984**, *19* (6), 952−953.

(13) Stibbs, H. H.; Ongerth, J. E. Immunofluorescence detection of *Cryptosporidium* oocysts in fecal smears. *J. Clin. Microbiol.* **1986**, *24*, 517−521.

(14) Ongerth, J. E.; Stibbs, H. H. Identification of *Cryptosporidium* in river water. *Appl. Environ. Microbiol.* **1987**, *53*, 672−676.

(15) Rose, J. B.; Gerba, C. P.; Jakubowski, W. Survey of potable water supplies for *Cryptosporidium* and *Giardia*. *Environ. Sci. Technol.* **1991**, *25* (8), 1393−1400.

(16) LeChevallier, M. W.; Norton, W. D.; Lee, R. G. Occurrence of *Giardia* and *Cryptosporidium* spp. in Surface Water Supplies. *Appl. Environ. Microbiol.* **1991**, *57* (9), 2610−2616.

(17) Karanis, P.; Schonen, D.; Seitz, H. M. Distribution and removal of *Giardia* and *Cryptosporidium* in water supplies in Germany. *Water Sci. Technol.* **1997**, *37* (2), 9−18.

(18) U.S. EPA. *Method 1623: Cryptosporidium and Giardia in Water by Filtration/IMS/FA*; 2005. http://www.epa.gov/nerlcwww/online.html#protos.

(19) U.S. EPA. Method 1623.1: *Cryptosporidium* and *Giardia* in Water by Filtration/IMS/FA; 2012. http://water.epa.gov/scitech/drinkingwater/labcert/upload/epa816r12001.pdf.

(20) U.S. EPA. *Occurrence and Exposure Assessment for the Long Term 2 Enhanced Surface Water Treatment Rule (PDF)*; EPA 815-R-06-002; Washington, DC, December 2005.

(21) Ongerth, J. E. The concentration of *Cryptosporidium* and *Giardia* in water–the role and importance of recovery efficiency. *Water Res.* **2013**, DOI: http://dx.doi.org/10.1016/j.watres.2013.02.015.

(22) Saaed, F. M. A. Measurement of *Cryptosporidium* and *Giardia* Concentrations in Water by Conventional and Molecular Methods. Doctoral Thesis, University of Wollongong, Wollongong NSW, 2012.

(23) Ongerth, J. E.; Saaed, F. M. A. Distribution of *Cryptosporidium* oocysts and *Giardia* cysts in water above and below the normal limit of detection. *Parasitol Res.* **2013**, *112* (2), 467−471.

(24) U.S. EPA. *LT2/Stage 2 Data Collection and Tracking system*; 2006. http://water.epa.gov/lawsregs/rulesregs/sdwa/mdbp/tools-dcts.cfm.

(25) Ongerth, J. E. *Giardia* cyst concentrations in river water. *J. Am. Water Works Assoc.* **1989**, *81* (9), 81−86.

(26) Hansen, J.; Ongerth, J. E. Effects of time and watershed characteristics on the concentration of *Cryptosporidium* oocysts in river water. *Appl. Environ. Microbiol.* **1991**, *57* (10), 2790−2795.

(27) Messner, M. LT2 Round 1 *Cryptosporidium* Occurrence Presentation; U.S. EPA, 2011. http://water.epa.gov/lawsregs/rulesregs/sdwa/lt2/upload/lt2round1crypto.pdf. 12.7.11.

(28) Meybeck, M.; Friedrich, G.; Thomas, R.; Chapman, G. Ch. 6 in *Water Quality Assessment*, 2nd ed.; Chapman, D., Ed.; UNESCO/WHO/UNEP, 1996. http://www.who.int/water_sanitation_health/resourcesquality/wqachapter6.pdf .

(29) Stumm, W.; Morgan, J. J. *Aquatic Chemistry*; Wiley: New York, 1981; p 535.

(30) Messner, M. LT2 Round 1 *Cryptosporidium* Matrix Spike Recovery; U.S. EPA, 2011. http://water.epa.gov/lawsregs/rulesregs/sdwa/lt2/upload/lt2round1crypto.pdf. 12.7.11.

J

dx.doi.org/10.1021/es4006509 | *Environ. Sci. Technol.* XXXX, XXX, XXX−XXX