

Importance of Distributional Forms for the Assessment of Protozoan Pathogens Concentrations in Drinking-Water Sources

Sylvestre, Émile; Prévost, Michèle; Smeets, Patrick; Medema, Gertjan; Burnet, Jean Baptiste; Cantin, Philippe ; Villion, Manuela ; Robert, Caroline; Dorner, Sarah

DOI

[10.1111/risa.13613](https://doi.org/10.1111/risa.13613)

Publication date

2020

Document Version

Final published version

Published in

Risk Analysis

Citation (APA)

Sylvestre, É., Prévost, M., Smeets, P., Medema, G., Burnet, J. B., Cantin, P., Villion, M., Robert, C., & Dorner, S. (2020). Importance of Distributional Forms for the Assessment of Protozoan Pathogens Concentrations in Drinking-Water Sources. *Risk Analysis*, 41(8), 1396-1412. <https://doi.org/10.1111/risa.13613>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Impact of Hydrometeorological Events for the Selection of Parametric Models for Protozoan Pathogens in Drinking-Water Sources

Émile Sylvestre ,^{1,2,*} Jean-Baptiste Burnet,^{1,2} Sarah Dorner,² Patrick Smeets,³ Gertjan Medema,^{3,4} Manuela Villion,⁵ Mounia Hachad,^{1,2} and Michèle Prévost¹

Temporal variations in concentrations of pathogenic microorganisms in surface waters are well known to be influenced by hydrometeorological events. Reasonable methods for accounting for microbial peaks in the quantification of drinking water treatment requirements need to be addressed. Here, we applied a novel method for data collection and model validation to explicitly account for weather events (rainfall, snowmelt) when concentrations of pathogens are estimated in source water. Online *in situ* β -D-glucuronidase activity measurements were used to trigger sequential grab sampling of source water to quantify *Cryptosporidium* and *Giardia* concentrations during rainfall and snowmelt events at an urban and an agricultural drinking water treatment plant in Quebec, Canada. We then evaluate if mixed Poisson distributions fitted to monthly sampling data ($n = 30$ samples) could accurately predict daily mean concentrations during these events. We found that using the gamma distribution underestimated high *Cryptosporidium* and *Giardia* concentrations measured with routine or event-based monitoring. However, the log-normal distribution accurately predicted these high concentrations. The selection of a log-normal distribution in preference to a gamma distribution increased the annual mean concentration by less than 0.1-log but increased the upper bound of the 95% credibility interval on the annual mean by about 0.5-log. Therefore, considering parametric uncertainty in an exposure assessment is essential to account for microbial peaks in risk assessment.

KEY WORDS: *Cryptosporidium*; drinking water; *Giardia*; hydrometeorology; risk assessment

1. INTRODUCTION

Hydrometeorological events such as heavy rainfall and snowmelt can lead to short-term deteriora-

¹NSERC Industrial Chair on Drinking Water, Department of Civil, Geological, and Mining Engineering, Polytechnique Montreal, Montreal, Quebec, H3C 3A7, Canada.

²Canada Research Chair in Source Water Protection, Department of Civil, Geological, and Mining Engineering, Polytechnique Montreal, Montreal, Quebec, H3C 3A7, Canada.

³KWR Water Research Institute, Groningehaven 7, Nieuwegein, 3433 PE, The Netherlands.

⁴Sanitary Engineering, Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, P.O. Box 5048, Delft, 2600GA, The Netherlands.

⁵Centre d'expertise en analyse environnementale du Québec, Ministère de l'Environnement et de la Lutte contre les changements climatiques, Québec, Canada.

*Address correspondence to Émile Sylvestre, Polytechnique Montréal, Department of Civil, Geological, and Mining Engineering, 2900 boul. Édouard-Montpetit, Montréal, QC, H3T 1J4, Canada; emile.sylvestre@polymtl.ca

tion of source water quality and may pose a challenge for drinking water treatment. Peak concentrations of pathogens in source water have been recognized as causes of waterborne outbreaks associated with drinking water when synchronous with suboptimal or inadequate treatment performance (Hrudey & Hrudey, 2004). Over the last 20 years, event-based sampling strategies have been developed to assess variations in protozoan pathogens concentrations during rainfall-induced runoff conditions in tributaries of drinking water sources (Dorner et al., 2007; Kistemann et al., 2002; Swaffer et al., 2014; Swaffer, Abbott, King, van der Linden, & Monis, 2018), in reservoirs used as a drinking water source (Burnet, Penny, Ogorzaly, & Cauchie, 2014), and in raw water from surface drinking water systems (Astrom, Petterson, Bergstedt, Pettersson, & Stenstrom, 2007; Atherholt, LeChevallier, Norton, & Rosen, 1998; Dechesne & Soyeux, 2007; Signor, Roser, Ashbolt, & Ball, 2005). In these studies, the association between target pathogens, fecal indicator bacteria (FIB), and physical parameters (flow rate, water level, turbidity) were investigated. Although progress has been made to accelerate culture-based methods for the detection of FIB, these methods cannot be used as a trigger for event-based sampling because culture typically requires 6- to 24-hour incubation periods. Advances in rapid detection of enzyme activity that is associated with fecal contamination (Farnleitner, Hocke, & Beiwil, 2001; George, Petit, & Servais, 2000) and its recent automation (Koschelnik, Vogl, Epp, & Lackner, 2015; Ryzinska-Paier et al., 2014) allows for rapid detection of peak fecal contamination events and trigger for simultaneous collection of sample for pathogens. Commercially available prototypes for β -D-glucuronidase (GLUC) activity *in situ* monitoring in near real-time are now available to characterize fecal pollution temporal dynamics in environmental waters (Burnet et al., 2019; Burnet, Sylvestre et al., 2019; Ryzinska-Paier et al., 2014; Stadler et al., 2016). These automated measurement systems could further be used to design new sampling strategies targeting short-term fluctuations in microbial pathogen concentrations during hydrometeorological events.

Upon characterization of a critical contamination event, risk assessors need to integrate this information into a probabilistic risk assessment. Basic principles of probability theory need to be considered to adequately use results from event-based sampling to inform microbial risk assessment. Results from routine sampling (also known as systematic sampling) are independent and identi-

cally distributed (i.i.d.) random variables because routine samples are collected at a fixed periodic interval (e.g., monthly sampling). Mixed Poisson distributions, such as the Poisson–Gamma (negative binomial) distribution and the Poisson–log-normal distribution, have been used to infer microbiological data (Haas, Rose, & Gerba, 1999; Masago, Oguma, Katayama, Hirata, & Ohgaki, 2004; Teunis, Medema, Kruidenier, & Havelaar, 1997; Westrell et al., 2006), but these distributions have not yet been validated for the prediction of microbial peak events. Results from event-based sampling can also be considered as i.i.d. random variables if event-based samples are collected at a fixed periodic interval (e.g., hourly) during the event. However, routine and event-based samples cannot be combined for statistical inference because their periodic intervals differ. A potential solution to this problem could be to use results from event-based sampling campaigns to evaluate if mixed Poisson models fitted to routine monitoring data can accurately predict pathogen concentrations during peak events. The identification of the most appropriate distribution to predict these peak events would improve the assessment of health risk associated with the finished water and the selection of treatment requirements.

The first objective of this study is therefore to assess if online GLUC activity and turbidity measurements can indicate periods of high concentrations of protozoan pathogens in source water. The second objective is to determine if mixed Poisson distributions fitted to routine monitoring data accurately predict *Cryptosporidium* and *Giardia* concentrations during hydrometeorological events (snowmelt and rainfall episodes).

2. MATERIALS AND METHODS

2.1. Sample Site

Two drinking water treatment plants (DWTPs) introduced in Sylvestre et al. (2020) (DWTPs C6, A4) were selected for case studies.

2.1.1. Urban Site

DWTP C6 is supplied by surface water from a river in the Greater Montreal Area in Quebec, Canada (Table I). Raw and settled water were sampled during event-based campaigns. From February to April 2017, the raw water was processed by a sludge blanket clarifier dosed with aluminum sulfate

Table I. Summary of Catchment Information for DWTPs C6 and A4

DWTP	Mean River Flow Rate(m ³ /s) [min-max]	CatchmentSize (km ²)	Main Land Cover Type in the Intake Protection Zone ^a	WWTPs /CSOs in the Intake Protection Zone ^a
C6	300 [20-1000]	>50,000	Urban	4/26
A4	15 [3-100]	<100	Agricultural	1/4

^a10 km upstream and 100 m downstream from the withdrawal site. The distances include surface water, portions of tributaries and a 120 m strip of land measured from the high-water mark.

hydrate “alum” (Al₂(SO₄)₃; dosing rate: 50 mg L⁻¹) and silica (SiO₂; dosing rate: 2 mg L⁻¹) in 1 °C raw water at pH 6.0. The land use in this area is dominated by low to medium intensity urban residential areas. The air temperature during winter (January–March) averages –10 °C. The flow rate of the river is measured continuously at a gauging station 5 km downstream of the drinking water intake. Between 1970 and 2012, the average flow rate of the river during winter was around 200 m³/s. During the local snowmelt period, generally in March and April, the average flow rate peaks at approximately 600 m³/s. The flow rate typically peaks one week following the local snowmelt because of the large size of the catchment (146,334 km²). Up to 10 km upstream from DWTP C6, the river receives treated effluent discharges from four municipal wastewater treatment plants (WWTPs), as well as untreated sewage discharges from 37 combined sewer overflow (CSO) outfalls, and two tributaries draining agricultural lands of approximately 70 km². Limited catchment management practices are implemented to control the volume and the duration of CSO discharges during snowmelt periods (Gouvernement du Québec, 2015).

2.1.2. Agricultural Site

DWTP A4 is supplied by a small agricultural river in southern Quebec. The annual average flow rate of the river is 16 m³/s. A municipal WWTP and four CSO outfalls are located 10 km upstream from the drinking water intake. At the WWTP, wastewater is treated through aerated ponds, and around 10,000 m³/day of treated water is discharged into the river. The regional watershed protection plan indicates that intensive pig and cattle farming (>1500 animal units) occurs in this area and that 30 to 60% of the land is dedicated to agriculture. Cattle and swine manure are applied to agricultural lands from April to October, and a maximum of 35% of manure

produced onsite can be used for agricultural spraying. Buffer strips of at least 3 m from the river are required for source water protection (Gouvernement du Québec, 2018).

2.2. Monitoring Strategies

Raw water samples were collected monthly (from 2014 to 2017 at DWTP C6 and from 2014 to 2015 at DWTP A4) for the enumeration of *Cryptosporidium* and *Giardia*. No samples were collected between June and September from 2015 to 2017 at DWTP C6. The statistical characterization of these data sets is presented in Table II. An automated rapid onsite monitoring system (ColiMinderTM, VWMS GmbH, Vienna, Austria) was installed at each DWTP intake around 30 days before snowmelt or rainfall events for a preliminary investigation of the GLUC activity fluctuation ranges in each source water. Detailed technical information about the device can be found in Koschelnik et al. (2015). Analytical validation of the technology for source waters and challenging against established culture- and molecular-based assays has been recently performed by Burnet et al. (2019). The GLUC activity was measured every 1–3 hours during dry weather conditions and every 30–60 minutes during hydrometeorological events. Fifteen minutes after sample collection, results were reported online and expressed in modified Fishman units (MFU/100 mL) based on the enzyme unit definition for GLUC activity (Koschelnik et al., 2015).

Event-based samples were collected when two conditions were met: (1) cumulative rainfall exceeding 20 mm or air temperature higher than 5 °C (causing rapid snowmelt) were measured in 24 hours, and (2) an increase in GLUC activity of 5 mMFU/100 mL was observed within an hour. A trigger of 5 mMFU/100 mL was selected based on short-term increases in GLUC activity measured during previous hydrometeorological events at these DWTPs (Fig. 1). For sample collection, 1 L grab samples of raw water

Table II. Sample size, Sample Mean Concentration, and Relative Standard Deviation for *Cryptosporidium* and *Giardia* Concentrations (Uncorrected for Recovery) for DWTPs C6 and A4

DWTP	<i>Cryptosporidium</i>			<i>Giardia</i>		
	Sample Size	Sample Mean (oocysts/L)	Relative Standard Deviation	Sample Size	Sample Mean (cysts/L)	Relative Standard Deviation
C6	27	0.064	1.65	27	1.57	0.93
A4	24	0.181	1.63	24	1.54	1.01

were collected in autoclaved polypropylene bottles at a frequency of 4–6 hour for around 24 hours for the enumeration of *Escherichia coli*. Additionally, 10–40 L samples were simultaneously filtered onsite for the enumeration of *Cryptosporidium* and *Giardia* using Envirochek HV sampling capsules (Pall Gelman Laboratory, Ann Arbor, MI, USA) at DWTP C6, and Hemoflow F80A hollow-fiber ultrafilters (Fresenius Medical Care, Lexington, MA, USA) at DWTP A4. Ten liter-samples are not typical for surface water sampling, but the filtration of small volumes was necessary to avoid filter clogging due to high raw water turbidities during the rainfall event at DWTP A4. Sequential grab samples were collected for 24 hours to estimate the daily mean concentration. At DWTP C6, 50 L samples of settled water were also filtered during the first event-based sampling campaigns. Settled water samples were collected three hours after raw water samples to match the theoretical hydraulic residence time throughout coagulation/flocculation/sedimentation (C. Durivage, *personal communication*). The sampling capsules (Envirochek HV filtration) or concentrates (Hemoflow ultrafiltration) were shipped overnight in coolers at 4 °C to the Centre d'expertise en analyse environnementale du Québec (CEAEQ) in Québec City, QC, and eluted (Envirochek filters) and processed within 48 hours of sampling.

2.3. Microbial Enumeration Methods

Escherichia coli was enumerated using the defined substrate technology (IDEXX Quanti-Tray/2000) with Colilert reagents (Method 9223B, American Public Health Association, 2005). The enumeration of oocysts of *Cryptosporidium* and cysts of *Giardia* filtered with Envirochek HV sampling capsules was carried out following the USEPA method 1623.1 (USEPA 2012). The elution procedure was adapted for the enumeration of

(oo)cysts filtered with Hemoflow ultrafilters. Following Hemoflow-based concentration, volumes of filter eluates were approximately 500–700 mL. postconcentration was done by centrifugation to obtain a final volume between 20 mL and 50 mL and a packed pellet volume between 2 mL and 5 mL. Between 20% and 50% of the packed pellet volume was then processed by immunomagnetic separation (IMS), before sample staining and examination following USEPA method 1623.1.

Sample-specific analytical recoveries were not measured for routine monitoring samples, but ongoing precision recovery (OPR) samples prepared in tap water were done regularly, following standard method recommendations (USEPA, 2012). Mean analytical recovery rates of 0.46 (Standard Deviation [SD] = 0.14) and 0.50 (SD = 0.17) were measured for *Cryptosporidium* and *Giardia*, respectively, based on 43 *Cryptosporidium* and *Giardia* matrix spike recovery experiments. These experiments were carried out with flow-cytometry sorted fluorescently labeled (oo)cysts (Colorseed™, BTF, Australia) by spiking a target dose of 98–100 (oo)cysts in 10 L samples of raw water collected at 30 DWTPs in Québec over nine years. Additional recovery rates were measured for each sample collected during the event-based campaign at DWTP A4. The same fluorescently labeled controls (Colorseed™) were spiked at a target dose of 98–100 (oo)cysts in the raw water sample before careful manual mixing and onsite concentration using hollow-fiber ultrafiltration. Seeded oocysts and naturally occurring oocysts were enumerated in each event-based sample. At DWTP A4, some samples were partially analyzed because of the high turbidity of the raw water. For these samples, the distribution of seeded (oo)cysts was assumed to be homogenous at the time of subsampling and directly proportional to the analyzed volume. Sample-specific analytical recovery rates were not measured for samples collected during both events at DWTP C6.

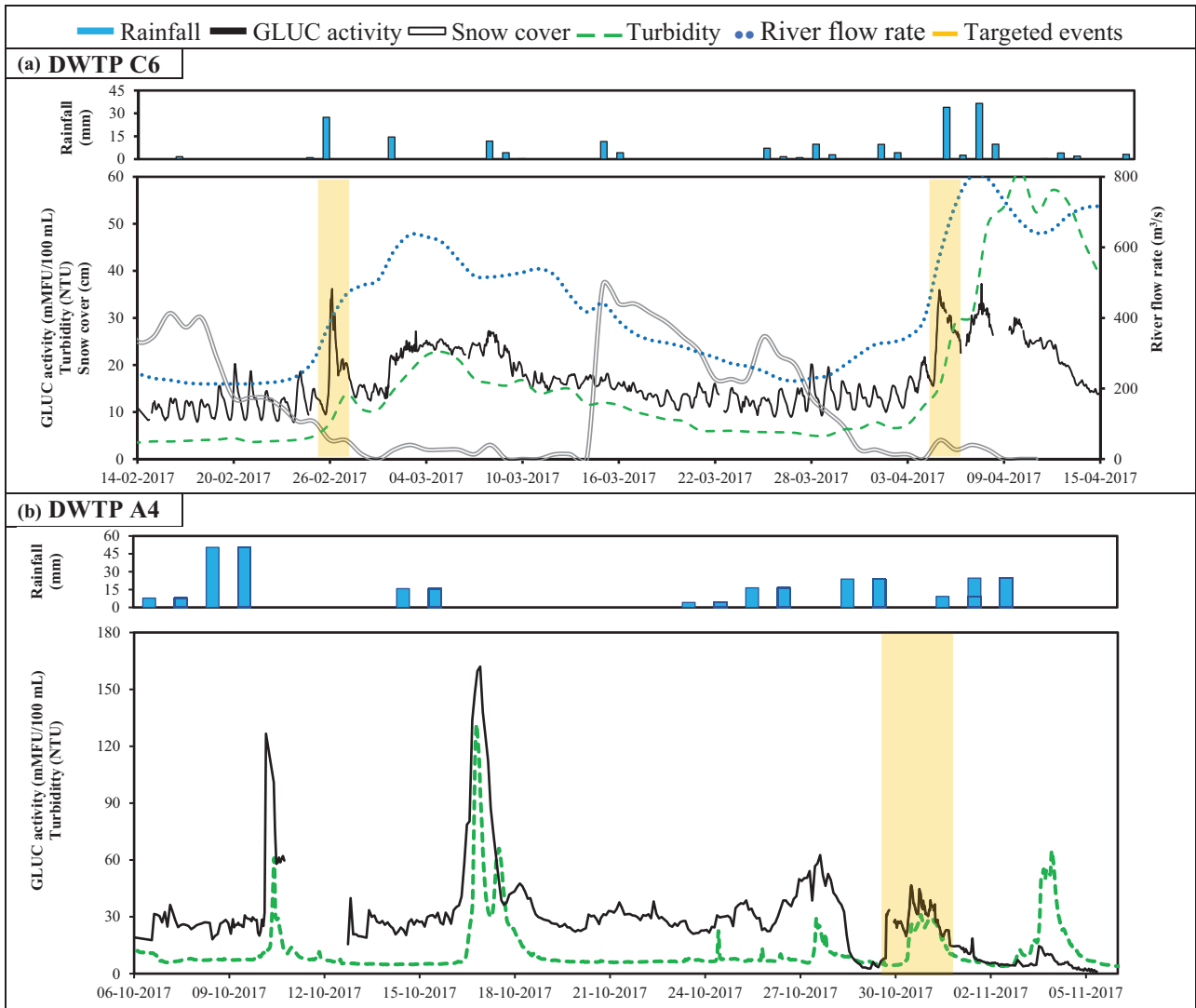


Fig 1. Time series of daily rainfall, GLUC activity, snow cover, turbidity, and river flow rate during sampling periods at DWTPs C6 and A4. Yellow rectangles indicate targeted events.

We conservatively assumed that all detected *Cryptosporidium* oocysts and *Giardia* cysts were human infectious.

2.4. Statistical Analysis

2.4.1. Model Parametrization and Implementation

The temporal variations in protozoan pathogen concentrations were evaluated with the three-level hierarchical Bayesian model presented in detail in Sylvestre et al. (2020). Briefly, at the first level, the analytical error of the enumeration method is binomially distributed:

$$y_i \sim \text{Binomial}(x_i, p_i), \tag{1}$$

where y_i is the number of (oo)cysts observed in each sample i ; x_i is the true number of (oo)cysts in the sample; and p_i is the probability of detection of each organism x_i . The nonconstant analytical recovery p_i (i.e., the sample-to-sample variability in recovery rate) was assumed to vary randomly according to a beta distribution with shape parameters α and β . Posterior means of the parameters were $(\hat{\alpha}, \hat{\beta}) = (6.48, 7.70)$ for *Cryptosporidium* and $(\hat{\alpha}, \hat{\beta}) = (3.80, 3.91)$ for *Giardia*. The second level of the hierarchical structure takes into consideration the sampling error. The true number of (oo)cysts x_i is Poisson distributed

with mean $\lambda_i = c_i V_i$, the product of the concentration (c_i) and the analyzed volume (V_i).

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(cV)^x e^{-(cV)}}{x!}. \quad (2)$$

At the third level, temporal variations in the concentration c_i are described by a continuous distribution. In this study, concentrations predicted by these models are assumed to be daily mean concentrations. The gamma and log-normal distributions were selected and compared because they have different upper tail behaviors at large coefficients of variation (Haas, 1997), and this property is preserved under mixed Poisson models (Kaas & Hesselager, 1995). The gamma distribution has a density

$$f(c) = \frac{\lambda^\alpha c^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda c}, \quad (3)$$

and an expectation (i.e., mean) $E(c) = \alpha/\lambda$, where $\alpha > 0$ is the shape parameter and $\lambda > 0$ is a scale parameter. The log-normal distribution has a density

$$f(c) = \frac{1}{\alpha c \sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{[\ln c - \lambda]^2}{\alpha^2}\right], \quad (4)$$

and an expectation $E(c) = \exp(\alpha + \frac{\lambda^2}{2})$, where the shape parameter $\alpha > 0$ and the scale parameter λ may take each real value.

Estimations and inferences were carried out in a Bayesian framework using Markov Chain Monte Carlo (MCMC). Gamma priors with hyperparameters set to Gamma (0.01, 0.01) were selected for the shape parameter α and the scale parameter λ of the gamma distribution. For the log-normal distribution, the shape parameter α was assigned a uniform prior with hyperparameters set to Uniform (-10, 10), and the scale parameter λ was assigned a weakly informative exponential prior with hyperparameters set to exp (1). The hyperparameter in the weakly informative prior was set to a conservative value based on evidence regarding the logarithm of the empirical standard deviation of *Cryptosporidium* and *Giardia* measured at 30 DWTPs (Sylvestre et al. (2020)). The rationale for the selection of the other priors is presented in Sylvestre et al. (2020). A sensitivity analysis was conducted in this study to investigate the influence of the hyperparameter value in the exponential prior of the scale parameter λ of the log-normal distribution. The hyperparameter value was adjusted upward (exp (0.1)) and downward (exp (3)), and the log-normal distribution was re-estimated with these varied priors.

Models were fitted using the MCMC technique with rjags (v4-6) (Plummer, 2013) in R (v3.4.1). Four Markov chains were run for 3×10^5 iterations after a burn-in phase of 10^4 iterations. The convergence of the four chains was monitored with the Brooks—Gelman—Rubin scale reduction factor (Gelman & Shirley, 2011).

2.4.2. Estimation of Daily Mean Concentrations During Events

The daily mean concentration was considered in this study because: (1) the exposure is usually characterized in terms of the arithmetic mean number of organisms in the dose (Haas, 1996), and (2) a 24-hour period is typically used to account for short-term exposures in microbial risk assessment (WHO, 2016). *Cryptosporidium* and *Giardia* concentrations in each event-based sample were estimated with a Poisson model (Equation 2). Counts were corrected with sample-specific recovery rates when available (DWTP A4). The daily mean concentration \bar{C}_{Event} was estimated by averaging concentrations C_i collected at regular intervals over 24 hours.

$$\bar{C}_{Event} = \frac{1}{n} \sum_{i=1}^n C_i. \quad (5)$$

The uncertainty of the daily mean concentrations was evaluated with Monte Carlo simulations. A random sample was drawn from the 95% credibility interval on the mean concentration c_i (Equation 2) of each sample i . The draws were summed and divided by the number of event-based samples N collected in 24 hours. The procedure was repeated 10,000 times to estimate the 95% predictive interval for the daily mean concentration. The R code used to calculate these daily mean concentrations is provided in the Supporting Information.

2.4.3. Model Validation

Distributions were illustrated with complementary cumulative distribution function (CCDF) curves. Each best fit distribution was generated for probabilities of exceedance between 100% and 0.27% (1 day per year) using the posterior mean of the parameters (α, λ). The predictive interval about each best fit distribution was created by simulating 1,000 CCDF curves parametrized by random values included in the 95% credibility interval for the parameters. To visually assess the capacity of the distributions to

predict high concentration observations, two vertical lines were juxtaposed with CCDF curves. These two lines represent (1) the sample maximum concentration measured with routine monitoring, and (2) the daily mean concentration during the hydrometeorological event. Only the highest event mean *Cryptosporidium* and *Giardia* concentrations (Event 1) were illustrated for DWTP C6. We assumed that these daily mean concentrations have probabilities of exceedance higher than 1 day per year.

2.4.4. Estimation of Annual Mean Concentrations

It is important to note that there may be a difference between the uncertainty on the *mean* of the distribution and the uncertainty on the *annual mean* predicted by a skewed distribution. If each value of the distribution represents a daily mean concentration, a difference will be observed if the upper tail of the distribution does not have an asymptotic behavior from a probability of exceedance smaller than 1 day per year. In other words, the occurrence of daily concentrations predicted to occur less than once a year may generate variations in the annual mean estimates. To investigate the importance of this difference, the upper bound of the 95% credibility interval on the mean of the distribution was compared to the upper bound of the 95% credibility interval on the annual mean of the distribution. The upper bound of the 95% credibility interval on the annual mean of the distribution was evaluated as follows:

- (1) the 95% credibility interval was calculated for the shape parameter α and the scale parameter λ of the distribution;
- (2) the pair of parameters contained in the 95% credibility interval that maximize the mean of the distribution was determined;
- (3) 365 samples were drawn randomly from the distribution generated with the pair of parameters determined in step (2). The average of these 365 samples (annual mean) was calculated;
- (4) Step 3 was repeated 10,000 times to produce a distribution of these annual means. The 97.5th percentile of the distribution of these annual means was determined.

This model was implemented using R (v3.4.1). The R code is provided in the Supporting Information.

3. RESULTS AND DISCUSSION

3.1. Short-Term Fluctuations in Microbial Contaminants

Short-term fluctuations in microbial contaminants were studied during two snowmelt events in an urban catchment and one rainfall event in an agricultural catchment. The collection of event-based samples was triggered by meteorological conditions (cumulative rainfall, change in air temperature) and rapid increases in GLUC activity (Section 2.2). This sampling strategy allowed us to characterize short-term variations in *Cryptosporidium* and *Giardia* concentrations in raw water.

At the urban DWTP C6, the amplitudes of *E. coli*, *Cryptosporidium*, and *Giardia* concentration peaks were 1.1-log, 0.7-log, and 1.4-log, respectively, during Event 1 (Fig. 2). Sample-specific recovery rates were not measured at DWTP C6; therefore, the intraevent variations in protozoan pathogen concentrations could be influenced by the difference in recovery rates among samples. The impact of source water turbidity on recovery rates could be small during Event 1 because turbidity was low and only ranged from 6 to 13 NTU (mean absolute deviation [MAD] = 1.5 NTU). However, other short-term changes in the composition of the water matrix could have influenced the recovery performance.

At the agricultural DWTP A4, the amplitudes of the protozoan pathogen concentration peaks were higher (0.8–1.1-log) than the amplitude of the *E. coli* concentration peaks (0.5-log) (Fig. 3). In 24 hours, sample-specific recovery rates varied between 22% and 70% for *Cryptosporidium* and between 8% and 70% for *Giardia* (Table III). Sample-specific recovery rates decreased during the contamination event, especially for *Giardia*. Negative correlations between turbidity and recovery rates were obtained for *Cryptosporidium* ($r = -0.50$) and *Giardia* ($r = -0.87$); however, these results should be interpreted with caution because the sample size was small ($n = 6$) and turbidity only ranged from 18 to 28 NTU (MAD = 2.8 NTU). Low recovery rates during peak events could be associated with the nature of the turbidity and the background matrix of the water (DiGiorgio, Gonzalez, & Huitt, 2002). Positive correlations between measured concentrations (i.e., uncorrected for the analytical recovery) and recovery rates were obtained for *Cryptosporidium* ($r = 0.83$) and *Giardia* ($r = 0.57$). Theoretical recovery rates of 30% were assumed for all event-based samples collected at urban

Table III. Characterization of Raw Water Samples Collected With Event-Based Sampling at DWTPs C6 and A4

DWTP	Event	Date/Hour	Raw Water Turbidity (NTU)	Count				Colorseed Count				Recovery rate ^a		Concentration ^b (oo)cysts/L	
				Oocysts	Cysts	Vol. Sampled(L)	Vol. Analyzed(L)	Oocysts	Cysts	Oocysts	Cysts	Cryptosporidium	Giardia	Cryptosporidium	Giardia
C6	February	25/20:11	6.1	1	8	15	15	-	-	-	-	0.30	0.30	0.22	1.8
		26/00:04	7.8	1	94	15	15	-	-	-	-	0.30	0.30	0.22	20.8
		26/3:11	8.4	5	200	15	15	-	-	-	-	0.30	0.30	1.11	44.4
		26/7:11	10.4	2	138	15	15	-	-	-	-	0.30	0.30	0.44	31.1
		26/10:18	10.7	0	102	15	15	-	-	-	-	0.30	0.30	0.00	22.2
		26/14:13	12.9	5	110	15	15	-	-	-	24-hour event mean	0.30	0.30	1.11	24.4
A4	April	4/16:49	14.0	1	40	14	14	-	-	-	-	0.30	0.30	0.15	2.9
		4/23:52	17.2	2	152	15	15	-	-	-	-	0.30	0.30	0.28	10.1
		5/3:47	20.7	1	43	15	15	-	-	-	-	0.30	0.30	0.15	2.9
		5/6:35	27.3	2	96	14	14	-	-	-	24-hour event mean	0.30	0.30	0.31	6.9
A4	October	30/12:00	18.4	7	94	40	40	10	14	14	14	0.50	0.70	1.80	16.8
		30/16:00	23.4	9	121	35	35	14	8	8	8	0.70	0.40	1.80	43.2
		30/20:00	28.0	6	142	30	15	20	5	5	5	0.40	0.10	1.00	94.7
		31/00:00	24.0	12	117	35	17.5	12	0	0	0	0.24	<0.08	2.90	83.6
		31/4:00	28.0	1	202	32	16	14	9	9	9	0.28	0.18	0.20	70.1
		31/8:00	27.0	4	102	30	15	11	4	4	4	0.22	0.08	1.20	85.0
											24-hour event mean		1.48	65.5	

^aTHEORETICAL recovery rates at DWTP C6 and sample-specific recovery rates at DWTP A4.

^bConcentrations corrected for the analytical recovery.

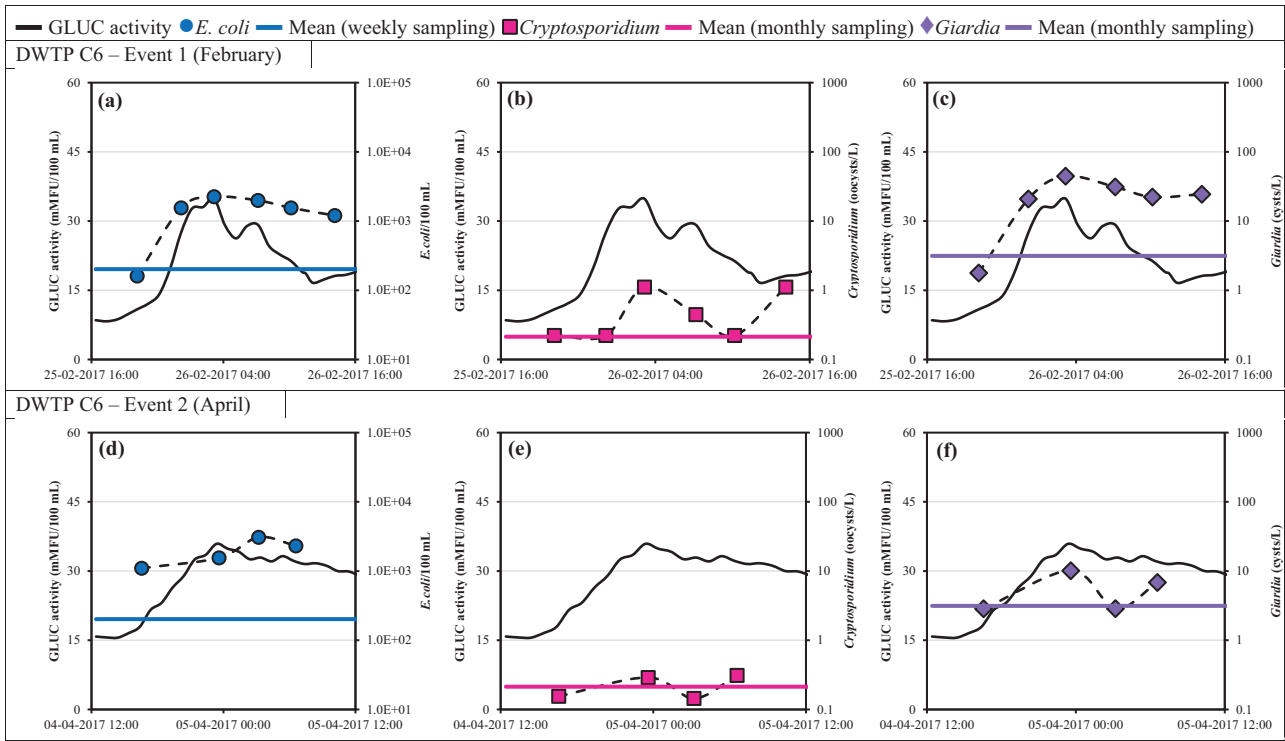


Fig 2. Short-term variations in *E. coli*, *Cryptosporidium*, and *Giardia* concentrations for the first 24 hours of two hydrometeorological events (snowmelt and rainfall) in February (Event 1) and April (Event 2) 2017 at DWTP C6.

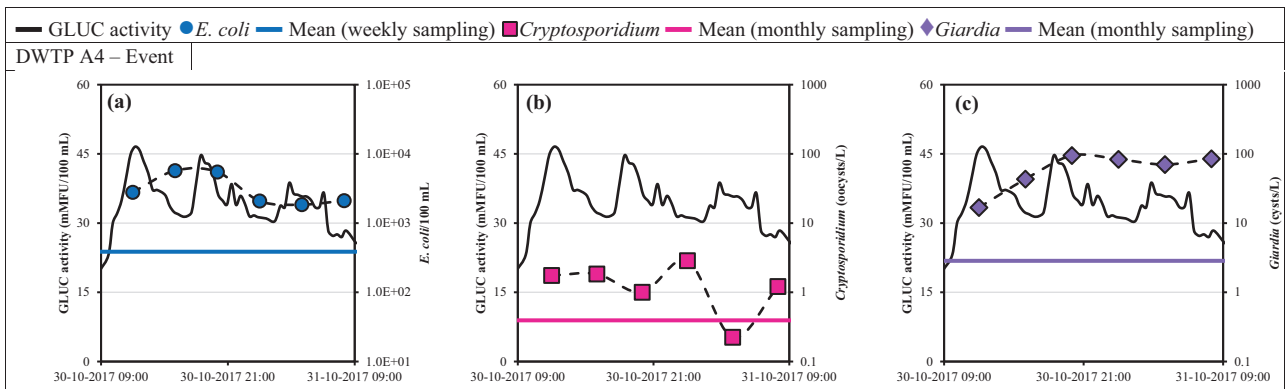


Fig 3. Short-term variations in *E. coli* (a), *Cryptosporidium* (b), and *Giardia* (c) concentrations for the first 24 hours of an hydrometeorological event (rainfall) in October 2017 at DWTP A4.

DWTP C6 based on average recovery rates measured during the rainfall event at DWTP A4.

During these three hydrometeorological events, the GLUC activity rapidly increased for about 12 hours and then slowly decreased over several days to return to the baseline level (Fig. 1). *Cryptosporidium* and *Giardia* concentrations also increased during the first 12 hours but did not decrease in the

12 hours following the GLUC activity peak. Therefore, a decrease in GLUC activity may not indicate a decrease in protozoan pathogen concentrations during snowmelt/rainfall episodes. The 24-hour sampling strategy did not allow us to determine the full duration of protozoan pathogen peaks. Consequently, measured 24-hour mean *Cryptosporidium* and *Giardia* concentrations could be lower than the maximum

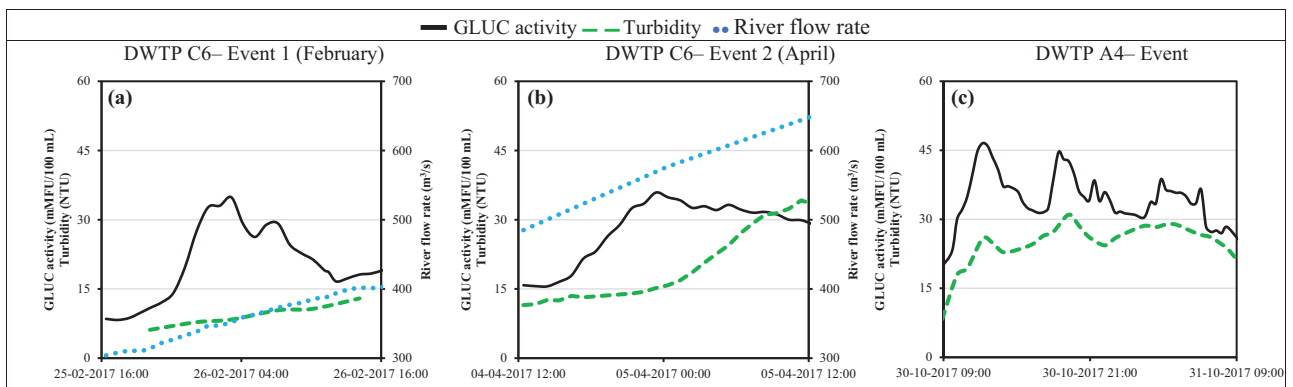


Fig 4. Short-term fluctuations in GLUC activity, raw water turbidity, and river flow rate during event conditions (first 24 hours) at DWTP C6 in February 2017 (a), April 2017 (b) and DWTP A4 in October 2018 (c). River flow rate measurements were not available at DWTP A4.

24-hour mean concentrations for these events. Nevertheless, at the two DWTPs, the 24-hour event mean *Giardia* concentration was higher than the sample maximum measured with routine monitoring (corrected for recovery) at DWTP C6 (Event 1; +0.3-log) and DWTP A4 (+0.7-log) (Tables III).

During the two events at DWTP C6, the turbidity did not increase simultaneously with protozoan pathogen concentrations (Figs. 4a and b). The lack of systematic association between protozoan pathogen concentrations and turbidity has been reported for large datasets (USEPA, 2005). Differences in protozoan pathogens and turbidity dynamics may be associated with the varying contributions of multiple sources, including watershed-scale nonpoint source pollution during snowmelt- and rainfall-runoff and local point source discharges of fecal contamination. Local sewer discharges can increase fecal contamination loads in the river without increasing total suspended solids (TSS) because correlations between these parameters are not expected during the snowmelt period (Madoux-Humery et al., 2013). At agricultural DWTP A4, GLUC activity, *Cryptosporidium*, and *Giardia* concentrations increased with turbidity, suggesting that turbidity could be a valid surrogate to trigger the sampling of peak protozoan pathogen concentrations in agricultural catchments. Additional event-based sampling campaigns could be designed to assess if the magnitude of turbidity and microbial peaks are associated or not. However, recovery rates for protozoan pathogens may be very low at the high raw water turbidities (>100 NTU) that can be measured at the drinking water intake.

3.2. Model Validation

It was demonstrated in Sylvestre et al. (2020) that, as only a few samples informed on the behavior of the upper tail, the differences in marginalized deviance information criterion (mDIC) between candidate parametric distributions (gamma, log-normal) were too small (less than four points) for model selection based on mDIC alone. Results from the sensitivity analysis of the influence of hyperparameter values in the exponential prior of the scale parameter λ of the log-normal distribution are shown in the Supporting Information (Figure S1). Changes in hyperparameter values had a small effect on the behavior of the upper tail of the distribution for *Cryptosporidium* and a negligible effect for *Giardia* (Figure S1).

The present study investigated whether results from event-based sampling of protozoan pathogens can be predicted by a parametric distribution fitted to routine monitoring data. The CCDF curves of the gamma and the log-normal distributions fitted to routine monitoring *Cryptosporidium* and *Giardia* data are presented in Fig. 5. The capacity of each distribution to predict a fixed concentration (e.g., event mean concentration) can be visually assessed for probabilities of exceedance varying between 1.0 (all the time) and 0.002 (about 1 day per year). For the agricultural DWTP A4, the gamma and the log-normal distribution predicted the 24-hour event mean *Cryptosporidium* concentration at a probability of exceedance of 0.002 (Fig. 5a). However, only the log-normal distribution predicted the 24-hour event mean *Giardia* concentration; the upper tail of the gamma distribution did not predict these high

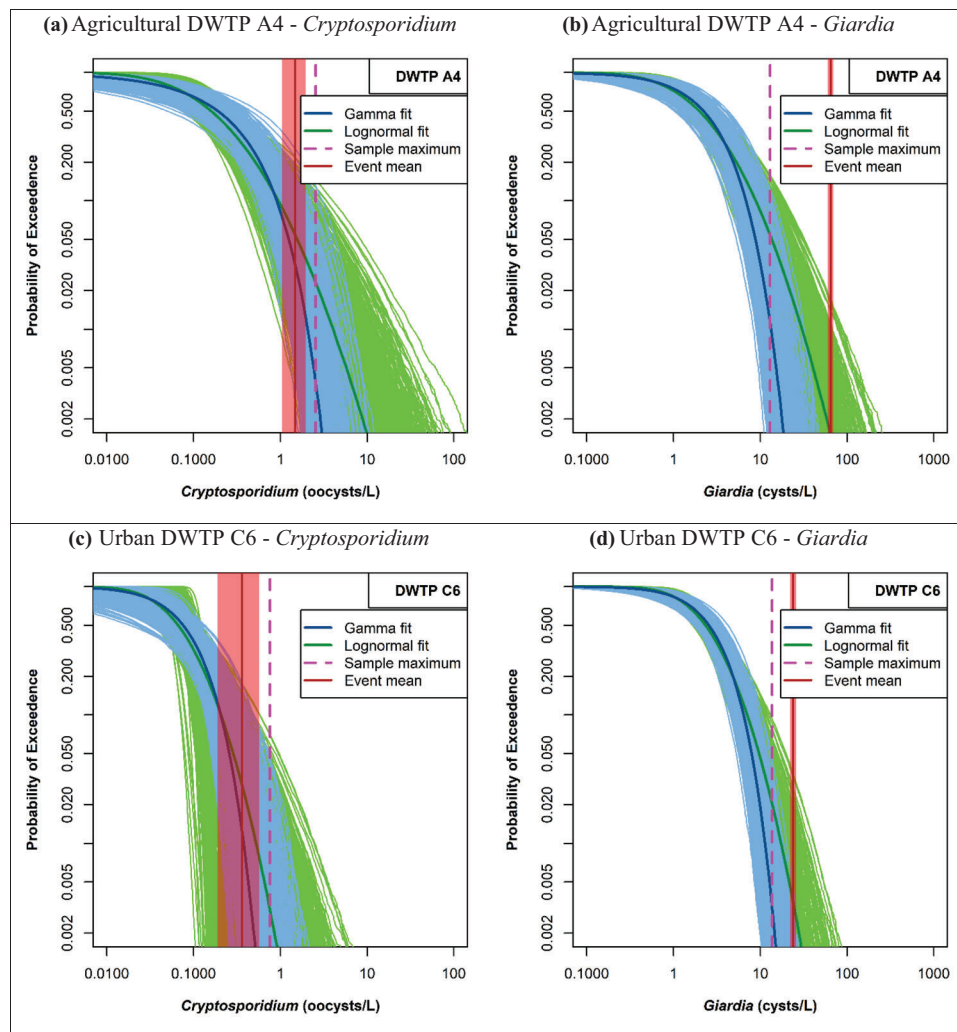


Fig 5. CCDF of *Cryptosporidium* and *Giardia* concentrations in raw water at DWTPs A4 and C6. Blue and green lines are the best fit gamma and log-normal distributions, respectively. Light blue and light green areas are the 95% predictive intervals about the best fit gamma and log-normal distributions, respectively. Parametric distributions were generated assuming beta distributed recovery rates. Sample maximum concentrations measured with the routine monitoring strategy (pink dashed line) were corrected with the mean recovery rate of the beta distribution. 24-hour event mean concentrations (red lines) were corrected with a theoretical recovery rate of 30% for all event-based samples at DWTP C6, and sample-specific recovery rates at DWTP A4. Pale red boxes represent the 95% predictive interval on the 24-hour event mean. Only the largest 24-hour event mean *Cryptosporidium* and *Giardia* concentrations (Event 1) were illustrated for DWTP C6.

concentrations (Fig. 5b). For DWTP C6, at a probability of exceedance of 0.002, only the log-normal distribution conservatively predicted the sample maximum concentration for *Cryptosporidium* measured by routine monitoring (Fig. 5c) and the 24-hour event mean concentration for *Giardia* (Fig. 5d). Sylvestre *et al.* (2020) recently demonstrated, using raw water *E. coli* concentration data collected at DWTP C6, that the log-normal distribution better predicted peak *E. coli* concentrations than the gamma distribution during snowmelt events. Hence,

care needs to be taken when a distribution is selected to describe temporal variations in source water concentrations because its upper tail may be too light to account for peak contamination levels. Quantifying the maximum concentration of a distribution might also be of interest to evaluate worst-case scenarios in a quantitative risk assessment. If so, extreme value theory may be used to evaluate the expected maximum concentration of a distribution based on observations (Embrechts, Klüppelberg, & Mikosch, 2013).

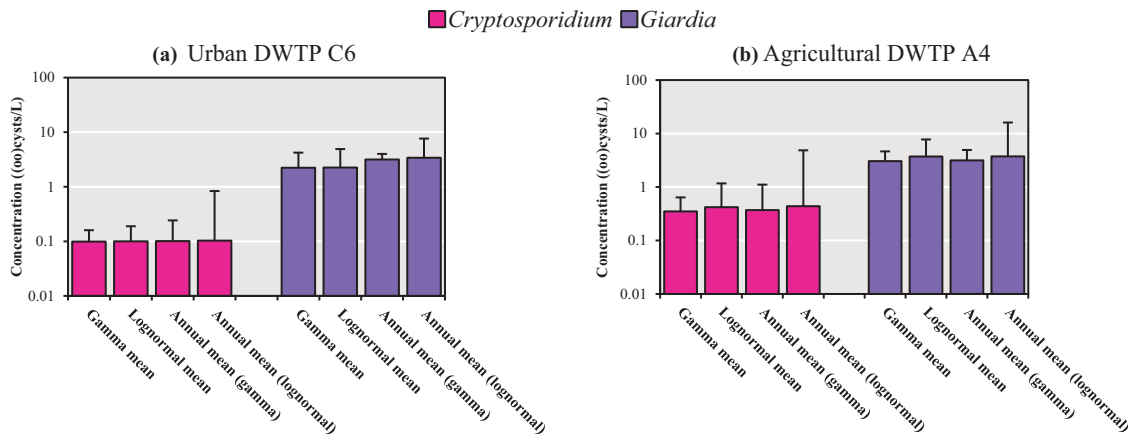


Fig 6. Mean and annual mean of the gamma and log-normal distributions at DWTPs C6 (a) and A4 (b). Whiskers indicate the upper bound on the 95% credibility interval. For the annual mean, the 95% credibility interval represents the year-to-year variation (365 daily mean concentrations per year) of the upper bounds of the 95% predictive interval about the best fit distribution.

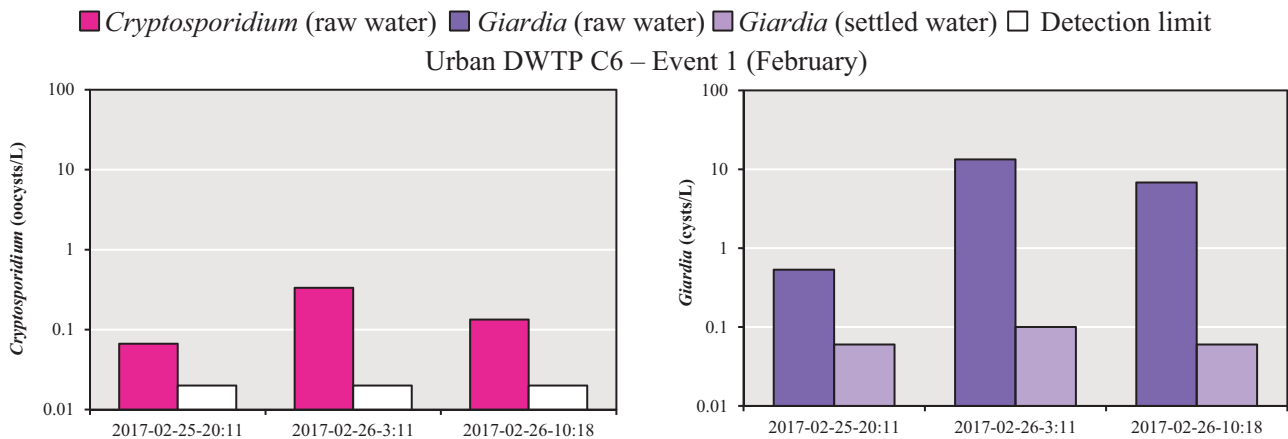


Fig 7. Short-term variations in *Cryptosporidium* (a) and *Giardia* (b) concentrations in raw water and settled water for the first 24 hours of an hydrometeorological event (snowmelt and rainfall) in February (Event 1) at DWTP C6.

3.3. Implications for Risk Assessment

The selection of a log-normal distribution in preference to a gamma distribution had a minor effect on the estimate of the annual mean concentration but increases the upper bound of the 95% credibility interval on the annual mean from 0.5-log for *Cryptosporidium* and 0.3-log for *Giarda* at DWTP C6, and from 0.6-log for *Cryptosporidium* and 0.4-log for *Giarda* at DWTP A4 (Fig. 6). Treatment requirements for the reduction of microbial pathogens at DWTPs are commonly scaled to log₁₀-reduction; therefore, the choice of parametric distribution for source water characterization could result in different treatment requirements. It should be noted that the upper bound of the 95% credibility interval on

the *annual mean* is higher than the upper bound of the 95% credibility interval on the *mean* of the distribution for the log-normal but not for the gamma (Fig. 6). This difference indicates that, for the log-normal distribution, daily mean concentrations having a very small probability of exceedance (e.g., once every 10 years) can have a significant impact on the annual mean concentration.

Improved knowledge of the dependencies between source water concentrations and removal/inactivation efficiencies of treatment processes could also reduce uncertainties on exposure estimates. In this study, stable *Giardia* concentrations (0.08 ± 0.02 cyst/L) were measured in settled water at DWTP C6 during Event 1 regardless of an increase in source water concentrations of 1.4-log (Fig. 7).

These results must be interpreted with care because the sample size is small, and recovery rates in raw and settled water matrices were not measured. It is worth noting that, according to the Smoluchowski theory of flocculation, a higher flocculation rate should be observed at higher particle concentrations (Benjamin & Lawler, 2013). Basic research on the mechanisms of aggregation of microorganisms during coagulation/flocculation and the evaluation of full-scale performances of treatment processes during periods of microbial challenge in raw water could be valuable to improve the assessment and management of microbial peaks at DWTPs.

4. CONCLUSIONS

This article describes a methodology for data collection and model validation to explicitly account for hydrometeorological events when source water pathogen concentrations are characterized. An event-based sampling strategy triggered by meteorological conditions and rapid increases in GLUC activity was implemented at two drinking water treatment plants to investigate the impact of snowmelt and rainfall events on source water contamination. These event-based campaigns allowed us to find that:

- Increase in GLUC activity was indicative of an increase in *Cryptosporidium* and *Giardia* concentrations in source water, which varied over about 1.0-log over 24 hours;
- At the urban site, GLUC activity level was a better surrogate than turbidity to identify transient peak contaminations by protozoan pathogens in source water during two snowmelt events.

The model validation approach using mixed Poisson distributions and results from event-based sampling demonstrated that:

- The gamma distribution underestimated high protozoan pathogen concentrations collected with routine and event-based monitoring, but the log-normal distribution accurately predicted these high protozoan pathogen concentrations;
- The selection of a log-normal distribution rather than a gamma distribution increased the uncertainty of the annual mean concentration by about 0.5-log, which can result in additional treatment requirements. Appropriately conservative parametric models should be carefully chosen to adequately manage human health

risks and avoid unnecessary costs for water utilities.

Additional studies confirming these findings in other catchments and for other hydrometeorological events would be relevant. Improved knowledge of full-scale reduction of protozoan pathogens and microbial surrogates during hydrometeorological events would be valuable to quantify the risk of microbial peaks at drinking water treatment plants.

ACKNOWLEDGMENTS

This work was funded by the NSERC Industrial Chair funded on Drinking Water, the Canadian Research Chair on Source Water Protection and the Canada Foundation for Innovation. A part of the outcomes presented in this paper was based on research financed by the Dutch-Flemish Joint Research Programme for the Water Companies. We thank the technical staff of the NSERC Industrial Chair and the technical staff of the biology and microbiology division at CEAEQ for their help with sample collection and analysis.

REFERENCES

- Astrom, J., Petterson, S., Bergstedt, O., Pettersson, T. J., & Stenstrom, T. A. (2007). Evaluation of the microbial risk reduction due to selective closure of the raw water intake before drinking water treatment. *Journal of Water and Health*, 5, 81–97.
- Atherholt, T. B., LeChevallier, M. W., Norton, W. D., & Rosen, J. S. (1998). Effect of rainfall of *Giardia* and *Crypto*. *Journal American Water Works Association*, 90, 66–80.
- Benjamin, M. M., & Lawler, D. F. (2013). *Water quality engineering: Physical/chemical treatment processes*. Hoboken, NJ: John Wiley & Sons, Inc.
- Burnet, J.-B., Dinh, Q. T., Imbeault, S., Servais, P., Dorner, S., & Prévost, M. (2019). Autonomous online measurement of β -D-glucuronidase activity in surface water: Is it suitable for rapid *E. coli* monitoring? *Water Research*, 152, 241–250.
- Burnet, J.-B., Penny, C., Ogorzaly, L., & Cauchie, H.-M. (2014). Spatial and temporal distribution of *Cryptosporidium* and *Giardia* in a drinking water resource: Implications for monitoring and risk assessment. *Science of the Total Environment*, 472, 1023–1035.
- Burnet, J.-B., Sylvestre, É., Jalbert, J., Imbeault, S., Servais, P., Prévost, M., & Dorner, S. (2019). Tracking the contribution of multiple raw and treated wastewater discharges at an urban drinking water supply using near real-time monitoring of β -d-glucuronidase activity. *Water Research*, 164, 114869.
- Dechesne, M., & Soyeux, E. (2007). Assessment of source water pathogen contamination. *Journal of Water and Health*, 5, 39–50.
- DiGiorgio, C. L., Gonzalez, D. A., & Huitt, C. C. (2002). *Cryptosporidium* and *Giardia* recoveries in natural waters by using environmental protection agency method 1623. *Applied and Environmental Microbiology*, 68, 5952–5955.
- Dorner, S., Anderson, W. B., Gaulin, T., Candon, H. L., Slawson, R. M., Payment, P., & Huck, P. M. (2007). Pathogen and indicator variability in a heavily impacted watershed. *Journal of Water and Health*, 5, 241–257.

- Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). *Modelling extreme events: For insurance and finance*. Berlin, Germany: Springer Science & Business Media.
- Farnleitner, A. H., Hocke, L., & Beiwil, C. (2001). Rapid enzymatic detection of *Escherichia coli* contamination in polluted river water. *Letters in Applied Microbiology*, 33, 246–250.
- Gelman, A., & Shirley, K. (2011). Inference from simulations and monitoring convergence. *Handbook of Markov Chain Monte Carlo*, 6, 163–174.
- George, I., Petit, M., & Servais, P. (2000). Use of enzymatic methods for rapid enumeration of coliforms in freshwaters. *Journal of Applied Microbiology*, 88, 404–413.
- Gouvernement du Québec. (2018). Chapitre Q-2, r. 26. Règlement sur les exploitations agricoles. Loi sur la qualité de l'environnement (chapitre Q-2, a. 31, 53.30, 70, 115.27, 115.34 et 124.1).
- Gouvernement du Québec. (2015). Règlement sur les ouvrages municipaux d'assainissement des eaux usées. Page 17, Québec, Canada.
- Haas, C. N. (1996). How to average microbial densities to characterize risk. *Water Research*, 30, 1036–1038.
- Haas, C. N. (1997). Importance of distributional form in characterizing inputs to Monte Carlo risk assessments. *Risk Analysis*, 17, 107–113.
- Haas, C. N., Rose, J. B., & Gerba, C. P. (1999). *Quantitative microbial risk assessment*. New York: John Wiley and Sons, Inc.
- Hrudey, S. E., & Hrudey, E. J. (2004). *Safe drinking water. Lessons from recent outbreaks in affluent nations*. London, UK: International Water Association Publishing.
- Kaas, R., & Hesselager, O. (1995). Ordering claim size distributions and mixed Poisson probabilities. *Insurance: Mathematics and Economics*, 17, 193–201.
- Kistemann, T., Classen, T., Koch, C., Dangendorf, F., Fischeder, R., Gebel, J., ... Exner, M. (2002). Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. *Applied and Environmental Microbiology*, 68, 2188–2197.
- Koschelnik, J., Vogl, W., Epp, M., & Lackner, M. (2015). Rapid analysis of β -D-glucuronidase activity in water using fully automated technology. *WIT Transactions on Ecology and the Environment*, 196, 471–481.
- Madoux-Humery, A.-S., Dorner, S., Sauvé, S., Aboufadi, K., Galarneau, M., Servais, P., & Prévost, M. (2013). Temporal variability of combined sewer overflow contaminants: Evaluation of wastewater micropollutants as tracers of fecal contamination. *Water Research*, 47, 4370–4382.
- Masago, Y., Oguma, K., Katayama, H., Hirata, T., & Ohgaki, S. (2004). *Cryptosporidium* monitoring system at a water treatment plant, based on waterborne risk assessment. *Water Science and Technology*, 50, 293–299.
- Plummer, M. (2013). JAGS: Just another Gibbs sampler (Version 3.4.0)[Computer software].
- Ryzinska-Paier, G., Lendenfeld, T., Correa, K., Stadler, P., Blaschke, A. P., Mach, R. L., ... Farnleitner, A. H. (2014). A sensitive and robust method for automated on-line monitoring of enzymatic activities in water and water resources. *Water Science & Technology*, 69, 1349–1358.
- Signor, R. S., Roser, D. J., Ashbolt, N. J., & Ball, J. E. (2005). Quantifying the impact of runoff events on microbiological contaminant concentrations entering surface drinking source waters. *Journal of Water and Health*, 3, 453–468.
- Stadler, P., Blöschl, G., Vogl, W., Koschelnik, J., Epp, M., Lackner, M., ... Zessner, M. (2016). Real-time monitoring of beta-D-glucuronidase activity in sediment laden streams: A comparison of prototypes. *Water Research*, 101, 252–261.
- Swaffer, B., Abbott, H., King, B., van der Linden, L., & Monis, P. (2018). Understanding human infectious *Cryptosporidium* risk in drinking water supply catchments. *Water Research*, 138, 282–292.
- Swaffer, B. A., Vial, H. M., King, B. J., Daly, R., Frizenschaf, J., & Monis, P. T. (2014). Investigating source water *Cryptosporidium* concentration, species and infectivity rates during rainfall-runoff in a multi-use catchment. *Water Res*, 67, 310–320.
- Sylvestre, É., Burnet, J. B., Smeets, P., Medema, G., Prévost, M., & Dorner, S. (2020). Can routine monitoring of *E. coli* fully account for peak event concentrations at drinking water intakes in agricultural and urban rivers? *Water Research*, 170, 115369.
- Sylvestre, É., Prévost, M., Smeets, P., Medema, G., Burnet, J.-B., Cantin, P., Villion, M., Robert, C., & Dorner, S. (2020). Importance of distributional forms for the assessment of Protozoan Pathogens concentrations in drinking-water sources. *Risk Analysis*. <https://doi.org/10.1111/risa.13613>
- Teunis, P. F. M., Medema, G. J., Kruidenier, L., & Havelaar, A. H. (1997). Assessment of the risk of infection by *Cryptosporidium* or *Giardia* in drinking water from a surface water source. *Water Research*, 31, 1333–1346.
- USEPA. (2005). Occurrence and exposure assessment for the final long term 2 enhanced surface water treatment rule. EPA 815-R-06-002, Office of Water (4606-M), Washington, DC.
- USEPA. (2012). Method 1623.1: *Cryptosporidium* and *Giardia* in water by filtration/IMS/FA. EPA 816-R-12-001, Office of Water (MS-140), Washington, DC.
- Westrell, T., Teunis, P., van den Berg, H., Lodder, W., Ketelaars, H., Stenstrom, T. A., & de Roda Husman, A. M. (2006). Short- and long-term variations of norovirus concentrations in the Meuse river during a 2-year study period. *Water Research*, 40, 2613–2620.
- WHO. (2016). *Quantitative microbial risk assessment: Application for water safety management*. Geneva, Switzerland: WHO.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1: Complementary cumulative distribution function (CCDF) curves for the log-normal distribution of *Cryptosporidium* and *Giardia* concentrations using different values of the hyperparameter based on its prior scale parameter λ at drinking water treatment plants A and B.

Supplementary Material
Supplementary Material