## Supplemental Content

Bacteria, viruses, and parasites in an intermittent stream protected from and exposed to pasturing cattle: prevalence, densities, and quantitative microbial risk assessment

by

Wilkes, G., Brassard, J., Edge, T.A., Gannon, V., Jokinen, C.C., Jones, T.H., Neumann, N., Pintar, K.D.M., Ruecker, N., Schmidt, P.J., Sunohara, M., Topp, E., and Lapen, D.R.

QMRA Method: Probabilistic modelling and OpenBUGS code

This section provides the OpenBUGS code used in the evaluation of mean risk for *Cryptosporidium* spp., *Giardia*, and *E. coli* O157:H7, as well as some additional information about the utilized models. Details on the following modules are provided:

- A) Non-constant analytical recovery analysis for Cryptosporidium/Giardia matrix spike data
- B) Temporal concentration variability analysis for Cryptosporidium/Giardia enumeration data
- C) Hierarchical infectivity analysis for Cryptosporidium
- D) Temporal concentration variability and infectivity analysis for E. coli and E. coli O157:H7
- E) Strain-to-strain variability analysis for the exponential Cryptosporidium dose-response model

Bayesian procedures used to evaluate parameter uncertainty for the exponential dose-response model (*Giardia*) and the actual beta-Poisson dose-response model (pathogenic *E. coli*) are provided in Schmidt *et al.* (2013).

The numerical integration approach used to evaluate mean risk  $(P^*)$  at each site uses 1) parameters of a gamma distribution  $(\rho, \lambda)$  describing temporal concentration variability for the target microorganism, 2) a constant fraction of target microorganisms that are infectious (I), 3) a constant volume of surface water that is consumed during a single swimming exposure (V), and 4) the parameter(s) of the exponential (r) or actual beta-Poisson dose-response model  $(\alpha, \beta)$ . Schmidt *et al.* (Accepted) provided details on a similar numerical integration and how uncertainty in the specified parameter values was incorporated into a second-order risk assessment. This work differs only in the addition of a constant infectivity value (except in the case of *Giardia*) and use of an exponential dose-response model for *Cryptosporidium* spp. and *Giardia*. The numerical integration equations with exponential and actual beta-Poisson doseresponse models, respectively, are as follows:

$$P^{*}(\rho,\lambda,I,V,r) = \sum_{d=1}^{\infty} \left[ \frac{\Gamma(d+\rho)}{d!\Gamma(\rho)} \left( \frac{\lambda PV}{\lambda PV+1} \right)^{d} \left( \frac{1}{\lambda PV+1} \right)^{\rho} \right] \left[ 1 - (1-r)^{d} \right]$$
$$P^{*}(\rho,\lambda,I,V,\alpha,\beta) = \sum_{d=1}^{\infty} \left[ \frac{\Gamma(d+\rho)}{d!\Gamma(\rho)} \left( \frac{\lambda IV}{\lambda IV+1} \right)^{d} \left( \frac{1}{\lambda IV+1} \right)^{\rho} \right] \left[ 1 - \frac{\Gamma(\alpha+\beta)\Gamma(d+\beta)}{\Gamma(\beta)\Gamma(d+\alpha+\beta)} \right]$$

```
Non-constant Analytical Recovery Analysis for Cryptosporidium/Giardia Matrix Spike Data
A)
loga: logarithm (base 10) of shape parameter of beta dist. describing non-constant analytical recovery
     - STOCHASTIC (unknown)
logb: logarithm (base 10) of shape parameter of beta dist. describing non-constant analytical recovery
     - STOCHASTIC (unknown)
a: shape parameter of beta distribution describing non-constant analytical recovery
     - DETERMINISTIC (parameter of interest)
b: shape parameter of beta distribution describing non-constant analytical recovery
     - DETERMINISTIC (parameter of interest)
logC[i]: logarithm (base 10) of (oo)cyst conc. during i<sup>th</sup> sampling event - STOCHASTIC (unknown) C[i]: (oo)cyst concentration during i<sup>th</sup> sampling event in (oo)cysts/L- DETERMINISTIC
Pf[i]: analytical recovery of ith field sample - STOCHASTIC (nuisance parameter)
VEf[i]: enumerated volume of ith field sample in litres - CONSTANT
EXf[i]: expected number of observed (oo) cysts in enumerated volume of i<sup>th</sup> field sample
     - DETERMINISTIC
Xf[i]: number of observed (oo)cysts in enumerated volume of i<sup>th</sup> field sample
     - STOCHASTIC (measured)
Pm[i]: analytical recovery of i<sup>th</sup> matrix spike sample - STOCHASTIC (nuisance parameter)
VTm[i]: total volume of ith matrix spike sample in litres - CONSTANT
ENative[i]: expected number of native (oo) cysts contained in total volume of ith matrix spike sample
     - DETERMINISTIC
Native[i]: number of native (oo)cysts contained in total volume of i<sup>th</sup> matrix spike sample
     - STOCHASTIC (nuisance parameter)
Seed[i]: number of seeded (oo) cysts added to total volume of ith matrix spike sample - CONSTANT
N[i]: number of total (oo)cysts contained in total volume of ith matrix spike sample - DETERMINISTIC
Q[i]: enumerated fraction of ith matrix spike sample - CONSTANT
PEm[i]: effective analytical recovery of ith matrix spike sample - DETERMINISTIC
Xm[i]: number of observed (oo)cysts in enumerated volume of i<sup>th</sup> matrix spike sample
     - STOCHASTIC (measured)
model {
     # Prior q(a,b) = 1/ab with 0.000001 < a < 1000000, 0.000001 < b < 1000000
     loga \sim dunif(-6, 6)
     logb ~ dunif (-6, 6)
     # Transform loga and logb to a and b
     a \le pow(10, loga)
     b \le pow(10, logb)
     for (i in 1:m) {
          # Matrix (oo)cyst concentrations
          logC[i] ~ dunif(-6, 6) # Prior g(C[i]) = 1/C[i] with 0.000001 < C[i] < 1000000
          C[i] \le pow(10, logC[i])
          # Model for field sample count
          Pf[i] \sim dbeta(a, b)
          EXf[i] <- C[i] * VEf[i] * Pf[i]
          Xf[i] ~ dpois(EXf[i])
          # Model for matrix spike sample count
          Pm[i] \sim dbeta(a, b)
          ENative[i] <- C[i] * VTm[i]
          Native[i] ~ dpois(ENative[i])
          N[i] <- Native[i] + Seed[i]
          PEm[i] <- Pm[i] * Q[i]
          Xm[i] ~ dbin(PEm[i], N[i])
     }
```

In the analysis of *Cryptosporidium* and *Giardia* matrix spike recovery, it was assumed that there was a single beta distribution describing sample-to-sample variation in analytical recovery for each type of protozoan among all surface water monitoring sites. Matrix spike recovery data from 13 sites throughout the South Nation River basin were used in this analysis. The modelling approach is modified from Teunis and Havelaar (1999) and Schmidt (2010). It accounts for random measurement errors in the enumeration of both the field and matrix spike samples, which are not assumed to have equal analytical recovery. A uniform prior on the base 10 logarithm of concentration was used for the concentration of indigenous (oo)cysts in each matrix spike analysis. The numbers of seeded (oo)cysts were assumed to be precisely known.

## B) Temporal Concentration Variability Analysis for Cryptosporidium/Giardia Enumeration Data

```
logmu: logarithm (base 10) of mean concentration - STOCHASTIC (unknown)
logsigma: logarithm (base 10) of standard deviation of concentration - STOCHASTIC (unknown)
mu: mean concentration - DETERMINISTIC
sigma: standard deviation of concentration - DETERMINISTIC
rho: shape parameter of gamma distribution describing temporal concentration variability
        - DETERMINISTIC (parameter of interest)
lambda: scale parameter of gamma distribution describing temporal concentration variability
        - DETERMINISTIC (parameter of interest)
tau: reciprocal of scale parameter of gamma distribution describing temporal concentration variability
        - DETERMINISTIC
a: shape parameter of the beta distribution describing non-constant analytical recovery - CONSTANT
b: shape parameter of the beta distribution describing non-constant analytical recovery - CONSTANT
C[i]: concentration during i<sup>th</sup> sampling event in (oo)cysts/L - STOCHASTIC (nuisance parameter)
P[i]: analytical recovery of enumeration method for i<sup>th</sup> sample - STOCHASTIC (nuisance parameter)
V[i]: enumerated volume of ith sample in litres - CONSTANT
theta[i]: expected number of (oo)cysts in i<sup>th</sup> sample - DETERMINISTIC
X[i]: number of (oo)cysts observed in ith sample - STOCHASTIC (measured)
model {
     # Standard normal prior on logarithm (base 10) of mean and std. dev. of conc. in (oo)cysts/L
    \log mu \sim dnorm(0, 1)
    \log \alpha \sim dnorm(0, 1)
     # Transform logmu and logsigma to rho and lambda
    mu <- pow(10, logmu)
     sigma <- pow(10, logsigma)
     rho <- mu * mu / (sigma * sigma)
    lambda <- sigma * sigma / mu
    tau <- 1 / lambda
         Probabilistic model
     #
    for (i in 1:m) {
         C[i] ~ dgamma(rho, tau)
         P[i] \sim dbeta(a, b)
         theta[i] <- C[i] * V[i] * P[i]
         X[i] ~ dpois(theta[i])
    }
```

This procedure is based upon Schmidt and Emelko (2011). The principal difference is that independent standard normal priors are assumed for the mean and standard deviation of concentration in units of (oo)cysts  $L^{-1}$ .

The mean and standard deviation of analytical recovery were approximately 40% and 10% respectively for both *Cryptosporidium* and *Giardia*. The base-10 logarithms of the mean and standard deviation were assigned independent normal priors, as per Schmidt *et al.* (2013). The priors for *Cryptosporidium* and *Giardia* suggest that the mean and standard deviation of the (oo)cyst concentration were most likely near 1 (oo)cyst  $L^{-1}$  and were unlikely to exceed 1000 (oo)cysts  $L^{-1}$  or to fall below 0.001 (oo)cysts  $L^{-1}$ . Cumulative *Cryptosporidium* oocyst counts per cumulative enumerated volume at monitoring sites throughout the South Nation River watershed (with adjustment for 40% mean analytical recovery) ranged from 0.068-0.885 oocysts  $L^{-1}$ , so this prior is somewhat conservatively high. It was assumed that the standard deviations and means would have similar ranges of values and also that *Giardia* should have the same priors as were used for *Cryptosporidium*.

## C) Hierarchical Infectivity Analysis for Cryptosporidium

```
loga: logarithm (base 10) of shape parameter of beta dist. describing infectivity variation among sites
    - STOCHASTIC (unknown)
logb: logarithm (base 10) of shape parameter of beta dist. describing infectivity variation among sites
    - STOCHASTIC (unknown)
a: shape parameter of beta distribution describing infectivity variation among sites - DETERMINISTIC
b: shape parameter of beta distribution describing infectivity variation among sites - DETERMINISTIC
I[i]: Fraction of oocysts that are pathogenic at ith surface water monitoring site
    - STOCHASTIC (parameter of interest - selected sites)
X[i]: number of sequences from ith surface water monitoring site classified as pathogenic
    - STOCHASTIC (measured)
N[i]: number of sequences from ith surface water monitoring site - CONSTANT
model {
         Prior g(a,b) = 1/ab with 0.001 < a < 1000, 0.001 < b < 1000
    #
    loga \sim dunif(-3,3)
    logb \sim dunif(-3,3)
    # Transform loga and logb to a and b
    a \le pow(10, loga)
    b \le pow(10, logb)
         Probabilistic model for fraction of sequences that are pathogenic
    #
    for (i in 1:m) {
         I[i] ~ dbeta(a,b)
         X[i] \sim dbin(I[i],N[i])
    }
```

In the analysis of *Cryptosporidium* infectivity, it was assumed that each monitoring site has a constant fraction of total oocysts that are pathogenic and that this fraction varies from site to site according to a beta distribution. Data from 17 sites throughout the South Nation River basin (of which only four are addressed in this study) were used in the analysis.

D) Temporal Concentration Variability and Infectivity Analysis for E. coli and E. coli O157:H7
logmu: logarithm (base 10) of mean concentration - STOCHASTIC (unknown) logsigma: logarithm (base 10) of standard deviation of concentration - STOCHASTIC (unknown) mu: mean concentration - DETERMINISTIC sigma: standard deviation of concentration - DETERMINISTIC rho: shape parameter of gamma distribution describing temporal concentration variability - DETERMINISTIC (parameter of interest) lambda: scale parameter of gamma distribution describing temporal concentration variability - DETERMINISTIC (parameter of interest) tau: reciprocal of scale parameter of gamma distribution describing temporal concentration variability - DETERMINISTIC I: constant fraction of E. coli that are O157:H7 C[i]: concentration during i <sup>th</sup> sampling event in CFU/100mL - STOCHASTIC (measured) theta[i]: expected number of CFU in i <sup>th</sup> sample - DETERMINISTIC X[i]: O157:H7 presence/absence - STOCHASTIC (measured)
<pre>model {     # Normal prior on logarithm (base 10) of mean and std. dev. of concentration in CFU/100mL     logmu ~ dnorm(3, 1)     logsigma ~ dnorm(3, 1)     # Transform logmu and logsigma to rho and lambda     mu &lt;- pow(10, logmu)     sigma &lt;- pow(10, logsigma)     rho &lt;- mu * mu / (sigma * sigma)     lambda &lt;- sigma * sigma / mu     tau &lt;- 1 / lambda     # Uniform prior on I     I ~ dunif(0, 1)     # Probabilistic model     for (i in 1:m) {         C[i] ~ dgamma(rho, tau)         theta[i] &lt;- 1 - exp(-C[i] * 5 * I)         X[i] ~ dbern(theta[i])     } }</pre>
}

The distributions for *E. coli* were fit directly to colony forming unit (CFU) concentration estimates because raw data (i.e., counts, volumes, dilutions) were unavailable. A gamma distribution was fit to the concentration estimates without consideration of random measurement error in the plate count enumerations. For *E. coli*, it was assumed that the mean and standard deviation of the concentration were most likely near 1000 CFU 100 mL<sup>-1</sup> and were unlikely to exceed  $10^6$  CFU 100 mL<sup>-1</sup> or to fall below 1 CFU 100 mL<sup>-1</sup>. These priors are considered to be relatively uninformative due to the broad range of values that are considered to be highly plausible (6 orders of magnitude). In the case of *E. coli*, mean concentration estimates in the range of 73-1119 CFU 100 mL<sup>-1</sup> were obtained from the sites considered in Schmidt *et al.* (2013), and the prior on the mean is centred around 1000 CFU 100 mL<sup>-1</sup>.

Each monitoring site was assumed to have a constant fraction of total *E. coli* that were *E. coli* O157:H7. A uniform prior on the interval (0,1) was used. A truncated uniform prior on the interval (0,0.5) was considered, but would not have changed the outcome given that values exceeding 0.5 were not supported by the available data. In this model, the probability of a positive *E. coli* O157:H7 presence-absence result is linked to the corresponding total *E. coli* concentration estimate.

E) Strain-to-Strain Variability Analysis for the Exponential <i>Cryptosporidium</i> Dose-response Model
loga: logarithm of shape parameter of beta distribution describing variation among strains - STOCHASTIC (unknown)
logb: logarithm of shape parameter of beta distribution describing variation among strains - STOCHASTIC (unknown)
a: shape parameter of beta distribution describing variation among strains - DETERMINISTIC
b: shape parameter of beta distribution describing variation among strains - DETERMINISTIC
R[i]: probability of infection by a single strain i oocyst - STOCHASTIC (nuisance parameter)
I[j]: strain index variable for jth consumer - CONSTANT
D[j]: mean dose consumed by jth consumer - CONSTANT
X[i]: infection indicator variable for ith consumer - STOCHASTIC (measured)
mu: mean of beta dist, describing variation among strains - DETERMINISTIC (parameter of interest)
<pre>model {     # Prior g(a,b)=1/ab with each parameter constrained to [0.000001,1000000]     loga ~ dunif(-6, 6)     logb ~ dunif(-6, 6)     a &lt;- pow(10, loga)     b &lt;- pow(10, logb)     mu &lt;- a / (a + b)     for (i in 1:m) {         R[i] ~ dbeta(a, b)     }     for (j in 1:n) {             P[j] &lt;- 1 - exp(-D[j] * R[I[j]])             X[j] ~ dbern(P[j])     } }</pre>
}

This procedure is modified from a modelling approach considered in U.S. EPA (2005). It assumes that each of 6 *Cryptosporidium* dose-response studies using different strains/species can be modelled by an exponential dose-response model, and that the parameters of these models are random values from a beta-distribution describing overall variation in dose-response among pathogenic *Cryptosporidium* strains/species. The mean of this distribution is then used as the parameter of an exponential dose-response model because variation among individual microorganisms (*i.e.* different strains/species of oocysts in a water body to which a consumer might be exposed) is averaged out in an exponential dose-response model (Schmidt *et al.*, 2013).

Supplemental Content: References for QMRA Method

- Schmidt, P.J., 2010. Addressing the Uncertainty Due to Random Measurement Errors in Quantitative Analysis of Microorganism and Discrete Particle Data. PhD Thesis, University of Waterloo, Waterloo, Ontario, Canada, 2010.
- Schmidt, P.J., Emelko, M.B., 2011. QMRA and decision-making: Are we handling measurement errors associated with pathogen concentration data correctly? *Wat. Res.* 45, 427-438.
- Schmidt, P.J., Pintar, K.D.M, Fazil, A.M., Topp, E., 2013. Harnessing the theoretical foundations of the exponential and beta-Poisson dose-response models to quantify parameter uncertainty using Markov Chain Monte Carlo. *Risk Analysis* http://dx.doi.org/10.1111/risa.12006.
- Schmidt, P.J., Pintar, K.D.M., Fazil, A.M., Flemming, C.A., Lanthier, M., Laprade, N., Sunohara, M., Simhon, A., Thomas, J.L., Topp, E., Wilkes, G., Lapen, D.R., Accepted. Using *Campylobacter* spp. and *Escherichia coli* data and Bayesian microbial risk assessment to examine public health risks in agricultural watersheds under tile drainage management. *Wat. Res.* 47, 3255-3272.
- Teunis, P.F.M., Havelaar, A.H., 1999. *Cryptosporidium* in drinking water: Evaluation of the ILSI/RSI quantitative risk assessment framework. RIVM Report no. 284 550 006, RIVM, Bilthoven, The Netherlands.
- United States Environmental Protection Agency (U.S. EPA), 2005. Economic Analysis for the Final Long Term 2 Enhanced Surface Water Treatment Rule; EPA 815-R-06-001, U.S. Environmental Protection Agency, Office of Water, Washington, DC.



Supp. Cont. Fig. 1. These plots illustrate uncertainty in the mean concentration at each site for a) total and pathogenic *Cryptosporidium* oocysts, b) total *Giardia* cysts, and c) total *E. coli* and *E. coli* O157:H7. Uncertainty in each mean concentration was evaluated using 30,000 sets of plausible parameter values generated by Markov Chain Monte Carlo.



Supp. Cont. Fig. 2. These plots illustrate the uncertainty in the relative mean risk (relative to the RCA<sub>in</sub> site) at each site for a) *Cryptosporidium*, b) *Giardia*, and c) *E. coli* O157:H7. The points show the posterior mean (of the relative mean risk) and the error bars show the 95% equal-tailed credible intervals. Results for both total *Cryptosporidium* (black). Relative risks, relative to the mean risk for each pathogen at RCAin, were calculated using the same dose-response model parameters for each site. Uncertainty in relative mean risk was evaluated using 30,000 iterations, in which the same dose-response parameters were used for each site within a single iteration, but the values of these parameters varied among iterations. This incorporates dose-response parameter uncertainty in a way that acknowledges that dose-response does not differ among sites.

	Target	Limit of quantification [Laboratory]
Fecal indicators	E. coli	1 CFU 100 mL <sup>-1</sup> [AAFC LONDON, ON]
	Enterococcus spp.	1 CFU 100 mL <sup>-1</sup> [AAFC LONDON, ON]
	C. perfringens	<2 CFU 100 mL <sup>-1</sup> [AAFC LONDON, ON]
	Fecal Coliforms	<2 CFU 100 mL <sup>-1</sup> [AAFC LONDON, ON]
	Total Coliforms	<2 CFU 100 mL <sup>-1</sup> [AAFC LONDON, ON]
	Aeromonas	10 CFU 100 mL <sup>-1</sup> [AAFC LONDON, ON]
Pathogens	Cryptosporidium spp.	1 oocyst 0.63 to 22.75 L <sup>-1</sup> [CALGARY, AB]
	Giardia spp.	1 cyst 0.63 to 22.75 L <sup>-1</sup> [CALGARY, AB]
	Campylobacter spp.	1-10 CFU 500 mL <sup>-1</sup> [LETHBRIDGE, AB]
	Salmonella spp.	1-10 CFU 500 mL <sup>-1</sup> [LETHBRIDGE, AB]
	<i>E. coli</i> O157:H7	1-10 CFU 500 mL <sup>-1</sup> [LETHBRIDGE, AB]
Viruses	Hepatitis E	48 viral genomic copies 500 mL <sup>-1</sup> [AAFC ST. HYACINTHE, QC]
	Norovirus GII	48 viral genomic copies 500 mL <sup>-1</sup> [AAFC ST. HYACINTHE, QC]
	Rotavirus	48 viral genomic copies 500 mL <sup>-1</sup> [AAFC ST. HYACINTHE, QC]
Coliphage	F-RNA	0.5 PFU 100 mL <sup>-1</sup> [AAFC LACOMBE, AB]
	F-DNA	0.5 PFU 100 mL <sup>-1</sup> [AAFC LACOMBE, AB]

Supp. Cont. Table 1. Detection limits of various pathogens, parasites, viruses and indicator bacteria targeted in this experiment.

Supp. Cont. Table 2. Presence and absence of pathogens associated with final CART classification models, using as independent variables sample site (pasture treatment), season, and flow regime for <u>*Temporally Concurrent Data*</u>. High flow  $\ge 0.018 \text{ m}^3 \text{ s}^{-1}$ ; low flow  $\ge 0.002 \text{ m}^3 \text{ s}^{-1}$  and  $< 0.018 \text{ m}^3 \text{ s}^{-1}$ ; and no flow  $< 0.002 \text{ m}^3 \text{ s}^{-1}$ .

					Detections	
					as percent	
					of total	Presence:
Microorganism	Criteria	Ν	Presence	Absence	samples	absence ratio
Campylobacter spp.	All data	192	84	108	44%	0.78
Campylobacter spp.	No flow	39	10	29	26%	0.35
Campylobacter spp.	(High flow) AND (Low flow)	153	74	79	48%	0.94
Salmonella spp.	All data	195	12	183	6%	0.07
Salmonella spp.	(Low flow) AND (No flow)	117	3	114	3%	0.03
Salmonella spp.	High flow	78	9	69	12%	0.13
Cryptosporidium oocysts	All data	74	38	36	51%	1.06
Cryptosporidium oocysts	(Summer, Fall) AND (High flow)	26	20	6	77%	3.33
Cryptosporidium oocysts	(Summer, Fall) AND (No flow, low flow)	36	17	19	47%	0.90
Cryptosporidium oocysts	Spring	12	1	11	8%	0.09
Giardia cysts	All data	74	19	55	26%	0.35
Giardia cysts	(Spring) AND (Fall)	36	5	31	14%	0.16
Giardia cysts	Summer	38	14	24	37%	0.58

Supp. Cont. Table 3. Median and mean absolute deviation (MAD) of fecal indicator bacteria (CFU 100 mL<sup>-1</sup>) and parasites ((oo)cysts 100 L<sup>-1</sup>) associated with site, season and flow regime, as determined using least absolute deviation regression tree analysis using CART. Analysis uses <u>*Temporally Concurrent Data*</u>. High flow  $\geq 0.018 \text{ m}^3 \text{ s}^{-1}$ ; low flow  $\geq 0.002 \text{ m}^3 \text{ s}^{-1}$  and  $< 0.018 \text{ m}^3 \text{ s}^{-1}$ ; and no flow  $< 0.002 \text{ m}^3 \text{ s}^{-1}$ . N, number of samples in CART grouping.

, und no no		<u>8 </u>
Microorganism	Criteria	Median ± MAD, N
Total Coliform	All data	500 ± 19127, 255
Total Coliform	(Summer) AND (No flow)	1800 ± 132402, 30
Total Coliform	(Summer) AND (Low flow)	1000 ± 6429, 48
Total Coliform	(Summer) AND (High flow)	3700 ± 9057, 12
Total Coliform	Fall	430 ± 1033, 72
Total Coliform	Spring	150 ± 3944, 93
Fecal Coliform	All data	370 ± 3910, 246
Fecal Coliform	(Summer) AND (No flow)	1200 ± 16467, 30
Fecal Coliform	(Summer) AND (High flow, Low flow)	650 ± 493, 57
Fecal Coliform	Fall	360 ± 976, 69
Fecal Coliform	Spring	108 ± 3954, 90
E. coli	All data	128 ± 1621, 258
E. coli	Summer	320 ± 2996, 90
E. coli	(Spring) AND (Fall)	88 ± 860, 168
Enterococcus	All data	184 ± 1424, 258
Enterococcus	(Summer) AND (No flow)	370 ± 8245, 30
Enterococcus	(Summer) AND (High flow, Low flow)	530 ± 780, 60
Enterococcus	Fall	174 ± 284, 72
Enterococcus	Spring	50 ± 429, 96
C. perfringens	All data	6 ± 121, 252
C. perfringens	(No flow) AND (Summer)	11 ± 23, 30
C. perfringens	(No flow) AND (Spring, Fall)	30 ± 25, 18
C. perfringens	(High flow) AND (Low flow)	6 ± 143, 204
Aeromonas	All data	8700±15146, 84
Aeromonas	Summer	22000 ± 18663, 27
Aeromonas	(High Flow) AND (Spring)	440 ± 639, 12
Aeromonas	(High Flow) AND (Fall)	4600 ± 3653, 18
Aeromonas	(Spring, Fall) AND (Low flow, No flow)	6600 ± 18911, 27
Cryptosporidium oocysts	All data	6 ± 134, 74
Cryptosporidium oocysts	Fall	27 ± 39, 24
Cryptosporidium oocysts	(Spring) AND (Summer)	0 ± 176, 50
Giardia cysts	No useful split was found. No tree created	
F-RNA coliphage	No useful split was found. No tree created	
F-DNA coliphage	No useful split was found. No tree created	

Supp. Cont. Table 4. Associations between the densities of fecal indicator microorganisms and the presence and absence of pathogens. CART was used in classification mode defining indicator densities that best split target pathogen observations into homogeneous presence and absence groups (for *All Available Data*) (see Wilkes et al., 2009).

			For fecal indicator							
				microorga	nism dens	ities ≤	For fecal indicator microorganism densities >			
				CART split	value		CART Split	value		
Pathogen	Fecal indicator microorganism	Optimal fecal indicator split value	CART cross validated? Yes(Y)/ No(N)	Presence of pathogen	Absence of pathogen	Positive samples (%)	Presence of pathogen	Absence of pathogen	Positive samples (%)	Is the percentage positive in this grouping greater than ≤ grouping?
Campylobacter spp.	F-RNA coliphage *	3.35	Y	91	151	38	20	7	74	Yes, by 36%
Campylobacter spp.	F-DNA coliphage *	0.65	Y	81	142	36	30	16	65	Yes. by 29%
Campylobacter spp.	Total Coliforms **	245	Y	24	64	27	87	94	48	Yes. by 21%
Campylobacter spp.	Fecal Coliforms **	17	Y	2	19	10	109	139	44	Yes. by 34%
Campylobacter spp.	E. coli **	57	Y	17	52	25	94	106	47	Yes, by 22%
Campylobacter spp.	Enterococcus **	45	Y	8	36	18	103	122	46	Yes, by 28%
Campylobacter spp.	C. perfringens **	7.5	Y	49	97	34	62	61	50	Yes. by 16%
Campylobacter spp.	Aeromonas **	49500	Y	102	158	39	9	0	100	Yes, by 61%
Cryptosporidium	F-RNA coliphage *	300	Ŷ	58	46	56	2	0	100	Yes by 44%
Cryptosporidium	F-DNA coliphage *	27.5	Ŷ	55	44	56	5	5	50	No. lower by 6%
Cryptosporidium	Total Coliforms **	35	Y	0	7	0	60	39	61	Yes. by 61%
Crvptosporidium	Fecal Coliforms **	43	Y	0	10	0	60	36	63	Yes. by 63%
Cryptosporidium	E. coli **	3	Y	1	6	14	59	40	60	Yes, by 46%
Cryptosporidium	Enterococcus **	15.5	Y	4	8	33	56	38	60	Yes, by 27%
Cryptosporidium	C. perfringens **	0.5	Y	8	14	36	52	32	62	Yes, by 26%
Cryptosporidium	Aeromonas **	-	N	-	-	-	-	-	-	
Giardia	F-RNA coliphage *	100	Y	23	80	22	0	3	0	No, lower by 22%
Giardia	F-DNA coliphage *	-	N	-	-	-	-	-	-	· · ·
Giardia	Total Coliforms **	3800 <sup>i</sup>	Y	16	76	17	7	7	50	Yes, by 33%
Giardia	Fecal Coliforms **	285	Y	4	35	10	19	48	28	Yes, by 18%
Giardia	E. coli **	919	Y	19	79	19	4	4	50	Yes, by 31%
Giardia	Enterococcus **	213	Y	6	50	11	17	33	34	Yes, by 23%
Giardia	C. perfringens **	83	Y	23	80	22	0	3	0	No, lower by 22%
Norovirus GII	F-RNA coliphage *	-	N	-	-	-	-	-	-	
Norovirus GII	F-DNA coliphage *	1450	Y	14	93	13	1	0	100	Yes, by 87%
Norovirus GII	Total Coliforms **	-	N	-	-	-	-	-	-	
Norovirus GII	Fecal Coliforms **	-	N	-	-	-	-	-	-	
Norovirus GII	E. coli **	1075	Y	15	82	15	0	11	0	No, lower by 15%
Norovirus GII	Enterococcus **	-	Ν	-	-	-	-	-	-	
Norovirus GII	C. perfringens **	49 <sup>i</sup>	Y	15	74	17	0	19	0	No, lower by 17%
Norovirus GII	Aeromonas **	-	Ν	-	-	-	-	-	-	
Rotavirus	F-RNA coliphage *	1.75	Y	4	78	5	0	26	0	No, lower by 5%
Rotavirus	F-DNA coliphage *	0.25	Y	4	72	5	0	32	0	No, lower by 5%
Rotavirus	Total Coliforms **	-	Ν	-	-	-	-	-	-	
Rotavirus	Fecal Coliforms **	-	Ν	-	-	-	-	-	-	
Rotavirus	E. coli **	-	Ν	-	-	-	-	-	-	
Rotavirus	Enterococcus **	-	Ν	-	-	-	-	-	-	
Rotavirus	C. perfringens **	19 <sup>i</sup>	Y	0	61	0	4	43	9	Yes, by 9%
Rotavirus	Aeromonas **	6000	Y	0	38	0	4	66	6	Yes, by 6%
Salmonella spp.	F-RNA coliphage *	0.25	Y	15	208	7	0	47	0	No, lower by 7%
Salmonella spp.	F-DNA coliphage *	-	N	-	-	-	-	-	-	
Salmonella spp.	Total Coliforms **	65	Y	0	35	0	15	220	6	Yes, by 6%
Salmonella spp.	Fecal Coliforms **	1545	Y	15	214	7	0	41	0	No, lower by 7%
Salmonella spp.	E. coli **	-	Ν	-	-	-	-	-	-	
Salmonella spp.	Enterococcus **	115	Y	9	85	10	6	170	3	No, lower by 7%
Salmonella spp.	C. perfringens **	F	Ν	-	-	-	-	-	-	
Salmonella spp.	Aeromonas **	950'	Y	5	10	33	10	245	4	No, lower by 29%

i, CART selected primary split criteria amongst all possible indicator microorganism (F-RNA coliphage, F-DNA coliphage, total coliform, fecal coliform, *E. coli*, Enterococcus, *C. perfringens*) and *Aeromonas* criteria that cross validated.

\*PFU 100 mL<sup>-1</sup>.

\*\*CFU 100mL<sup>-1</sup>.

Supp. Cont. Table 5. Number of observations for <u>*Temporally Concurrent Data*</u> for different season and flow regimes.

Microorganism	Site	All data	Spring	Summer	Fall	High flow	Low flow	No flow
-		N	N	N	N	N	N	N
Campylobacter spp.	RCA <sub>in</sub>	64	18	26	20	25	26	13
Campylobacter spp.	RCA <sub>out</sub> /URCA <sub>in</sub>	64	18	26	20	25	26	13
Campylobacter spp.	URCA <sub>out</sub>	64	18	26	20	25	26	13
Salmonella spp.	RCA <sub>in</sub>	65	19	26	20	26	26	13
Salmonella spp.	RCA <sub>out</sub> /URCA <sub>in</sub>	65	19	26	20	26	26	13
Salmonella spp.	URCA <sub>out</sub>	65	19	26	20	26	26	13
E. coli 0157:H7	RCAin	65	19	26	20	26	26	13
E. coli 0157:H7	RCA <sub>out</sub> /URCA <sub>in</sub>	65	19	26	20	26	26	13
E. coli 0157:H7	URCA <sub>out</sub>	65	19	26	20	26	26	13
Cryptosporidium oocysts	RCA <sub>in</sub>	37	6	19	12	15	11	11
Cryptosporidium oocysts	URCA <sub>out</sub>	37	6	19	12	15	11	11
Giardia cysts	RCA <sub>in</sub>	37	6	19	12	15	11	11
Giardia cysts	URCA <sub>out</sub>	37	6	19	12	15	11	11
Norovirus GII	RCA <sub>in</sub>	27	9	9	9	11	13	3
Norovirus GII	RCA <sub>out</sub> /URCA <sub>in</sub>	27	9	9	9	11	13	3
Norovirus GII	URCA <sub>out</sub>	27	9	9	9	11	13	3
Rotavirus	RCA <sub>in</sub>	27	9	9	9	11	13	3
Rotavirus	RCA <sub>out</sub> /URCA <sub>in</sub>	27	9	9	9	11	13	3
Rotavirus	URCA <sub>out</sub>	27	9	9	9	11	13	3
F-RNA coliphage	RCA <sub>in</sub>	38	13	15	10	16	19	3
F-RNA coliphage	RCA <sub>out</sub> /URCA <sub>in</sub>	38	13	15	10	16	19	3
F-RNA coliphage	URCA <sub>out</sub>	38	13	15	10	16	19	3
F-DNA coliphage	RCA <sub>in</sub>	38	13	15	10	16	19	3
F-DNA coliphage	RCA <sub>out</sub> /URCA <sub>in</sub>	38	13	15	10	16	19	3
F-DNA coliphage	URCA <sub>out</sub>	38	13	15	10	16	19	3
Total Coliform	RCA <sub>in</sub>	85	31	30	24	37	32	16
Total Coliform	RCA <sub>out</sub> /URCA <sub>in</sub>	85	31	30	24	37	32	16
Total Coliform	URCA <sub>out</sub>	85	31	30	24	37	32	16
Fecal Coliform	<b>RCA</b> in	82	30	29	23	37	29	16
Fecal Coliform	RCA <sub>out</sub> /URCA <sub>in</sub>	82	30	29	23	37	29	16
Fecal Coliform	URCA <sub>out</sub>	82	30	29	23	37	29	16
E. coli	RCA <sub>in</sub>	86	32	30	24	38	32	16
E. coli	RCA <sub>out</sub> /URCA <sub>in</sub>	86	32	30	24	38	32	16
E. coli	URCA <sub>out</sub>	86	32	30	24	38	32	16
Enterococcus	RCA <sub>in</sub>	86	32	30	24	38	32	16
Enterococcus	RCA <sub>out</sub> /URCA <sub>in</sub>	86	32	30	24	38	32	16
Enterococcus	URCA <sub>out</sub>	86	32	30	24	38	32	16
Clostridium perfringens	RCA <sub>in</sub>	84	31	30	23	36	32	16
Clostridium perfringens	RCA <sub>out</sub> /URCA <sub>in</sub>	84	31	30	23	36	32	16
Clostridium perfringens	URCA <sub>out</sub>	84	31	30	23	36	32	16
Aeromonas	RCA <sub>in</sub>	28	11	9	8	12	13	3
Aeromonas	RCA <sub>out</sub> /URCA <sub>in</sub>	28	11	9	8	12	13	3
Aeromonas	URCA <sub>out</sub>	28	11	9	8	12	13	3

		All				High	Low	No
Microorganism	Site	data	Spring	Summer	Fall	flow	flow	flow
		Ν	Ν	N	Ν	Ν	N	N
Campylobacter spp.	RCA <sub>in</sub>	79	23	33	23	28	29	22
Campylobacter spp.	RCA <sub>out</sub> /URCA <sub>in</sub>	65	19	26	20	26	26	13
Campylobacter spp.	URCA <sub>mid</sub>	50	13	21	16	21	23	6
Campylobacter spp.	URCA <sub>out</sub>	75	25	28	22	28	29	18
Salmonella spp.	RCA <sub>in</sub>	79	23	33	23	28	29	22
Salmonella spp.	RCA <sub>out</sub> /URCA <sub>in</sub>	66	20	26	20	27	26	13
Salmonella spp.	URCA <sub>mid</sub>	50	13	21	16	21	23	6
Salmonella spp.	URCA <sub>out</sub>	75	25	28	22	28	29	18
E. coli 0157:H7	RCA <sub>in</sub>	79	23	33	23	28	29	22
E. coli 0157:H7	RCA <sub>out</sub> /URCA <sub>in</sub>	66	20	26	20	27	26	13
E. coli 0157:H7		50	13	21	16	21	23	6
E. coli 0157:H7	URCAout	75	25	28	22	28	29	18
Cryptosporidium oocysts	RCAin	40	7	19	14	15	11	14
Cryptosporidium oocysts	RCA <sub>out</sub> /URCA <sub>in</sub>	8	0	6	2	4	4	0
Cryptosporidium oocysts		0	0	0	0	0	0	0
Cryptosporidium oocysts		58	17	22	19	25	21	12
Giardia cysts	RCA:	40	7	19	14	15	11	14
Giardia cysts		8	0	6	2	4	4	0
Giardia cysts		0	0	0	0	0	0	0
Giardia cysts		58	17	22	19	25	21	12
Henatitis F	BCA	31	9	13	9	11	15	5
Henatitis E		27	9	9	9	11	13	3
Henatitis E		23	6	8	9	9	11	3
Hepatitis F		27	9	9	9	11	13	3
Norovirus GII	BCA:	31	9	13	9	11	15	5
Norovirus GII		27	9	9	9	11	13	3
Norovirus GII		23	6	8	9	9	11	3
Norovirus GII	URCA	27	9	9	9	11	13	3
Rotavirus	RCAin	31	9	13	9	11	15	5
Rotavirus	RCA <sub>sut</sub> /URCA <sub>in</sub>	27	9	9	9	11	13	3
Rotavirus	URCAmid	23	6	8	9	9	11	3
Rotavirus	URCA <sub>out</sub>	27	9	9	9	11	13	3
F-RNA coliphage	RCAin	44	13	18	13	16	20	8
F-RNA coliphage	RCA <sub>out</sub> /URCA <sub>in</sub>	39	13	15	11	17	19	3
F-RNA coliphage		31	8	13	10	14	16	1
F-RNA coliphage	URCA <sub>out</sub>	39	13	15	11	17	19	3
F-DNA coliphage	RCAin	44	13	18	13	16	20	8
F-DNA coliphage	RCA <sub>out</sub> /URCA <sub>in</sub>	39	13	15	11	17	19	3
F-DNA coliphage		32	9	13	10	14	16	2
F-DNA coliphage	URCA <sub>out</sub>	39	13	15	11	17	19	3
Total Coliforms	RCAin	108	36	42	29	41	35	32
Total Coliforms	RCA <sub>out</sub> /URCA	86	32	30	24	38	32	16
Total Coliforms	URCAmid	65	24	23	18	28	29	8
Total Coliforms	URCAout	104	32	36	36	43	40	21
Fecal Coliforms	RCAin	105	35	40	29	41	32	32

Supp. Cont. Table 6. Number of observations for <u>All Available Data</u> for different season and flow regimes.

		All				High	Low	No
Microorganism	Site	data	Spring	Summer	Fall	flow	flow	flow
Fecal Coliforms	RCA <sub>out</sub> /URCA <sub>in</sub>	83	31	29	23	38	29	16
Fecal Coliforms	URCA <sub>mid</sub>	62	23	23	16	28	27	7
Fecal Coliforms	URCA <sub>out</sub>	103	32	36	35	43	39	21
E. coli	RCA <sub>in</sub>	109	37	42	29	42	35	32
E. coli	RCA <sub>out</sub> /URCA <sub>in</sub>	87	33	30	24	39	32	16
E. coli	URCA <sub>mid</sub>	66	25	23	18	29	29	8
E. coli	URCA <sub>out</sub>	105	33	36	36	44	40	21
Enterococcus	RCA <sub>in</sub>	109	37	42	29	42	35	32
Enterococcus	RCA <sub>out</sub> /URCA <sub>in</sub>	87	33	30	24	39	32	16
Enterococcus	URCA <sub>mid</sub>	66	25	23	18	29	29	8
Enterococcus	URCA <sub>out</sub>	105	33	36	36	44	40	21
Clostridium perfringens	RCA <sub>in</sub>	107	36	42	28	40	35	32
Clostridium perfringens	RCA <sub>out</sub> /URCA <sub>in</sub>	85	32	30	23	37	32	16
Clostridium perfringens	URCA <sub>mid</sub>	64	24	23	17	27	29	8
Clostridium perfringens	URCA <sub>out</sub>	104	32	36	36	43	40	21
Aeromonas	RCA <sub>in</sub>	32	11	13	8	12	15	5
Aeromonas	RCA <sub>out</sub> /URCA <sub>in</sub>	28	11	9	8	12	13	3
Aeromonas	URCA <sub>mid</sub>	22	7	7	8	9	11	2
Aeromonas	URCA <sub>out</sub>	28	11	9	8	12	13	3