Journal of Water and Health | 13.3 | 2015

Characterizing the concentration of *Cryptosporidium* in Australian surface waters for setting health-based targets for drinking water treatment

APPENDIX A

The statistical models for characterizing *Cryptosporidium* concentration from discrete counts and volumes including recovery data were as follows:

Constant concentration

When counts are assumed to be generated from a Poisson (random) process, then the probability of counting n organisms in a sample of volume (V) given a mean concentration (∓) is:

$$P(n|\mu, V) = \frac{(\mu.V)^n \mathrm{e}^{-(\mu.V)}}{n!} \tag{A1}$$

Accounting for recovery: when recovery is assumed to be a binomial process, where each organism has a certain probability of being detected (π) or not being detected ($1 - \pi$), then the counts are assumed to follow a Poisson distribution with parameter $\pi\mu V$.

$$P(n|\pi, \mu, V) = \frac{(\pi.\mu.V)^n \mathrm{e}^{-(\pi.\mu.V)}}{n!} \tag{A2}$$

The log-likelihood function for concentration, ∓, given by Equation (A2) and a data set of m counts (n) with paired volume (V) and recovery (π) measurements is given by:

$$L(\mu\,|n_i, V_i, \pi_i) = \prod_{i=1}^{m} P(\mu|n_i, V_i, \pi_i) \tag{A3}$$

For each data set, the log-likelihood function was maximized to find the maximum likelihood estimator (MLE) for μ (mean concentration).

Variable concentration

If, rather than being constant, the mean concentration (∓) is assumed to follow a gamma distribution with parameters (→ and ℓ), then the probability distribution of counts $g(n|\lambda, \rho, V)$ is given by:

$$g(n|\mu, \lambda, \rho, V) = \int \frac{(\mu.V)^n e^{-(\mu.V)}}{n!} \frac{\lambda^\rho}{\Gamma(\rho)} \mu^{\rho-1} \mathrm{e}^{-\lambda\mu} \mathrm{d}\mu \tag{A4}$$

The solution to the integral can be rearranged into the form of the negative binomial count distribution described by gamma parameters → and ℓ:

$$g(n|\lambda, \rho, V) = \frac{\Gamma(\rho + n)}{n!\Gamma(\rho)} \frac{\lambda^\rho V^n}{(\lambda + V)^{\rho+n}} \tag{A5}$$

Accounting for recovery

When recovery is assumed to be a binomial process, then the counts are assumed to follow a Poisson distribution with parameter $\pi\mu V$. If μ is assumed to follow a gamma distribution with parameters λ and ρ, the $\pi\mu$ will also have a gamma distribution with parameters λ/π and ρ. The distribution of the number of organisms after passage of the process can be simplified to:

$$g(n|\pi, \lambda, \rho, V) = \frac{\Gamma(\rho + n)}{n!\Gamma(\rho)} \frac{\lambda^\rho (\pi V)^n}{(\lambda + \pi V)^{\rho+n}} \tag{A6}$$

The log-likelihood function for → and ℓ for a data set of m counts (n) with paired volume (V) and recovery (π)

measurements is then given by:

$$L(\lambda,\ \rho\ |n_i, V_i, \pi_i) = \prod_{i=1}^{m} g(\lambda, \rho|n_i, V_i, \pi_i) \tag{A7}$$

For each data set, the log-likelihood function was maximized to find the MLE for parameters λ and ρ describing the gamma distribution for variability in concentration.