Incorporating parameter uncertainty into Quantitative Microbial Risk Assessment (QMRA)
Margaret Donald, Kerrie Mengersen, Simon Toze, Jatinder P.S. Sidhu and Angus Cook

ABSTRACT
Modern statistical models and computational methods can now incorporate uncertainty of the parameters used in Quantitative Microbial Risk Assessments (QMRA). Many QMRAs use Monte Carlo methods, but work from fixed estimates for means, variances and other parameters. We illustrate the ease of estimating all parameters contemporaneously with the risk assessment, incorporating all the parameter uncertainty arising from the experiments from which these parameters are estimated. A Bayesian approach is adopted, using Markov Chain Monte Carlo Gibbs sampling (MCMC) via the freely available software, WinBUGS. The method and its ease of implementation are illustrated by a case study that involves incorporating three disparate datasets into an MCMC framework. The probabilities of infection when the uncertainty associated with parameter estimation is incorporated into a QMRA are shown to be considerably more variable over various dose ranges than the analogous probabilities obtained when constants from the literature are simply ‘plugged’ in as is done in most QMRAs. Neglecting these sources of uncertainty may lead to erroneous decisions for public health and risk management.

Key words | MCMC, parameter uncertainty, Quantitative Microbial Risk Assessment (QMRA) recycled water, risk assessment, Salmonella spp.

INTRODUCTION
In Australia, Quantitative Microbial Risk Assessment (QMRA) is recommended as the method of choice for assessing health risks from exposure to pathogens in recycled water, e.g. NRMMC (2006). The particular application examined in this paper is the risk of microbial infections associated with exposure to recycled water.

This paper presents a modification to the standard QMRA methodology, in which the risk assessor typically finds various quantities of interest, such as dose–response, die-off and/or log-reduction parameters and plugs these quantities into the risk assessment model. There is often little acknowledgement of the fact that these quantities are uncertain. We contrast this ‘plug-in’ approach with an approach based on a Bayesian risk assessment model, in which all the data which have been used to produce the quantities of interest necessary to the risk assessment are included. The uncertainty associated with the model parameters is therefore propagated throughout the analysis. This may be considered an extension of the standard QMRA model.

To illustrate the approach, we consider the probability of a person becoming infected with Salmonella spp. after being exposed to recycled wastewater. The scenario is not drawn
from actuality but is designed to illustrate the extension of the standard QMRA methodology. In the illustration, we ignore the problems of dose estimation, and investigate the part of risk estimation for which we have data available.

This paper is arranged as follows. First a standard QMRA method is outlined, followed by a brief description of the extended method, together with the datasets which will be used to illustrate it. The conceptual and statistical models into which these data are incorporated are then detailed, and the results are compared with those one would obtain without the incorporation of parameter uncertainty. This case study demonstrates that considerable uncertainty is induced in the probability of infection when this Bayesian approach is adopted. In the discussion, we elaborate the differences seen between the two methods and note the simplicity of our method.

**METHODS**

**Standard QMRA methodology**

A QMRA requires a knowledge of pathogen numbers at some stage of the treatment process, generally in the influent. Estimates for log reductions for various water treatment processes from pilot or other studies are then needed to estimate pathogen numbers in the treated water. A mechanism of ingestion and an amount of the treated recycled water ingested must be postulated or found. This, together with the pathogen numbers in the treated water, allows estimation of possible microbial doses. Finally, specification of a dose–response curve for the microbe of interest is needed to allow estimation of the probability of infection given a particular dose. A natural representation for a QMRA is via a graphical model such as Figure 1.

The QMRA of Figure 1 shows the steps for assessing the risk associated with eating a crop irrigated with recycled water. In such a figure, nodes without parents need information in order to run the risk assessment. Thus, for a standard QMRA, reading down the figure and from left to right, we need:

1. A description of the microbe numbers in either the wastewater or in the final treated water. Typically, if *Salmonella* spp. is sampled at all, it is sampled in the wastewater and may be described as coming from a log normal distribution with mean, $\mu$, and possibly a standard deviation $\sigma$.

2. ‘Log reductions’ in order to estimate the microbial numbers in the treated water. Water treatments are generally thought to reduce the numbers of pathogens at a rate proportional to the influent numbers of the pathogen in the water. This may be expressed in terms of log base 10, when it may be referred to as ‘log reduction’ or a decimal elimination capacity (DEC); see, for example, Hijnen et al. (2004, 2007). However, the DEC is typically given by a single number, e.g. 3, which would mean that $\log_{10}C_{\text{influent}} = \log_{10}C_{\text{effluent}} = 3$, where $C_{\text{influent}}$ is the number/L in the influent and $C_{\text{effluent}}$ is the number/L in the effluent. Such a log reduction would imply that the effluent numbers are

![Figure 1](image-url)
one thousandth those of the influent. To find these, published or grey literature involving the particular treatment type for a particular plant is searched.

3. A die-off constant $k$ or $T_{90}$ (time to 90% die-off). In the case study, where a field is irrigated with recycled wastewater, it is expected that sunlight will kill particular microbes at a rate proportional to their number, i.e. $dN/dt \propto N$ or $N_t = N_0 e^{-kt}$, where $k$ is sometimes referred to as the die-off constant. Other equations may be used, but this is a reasonably common approximation to die-off for some organisms, and is a good fit for the data used in this case study. Sinton et al. (2007) use a ‘shoulder’ equation (100/[1/(C_0/[1/(C_0)exp(-kT)])^(1/n)]), but, as is common, the quantities in their various equations are given as constants, with no error indicated.

4. Sunlight and shade hours, for the locality in which the recycled water is to be used.

5. A suitable amount of crop/water ingested by a person. One may use survey data if available, or use choices made by other researchers, for example, Tanaka et al. (1998). (Typically such data are supplied as constants.)

6. An equation and the parameters which describe the dose–response, i.e. the probability of becoming infected, having ingested a particular dose of the microbe. For Salmonella, the equation usually used is beta-Poisson, and from p 401 of Haas et al. (1999), the risk assessor would select $x = 0.3126$ and $N_{50} = 2.36 \times 10^4$, to give the probability of infection, $P$, from a given dose $D$, where $D$ is the number of microbes ingested, as: $P = 1 - [1 + (D/N_{50})(2^{1/2} - 1)]^{-x}$. In an alternative parameterization, we have $P = 1 - [1 + (D/\beta)]^{-x}$, where $\beta = N_{50}(2^{1/2} - 1) \approx 193$ 120, and $N_{50}$ is the number of microbes which give a 50% probability of infection.

Thus, to perform a risk assessment, the risk assessor performs a Monte Carlo simulation, working through the graphical model (Figure 1). Starting at some stage in the water processing cycle, an initial number/L of the pathogen is drawn from the water treatment distribution described by constants $\mu_0, \sigma_0$. This number is then reduced by either the value obtained by drawing a log-reduction value from the DEC distribution, described by $\mu_1, \sigma_1$, or, if no distribution is given or able to be inferred, then reduced by the DEC, $\mu_3$, for the process or processes. In the scenario considered, sunlight is expected to reduce pathogen numbers, so the die-off equation is used to give a final pathogen number in water which, then, together with a draw for the quantity of water ingested, gives the number of pathogens ingested. Finally, the probability of infection is calculated, via a dose–response equation, and a final draw made from a Bernoulli distribution to simulate the person’s infection status. This is repeated many times to simulate the risk, resulting in a distribution of the simulated endpoint risk.

For the case study, we consider an abbreviated version of the QMRA of Figure 1. This is represented by Figure 2. In this version, the information requirements enumerated above are limited to 1 (distribution for treated water, not influent), 3 (die-off constants), 4 (sunlight hours) and 6 (dose–response equation parameter constants). Table 1 shows the fixed constants used in the risk simulation of the QMRA of Figure 2.

As can be seen, this is not a risk assessment, since we abstract just a part of the full model in order to illustrate more clearly that much uncertainty may fail to be incorporated into risk assessments. In partial justification, we note that it is generally not thought worthwhile to monitor the end-use water for the pathogens of interest as it is believed that they
will be present in such small quantities and will be so diffuse within the water body that substantive positive results would only be obtained by processing impractically large samples. Data on pathogen reductions or log reduction studies exist but have been collected from typically small-scale, short-run experiments, usually in countries with very different climatic conditions. Moreover, such data may be owned by private utilities and are either not publicly available, or provided with minimal details. Thus, often only summary statistics or incomplete statistics (at best) filter into the public domain.

If the risk of a particular health outcome needs to be estimated, available data are even more limited. For example, *Salmonella* spp. have been linked with a number of outbreaks in the USA, Europe and Japan (Marks et al. 1998). In Australia, limited data for *Salmonella* spp. numbers and their inactivation by various wastewater treatment processes are available. The few studies available are those of Gibbs and others, and these focus on *Salmonella* spp. in sludge, rather than within the water fraction (Gibbs & Ho 1993; Gibbs 1995; Gibbs et al. 1995).

Thus, this study takes a small part of the risk assessment process and shows how it may be extended, by embedding data within a Bayesian framework to estimate the corresponding parameters, and thereby give better uncertainty estimates. The extended model for doing this is described in the next section.

### The extended QMRA model

In the extended model, the small experiments which lead to the various required constants are incorporated directly into the risk assessment process, allowing the uncertainty associated with the estimates to be automatically incorporated into the risk assessment. Thus, Figure 3, the extended model, contains two additional nodes (1 and 7) in comparison with Figure 2, the standard QMRA. These nodes represent the data which give rise to the ‘constants’ fed into the QMRA assessment of Figure 2, but in the extended model are used to derive estimates of the random quantities that describe these data.

The starting point for the extended model is the graphical representation of the QMRA which is seen to be a directed acyclic graph or DAG. Thus, a risk assessment may be embedded in a Bayesian framework, thereby allowing parameters to be estimated simultaneously with the risk assessment. In the extended model (Figure 3), parameters (supplied as constants under Figure 2) are both estimated and used for the derivation of other quantities. Thus, the model descriptions at nodes (2) and (6) are the explanatory models of the data at the new nodes (1) and (7), and also the means for estimation of dose after die-off (node 5) and estimation of the probability of a person becoming infected (node 8). In this Bayesian framework ‘prior’ probabilities (prior beliefs) for the parameters of the explanatory models are needed and uninformative priors are used in order that the parameter estimates and the uncertainty associated with them will closely approximate the maximum likelihood solutions for each set of parameters and data.

As with the standard QMRA model, a Monte Carlo approach is taken to analyse the extended model. Here, however, given the Bayesian setup and additional information, a more formal Markov chain Monte Carlo approach is used to estimate posterior distributions of the various quantities of interest such as dose–response parameters, die-off parameters and the risk of infection. The Bayesian framework

<table>
<thead>
<tr>
<th>Constant</th>
<th>Description</th>
<th>Value</th>
<th>Derivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)</td>
<td>(P(\text{infection}) = 1 - (1 + \text{dose}/\beta)^{-\alpha})</td>
<td>0.451</td>
<td>From Teunis et al. (1996)</td>
</tr>
<tr>
<td>(\beta)</td>
<td></td>
<td>15177</td>
<td>(as above)</td>
</tr>
<tr>
<td>(k_1)</td>
<td>Winter die-off constant (sunlight)</td>
<td>–0.3010</td>
<td>An earlier run</td>
</tr>
<tr>
<td>(k_2)</td>
<td>Winter die-off constant (shade)</td>
<td>–0.1237</td>
<td>(as above)</td>
</tr>
<tr>
<td>(k_3)</td>
<td>Summer die-off constant (sunlight)</td>
<td>–1.0390</td>
<td>(as above)</td>
</tr>
<tr>
<td>(k_4)</td>
<td>Summer die-off constant (shade)</td>
<td>–0.6457</td>
<td>(as above)</td>
</tr>
<tr>
<td>(S_W)</td>
<td>Sunlight hours (winter)</td>
<td>6.584</td>
<td>Mean June 2008 (Perth)</td>
</tr>
<tr>
<td>(S_S)</td>
<td>Sunlight hours (summer)</td>
<td>11.625</td>
<td>Mean January 2008 (Perth)</td>
</tr>
</tbody>
</table>
used was the freely available WinBUGS (Lunn et al. 2000). This is described in more detail later in the context of the case study.

We now give a more detailed description of the data and the models used to explain them.

Data for the extended model

Three disparate sources of information are integrated into the model described above: die-off data for *S. typhimurium*, dose–response data for *S. anatum* (Teunis et al. 1996) and a short run of weather data from an Australian city, Perth (Bureau of Meteorology 2010, pers. comm), giving the number of hours of sunlight in a summer and a winter month. We also use a fictitious pathogen distribution for the treated water with a range which allows the possibility of a 100% infection rate. These datasets and distributions are now described in more detail.

**Salmonella dose-response data (Figure 3, node 7)**

In considering the risks of *Salmonella* spp. poisoning, we chose to use the *S. anatum* data presented in the report by Teunis et al. (1996), in which infection curves were fitted by strain and species. These authors concluded that for *S. anatum* the three strains could be grouped together to determine a single dose–response curve, using a likelihood ratio test. Others have made different choices; thus Haas et al. (1999) used all 13 species and strains detailed in Teunis et al. (1996) and discarded some ‘outliers’, after similar testing, as did Oscar (2004). Each of these authors’ strategies gives a different set of quantities which the risk-assessor may use, but whatever parameter estimates he/she uses, they are used without any error being associated with them. Our purpose was not to determine a best model for dose–response for *Salmonella*, but to show how to incorporate the uncertainty associated with the estimation of such model parameters into a risk assessment.

A further reason for using *Salmonella* dose–response data is that the data are best summarised by a beta-Poisson dose–response curve, where the probability of infection for a given dose of *D* microbes, is given as

\[ P(\text{infection}) = 1 - (1 + D/\beta)^{-\alpha} \]

which is characterised by two parameters *(\alpha, \beta)* which are highly correlated.

Hence, there are two issues here in using these parameters in a risk assessment. Firstly, they are typically included as point estimates without acknowledgment of uncertainty of specification or of their correlation. However, even if this uncertainty were included, it is preferable to use the posterior distribution of the two parameters directly instead of making the standard assumption of bivariate normality, which Teunis et al. (1996) have shown does not hold.
Salmonella typhimurium die-off data (Figure 3, node 1)

Salmonella typhimurium die-off data from Sidhu et al. (2008) were supplied by the authors. These data allow the estimation of die-off rates with uncertainty for S. typhimurium under several conditions, winter/summer, sun/shade and grass/thatch. The available dataset consists of 34 observations. The summer observations were taken over all the combinations of conditions, but the winter data were for grass only and measured die-off in light and shade, thereby giving six sets of experimental conditions, and potentially six die-off constants. In the experiment, grass irrigated with sterile effluent was seeded with known numbers of S. typhimurium and samples of the grass and thatch were then harvested at 1, 2, 4, 6, 7.3 and 9.3 h after the initial seeding (in summer). Microbial numbers were counted and averaged for the samples taken at each harvest time and sample type (sun/shade, grass/thatch). For the winter samples, harvest times were 1, 2, 4, 6, and 8 h and grass and thatch were not separated. For further details of the experiment, see Sidhu et al. (2008).

Die-off over time is expected to be proportional to the number of organisms. Thus, \[ \frac{dN}{dt} = -kN \]. This equation has the solution \[ N_t = N_0e^{-kt} \], where \( k \) is positive. One can use any base for exponentiation and this changes the constant, \( k \), often referred to as the ‘die-off’ constant. To avoid confusion about bases for exponentiation, the constant used to express this equation may be given as \( T_{90} \) or the time to 90% die-off. Solving \( 0.10 = \exp(-kT_{90}) \) gives \( T_{90} \) in terms of \( k \) and vice versa.

Microbial count numbers are usually thought to be log normally distributed. Hence, the die-off distribution equation takes the form \( \log(N_t) \sim N(\log(N_0) - kt, \sigma^2) \) or, alternatively, \( \log(N_t/N_0) \sim N(-kt, \sigma^2) \). We fit the second version of this equation for each set of experimental conditions. This was done because greater effort (in terms of replicates) had gone into finding the value of the initial seedings. The original complete data had shown differences in the die-off rates for all combinations of sun/shade and winter/summer, but none for thatch. Hence, we fit four die-off constants, \( k_1, \ldots, k_4 \) to the model \( \log(N_{ti}/N_{i0}) \sim N(-k_{ti}, \sigma^2) \), where \( i \) references the summer/winter, sun/shade combination, and a common (pooled) variance \( \sigma^2 \) is used. Natural logs were used and the die-off value was not constrained to be negative; indeed, 14% of the posterior estimates for die-off in winter and in the shade were positive. When this occurred, a zero was substituted in the corresponding decay equation in the risk modelling (see below), although some evidence exists for the regrowth of Salmonella spp. on lettuce leaves under the right circumstances (Brandl & Amundson 2008).

The posterior estimates for die-off act on the treated water pathogen number node (node 3 of Figure 3), together with the sunlight hours of node 4, to produce the dose after die-off at node 5. The die-off calculation uses the maximum 17 h period for winter and summer available for die-off, based on the irrigation regime for the sports ovals in Perth where the experimental die-off data were collected. Note that the various values of \( k \) are the die-off constants of item 3 in the description of standard QMRA methodology.

Sunlight hours (Figure 1, Figure 2 and Figure 3, node 4)

Die-off is a function of sun/shade, summer/winter. The daily sunshine hours at Perth Airport, for January 2008 and June 2008 were supplied by the Australian Bureau of Meteorology (2010 pers.comm.). Rather than work with summary statistics, or fit a distribution to these data, they were resampled. These data are clearly not normally distributed (see Figure 4) nor are they expected to be, since the number of sunlight hours is bounded by 0 and the number of possible hours of sunlight on a particular day at the latitude of Perth. Figure 4 indicates a mixture of rainy and sunny days and discretization. Given that the data are bounded and possibly a mixture of distributions, it seemed more sensible to resample, rather than to fit and sample from an arbitrary distribution.
Doses (item 5, standard QMRA methodology section and Figure 1)

No data were used for the person’s dose. This node is not included in the case study (Figures 2 and 3).

The Salmonella dose–response curve (Figure 5) shows that very high doses of S. anatum are required for infection. (Using the point estimates found by Teunis et al. (1996), a dose of 400 S. anatum gives a probability of infection of 0.01, while for a dose of 1000 the probability of infection becomes 0.03.) Given the probabilities of infection, it appears that the dose–response curve applies largely to healthy adults. Since small children and the elderly are more likely to become ill under the same dosing regime, a range of doses was induced (via the treated water pathogen numbers distribution) in order to see the effect of parameter uncertainty over the full dose–response curve.

Under the models used in this case study, the node ‘person’s microbe dose’ of Figure 1 is equated to ‘dose after die-off’ (Figure 2 and node 5 of Figure 3).

Treated water numbers

This node is a derived node in Figure 1 but an initial node in Figures 2 and 3. The pathogen numbers’ distribution in treated water was ascribed a (natural) log-uniform distribution over the range (-1, 30). Thus, in this study, both consumption rates and numbers per L of the pathogen in the recycled or the influent source water were ignored. Instead, an arbitrary distribution was chosen for the Salmonella spp. numbers distribution in treated water, to allow the possibility of seeing the effect of the uncertainty in die-off rates and the uncertainty in the dose–response parameters on the estimate of the probability of infection, under many possible scenarios.

Putting it all together

Conceptual model

The directed acyclic graph for the extended model (the ‘conceptual model’) is given by Figure 3. Here, the Sidhu et al. (2008) data (node 1) are explained by the regression model (Equation (3)) of node 2 which estimates the die-off parameters. For each iteration of the MCMC algorithm, the four die-off rates are estimated; a dose sample is drawn from node 3; and a sample is drawn from the sunlight hour data of node 4. At node 5, the die-off constants for this iteration are applied using Equations (4)–(8) for a 17 h day with the sunlight and shade hours from node 4. When the draw for the winter shade or sunlight die-off parameter is negative, it is replaced by zero.

Independently, the Teunis et al. (1996) dose–response data for S. anatum (node 7) are explained by the current estimates of \((\alpha, \beta)\) (node 6, and fitted using Equations (1) and (2)), which are used in the same MCMC iteration (at node 8, using Equations (9) and (10)) to calculate the probability of infection, thus allowing a single estimate of the probability of infection (and the infection status of an individual) at each iteration.

Statistical model

Node 7 contains the dose–response data from Teunis et al. (1996) which may be represented as \((D_i, N_i, X_i), i = 1, \ldots, 19\), where \(D_i\) is the \(i\)th dose, \(N_i\) is the number of subjects given the \(i\)th dose and \(X_i\) is the number of subjects infected by the \(i\)th dose.

These are explained by the dose–response equation (node 6, Equations (1) and (2)) with parameters \((\alpha, \beta)\). Uninformative log-uniform priors are given for \((\alpha, \beta)\), and after burn-in, the posteriors for \((\alpha, \beta)\) are essentially identical to the
maximum likelihood estimates. Thus, nodes 6 and 7 are described by

\[ X_i \sim \text{Bin}(p_i, N_t) \]  
\[ p_i = 1 - \left(1 + \frac{D_i}{\beta}\right)^{-z}. \]  

(1) (2)

with priors for \((x, \beta)\) given by

\[ \ln(x) \sim U(-10, 15) \]  
\[ \ln(\beta) \sim U(-6, 20). \]

The current MCMC simulation of the posteriors for \((x, \beta)\) is passed to node 8, again using Equations (1) and (2) (but now in the form of (9) and (10)), to give a value for the probability of infection (and whether an individual is infected) after sunlight die-off.

Node 1 represents the die-off data, which may be considered as \((L_j, t_j, N_{0(j)}, N_{t(j)}), j = 1, \ldots, 34\). \(L_j\) references each data point, while \(L_j = 1, \ldots, 6\), represents the line and experimental condition to which the \(j\)th point belongs, and there are 6 of these corresponding to the number of different conditions of the experiment, \(t_j\) is the number of hours elapsed from the initial seeding (with count \(N_{0(j)}\) on line \(L_j\)) and \(N_{t(j)}\) is the count at time \(t_j\) for line \(L_j\). The die-off constants \(k_{L(j)} = k_{L_0}\), \(L_j = 1, \ldots, 6\), \(j = 1, \ldots, 34\).

However, as discussed earlier, different values of \(k\) are fitted for summer/winter and sun/shade, since the grass/thatch in combination with sun/shade for summer did not need separate fits. The die-off regression equations (node 2) which explain the die-off data are given by

\[ \ln\left(\frac{N_{0(j)}}{N_{t(j)}}\right) \sim N(-t_j k_1, \sigma^2) \]  

(3)

with uninformative priors for \(k_1 (l = 1, \ldots, 4)\) and \(\sigma^2\), given by

\[ k_1 \sim N(0, 1000) \]  
\[ \sigma^2 \sim IG(0.01, 0.01). \]

The posterior estimates for \(k_1\) and \(\sigma^2\) in the MCMC simulation are used at node 5, to estimate the dose after die-off from sunlight, based on the dose from node 3, and the sunlight hours from node 4.

At node 4 the season sunlight hours are sampled directly from the data \((S_m, h_m)\), where \(S_m\) is the season (winter/summer) and \(h_m\) are the sunlight hours for the day of that season.

Let \(D_0\) be the initial number of pathogens drawn from the treated water distribution (node 3) and the number of hours of sunlight drawn in winter/summer be \(h\) (node 4). Then \(D_{17}\), the number of pathogens 17 h after irrigation, is drawn from

\[ \log(D_{17}/D_0) \sim N(-k_1 h - k_2(17 - h), \sigma^2) \]

Winter, \(k_1, k_2 \geq 0\)  
\[ \sim N(-k_2(17 - h), \sigma^2) \]  
where \(k_1 < 0, k_2 \geq 0\)  
\[ \sim N(-k_1 h, \sigma^2) \]  
where \(k_1 \geq 0, k_2 < 0\)  
\[ \sim N(0, \sigma^2) \]  
where \(k_1, k_2 < 0\)

(4) (5) (6) (7)

\[ \log(D_{17}/D_0) \sim N(-k_3 h - k_4(17 - h), \sigma^2) \]

Summer  

(8)

where \(k_1, \ldots, k_4\) and \(\sigma^2\) are posterior draws from node 2. (Note that, although there may be a possibility of bacterial growth, this possibility was not permitted in the risk estimation since where an estimate for any winter die-off \(k\) value was negative, it was replaced by zero.) \(D_0\), the initial dose (node 3: treated water/effluent distribution), is drawn from a log uniform distribution which allows the full curve for the probability of infection to be seen:

\[ \ln(D_0) \sim U(-1, 1000). \]

\(D_{17}\) then passes to node 8, where the probability of infection is calculated using the current posterior estimates for \(x\) and \(\beta\) (from node 6). Then \(p_{\text{inf}}\) the probability of infection, and \(I\) (whether an individual is infected or not, taking a value of 1 for infected, 0 for not infected), are given by

\[ p_{\text{inf}} = 1 - \left(1 + \frac{D_{17}}{\beta}\right)^{-z} \]

(9)

\[ I \sim \text{Bin}(p_{\text{inf}}, 1). \]

(10)

As noted earlier, the model described above and in Figure 3 were implemented in WinBUGS (Lunn et al. 2000). The initial distribution of the dose is drawn from a log uniform distribution to allow the consequences of parameter uncertainty at any dose to be explicitly included. In the simulation, for the draw of each dose, each parameter is drawn conditional on the data and all other associated parameters.
For the final results, a burn-in of 30 000 was used to reach the target distributions for dose and die-off, with a further 10 000 iterations used for the ‘risk’ estimation. Two chains and Gelman–Rubin statistics (Lunn et al. 2000) for each of the quantities of interest were used to verify convergence to the stationary distribution.

Further extensions to the model

The model described above can be extended in a number of ways. We present here two further conceptual models, which again can be expressed as DAGs: an errors-in-variables (Fuller 1987; Wand 2009) model for the estimation of the dose–response model (Figure 6) and a DAG for the incorporation of the errors-in-variables model into the QMRA presented here (Figure 7). The errors-in-variables model estimates the parameters of the dose–response equation on the assumption that the doses are measured with error, and is detailed below. Not surprisingly, the additional uncertainty postulated in this model increases the uncertainty associated with the estimation of the probability of infection. This approach is appropriate: dose is measured with error and this should be taken into account when estimating the dose–response curve, though this example is intended to be illustrative rather than definitive. The second model, Figure 7, expands node 7 of Figure 3 and shows how the errors-in-variables dose–response model would be integrated into the ‘risk assessment’ carried out in this paper. That such a model can be easily fit and incorporated into a risk assessment further justifies the data-based risk assessment approach used here.

Estimating dose–response assuming errors in dose

McCullough & Eisele (1951) prepared batches of *S. anatum* for which the *S. anatum* count was measured. In the model we present below, we recognise the difficulty of ascertaining such dosages. Thus, we assume that the batch dose is measured with error and that the individual’s true dose from the batch is not the true batch dose. The individual then becomes infected or not infected. In the model, the status of infected/not infected has been assumed to be measured with no error. Figure 6 shows a schematic directed acyclic graph for this model.

Let the unobserved true dose of batch $b$ be $Z_b$, the unobserved true dose for individual $i$ subjected to batch $b$ be $Y_{i(b)}$, the observed dose for batch $b$ be $X_b$ and the infected
status of individual $i$ be $I_i$. There were 19 dosage batches, and 114 individuals each received a dose from a particular batch. Then letting $i(b)$ reference the individual $i$ receiving a dosage from batch $b$ and $p_i$ be the probability of individual $i$ becoming infected, we have

$$\log(Z_b) \sim N(0, \sigma^2)$$
$$\log(X_b) \sim N(\log(Z_b), 0.001)$$

where $b = 1, ..., 19$ and $1/\sigma^2 \sim \Gamma(0.1, 0.1)$. 

$$\log(Y_{i(b)}) \sim N(\log(Z_b), 0.001)$$

$$p_i = 1 - (1 + Y_{i(b)}/\beta)^{-2}$$

$I_i \sim \text{Bernoulli}(p_i)$

where $i = 1, ..., 114$. (All measured batch doses were divided by 1000 prior to fitting.)

The assumed errors in measurement used here are possibly unrealistic, with the variance of the errors for the true batch dose, and for the true individual dose being set at 0.001, but they do affect the model and particularly the width of its credible intervals. (When we consider the rounding of McCullough and Eisele’s $S. anatum$ numbers, it is, however, more likely that we have underestimated the variance.)

**RESULTS**

Figure 8 shows the bivariate posterior distribution of the dose–response parameters ($\alpha$, $\beta$); as indicated earlier, this is unlikely to be bivariate normal and indeed this is apparent from the figure. The parameters are highly correlated and the surface of the loglikelihood at the point of convergence is fairly flat (not shown), which means that the values of the parameters estimated using conventional maximum likelihood methods are somewhat dependent on the stopping rule for convergence. In terms of the methods advocated in this paper, it would seem that the dose–response curve parameters are not distributed as a bivariate normal, and that to simulate such a distribution via some summary parameters would be relatively difficult.

Figure 5 shows the dose–response curve distribution given by these parameters’ posterior distribution. This curve is created from the outputs of nodes 1 and 2 and shows the estimates of the probability of infection based solely on the Teunis et al. (1996) data. The considerable variation of the probability at low doses should be noted (not shown in the graph, but noted from the MCMC data). $P(\text{infection})$ between $e^0$ and $e^{10}$ ($2 \times 10^5$) ranges from almost zero to occasionally 0.5 for the same dose. Even for a dose of 20 bacteria ($e^2$) some realisations show a probability of infection of 0.2. For extremely high dose values, the majority of probabilities of infection are close to one, but occasionally the probability is considerably less.

Figure 9 shows the distributions for the die-off parameters for summer/winter and sun/shade. These are fairly symmetric, reflecting in part the model assumptions of normality, given the few data. Note that, for shade in winter, a
substantial proportion (14%) of the posterior die-off values are negative, which could lead to the inference of no die-off under these conditions.

**Figure 4** shows the daily sunlight data for winter or summer, and for both periods it can be seen that the majority of days were neither cloudy, overcast or rainy, but attained the maximum possible number of sunlight hours.

In **Figure 10**, the probabilities of infection for summer and winter (estimated with parameter uncertainty – ‘Varying’) are contrasted with the corresponding estimates where all needed parameter uncertainty is accounted for.

**Figure 10** | Summer and winter: probability of infection – constant vs varying.
values have been plugged in as constants (‘Constant’). This figure shows the addition of considerable variation when the underlying data and their model are incorporated in the risk assessment model. When most of the probabilities lie below 0.5, the additional uncertainty increases the range of probabilities, thereby giving an increased likelihood of infection. When most of the probabilities lie above 0.5, the added uncertainty again gives a wider range and therefore includes lower probabilities of infection in comparison with the model using a constant.

Box plots for the probability of infection for 15 initial dose groups (Figure 11) indicate that, if the infection probabilities are not close to zero or one, the uncertainty is very greatly increased. Thus, including parameter uncertainty could make a very great difference to conclusions about risk. Table 3 gives summary statistics for the initial doses by grouping. Table 2 gives summary statistics for the probability of infection for each of these groupings shown in the graphs (Figure 11). Thus, in Table 2, in dose grouping 8, the mean probability of infection for winter when the constants are used is 0.82 with a 90% CI (0.75, 0.89), compared with 0.78 (0.43, 0.96) for the varying parameters, again more than double the spread. Table 3 shows that the initial dose range for dose grouping 8 is $6.4 \times 10^6$ to $5.52 \times 10^7$ with a median dose of $1.89 \times 10^7$ cells. Looking at the summer scenarios for dose grouping 12 (Table 2), the mean probability of infection is 0.25 with 90% CI (0.12, 0.40) using constants, compared with a mean probability of infection of 0.30 (0.04, 0.72), when parameters are drawn with uncertainty from their distributions. This equates to a difference in interval width of 0.68 versus 0.28. That is, using varying parameters the credible interval covers two-thirds of the probability scale, whereas using constants the interval is one-third of the scale, constituting a very substantial difference.

The effect of the uncertainty induced by the uncertainty of the die-off parameters was so great that it seemed useful to estimate the probability of infection for the initial dose (divided by 1000) allowing no die-off, to show the effect of the uncertainty induced by the dose–response parameters alone. Figure 12 shows the effect of including the uncertainty of the parameter estimates for dose–response when no die-off is considered. The results for each grouping are shown in Tables 2 and 3. Thus, for dose grouping 8, using constants only, the 90% interval is (0.16, 0.48), but (0.12, 0.56) with varying parameters, a difference in width of 0.32 to 0.44. This is a considerable difference when the response lies between 0 and 1. From Table 3, this group is seen to encompass doses from $6.4 \times 10^3$ to $5.5 \times 10^4$ cells.

Figure 13 shows the additional uncertainty in the probabilities of infection from the dose–response model which results from the errors-in-variables model. The 95% credible intervals for the probability of infection when error is assumed in the dose are considerably wider, but also the mean probability of infection has moved to the left, indicating a higher probability of infection for a dose than when the probability is estimated without recognising that the dose is measured with error. The additional uncertainty operates to make the probability of infection markedly higher at lower doses. (Note that none of the results discussed in the ‘risk
<table>
<thead>
<tr>
<th>Group</th>
<th>Period</th>
<th>Type</th>
<th>Mean</th>
<th>Std.</th>
<th>Median</th>
<th>q5</th>
<th>q95</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>No die-off</td>
<td>Constant</td>
<td>0.07</td>
<td>0.001</td>
<td>0.06</td>
<td>0.03</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.10</td>
<td>0.003</td>
<td>0.07</td>
<td>0.02</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>Constant</td>
<td>0.00</td>
<td>0.000</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.00</td>
<td>0.000</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Winter</td>
<td>Constant</td>
<td>0.57</td>
<td>0.004</td>
<td>0.58</td>
<td>0.41</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.54</td>
<td>0.009</td>
<td>0.56</td>
<td>0.11</td>
<td>0.85</td>
</tr>
<tr>
<td>8</td>
<td>No die-off</td>
<td>Constant</td>
<td>0.31</td>
<td>0.004</td>
<td>0.31</td>
<td>0.16</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.34</td>
<td>0.005</td>
<td>0.33</td>
<td>0.12</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>Constant</td>
<td>0.00</td>
<td>0.000</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.00</td>
<td>0.000</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Winter</td>
<td>Constant</td>
<td>0.82</td>
<td>0.002</td>
<td>0.83</td>
<td>0.73</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.78</td>
<td>0.006</td>
<td>0.82</td>
<td>0.43</td>
<td>0.96</td>
</tr>
<tr>
<td>9</td>
<td>No die-off</td>
<td>Constant</td>
<td>0.66</td>
<td>0.005</td>
<td>0.67</td>
<td>0.52</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.67</td>
<td>0.004</td>
<td>0.68</td>
<td>0.49</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>Constant</td>
<td>0.00</td>
<td>0.000</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.01</td>
<td>0.001</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Winter</td>
<td>Constant</td>
<td>0.93</td>
<td>0.001</td>
<td>0.93</td>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.91</td>
<td>0.003</td>
<td>0.93</td>
<td>0.76</td>
<td>0.99</td>
</tr>
<tr>
<td>10</td>
<td>No die-off</td>
<td>Constant</td>
<td>0.86</td>
<td>0.002</td>
<td>0.86</td>
<td>0.79</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.85</td>
<td>0.002</td>
<td>0.86</td>
<td>0.76</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>Constant</td>
<td>0.01</td>
<td>0.000</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.04</td>
<td>0.004</td>
<td>0.01</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Winter</td>
<td>Constant</td>
<td>0.97</td>
<td>0.000</td>
<td>0.97</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.95</td>
<td>0.002</td>
<td>0.97</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>11</td>
<td>No die-off</td>
<td>Constant</td>
<td>0.94</td>
<td>0.001</td>
<td>0.95</td>
<td>0.92</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.94</td>
<td>0.001</td>
<td>0.95</td>
<td>0.87</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>Constant</td>
<td>0.05</td>
<td>0.001</td>
<td>0.04</td>
<td>0.02</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.12</td>
<td>0.006</td>
<td>0.05</td>
<td>0.01</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Winter</td>
<td>Constant</td>
<td>0.99</td>
<td>0.000</td>
<td>0.99</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.98</td>
<td>0.001</td>
<td>0.99</td>
<td>0.93</td>
<td>1.00</td>
</tr>
<tr>
<td>12</td>
<td>No die-off</td>
<td>Constant</td>
<td>0.98</td>
<td>0.000</td>
<td>0.98</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.97</td>
<td>0.001</td>
<td>0.98</td>
<td>0.95</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>Constant</td>
<td>0.25</td>
<td>0.003</td>
<td>0.24</td>
<td>0.12</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.30</td>
<td>0.008</td>
<td>0.26</td>
<td>0.04</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Winter</td>
<td>Constant</td>
<td>1.00</td>
<td>0.000</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varying</td>
<td>0.99</td>
<td>0.001</td>
<td>1.00</td>
<td>0.96</td>
<td>1.00</td>
</tr>
</tbody>
</table>

q5: 5th percentile of posterior distribution.
q95: 95th percentile of posterior distribution.
The extended QMRA model was expressed as a Bayesian model and analysed using a simulation-based approach, namely Markov chain Monte Carlo (MCMC) to estimate distributions of the probability of infection, thereby taking into account the uncertainty associated with parameter estimates needed in the risk assessment, automatically and more satisfactorily. In general, when parameter uncertainty is taken into account, it is typical to assume that the parameter estimate is normally distributed, which it may well not be. The manner in which uncertainty is incorporated in

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean pathogens</th>
<th>Std. pathogens</th>
<th>Median pathogens</th>
<th>Min. pathogens</th>
<th>Max. pathogens</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$1.21 \times 10^0$</td>
<td>$2.70 \times 10^{-2}$</td>
<td>$1.06 \times 10^0$</td>
<td>$3.70 \times 10^{-1}$</td>
<td>$2.81 \times 10^0$</td>
</tr>
<tr>
<td>1</td>
<td>$8.53 \times 10^0$</td>
<td>$1.80 \times 10^{-1}$</td>
<td>$7.37 \times 10^0$</td>
<td>$2.82 \times 10^1$</td>
<td>$1.99 \times 10^1$</td>
</tr>
<tr>
<td>2</td>
<td>$7.41 \times 10^1$</td>
<td>$1.71 \times 10^0$</td>
<td>$6.17 \times 10^1$</td>
<td>$1.99 \times 10^1$</td>
<td>$1.74 \times 10^2$</td>
</tr>
<tr>
<td>3</td>
<td>$5.68 \times 10^2$</td>
<td>$1.23 \times 10^1$</td>
<td>$4.81 \times 10^2$</td>
<td>$1.75 \times 10^2$</td>
<td>$1.28 \times 10^3$</td>
</tr>
<tr>
<td>4</td>
<td>$4.29 \times 10^3$</td>
<td>$9.65 \times 10^1$</td>
<td>$3.59 \times 10^3$</td>
<td>$1.30 \times 10^3$</td>
<td>$1.03 \times 10^4$</td>
</tr>
<tr>
<td>5</td>
<td>$3.89 \times 10^4$</td>
<td>$9.36 \times 10^2$</td>
<td>$3.20 \times 10^4$</td>
<td>$1.03 \times 10^4$</td>
<td>$9.64 \times 10^4$</td>
</tr>
<tr>
<td>6</td>
<td>$3.38 \times 10^5$</td>
<td>$7.61 \times 10^3$</td>
<td>$2.87 \times 10^5$</td>
<td>$9.70 \times 10^4$</td>
<td>$7.97 \times 10^5$</td>
</tr>
<tr>
<td>7</td>
<td>$2.75 \times 10^6$</td>
<td>$6.19 \times 10^4$</td>
<td>$2.31 \times 10^6$</td>
<td>$8.00 \times 10^5$</td>
<td>$6.40 \times 10^6$</td>
</tr>
<tr>
<td>8</td>
<td>$2.29 \times 10^7$</td>
<td>$5.32 \times 10^5$</td>
<td>$1.89 \times 10^7$</td>
<td>$6.40 \times 10^6$</td>
<td>$5.52 \times 10^7$</td>
</tr>
<tr>
<td>9</td>
<td>$1.83 \times 10^8$</td>
<td>$3.96 \times 10^6$</td>
<td>$1.59 \times 10^8$</td>
<td>$5.53 \times 10^7$</td>
<td>$4.14 \times 10^8$</td>
</tr>
<tr>
<td>10</td>
<td>$1.52 \times 10^9$</td>
<td>$3.63 \times 10^7$</td>
<td>$1.27 \times 10^9$</td>
<td>$4.14 \times 10^8$</td>
<td>$3.67 \times 10^9$</td>
</tr>
<tr>
<td>11</td>
<td>$1.16 \times 10^{10}$</td>
<td>$2.50 \times 10^8$</td>
<td>$9.88 \times 10^9$</td>
<td>$3.68 \times 10^9$</td>
<td>$2.69 \times 10^{10}$</td>
</tr>
<tr>
<td>12</td>
<td>$8.35 \times 10^{10}$</td>
<td>$1.72 \times 10^9$</td>
<td>$7.28 \times 10^{10}$</td>
<td>$2.70 \times 10^{10}$</td>
<td>$1.89 \times 10^{11}$</td>
</tr>
<tr>
<td>13</td>
<td>$6.09 \times 10^{11}$</td>
<td>$1.35 \times 10^{10}$</td>
<td>$5.01 \times 10^{11}$</td>
<td>$1.90 \times 10^{11}$</td>
<td>$1.58 \times 10^{12}$</td>
</tr>
<tr>
<td>14</td>
<td>$4.57 \times 10^{12}$</td>
<td>$1.0 \times 10^{11}$</td>
<td>$3.93 \times 10^{12}$</td>
<td>$1.38 \times 10^{12}$</td>
<td>$1.07 \times 10^{13}$</td>
</tr>
</tbody>
</table>

For the no die-off results, the doses are 1/1000th of these.

DISCUSSION

The extended QMRA model was expressed as a Bayesian model and analysed using a simulation-based approach, namely Markov chain Monte Carlo (MCMC) to estimate distributions of the probability of infection, thereby taking into account the uncertainty associated with parameter estimates needed in the risk assessment, automatically and more satisfactorily. In general, when parameter uncertainty is taken into account, it is typical to assume that the parameter estimate is normally distributed, which it may well not be. The manner in which uncertainty is incorporated in

![Figure 12](image1.png)  ![Figure 13](image2.png)

Figure 12 | Probability of infection (no die-off): using constant vs varying parameters for beta-binomial distribution.

Figure 13 | Comparison of the dose-response curves for S. anatum with 95% credible intervals, estimated with and without 'errors-in-variables'.

assessment' include this modification to nodes 6 and 7 of Figure 3.)
the extended model allows the experimental data to dictate the distribution of the parameter uncertainty and allows the possibility of asymmetry and long tails. The Bayesian framework permits embedding of several unrelated models in a single risk assessment, via directed acyclic graphs (DAGs), and may be compared with more conventional risk assessments, where parameter estimates and their associated distributional assumptions are used. As examples of such risk assessments, see, for example, Whiting & Buchanan (1997), Tanaka et al. (1998), Pouillot et al. (2004) and Gerba et al. (2008).

This approach may be compared with that proposed by other researchers (Teunis et al. 1997; Haas et al. 1999) who prepare a bootstrap sample of parameter estimates for the dose–response curve, thereby allowing for non-normality, prior to running the risk assessment. However, despite this being the method recommended in Haas et al. (1999), most risk assessments take their dose–response parameters as constants. It would seem that bootstrapping, choosing a size for the bootstrap sample and incorporating the resultant bootstrap sample into the simulation framework is generally discouraging for most practitioners. When the interest in a risk assessment involves tails of distributions or upper percentiles, it is critical not to ignore the tail behaviours of the fitted distributions. The method we propose permits all such asymmetries and uncertainties to be easily incorporated. In this particular case, where the dose–response parameters are correlated, the problem of simulation is particularly difficult, since the two parameters are not bivariate normal. Hence, when the dose–response curve is a beta-Poisson, some appropriate method must be used to capture the bivariate behaviour. Haas (1999) proposes a further method based on rank coefficients. This method is again complex to implement, whereas here we argue that using MCMC via Win-BUGS is not.

There are so little data used in the estimation of the die-off coefficients that there is little evidence of asymmetry. Nonetheless, using data rather than previously estimated constants ensures that uncertainty is propagated properly throughout the simulation. Had the ‘shoulder’ equation of Sinton et al. (2007) been considered appropriate, the same problem of correlated parameter estimates for the curve fit would again be as strongly evident as they are for the dose–response equation.

MCMC simulation and estimation has been available for some time, but is rarely used in the context of risk assessment as described here. Kelly & Smith (2009) present a simple primer of MCMC methods for this purpose and, in particular, discuss its use in fitting hierarchical models and in dealing appropriately with missing and uncertain data. Messner et al. (2001) use an MCMC approach to perform a meta-analysis using hierarchical MCMC modelling to develop a dose–response curve for C. parvum. The same approach is taken by Qian et al. (2004) who use MCMC to fit a hierarchical model to perform a meta-analysis for various studies of protozoan inactivation by UV light. Delignette-Muller et al. (2006) used a complex hierarchical model to describe the growth of Listeria in cold-smoked salmon and then used this to develop a further model for the time necessary to reach particular pathogen numbers, with the second model importing the uncertainty implicit in the original data.

Paulo et al. (2005) undertook a risk assessment, which closely parallels our approach: the parameter estimation for various submodels and the assessment of dietary exposure to pesticides (as a final node) was accomplished in the one MCMC model. Like the models of Paulo et al. (2005), the model presented in this paper differs substantially from the majority of risk assessments in two main ways. Firstly, it embeds the primary data (for dose–response, Salmonella die-off and doses) within the simulation itself, thus incorporating directly the uncertainties of the data, together with the (unknown) correlation structures. That is, no summary data are used and no process is represented by a constant. Secondly, by putting these together in a directed acyclic graph (DAG) and using MCMC, the model allows us to simultaneously estimate all the parameters currently used in a QMRA, together with the risks in which we are interested.

Here, every parameter, however disparate, is estimated simultaneously with the risk simulation. This permits all parameter uncertainty to be propagated throughout the risk assessment by incorporating all relevant data seamlessly into the one directed acyclic graph. Thus, ideally, the data nodes might consist of microbial cell numbers post-treatment, dose–response data and microbial numbers prior to treatment (thereby allowing estimation of the log-reduction constants and a potential comparison of the two methods of estimation), die-off data and users’ consumption behaviour data. This method means that there is no necessity for prior boot-
strap simulations, as in Cullen & Frey (1999)’s ‘two-dimensional’ approach to fitting ‘uncertainty’ and ‘variability’; the models and the methods are explicit and transparent.

In summary, we have demonstrated a method for incorporating parameter uncertainty, which does not require complex simulation methods. Where a risk assessor is trying to do more than arrive at a point estimate, and is running Monte Carlo simulations such as offered by @Risk (Palisade 2008), this method allows risk uncertainty to be satisfactorily described without resorting to two-step estimation procedures. It is also far more transparent than a spreadsheet approach where operations and their sequencing can be difficult to discern. This method incorporates all the original data used to derive the required parameters for a QMRA into the QMRA, whereas in the more traditional approach these parameters are derived prior to undertaking the risk assessment and are ‘plugged’ into the assessment. We would recommend it as a simple, transparent method which should be incorporated into a risk assessor’s armoury.

CONCLUSIONS

The aim of the study was twofold: (i) to indicate the potential problems arising from failure to include the uncertainty of parameter estimates in risk assessments and (ii) to illustrate the superiority of estimating the parameters to be used in the risk assessment simultaneously with the risk assessment. When one considers the ‘banana-shaped’ bivariate graph for the dose–response parameters (α, β) and its long left tail presented in Figure 8, there is little doubt that the simultaneous estimate of all parameters of interest is a better methodology to use. The techniques and programs used to derive such estimates are now readily available.

Our analysis indicated that, where dose ranges are either extremely large or small, estimating risk by including the uncertainty in the underlying parameters makes little difference in the possible ranges for the probability of infection. However, when the dose is within the range where the risk is neither very close to one nor zero, the inclusion of uncertainty in the parameters may make a marked difference in the possible ranges for the probability of infection.

However, the results of this study highlight the superiority of models developed directly from data for finding more realistic estimates of uncertainty. In practical terms, we would advocate that workers in this field report comprehensive data. Commonly, reported results only include a range and a mean, occasionally a standard deviation, and often not even the number of observations used. These are generally insufficient to permit adequate estimation of risk. In addition, there is a failure to acknowledge, let alone include, the uncertainty which results from small experiments. For the methodology advanced in this paper, we would recommend, firstly, that all data from experiments leading to parameters needed in a risk assessment be in the public domain, particularly when their interpretation may have important implications for public health. A major limitation imposed on this study was the inability to access data collected by, or on behalf of, any Australian water utility, much of which is mandated by law or regulation. Thus, our final recommendation is that such data be made publicly available. Journals may make a difference in the short term by insisting on this for data forming the basis of a published paper.

REFERENCES


Bureau of Meteorology 2010 (e-mail) 2010jr2235*** student/request for data, forecasts or other services/wa/climate and past weather*** (jr- [sec = unclassified]. e-mail: climate.wa@bom.gov.au. http://www.bom.gov.au.


Palisade Corporation 2008 @Risk5.0. Available at: http://www.palisade.com/risk/.


First received 20 July 2009; accepted in revised form 18 May 2010. Available online 20 October 2010