Notes on the need for microsimulation modelling for quantitative risk assessment for Norovirus in foods

Paul R Hunter

Quantitative Microbial Risk Assessment (QMRA) has become a commonly used tool to assess the risks associated with consumption of microbially contaminated food and water. In a previous European Framework Programme "MicroRisk" we set the framework for QMRA for European drinking water (https://www.kwrwater.nl/en/projecten/microrisk-microbiological-risk-assessment/) and directly led to the World Health Organization's current guidance on QMRA [1]. According to the current WHO guidance QMRA has four key steps:

- **Problem formulation**: where the pathogens, exposure pathways and health outcomes of interest are agreed along with any particular hazardous events to be considered.
- **Exposure assessment**: where the dose of each chosen pathogen likely to be consumed by an individual is estimated. For drinking water this will include an estimate of the concentration of pathogen in the drinking water/food and an estimate of the volume of unboiled water/amount of food product to be consumed. Exposure per day is effectively the product of the concentration of pathogen and the volume/amount of water/food.
- **Health effects**: where the probability of the chosen adverse health effect under varying exposures is determined by identifying an appropriate dose-response model.
- **Risk characterization**: where a quantitative estimate of risk is derived from the exposure assessment and the health effects dose-response model.

Since the completion of the MicroRisk project there have been many advances in risk assessment most notably in improved detection methods for pathogens as well as better understanding of the appropriate dose-response models. As well as the recent WHO document, a QMRA Wiki website has been established at the University of Michigan (http://qmrawiki.canr.msu.edu/).

Any literature search on QMRA is likely to find a considerable number of papers on appropriate dose response curves for different pathogens but also on issues relating to estimates of exposure, most notably uncertainty in the concentration of pathogen in the matrix. However, the initial risk characterization produces an estimate of daily risk. However, most risk assessors and policy makers are primarily interested in risk per annum as this then is comparable to epidemiological measures of disease risk. Such annualization of risk is essential for estimating disease burden. In contrast with other aspects of QMRA methodology the processes of using daily outputs of risk estimates has not been studied in that much detail. Over the past two decades the most frequently used method has been to simply annualize each daily risk and then

take the arithmetic mean of these annual risks to equate to population risk. This is shown in equation 1.

Equation 1: $$P_T = 1 - (1 - p)^{365}$$

However, this approach to annualising risk was called naïve by Karavarsamis and Hamilton [2]. The reason why Karavarsamis and Hamilton called this approach naïve, was that this equation effectively assumes that daily risk is constant for any individual for every day of the year. Given that pathogen concentration in food product or drinking water is not constant but varies from one day to the next such an assumption is clearly wrong. Karavarsamis and Hamilton then went onto propose the use of equation 2 which estimates annual risk by taking a random selection of 365 daily risks and then repeating this 10,000 times to generate a mean annual risk and its distribution [2]. They called this approach the gold standard approach.

Equation 2:

$$P_K = 1 - \prod_{k=1}^{365}(1 - p)$$

However, Karavarsamis and Hamilton were not the first to propose and alternate to equation 1. Teunis et al had some years previously used a bootstrap approach to estimate annual risk (3). However, this approach seems not have become popular until relatively recently.

Both these latter approaches assume that daily risk will vary randomly from one day to the next even within an individual. It was clear to us that far from being a “gold standard”, this latter approach was equally naïve. Consider the situation of two people, one an elderly lady who only uses tap water to brush her teeth and never drinks tap water, preferring tea, and a the second a professional athlete who drinks 3 to 4 liters of tap water per day. Daily risks of waterborne illness will not be randomly between the two but there will be substantial autocorrelation of risks within an individual. Consider risk associated with take-home meals, clearly risks will not be random between someone who only never buys a take-home meal from one year to the next and someone who has a take-home two or three times a week. Now in any QMRA model some variables will be random for everyone such as the concentration of pathogen in the food, other variables such as amount of water drunk, or size of meal, or number of meals eaten each year that will vary between people but not so much within the same individual.

So why should this matter? In both approaches, the assumption is made that the annual disease burden is at most one infection per person per year. This is a reasonable assumption, as multiple symptomatic infections due to the same

pathogen within the same year are uncommon, though not unheard of. So, consider again our high and low exposure individuals. In a low exposure individual (say one exposure per very 10 years) each exposure would give rise to one episode of illness. Whereas in the high exposure individual (say 10 exposures every year) each exposure would on average give rise to 0.1 episodes of illness.

Now consider a hypothetic town composed of 3,000 individuals. Let's say that a particular food was rationed depending on to which social class an individual belonged. The lowest 1000 got 10 meals a year, the next 1000 got 50 meals a year and the top 1000 got 200 meals a year. Let us assume that a serving of this food is contaminated randomly with norovirus but defined by a log normal distribution with mean of 10 viable gene copies per meal and a standard deviation of 10.

Doing a simple QMRA on this population would yield a daily risk of 2.23×10^{-3} with a standard deviation of 7.13×10^{-3}. However, the annual disease burden estimates would be very different depending on which annualization method was used. For equation 1, the annual risk would be 1.70×10^{-1} but using equation 2 the estimated annual risk is more than 3-fold greater at 5.55×10^{-1}. So which one is correct. Given that we know that one variable is constant within an individual (number of meals eaten each year) we can repeat the risk assessments for each of the three social groups. The annual risks for low, medium and high consumption rates by equation 1 are 2.29×10^{-2}, 1.10×10^{-1} and 3.69×10^{-1} and by equation 2 are 8.96×10^{-2}, 3.75×10^{-1} and 8.44×10^{-1}. Equation 2 still generates estimates that are quite a lot higher than equation 1. However, in this context the assumptions underpinning equation 2 are sound within a social class and equation 1 estimates are still not. A more valid way to estimate the mean risk for the population as a whole would be to average the risks for each social class which gives 4.36×10^{-1} which is between the original estimates of annual risk given by equations 1 and 2.

In this simple hypothetical population, it is possible to estimate risk by looking at risk in subpopulations. However, in more complex models this is no longer possible. For these reasons we developed a microsimulation approach to annualising risk in QMRA. A more formal definition is that "Microsimulation modelling is a simulation-based tool with a micro unit of analysis that can be used for ex-ante analysis" [4]. Microsimulation modelling has gained widespread use across multiple disciplines including economics, transport, health and the environment. In the context of QMRA, the unit of the modelling process is the individual rather than the population. As the modelling is done at the level of the individual this allows a much greater flexibility to deal with variables that are correlated within an individual. So, for example rather than each day randomly selecting a volume of water to be consumed one can specify the amount of water likely to be drunk by that person or the probability that a certain meal will be eaten. Using the microsimulation approach on the above hypothetical example we get an estimate of annual risk of 4.37×10^{-1} which is extremely close to the mean of the risks in the three classes.

Essentially the model would estimate risk for each of 365 runs for a single 'individual'. In a food consumption assessment with a microsimulation model with two food types, exposure would be modelled by initially randomly allocating a mean number of meals with each food type for that 'individual'. This would then be used to estimate the mean daily probability of each consuming a meal of each of the two food types. For each of day of the year simulation, the number of meals consumed would be randomly allocated give the known mean probability for that 'individual'. The size of the meal and the concentration of virus on each meal would also be randomly allocated. After 365 iterations the mean annual risk for that 'individual' would then be estimated. This would then be repeated 10,000 times to come up with a mean population risk.

As a real-world example, we present results of a reanalysis of a QMRA of cryptosporidiosis in private water supplies. This model is based on cryptosporidium counts obtained from six different private water supplies (that varied in quality from one to another) and water consumption behavior from the Drinking Water Inspectorate. We assumed that the amount of water consumed each day was constant for an individual and that each individual only drank water from his or her home water supply. By contrast we assumed that within a single supply the concentration of cryptosporidium in the water varied randomly from day to day but within a predefined distribution. Finally, we repeated the analysis assuming a treatment step was in place that reduced the counts by one or more logs. The results are shown in table 1.

Table 1. Estimated mean annual risks for Cryptosporidiosis in private water supplies using different approaches.

Log Reduction in count due to treatment	Equation 1	Equation 2	Microsimulation
0 (Raw)	5.76E-01	1.00E+00	8.52E-01
-1	3.26E-01	9.99E-01	6.38E-01
-2	1.39E-01	7.55E-01	3.47E-01
-3	4.31E-02	1.92E-01	1.04E-01
-4	1.04E-02	2.68E-02	1.60E-02
-5	1.85E-03	3.09E-03	1.71E-03
-6	2.45E-04	2.89E-04	1.82E-04

It is clear both from the hypothetical example and from the private water supply that equations 1 and 2 can give very different estimates of annual risk with equation 2 (Karavarsamis and Hamilton's "gold standard" method giving much higher estimates. However, as can been seen in table 1 the two estimates converge as annual risk gets small (less than 1 per 1000-person years). Given that for drinking water supplies a generally accepted tolerable risk of infection level is $1.0x10^{-4}$ differences in equation 1 and 2 would not have any meaningful impact on risk estimates. However, at risks greaten than $1.0x10^{-2}$ (10/1000 per person years) equation 1 would appear to systematically under-estimate risk and equation 2 over-estimate risk and a microsimulation model is most appropriate.

An additional advantage of microsimulation modelling is that it is easier to model multiple transmission pathways such as when one is interested in multiple food sources. Currently published QMRA models have focused almost exclusively on single transmission pathways. Indeed, it is not clear how well standard QMRA approaches would handle such multiple transmission pathways

The disadvantage of microsimulation modelling is that it is substantially more demanding on computing power than standard approaches, especially when multiple transmission pathways are included. In most QMRA models there are relatively few variables (nodes) to be estimated typically less than 12. However, in microsimulation models with multiple transmission pathways there may be several thousand such nodes. A relatively recent development within QMRA has been the rise of Bayesian Belief Networks which do have a number of advantages over standard Monte Carlo modelling [5]. However, Bayesian Belief Networks can be very computationally intensive [5]. The integration of Bayesian Belief Networks with microsimulation models is likely to be so computationally intensive that results would not be generated within a reasonable time. Bayesian Belief Networks may also suffer from other issues especially when large number of variables are included in the model [5].

References

1. World Health Organization. Quantitative microbial risk assessment: application for water safety management. http://apps.who.int/iris/bitstream/10665/246195/1/9789241565370-eng.pdf, accessed February 2018: 2016.

2. Karavarsamis N, Hamilton AJ. Estimators of annual probability of infection for quantitative microbial risk assessment. *J Water Health*. 2010;8(2):365-73.

3. Teunis PF, Medema GJ, Kruidenier L, Havelaar AH. Assessment of the risk of infection by Cryptosporidium or Giardia in drinking water from a surface water source. Water Research. 1997 Jun 1;31(6):1333-46.

4. O'Donoghue C. Introduction. In: Cathal O'Donoghue, editor. Handbook of Microsimulation Modelling (Contributions to Economic Analysis). 293: Emerald Group Publishing Limited; 2014. p. 1-21.

5. Smid JH, Verloo D, Barker GC, Havelaar AH. Strengths and weaknesses of Monte Carlo simulation models and Bayesian belief networks in microbial risk assessment. International Journal of Food Microbiology. 2010 May 30;139:S57-63.