National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

Generic Guidance to Quantitative Microbial Risk Assessment *for Food and Water*

Generic Guidance to Quantitative Microbial Risk Assessment for Food and Water

Authors
P. Teunis
J.F. Schijven

Colophon

RIVM report number: 2017-0188

DOI 10.21945/RIVM-2017-0188

Report prepared by:
WHO Collaborating Centre Risk Assessment of Pathogens in Food and Water, National Institute for Public Health and the Environment

A publication by the
National Institute for Public Health and the Environment, RIVM
P.O. Box 1
3720 BA Bilthoven
The Netherlands
www.rivm.nl

RIVM report number: 2017-0188

Synopsis

Generic guideline for Quantitative Microbial Risk Assessment of food and water

Pathogens in food or water can make people ill. The probability of this happening is calculated by means of risk assessments. The World Health Organization (WHO) has drawn up guidelines stipulating how to carry out these risk assessments. RIVM has written this background document with the most recent knowledge on these risk assessments and the newest calculation methods. The WHO and the Food and Agricultural Organization (FAO) will use the information to update the guidelines for Quantitative Microbial Risk Assessment (QMRA) of food and water. The information is useful for those working with risk assessment of food and water safety.

An example of the recent developments is a method that enables standard, uniform risk assessment if there are few data available or if the data available are uncertain (Bayesian analysis). RIVM also puts forward alternatives for the use of DALYs (Disability Adjusted Life Year) as a measure in the event of outbreaks of disease. DALYs express the harm or inconvenience (disease burden) due to various diseases, including those caused by microorganisms in food and water. This method is not suitable for outbreaks in which many people become ill due to the same pathogen. Furthermore, computer software is available as a tool to conduct risk assessments and for training and education. RIVM gives criteria for such software tools and examples of existing, generally-available QMRA tools.

RIVM is the WHO's collaborating centre for Risk Assessment of Pathogens in Food and Water.

Keywords: Guideline, QMRA, developments, pathogens, food, water

Publiekssamenvatting

Generieke richtlijn voor kwantitatieve microbiologische risicoschatting voor voedsel en water

Ziekteverwekkers, die in water en voedsel voorkomen, kunnen mensen ziek maken. De kans hierop wordt met risicoschattingen berekend. De Wereldgezondheidsorganisatie (WHO) heeft richtlijnen opgesteld waarin staat hoe deze risicoschattingen moeten worden uitgevoerd. Het RIVM heeft dit achtergronddocument gemaakt met de meest recente kennis over deze risicoschattingen en de nieuwste rekenmethoden. De WHO en de Food and Agricultural Organization (FAO) zullen de informatie gebruiken om de richtlijnen voor microbiologische risicoschattingen (Quantitative Microbial Risk Assessment of QMRA) voor water en voedsel te actualiseren. De informatie is bruikbaar voor mensen die werken met risicoschattingen voor water- en voedselveiligheid.

Een voorbeeld van de recente ontwikkelingen is een methode die het mogelijk maakt om risicoschattingen op een standaard en uniforme manier uit te voeren als er weinig of onzekere gegevens beschikbaar zijn (Bayesiaanse analyse). Verder reikt het RIVM alternatieven aan voor het gebruik van DALY's (Disability Adjusted Life Year) als maat bij 'uitbraken' van ziekten. DALY's drukken de schade of het ongemak uit (ziektelast) van verschillende ziektes, inclusief ziekten die micro-organismen via voedsel en water veroorzaken. Bij uitbraken, wanneer meerdere mensen ziek worden van dezelfde ziekteverwekker, is deze methode minder geschikt. Ook is computersoftware beschikbaar als hulpgereedschap voor het uitvoeren van risicoschattingen en ook voor training en onderwijs. Het RIVM geeft criteria voor zulke software tools en voorbeelden van bestaande, algemeen beschikbare QMRA-tools.

Het RIVM is collaborating centre for Risk Assessment of Pathogens in Food and Water van de WHO.

Kernwoorden: Richtlijn, QMRA, ontwikkelingen, pathogenen, voedsel, water

Contents

1 Introduction

1.1 Scope

The Food and Agriculture Organization of the United Nations and the World Health Organization have developed and published several guidance documents pertaining to Microbiological Risk Assessment (MRA) (FAO/WHO, 2003, 2008, 2009). However, some of these documents are over a decade old and like in many areas the science and technology around microbiological risk assessment (MRA) has continued to evolve since then. In this context this document aims to bring together the most recent developments in the area of MRA, and in this way, facilitate any updates and/or expansions of existing guidance documents considering technical and method descriptions, and new developments in assays for microbial detection and identification, information on host responses to infection, methods for mathematical modelling, and characterization of risk at the population level. Taking into consideration that there are additional WHO guidance documents on MRA for both food and water (FAO/WHO 2003,2008,2009; WHO, 2009), this document is also aimed at promoting harmonisation of MRA between these areas. This document too should be considered a snapshot overview: Quantitative Microbial Risk Assessment (QMRA) is subject to rolling revision because new data, methods and insights are generated continually in ongoing research.
This document, like the existing guidance, aims to provide generic guidance to conducting QMRA for food and water.
We adhere to the standard risk assessment framework by successively discussing hazard identification (HI), exposure assessment (EA), hazard characterization (HC), and risk characterization (RC) as defined by Codex Alimentarius Commission (1999).
A major development in risk science is the availability of powerful statistical methods for Bayesian analysis that allow dealing with small data sets, in particular uncertain categorical data in a standard, uniform manner. Another important development is the acceptance of the Disability Adjusted Life Year (DALY) as an endpoint metric in risk studies. It is instructive to show how the DALY is a special case of a mixed utility function, suitable for comparing disease burden across populations, or for different hazards within an exposed population. The DALY is less suitable for large-scale outbreaks or disasters, and the present document provides guidance on alternatives.
Details about models for reduction or increases of pathogens in food and water production and distribution chains are beyond the scope of this report, but references are given. There is a large and growing body of literature on models for transport and fate of (pathogenic) microbes in the environment (including food and drink). Guidance for specific problem areas in exposure modelling should therefore be commissioned to an expert panel, and documented in a separate report.

Finally, computer software packages are becoming increasingly important for QMRA, both to support numerical calculations (parameter estimation, Monte Carlo simulation, analysis of uncertainty and variability) and as guidance tools (expert system) for training and education. We provide criteria for such software tools and examples of existing (and generally available) QMRA tools.

1.2 Statement of purpose

Every risk assessment starts with a statement of purpose, specifying a hazard and setting a risk for a particular time frame or event. What is the problem that needs to be studied and which decisions are at stake?
The purpose of a risk assessment is to provide support for decision making. Risk assessment encompasses gathering information (data collection), interpreting the information (model building) and calculating expected losses or gains (risk characterization). The risk metric, the units the risk is expressed in, should allow for compliance testing (by comparison with standard levels), or selecting between alternative strategies (comparison among different intervention scenarios), for instance.
Examples: Food safety testing, checking whether the presence of enteric pathogens on ready-to-eat salads causes health risks in excess of a predefined risk level; Testing whether high pressure treatment decreases enteric virus concentrations in shellfish to acceptable levels; Select whether cleaning and disinfection decrease infection risks sufficiently to prevent outbreaks in nursing home settings.
It is important to note whose interests are at stake: for public health, the disease burden in the population may be the appropriate outcome to guide decision making. For the manager of a food production process, the risk that any adverse outcomes may be associated with the produced foods, is more relevant than the disease burden.

1.3 Risk assessment team

One individual or a small team may carry out a risk assessment, but access to a range of other expertise usually is needed (WHO, 2008, 2009). The team should represent all relevant areas of expertise and address all stakeholder perspectives throughout all stages of the risk assessment. Ideally, therefore, such a team comprises multidisciplinary experts and stakeholders. Experts may be trained risk assessors, modellers, mathematicians, statisticians, medical, general microbiologists, food and water technologists, animal and plant health specialists, agriculture technologists, human and veterinary epidemiologists, public health specialists, and other specialists as identified for specific projects. Such a multidisciplinary team warrants the appropriate application of risk assessment methodologies, collection of relevant information and the quality of the data and models.

Data may be (a combination of) locally specific data as well as from additional sources, such as literature. Note that in section 3.3, we provide a Bayesian approach to combine qualitative and quantitative information that allows calculation of a quantitative risk. The risk assessment team is also crucial for evaluation of the risk outcomes, deciding whether a next risk assessment tier is required (section 3.3), and whether and how the risk assessment and risk outcomes are to be documented for reviewing and communication.

1.4 General introduction to risk assessment

What is risk? In microbial risk assessment, risk is usually defined as having two dimensions: (1) the probability that something unpleasant happens: the probability of an adverse outcome, and (2) the severity of that outcome. To calculate risk, these two dimensions need to be combined into a single quantity. Below we will give a generally applicable definition, but first, a simple example may be helpful. One of the oldest areas of research in probability is the analysis of games of chance. The example given below introduces all aspects of risk assessment and shows how to define risk in a very simple, yet statistically sound manner.

Example: throwing a dice
To introduce the concept of risk, we play a game. You may throw a dice once, with the following rules: when the result is 5 or 6 you win €1.00 (loss is – €1.00); in case the result is 1, 2, 3 or 4 you lose, and pay €0.50. This means that you have a 2/6 chance of winning €1.00 and a 4/6 chance of losing €0.50. How high is the risk you would be taking when playing this game? The risk is the expected loss, which is not difficult to calculate: 2/3 × (+ €0.50) + 1/3 × (- €1.00). Which adds up to: nothing! Therefore, you may conclude that this was a fair game.

Definition of risk
Various elements of risk need to be considered for a quantitative analysis:
Probability – How likely are all possible outcomes that may occur.
Severity – What are the potential losses associated with any of the possible outcomes?
Degree of control – To what extent can any outcome be prevented from happening or can the severity of the consequences be mitigated? Are there any alternatives to choose from?
Decision – How can knowledge of the risk be used to make decisions about strategies for intervention or prevention?

Box 1 Definition of risk.

Risk is expected loss

One or more elements of the above dimensions of risk have been used in microbial risk assessments. A generic definition of risk may be found in the statistical literature. Risk may be simply calculated as expected loss (e.g. Cox and Hinkley, 1974).
Loss can be expressed numerically with a "utility" function, assigning a quantitative measure of loss to each possible outcome. This loss function represents the severity of the outcome (e.g. expressed in costs, DALYS or infection risk). In order to calculate the risk, we need to have a list of all possible outcomes: anything that may happen (e.g. as consequences of an event or a decision). Then, one wants to know the probability (distributions) of all these outcomes, and a loss function (utility function) assigning magnitudes of loss to all possible outcomes.

Box 2 How to calculate expected (total) loss for quantitative risk assessment.

Defining the mathematical expectation of loss: for any possible outcome x_k there is a loss c_k while the probability that any outcome x_k occurs is p_k. Therefore, the risk R is

$$R = \sum_k c_k p_k \qquad (1)$$

For a continuous outcome x with probability density f(x) and loss function c(x), the risk can be calculated as

$$R = \int_x c(x) f(x) dx \qquad (2)$$

Our knowledge of each of these variables c_k and p_k is usually incomplete. For instance, not all possible consequences of exposure to an infectious micro–organism may be known. However, it is often not hard to define a set of categories which adds up to a complete description (for instance: exposed or not, infected or not, ill or not).
The above example throwing a dice was very simple with only two possible outcomes, and the losses were expressed in a single currency. In risk assessment, outcomes usually have different dimensions, and their quantification is not straightforward. This may be illustrated by a second example to explore the dimensions of risk, a bicycle ride:
This activity is healthy and enjoyable, but it also includes the risk of a fall, or a punctured tyre, or the bicycle could be stolen, or one might even be hit by a car. How different are the risks involved with these outcomes? The probabilities of any each of these outcomes occurring is different. In the event that any of the possible outcomes would occur, the severities or costs involved are also different.

In addition, the subject involved has a different degree of control over each of these outcomes, both with respect to probability and severity.
If the cyclist wants to control the risk, they could adjust their behaviour: by avoiding dangerous situations the probability of an accident can hopefully be reduced. Alternatively, one could also attempt to reduce the damage in case of an accident, for instance by wearing a helmet.

Table 1 Adverse events potentially associated with bike riding in a city. Dimensions of risk and their magnitudes, expressed as (arbitrary) financial losses. The column for costs also shows the reduced loss when wearing a helmet.

		Costs (€)		
Event	**Probability**	**No helmet**	**Helmet**	**Control**
Fall	0.01	5000	1000	a lot
Puncture	0.10	10	10	little
Theft	0.001	500	500	little
Car accident	0.001	10000	6000	some

For the risky bike ride probabilities and losses for each outcome may be collected (Table 1) and expected losses may be calculated.
Without helmet:
p(fall)c(fall) + p(punct.)c(punct.) + p(theft)c(theft) + p(acc.)c(acc.) =
0.01 × 5 000 + 0.10 × 10 + 0.001 × 500 + 0.001 × 10 000 = €61.50
With helmet:
p(fall)c(fall) + p(punct.)c(punct.) + p(theft)c(theft) + p(acc.)c(acc.) =
0.01 × 1 000) + 0.10 × 10 + 0.001 × 500 + 0 001 × 6 000 = €17.50
By buying a helmet, I can decrease my expected losses by €44. Given the price of a bicycle helmet this might be cost-effective...

1.5 Units of risk: Integrated metrics

In the previous two examples, risk has been expressed in monetary currency. This is common for many types of decision making (cost–effectiveness). While attractive, because monetary value is easy to understand and often cost is a limiting factor, it is not always straightforward to express loss on a monetary scale. For instance, how should a human life be valued, or how expensive should living with a chronic illness be considered? In many risk studies, only probability outcomes are considered. For instance, risk resulting from drinking water has been expressed as annual individual probability of infection (1 in 10 000 annual infection risk) (Regli et al., 1991; Teunis et al., 1997). Aside from ignoring severity in symptomatic cases, this may be appropriate as long as infectious enteric pathogens are concerned, so that infection may be considered a (worst case) proxy for illness.

More recently the World Health Organization has proposed the use of Disability Adjusted Life Years (DALYs) to quantify risk. DALYs incorporate both mortality and decreased quality of life resulting from periods of illness, by calculating the sum of numbers of life years lost and years lived with disability, weighted for severity. Severity scores are standardized (based on psychometric research in experts). The DALY approach can be used for infectious and non–infectious diseases, and different life histories depending on gender, socio-economic status, geographic location can be incorporated. A major outbreak may influence a community in many ways: as more people stay at home public services may break down (public transport, health services, social services) and companies will have losses due to the absence of laborers. This may be a nonlinear phenomenon: moderately sized outbreaks may have only a small effect on public life, but as an outbreak increases in size, human resources will become a limiting factor and the losses may increase briskly. Such effects can be modelled as nonlinear utility functions, linking outbreak size and severity of losses (Teunis and Havelaar, 2001). More about integrated risk metrics and their calculations can be found in chapter 7.

1.6 Conceptual framework

At the basis of quantitative risk assessment lies the notion of causality, of risk as the result of a, possibly complicated, chain of conditional events. As soon as we know or accept that exposure to a certain chemical compound or pathogenic microbe causes health effects, it is possible to study and quantify the occurrence of these health effects. Note that this seems slightly at odds with epidemiological studies where an (statistical) association between two observations, e.g. exposure to particulate matter and premature death is observed, and this association is used for predictions of risk. In such a case, the association is not strictly known to result from a causal relation, but the existence of such a relation is considered plausible, for instance because of other, supporting evidence.
Through the Risk Paradigm, the existence of the causal chain connecting cause and effect, it is possible to study and quantify risk. Early risk assessments for microbial pathogens used a general approach outlined for chemical and radiation risks (National Academy of Sciences, 1977; National Re- search Council, 1983; WHO, 1981). This framework defined four stages: (1) Hazard identification, (2) Exposure assessment, (3) Hazard characterization and (4) Risk characterization. The Food and Agriculture Organization of the United Nations and the World Health Organization have worked on standards for food safety based on quantitative risk assessment (Codex Alimentarius Commission, 1993; WHO, 1995; Codex Alimentarius Commission, 1999; Codex Committee on Food Hygiene, 2003) and guidance for microbial risk assessment (FAO/WHO, 2000, 2002, 2006). In a parallel development, the International Life Sciences Institute (ILSI) and the United States Environmental Protection Agency (USEPA) have formulated a framework for microbial risk assessment (ILSI/EPA, 2000; Benford, 2001).

1.7 Variability and uncertainty

Because risk is described as a probability, inherently, at all stages of a risk assessment, variability and uncertainty of models and data (microbiological data, environmental conditions, parameter values of distributions) need to be addressed. Variability and uncertainty may be described qualitatively or quantitatively, and are included in assumption on models and data. Variability and uncertainty may be estimated from the available (monitoring) data. Variability and uncertainty need to be part of the reported predictions of risk.

Variability entails differences attributable to true heterogeneity of diversity e.g. in a population or exposure parameter. When collecting data, observed variation may be attributable to true heterogeneity or diversity in a population or exposure parameter. Variability describes differences among the observed population. For example, different individuals have different food intakes and different susceptibility to infection and illness. Variation over time for a given individual is referred to as intra-individual variability. Variation over members of a population at a given time is referred to as inter-individual variability. Variability cannot be reduced by collecting more data, only more precisely characterized (WHO, 2003).

Uncertainty arises from lack of knowledge about specific factors, parameters or models.
In QMRA, one may be uncertain about the structure of the model, e.g. whether the concentration of pathogens in a series of food samples should be assumed fixed or variable. Lack of knowledge regarding the appropriate and adequate inference options to use to structure a model or scenario, are also referred to as model uncertainty and scenario uncertainty.
Given the model assumptions, a model may be fitted to available data, resulting in parameter estimates. With real world data sets of limited size, these parameter estimates are uncertain, i.e. they usually can be quantified within a specified range (confidence interval, or credible interval). Lack of knowledge regarding the true value of a quantity, such as a specific characteristic (e.g. mean, variance) of a distribution for variability, is referred to as statistical or parameter uncertainty.
These uncertain parameters cause the calculated risk to be uncertain also: uncertainty in predicted risk is a consequence of the joint uncertainties in all contributing factors.
Uncertainty can be quantified by obtaining information from scientific studies, such as through research on mechanisms, and data collection with appropriate sample sizes and representative sampling designs (WHO, 2003). In quantitative risk assessment, the chosen models, their most likely parameter values and attendant uncertainty should also be clearly communicated (WHO, 2003). In QMRA practice, uncertainty is also the reason for caution and for making conservative or worst case assumptions: “err on the safe side”.

Nauta (2000) demonstrated how risk estimates may depend on how variability and uncertainty are separated. But, as Nauta (2000) pointed out, it may be difficult, or even impossible to distinguish them. Variability and uncertainty may be correlated as well. Monte Carlo methods have been developed to address issues of propagation of variability and uncertainty in risk models (see eg. Rai, S. N. and Krewski, D. ,1998). Analysis of uncertainty and variability given a risk model and available data may be addressed in Bayesian hierarchical models (Clough et al. 2005, 2009; Teunis et al. 2008).

A basic understanding of uncertainty and variability are crucial for decision making. This is an educational issue: not only risk analysists and topical scientists, but also any stakeholder needs to understand how uncertainty and variability may change the decision they need to make.

In order to decide what action to take three choices are at hand:
1. Take a **precautionary** approach: Assume the worst (i.e., use the appropriate confidence limit), unless data are convincing to the contrary.
2. Take a **permissive** approach: Assume the best (i.e, use the other limit), unless data are convincing to the contrary).
3. Take data at **face-value** (i.e., ignore uncertainty intervals).

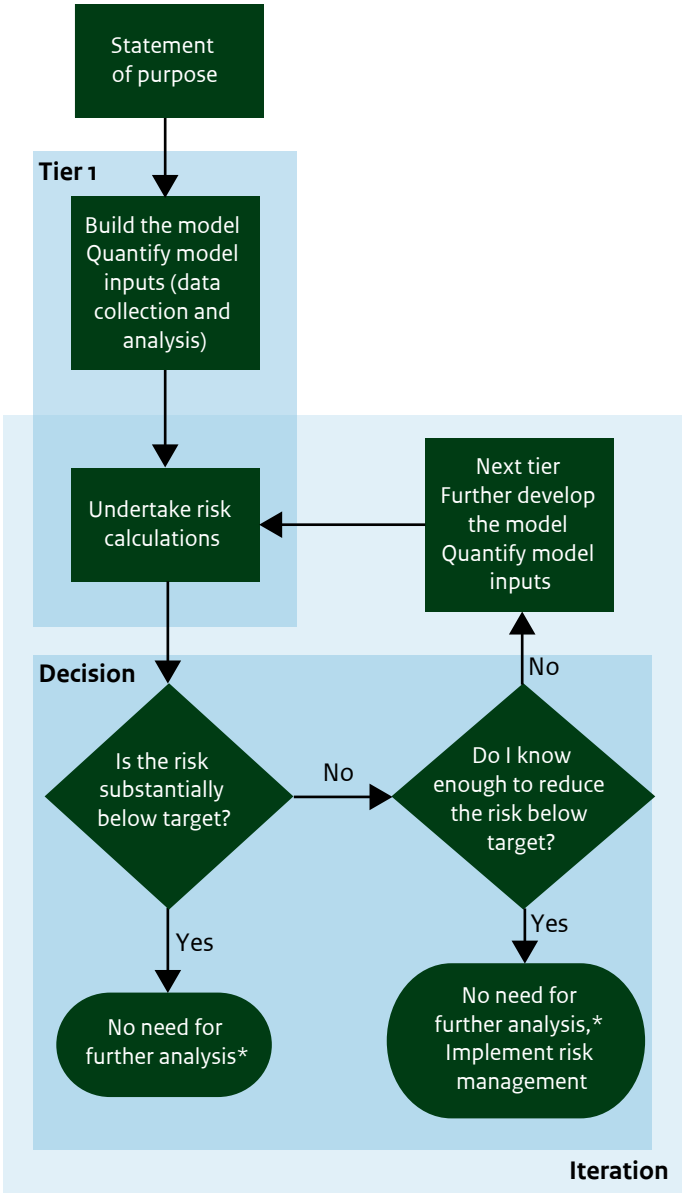
The choice should be guided by the consequences or cost of adopting the 'wrong' approach.

2 Tiered risk assessment

2.1 General

Risk assessment (RA) is a tiered process. In the first tier of risk assessment, an initial run through all stages (HC, EA, RC) is conducted using a priori available data and knowledge and the outcome is a first (and usually crude) quantitative estimate of risk. The expected outcome of this initial walkthrough of all stages of the risk assessment is understanding of the formulated problem, data requirements, and (possibly) efforts required to provide a suitable risk report. The conclusion may be that the risk is so small that further study is unnecessary, so that the first-tier RA is all that is needed. Usually, more iterations are needed, where following tiers require additional data and refined models (Figure 1) for HC, EA, and RC. In the following tiers, EA will commonly be repeated using more data, whereas HC may remain unchanged because mostly available data have been used. RC needs to be conducted in each new tier, because risks need to be calculated anew.
To reach the point that there is no need for further analysis requires testing of a hypothesis. The answer to the question in Figure 1, whether the risk is substantially below target, may not be simply confirmed when the risk estimate is ten or a hundred times lower than the target, because of uncertainty. An uncertainty estimate must be included to enable a decision.

Figure 1. Tiered (iterative) approach to implementing QMRA for risk management; increase in tier relates to gathering more/local data to reduce uncertainties (adopted from WHO, 2016). *The risk level of the system should be reviewed at a regular interval according to the relevant risk management protocol.



2.2 First tier risk assessment

The formulated problem must be translated into a risk model: a first, preliminary conceptual model describing the whole risk chain (network), from sources of microbial contamination to human exposure and health effects. The conceptual model tends to be based on general knowledge of pathways for contamination by faecal-oral pathogens. Topical experts may contribute specific knowledge related to e.g. human contact behaviour, food or water treatment processes, factors influencing pathogen survival and infectivity, and factors influencing host susceptibility and illness symptoms. In many cases, the conceptual model may be represented as a linear chain of (conditional) events: pathogen occurrence in raw materials, survival during processing, human consumption of contaminated food, infection, acute symptomatic illness. In the real world, exposure results from multiple concurrent pathways, with potential for cross-contamination (like preparing food with dirty hands), resulting in an exposure network, rather than a linear chain (Wang et al. 2017).
For most problems, a specific hazard must be selected: a pathogen of interest for which a risk assessment will be constructed (hazard identification). This pathogen may be identified from practical experience, e.g. outbreaks associated with the food of interest, or it may be an indicator pathogen with properties similar to the pathogen of interest. For instance, the indicator pathogen may have high infectivity in humans, and/or it may be resistant to environmental decay. Sometimes, it is not possible (or even desirable) to specify a pathogen, and risk may be thought to result from exposure to a compound source, e.g. faecal contamination due to lack of hygiene in food handling and/or preparation.
Once a conceptual model has been formulated, a list of factors (parameters) may be made, that must be specified in order to perform the risk calculations (Figure 1: Tier 1). Although determination of these factors (model parameters or risk variables) is the objective of the analysis stage of the risk assessment procedure (i.e. exposure assessment and hazard characterization), it is informative to start with crude, initial 'guestimates' for all factors in the conceptual model. Already at this stage, it is strongly recommended to collect quantitative data including information on variability and uncertainty as much as possible. When all elements of the risk chain (or the risk network) have been defined, a preliminary calculation of the risk can be done. Such a first-tier exercise provides vital information about the risk assessment: a comprehensive list of all information required to run the conceptual model; identification of gaps, in the structure of the conceptual model as well as in the data it needs; an impression of the computational effort required for the risk assessment; and an initial estimate of the resulting risk.

Box 3 Risk chain: a highly simplified example.

$$P_{ill} = C_{raw} \times 10^{-DR} \times V_{ing} \times P_{inf}(1) \times P_{ill|inf}(1) \quad (3)$$

where C_{raw} is the pathogen concentration in the raw source material; DR is the log (decimal) reduction in treatment/processing of the product; V_{ing} is the ingested amount of food or water; $P_{inf}(1)$ is the probability of infection for a single pathogen; and $P_{ill|inf}(1)$ is the (conditional probability of illness given infection, when exposed to a single pathogen. Each of these factors C_{raw}, DR, V_{ing}, $P_{inf}(1)$ and $P_{ill|inf}(1)$ in the risk model may result from a (sub)model, for instance a probability distribution describing its variation, and that sub-model is characterized by parameters. In the hazard identification stage, the model is set up and initial values for the factors are defined. In the analysis phase (exposure assessment and hazard characterization), each of the factors is studied in detail, sub-models are set up and parameters estimated, using information from any appropriate data sources.

After completion of the first-tier risk assessment, the model and risk outcomes must be reviewed (by the team) and the conclusions should be documented. In Figure 1, this is the conditional step in the decision stage: "is the risk substantially below target?" Most often, the conclusion is that the risk is insufficiently known, so that more information is needed. The data collected in the 1st tier RA now can be used to specify constraints for the analysis stage. Aside from more data collection, additional questions on model choice may be addressed, for example: is the conceptual model adequate for the task, or should it be modified? If so, how? What information is needed for exposure assessment? What is available (i.e. preliminary information is sufficient for a full-scale exposure assessment), and where is additional data collection necessary?
The same questions may be addressed for assessment of health effects: what is needed, how much was found in the preliminary stages, and where is additional research required? At this time, it is instructive to reconsider the first decision stage in Figure 1: how can we know whether the calculated risk is substantially above target? The adjective "substantially" suggests a qualification of how much the risk is above target: a category of risk considered important enough to warrant additional study.

2.3 Qualitative and quantitative assessment

In older documents on risk assessment (WHO, 2003, 2008, 2009), qualitative risk assessment is contrasted to quantitative risk assessment. Even, the term semi-quantitative risk assessment is used, as an intermediate stage between qualitative and quantitative risk assessments. Problems formulated in qualitative risk assessment are not necessarily different from those in quantitative risk assessment, nor do the conceptual models have to be different. The only difference is in the quantification of risk factors: when the available information does not allow straightforward quantification of a factor, it is characterized by a categorical outcome (high or low, strong or weak,...). This may happen when there is only data from a

sanitary survey, for instance. Obviously, this leads to difficulties in combining factors for risk calculations, as straightforward arithmetic does not apply. For that reason, a set of conventions is used to characterize risk when the basic data are categorical (WHO, 1981, 1993, 1995). In semi-quantitative risk assessment, risk factor categories are given a score (0: good, 1: medium, 2: bad) and calculations are based on score arithmetic (FAO/WHO, 2003, 2008, 2009). Unfortunately, such an approach tends to be arbitrary and cannot ensure that information is weighted according to its reliability.

Using a Bayesian approach to statistical analysis, uncertainty in any variable is expressed as a probability distribution. In other words, any variable is included in a quantitative manner, even when it is highly uncertain. Basic probability calculus can then be used to calculate the resulting uncertainty in outcomes. The Bayesian approach (Gelman et al., 2013) allows dealing with sparse data, and the information provided by observations that do not produce numerical data can be handled appropriately, even allowing for combing categorical data (like clean or dirty environment) with enumeration data (microbial indicator or pathogen counts). The advantage of such a statistically rigorous method for evaluating data from unequal sources is that any information that contributes to the risk may be included, while weighted appropriately by its content. Note that such unequal sources as observed cleanliness (from a sanitary survey) and microbe counts might previously have been considered incompatible for quantitative analysis.

This provides a framework for combining the information from different sorts of data, and defines a statistical basis for the matrix method used in semi-quantitative risk assessment [FAO/WHO, 2009: WHO guideline doc on Risk Char]. Combining two data sources that are consistent (e.g. both indicating high, or both indicating low risk) leads to more pronounced outcomes (either high or low risk) with decreased uncertainty. When the two data sets are inconsistent, i.e. one indicates that the risk is high, while the other implies low risk, the outcome of the combined analysis is not shifted to either high or low risk, but instead shows increased uncertainty, which is entirely appropriate [See section 3.3 Case 1]. Another advantage of a Bayesian approach is the opportunity for including statistical learning: when an additional observation is added to the data analysed previously, the effect on the (posterior) estimate may be calculated: shrinkage towards the mean in case of consistence, and expansion of the uncertainty in case the new observation conflicts with earlier data. The involved calculations are standard and can be done by software (BUGS, JAGS, Stan), without burdening the end user with technical detail, so that Bayesian methods have become feasible for general purpose analysis (Gilks et al. 1996, Gelman et al. 2013).
Instead of using the three different approaches (quantitative, qualitative, and semi-quantitative risk assessment) outlined in earlier guideline documents, any risk assessment can be done within the quantitative framework, without having to choose between alternative approaches. The subjective tables for combining categorical data are not needed, as we have a statistically sound procedure to combine variables with arbitrary precision. And finally, in a Bayesian framework, initial risk estimates as calculated in a first-tier assessment allow testing against a target risk (e.g. addressing a question like: how likely is it that the risk exceeds the target level?). Therefore, the decision in Figure 1: “is the risk substantially above target?” may now be made in a well-defined manner in every tier of risk assessment.

Box 4 Bayes' theorem: how observations change the probability that a hypothesis may be true.

$$P(H \mid O) = \frac{P(O|H)P(H)}{P(O)} \qquad (4)$$

$P(O|H)$ is the probability of observing O when hypothesis H is true (i.e. its likelihood)
$P(H)$ is the prior probability that hypothesis H is true, before any observations were made
$P(H|O)$ is the posterior probability that H is true given the observed data
$P(O)$ is the probability of observing O under any possible hypothesis
For any complete set of mutually exclusive hypotheses, $P(O) = \sum_i P(O|H_i)P(H_i)$

2.4 Case: Is the food safe or unsafe?

This is an example of a tier 1-risk assessment (the starting point) using only limited data: at a farm, fresh strawberries are harvested and packaged for shipping to retail. The produce may be contaminated by faecal matter through unsafe irrigation or cross-contamination, through handling or by dirty equipment. In order to investigate possible contamination of raw produce with faecal microbes, during an inspection survey, only one sample is taken and tested for faecal indicator organisms. The test only indicates presence or absence of the faecal indicator (categorical: positive/negative outcome). Now assume that extensive monitoring experience has shown that in poor hygiene conditions (categorical) the probability of a positive sample is 0.6 (hypothetical, quantitative value), in other words, when a farm is known to have low hygiene standards, there is a 60% chance that a produce sample tests positive for faecal indicators. Note that this extensive experience may be available as locally specific data, or may be taken from literature (to emphasize the usefulness of having default data as provided by literature databases and/or computational tools: see the Appendix with this report). In hygienic farm conditions, the probability of a positive sample is 0.05: when a farm is known to have good hygiene, there is only 5% chance that a produce sample tests positive for faecal indicators.

Now if the sample that was taken for our survey is positive, what is the probability that this farm has poor hygiene? We do not know beforehand, so that the probability that the hypothesis "this farm has poor hygiene" (H_0) is equally likely as the alternative "this farm has good hygiene" (H_1). Prior probabilities of either hypothesis are therefore 0.5. If the farm had poor hygiene, the likelihood of observing a positive sample would be 0.6. Using Bayes' theorem (Box), the posterior probability of the hypothesis "this farm has poor hygiene" then is

$$\frac{P(O|H_0)P(H_0)}{P(O|H_0)P(H_0)+P(O|H_1)P(H_1)} = \frac{0.6\text{x}0.5}{0.6\text{x}0.5+0.05\text{x}0.5} \approx 0.92 \qquad (5)$$

The denominator here is the probability of a positive sample, if either hypothesis H_0 is true or H_1 is true. This leads to the conclusion that, based on our general knowledge of test outcomes on many farms, we are confident that this is an unsafe farm, even with only a single sample. What if the produce sample had tested negative? In case of a negative sample, what would be the probability that the farm had good hygiene (H_1 is true)? The probability of a negative sample from a hygienic farm is 1-0.05=0.95 so that the probability that H1 is true becomes

$$\frac{P(O|H_1)P(H_1)}{P(O|H_1)P(H_1)+P(O|H_0)P(H_0)} = \frac{0.95\text{x}0.5}{0.95\text{x}0.5+0.4\text{x}0.5} \approx 0.70 \quad (6)$$

the denominator again calculating the probability of a negative sample under any of the hypotheses. When our sample tests negative, we are not nearly as confident that the meat came from a hygienic slaughterhouse. The reason of course is that negative samples are not only found in hygienic places: unsafe slaughterhouses are likely to produce negative samples as well (with 40% probability).

2.5 Iterative approach: beyond tier 1

When the outcome of the first-tier risk assessment warrants further study, the critical knowledge gaps must be identified. This is in fact a major goal in the first-tier assessment: use the risk calculations to find where the most influential contributions are least known. These knowledge gaps guide adaptations to the risk study: data collection to strengthen the basis for parameter estimation; adaptations to the design of the model, resulting from improved understanding of the system that is studied; improvements to the implementation of these adaptations, to allow higher precision and/or efficiency in calculations.
Once the adapted, improved model has been implemented and run, the results are again scrutinized for their usefulness in decision making (risk management). The same procedure that was used to decide whether the first-tier assessment provided sufficient knowledge of the risk for decision making, is applied again. And yet again, the result may not satisfy the requirements: the risk may still be insufficiently specified, e.g. to decide whether action should be taken or not. The reason may be that still more data are required, to achieve more precise outcomes, or the model needs more covariables to improve predictions. Adaptations are then needed, and another cycle of model (re-)formulation, data collection, parameter estimation, and risk prediction is required (Figure 1, iteration). The goal of the cyclic approach to risk assessment is to obtain ever better (i.e. more useful) risk estimates, with each new iteration learning from the shortcomings of the previous one. The stakeholders often also play an important role, not only in checking whether the goals formulated in the initial design stage are fulfilled, but also whether these goals may need correction. The iterative approach to risk assessment here models the development of science: as more knowledge is acquired, the questions that initiated the study are also better understood.

3 Exposure assessment

3.1 General

Exposure assessment simply involves estimating the ingested dose of a pathogen resulting from consumption of contaminated food or water. A generic model for exposure assessment may be formulated as

$$D = C \times 1/Se \times Sp \times Z_{1,\ldots,m} \times Y_{1,\ldots,n} \times V \quad (7)$$

where D is the ingested dose of a pathogen. C represents the concentration of the pathogen at the starting point of the risk assessment chain. In the case of food, this may be anywhere in the food production chain, where the pathogen is enumerated. In the case of drinking water, it is typically the concentration in the source water. Se represents the sensitivity of the method used for enumeration of the pathogen. Sp is the specificity of the enumeration method.
Z represents n consecutive treatments. A treatment may be aimed at reducing pathogen numbers, for example disinfection, an inactivation process, or otherwise physically removed by filtration or by taking out contaminated portions. Z may also represent mixing, by which pathogen concentrations are reduced (dilution). In that case, Z is the fraction of the pathogen, or its indicator organism, that passes a treatment. Y represents a growth (WHO, 2008). For Z predictive models may be used or distribution that describe changes in numbers (or concentrations) of pathogens. In food RA, increases and reductions in the food production and distribution chain are given by the ICMSF equation to test meeting a food safety objective (FSO) (Nauta, 2008; Zwietering et al., 2010). V is the ingested amount of food or water per intake event or within a particular period.

3.2 Enumeration of pathogens and indicator organisms

Exposure assessment (as part of QMRA for water or food) encompasses a quantitative evaluation of the intake of a pathogen. As direct enumeration in food or water as consumed is often not possible, because samples would rarely contain any pathogens, exposure assessment usually must be based on indirect estimation. In raw food stuffs pathogens may be more prevalent, resulting in countable numbers in reasonably sized samples. Therefore, the first step is to quantify pathogen occurrence, usually in terms of concentrations in the source material: raw food ingredients or source water.
It is important to realize that detection methods are imperfect: pathogens are usually present in low concentrations. A small sample of contaminated food or water may contain few

pathogens, or even none at all. And even when there are a few pathogens in a sample, they may be lost during processing for detection. The detection process may not reliably detect all pathogens present in the (processed) sample (Se<100%), or may produce stray positive outcomes, i.e. detect pathogens that were not present (Sp<100%. In some cases, surrogate microbes may be used instead of the pathogen of interest, but these surrogates may not mimic the real pathogen.
Microbiological detection methods can be complicated and fragile, depending on many factors that must be controlled in order to obtain precise (little random error) and reproducible results. Changing methodologies, particularly improvements in the selective media, need to be evaluated (WHO, 2008). In fact, of each detection method, the performance characteristics to detect a pathogen from a particular matrix must be known.
Documentation of all microbiological data (pathogens and indicator organisms) should be in the form of raw data. Raw data are unprocessed data. In the case of microbial enumeration, the raw data may consist of colony counts or plaque counts in combination with the equivalent size of the sample, corresponding to the counted numbers. For molecular assays, similar requirements may be formulated. For instance, gene copy counts resulting from PCR are inferred from fluorescent measurements, and depending on the assay, various data may be recorded: presence/absence with serial dilutions (conventional PCR), Ct values and standard curve data(qPCR) or numbers of cells and dilutions (dPCR).
The sample size is the amount of contaminated material, in case of foodstuffs usually a weight, in case of water usually a volume. The equivalent sample size is that part of a sample that was used for enumeration. Note that a microbe concentration, such as colonies per gram or per ml is not raw data. When data have been collected with presence/absence or most probable number testing, the raw data consists of zeroes and ones with sample sizes. Both microbial count and sample size contain information and determine accuracy. Awareness of the simple concept of raw microbial data is of fundamental importance in microbial risk assessment.
The following example is an illustration of what the raw data are in qPCR enumeration and how losses may of the original sample may accumulate. A large sample size is collected because of representativeness: for example, 10 liters of water or 10 portions of 10 g each of a food product. Adsorption-elution and ultrafiltration reduces the sample to 10 ml. Of the 10 ml, 0.1 to 1 ml is taken for extraction of nucleic acid. The nucleic acid is contained in 50 µl. For the qPCR reaction, 5 µl is used. So, finally, the gene copy count is determined for 0.1% - 1% of the original sample and 99% - 99.9% is discarded. The raw data may look like this: the copy count and the actual investigated sample size, which is 10 ml – 100 ml or 0.1 g to 1 g. The sensitivity may easily be improved by enumerating ten replicates.

Accuracy is the absence of systematic and random error – commonly known in metrology as trueness and precision, respectively (WHO, 2003, 2008). Sensitivity and specificity represent systematic error. Accuracy is the degree of agreement between average predictions of a model or the average of measurements and the true value of the quantity being predicted or measured (WHO, 2003).

For example, 100 colonies counted in 100 grams of sample and 10 colonies in 10 grams of sample have the same calculated concentration. However, the former observation is more accurate, because more colonies were observed, in a larger sample. When only the concentrations are reported, the information about accuracy is lost: it obviously does not matter whether 100 colonies were counted in 100 grams or 10 colonies were counted in 10 grams. However, when fitting a distribution to the raw data to describe the variability in concentration between samples, it does matter. The observations based on higher counts are weighted higher than those based on small numbers of counted microbes.
Especially in food microbiology it is common place to discard low counts, typically fewer than 3. But this is deletion of information. For example, the raw data are 3 colonies in 10 grams and 23 colonies in 100 grams from one food sample. The raw data of this sample for fitting a distribution should be 26 colonies and 110 grams (so there is some processing of raw data), not 23 colonies and 100 grams. Note that discarding high counts per sample, like more than 300 colonies per agar plate is justified because of systematic error. The measurement error as a function of the counted numbers should be described as part of the performance characteristics of a detection method.
Typically, in qPCR enumerations, data sets consist of many non-detects (zero gene copies per sample) in combination with high gene copy counts. Detection by qPCR is considered as a highly sensitive detection method, where sensitive is understood as positive detection in samples with low concentration (detection limit). In the context of risk assessment, it is necessary to quantify the limitations of the microbial detection methods, including qPCR.

3.3 Sensitivity and specificity

Presence or absence
The sensitivity of a detection method is the probability of detecting an organism when it is known to be present:

$$Se = P\,(1 \mid 1) \tag{8}$$

The specificity of a detection method is the probability of not detecting an organism when it is known to not be present:

$$\mathrm{Sp} = P\,(1 \mid 0) \tag{9}$$

The probability of a positive outcome p1 then is

$$p1 = (Se)p + (1 - Sp)(1 - p) \tag{10}$$

where p is the "true" probability that an organism is present in the sample. The first part (Se)p is the probability of detecting a "true positive"; the second part (1 – Sp)(1 – p) is the probability of a "false positive" (1 – Sp is the probability that the test produces a positive result in absence of an organism and 1 – p is the probability that the organism is absent).

The "true" probability of detection is

$$p = \frac{P1-(1-Sp)}{Se-(1-Sp)} \tag{11}$$

Therefore, either $Se > 1-Sp$ and $p1 \geq 1-Sp$
or $Se < 1-Sp$ and $p1 \leq 1-Sp$

Enumeration

Use of a detection method with low sensitivity will result in the occurrence of false negatives (WHO, 2008). In microbiological literature, sensitivity may be designated as recovery: the fraction recovered from any organisms that were present in the sample. Because of imperfections and losses, the observed counts are lower than the true number in the present sample. Not correcting for the sensitivity of the detection method, therefore, results in underestimation of risk. In many risk assessments, this correction is indeed omitted because sensitivity data are lacking.

Suppose a sample of volume V is taken and processed for analysis, using a detection method with known specificity and sensitivity. When a number k is detected, the number m of true (not false) positives is binomially distributed.

$$P(m \mid k) = \binom{k}{m} Sp^m (1-Sp)^{k-m} \quad (k \geq m) \tag{12}$$

Any microbe of the k detected in the sample is assumed to have the same probability Sp that it is the targeted pathogen. When m positives would have been detected, the number of false negatives ($k - m$ missed pathogens) in the sample has a negative binomial distribution, and the total number that must have been present is distributed as

$$P(n \mid k) = \binom{n-1}{m-1} (1-Se)^{n-m} Se^m \quad (m \leq n) \tag{13}$$

Where Se is the probability that a pathogen is detected. The (marginal) distribution of the true number n may be calculated

$$P(n \mid k) = \sum_{\substack{m=0 \\ m \leq n}}^{k} P(n \mid m) P(m \mid k) \tag{14}$$

which can be written as

$$P\,(n \mid \mathrm{k}) = (1-\mathrm{Se})^{n}(1-\mathrm{Sp})^{k} \times \sum_{m=0}^{min(k,n)} \left(\frac{(n-1)!k!}{m!(m-1)!(n-m)!(k-m)!} \right) \left(\frac{SeSp}{(1-Se)\,(1-Sp)} \right)^{m} \qquad (15)$$

where both (0 < Se, Sp < 1). This relation, despite from being somewhat cumbersome to evaluate analytically, is not difficult to calculate numerically.
In QMRA, specificity is not always dealt with: false positives may be neglected. Assuming that there are only false negatives, the numbers counted underestimate the numbers of microbes that are present. The correction may be found in the above equation for P(n|m). Ideally, an internal control is used with each sample, to allow per sample correction. Often, such a control is not available as this would make detection laborious and costly. Instead, in order to estimate the sensitivity of the detection method of the index pathogen, a limited number of samples are spiked with a known number of the same microorganism, or with a suitable indicator organism (presumed to have similar characteristics in the detection procedure). After processing of the samples, a fraction of the spiked microorganisms will be recovered. Thus, this produces paired data on the spiked numbers and the recovered numbers of microbes. The sensitivity (recovered fraction) may be assumed to vary between samples (e.,g. according to a Beta distribution with two positive shape parameters, denoted by α and β; Beta distributed random numbers vary between 0 and 1). Suitable models for estimation of the sensitivity and its distribution have been published, details may be found in Teunis et al. (1999b, 2005b, 2009).

3.4 Pathogen concentration at the start of a risk assessment chain

When they are processed in the lab, it may be assumed that samples are well homogenized and mixed, so that the assumption that microbes are Poisson distributed within a sample, may be correct. However, in QMRA one usually deals with series of samples resulting from repeated observations within a surveillance program, over time and/or location. For such sets of samples, it is not self-evident that the concentration would be fixed, and not variable. To accommodate the plausible assumption that the pathogen concentration is variable, it is usually modelled as a random variable, so that the sampling distribution (the numbers counted in successive samples) is a mixture distribution, for instance a Poisson distribution with random parameter (cV: concentration c and sample volume V). The concentration c may be any positive real-valued random number, but it is convenient (and not very restrictive) to use a gamma distribution. The negative binomial distribution is the (marginal) distribution when the counts in any single sample are Poisson (meaning that each sample is well mixed) while samples taken at different times have random concentrations, that in turn are Gamma distributed (Teunis et al., 1999a, 2009). The Gamma distribution has two parameters, a scale factor λ and a shape factor r. These parameters can be estimated by fitting the Negative Binomial distribution to the observed raw data (counts and volumes). A Gamma distribution is very flexible in shape. It may be right- or left skewed.

The relations between the mean u, variance v, r and λ are simple: $u = r\lambda$ and $u = r\lambda 2$. An important feature of fitting a Negative Binomial distribution to the raw data is that zero counts (non-detects) can be included. This is not possible when fitting a Gamma distribution to concentrations. An alternative that is also popular in literature is the lognormal distribution. The disadvantage is that a Poisson-lognormal mixture cannot be easily written in closed form (like the negative binomial distribution for the Poisson-gamma mixture).

3.5 Enumeration by culture (infectious) versus by PCR

Enumeration by PCR methods provides counts of genes (gene copies). A subset of these genome copies may be within an intact, infectious microorganism or virus. Therefore, the numbers of infectious microorganisms or viruses are smaller than the numbers of genome copies. Assuming that all pathogens determined by PCR are infectious may overestimate risk. In this regard, uncertainty is large: the fraction of infectious particles to gene copy numbers may vary over a wide range, and, especially in the case of viruses or parasites that cannot be grown in tissue culture, to detect infectious virus particles, the difference is unknown. If it is possible to count gene copies as well as infectious particles, as similar approach may be used as for determining sensitivity. After all, the number of infectious particles is a fraction of the number of gene copies, and, therefore, a Beta distribution describes the relation between paired numbers of gene copies and infectious particles.

3.6 Surrogates for pathogen occurrence

A quite different problem is the use of surrogate microbes for assessment of the occurrence of pathogens in food or water. Traditionally, faecal indicators are used in food and water safety testing, based on the assumption that presence of human faecal material is a necessary condition for the presence of human enteric pathogens. This may in general be true more often than not, on average over a long-term period, which is perhaps why bacterial faecal indicators are still widely used. However, we should be looking for an indicator whose concentration is proportional to the pathogen of interest, with reasonably high (not just bordering on significant) correlation. Presence or absence of the surrogate (microbe or other, chemical or physical variable) should be influenced by the same environmental conditions as the pathogen. While studies trying to link occurrence of pathogens to that of indicators often do find weak correlation, this does not mean that presence of the indicator predicts presence of pathogens. This becomes apparent when studying the time course of indicators and pathogens in environmental sources. In surface water sources, for instance, both pathogens and indicator organisms tend to occur in peaks: short term high concentrations, followed by periods without any detectable organisms (Westrell et al., 2006). Occurrence in peaks indicates that the presence of an organism is related to some discrete event. Unfortunately, peaks in indicators and pathogens do not match (LeChevallier et al. 2001). It therefore is not likely that pathogens and indicators are present because of the same event or process.

One should therefore be very cautious in the use of indicators as surrogates for the occurrence of pathogens in microbial risk assessment.

3.7 Treatment

Food production processes usually imply some form of treatment to remove microorganisms, and reduce human exposure. There is a huge variety of process models for predicting pathogen inactivation, die-off or removal in various treatment process (heating, irradiation, filtration, removal of contaminated units, mechanical cleaning,...). For example, for first order, two-rate and Weibull inactivation models, see Van Boekel (2002). Uncertainty in process model predictions usually represent model uncertainty given the data and model choice. Variability can be included by predicting treatment efficacy as a function of variable process conditions.

Generically, it may be assumed that micoorganisms passing treatment do so independently, with a certain probability or fraction passing or surviving the treatment process. This fraction may be modelled as a binomial process, either with paired or unpaired samples (Teunis et al., 1999b, 2009), or as the ratio of the Gamma distributed effluent concentrations / the Gamma distributed influent concentrations. Paired samples result from sampling before and after treatment at the same time. Collection of paired data from a treatment step requires exact timing of the sampling. The pairing may be lost if mixing occurs during treatment. Residence times in treatment may vary from a few hours to several days. In many cases, even with short residence times and samples of influent and effluent collected on the same day, pairing is not evident.
Dependent of these settings and on the raw data values, treatment fraction z may be equal to one, be described by a Beta distribution, or be described as the ratio of the Gamma distributed concentrations of influent and effluent. This ratio distribution is a so-called type II Beta distribution or F-distribution (Teunis et al., 2009) that often provides a convenient description of the variation in treatment efficiency.

3.8 Surrogates for removal in treatment

Surrogate microbes are frequently used for the characterization of reduction in treatment. Bacteriophage data instead of human pathogenic viruses, non-pathogenic bacterial indicators instead of pathogenic bacteria, and bacterial spores instead of survival stages (cysts, oocysts) of human pathogenic protozoa. A major reason is that it is usually not possible to study full–scale industrial processes by seeding them with human pathogens, as this would require their isolation and decontamination, which would be highly impractical and costly. Some of the indicator organisms that are used also are present in the environment in high numbers, so that they can often be detected prior as well as posterior to treatment, so that seeding is not necessary. Ideally, the survival of these indicators should be compared to that of the real pathogens in small scale (bench top) lab studies, to validate their use as indicators.

Occasionally there may be theoretical justification for use of a particular surrogate organism, for instance with processes that physically remove microbes, where similar physicochemical properties (similar sizes, similar surface charges) indicate similar expected behaviour. Note also that when characterizing removal or inactivation, a fraction is estimated. This is a relative measure, possibly not strongly dependent on absolute concentrations of the organisms. Models for estimating fractions of concentrations may be used for determining whether data for two microbes present in different concentrations would be consistent with equal removal or inactivation rates, thus allowing for testing the usefulness of surrogates in different conditions.

3.9 Growth

Growth of the pathogen of concern is one of the most important processes of food QMRA for bacteria. Growth is the process that fundamentally differentiates microbial exposure assessment in foods from chemical exposure assessment. According to Nauta (2002a), models of bacterial growth need to be expressed in terms of probability, for example to predict the probability that a critical concentration is reached within a certain amount of time. In contrast, available predictive models have been developed and validated to produce point estimates of population sizes and therefore do not accommodate this requirement. Nauta (2002a) suggested that a new type of predictive models needs to be developed that incorporates modelling of variability and uncertainty in growth. An example is provided by Matagaras et al. (2006)

There is a big and rapidly growing body of literature on predictive pathogen modelling for foodborne risk assessment. Covering this field in any meaningful sense is beyond the scope of this document. Therefore, a separate document should be developed to summarize existing literature, document state of the art concerning quantitative models of pathogen growth and survival, and list knowledge gaps and research needs. This could be a component of the "rolling revision" approach, to extend and actualize WHO (2008).

3.10 Human behaviour and intake

Exposure depends on human choices made during selection of ingredients, preparation in the kitchen, and eating behaviour. At the time of consumption, foods may be contaminated by pathogens present in the ingredients, passed down through the farm-to-fork chain. In addition, there may also be contamination that is deposited on the food during handling, in retail, or in the kitchen during preparation, or during eating. Preferences of consumers in their diet and preferences in cooking and consumption may therefore be of decisive importance in exposure to microbial pathogens. In addition, there is variation in hygiene behaviour among individuals (Curtis et al. 2003) and the numbers of faecal bacteria on fomites vary considerably (Rusin et al. 1998, Medrano-Félix et al. 2011). For this reason, in addition to predicting the numbers of pathogens downstream from a contaminated source in a farm-to-fork risk

assessment, one also needs to consider the behaviour in the home environment that influences the contamination of foods at consumption.
First, this concerns choice of ingredients and where to purchase (supermarket, small local retailers). Then transport and storage at home (e.g. duration of storage and refrigerator temperature). Then cooking behaviour, determining the risk of cross-contamination in the home kitchen (Mylius et al. 2007). Initial steps have been made to describe contact behaviour of human with sources of faecal contamination (Curtis et al. 2003), and to set up quantitative models of such behaviour, to estimate probabilities of activities that cause faecal exposure (Teunis et al. 2016). Such studies indicate that oral exposure to food- and waterborne pathogens is linked to contaminated fomites, through hand contact (Wang et al 2017). Studies into consumer behaviour, to obtain quantitative descriptions of choices in food consumption and probabilities, durations, and/or frequencies of contact behaviour are highly needed. Risk studies often focus on scenarios for describing contingencies: a particular chain of events that may lead to exposure. If the probability would be known that any risky scenario does occur, as well as the probability of alternative scenarios, this would result in a much better basis for decision making. In addition, Nauta et al. (2008) concluded that in developing risk communication regarding domestic food hygiene practices, the focus should be on activation of the knowledge that consumers already possess at the moment of food preparation rather than food safety education.

3.11 Case: Pathogen peaks

A drinking water production plant using a surface water source has a chain of treatment processes that reduces pathogen concentrations about 5 $\log_{10}$ units. Source waters are from a protected catchment where low contamination levels are expected. For that reason, a minimal monitoring effort is made: three samples of 10 l are taken per year, and analyzed for the presence of Cryptosporidium oocysts.

Tier 1 – Business as usual
In the first year of monitoring, only negative samples are found: oocyst counts are 0 in 10 l. What is the estimated oocyst concentration in the source water? Assuming Poisson distributed counts, 0 in 30 l means that the oocyst concentration is 0 (oocysts/l), with a 95% upper level of 0.064 (oocysts/l). That means that a pessimistic estimate of the concentration in finished water would be $0.064 \times 10^{-5} = 6.4 \times 10^{-7}$ (oocysts/l). Assuming the risk of infection after exposure to a single oocyst of Cryptosporidium parvum is 0.16, and daily water intake 1 l for anyone in the population, the infection risk for a single exposure (daily) is approximately 10^{-7}. For such low daily risk, the yearly risk may be approximated as $365 \times 10^{-7} = 0.4 \times 10^{-4}$. When the 1 in 10 000 rule for annual individual infection risk is used, would this water be considered safe for consumption? One cannot really decide, because we have no idea how reliable this point estimate is. But since the point estimate is based on "best" data for reduction in treatment, and not a worst case estimate, the fact that it is so close to 1 in 10 000 is reason for some concern.

Tier 2 – An unusual event

In the second year of monitoring the first two samples are again zero, but the third sample appears to contain 100 oocysts in 10 l. What is the estimated oocyst concentration now? We might still assume that the oocyst counts are Poisson distributed. Now there were 100 oocysts in 30 l so that the concentration would be 3.3 (oocysts/l). The yearly infection risk now can be calculated as $3.3 \times 10^{-5} \times 0.16 \times 365 \approx 19.3 \times 10^{-4}$. If the 95% upper confidence level of the Poisson concentration is used, as above, the risk is $4.0 \times 10^{-5} \times 0.16 \times 365 \approx 23.5 \times 10^{-4}$. The Poisson distribution however is not a good model when the data show much dispersion, as in this example.

A better model is the negative binomial distribution. The parameters of the negative binomial distribution can be estimated by fitting the negative binomial distribution to the observed counts: 0 in 10 l, 0 in 10 l and 100 in 10 l. Maximum likelihood estimates are $(\hat{r},\hat{\lambda}) = (0.076,44.03)$. The 95% upper level for this gamma distributed oocyst concentration is 19.3 (oocysts/l). Note that this is a lot higher than the 95% upper level for the Poisson model. The reason is that we have now included variation in oocyst concentration between samples into the model.

[A likelihood ratio test comparing goodness of fit shows that the deviance (-2 times the log–likelihood) is 226.17 - 16.82 = 209.35 which is highly significant (χ^2 (1) deviate). Therefore, the negative binomial model provides a significantly better description of the observed data than the Poisson model.]

Using this high upper 95% level in the risk assessment as a point estimate would increase the estimated yearly infection risk even higher than but doing this would produce an overly pessimistic risk estimate, because we have seen that the concentration is not always so high throughout the year. A Monte Carlo estimate of the risk using only the oocyst concentration as a random variable, leads to an estimated yearly risk of 19.3×10^{-4}; the 95% level is 25.5×10^{-4}. Not unexpectedly, the high count in one of the samples causes the infection risk to exceed the 1 in 10 000 level. Action will have to be taken.

Tier 3 – Better monitoring

The high count in one of the samples causes alarm and first action is to take another sample, to check if the oocyst concentration is still high. It is not: the next sample, 1 week later, is zero again. Now the treatment plant managers decide to change their monitoring schedule: they are going to take a 100 l sample every month of the year. At the end of that third year 11 samples have all led to zero counted oocysts, but there was one sample of 100 l river water where 100 oocysts were found. As it has been established that a Poisson distribution is not suitable to analyse risk for varying pathogen concentrations the negative binomial model is used, and best (maximum likelihood) parameter estimates are $(\hat{r},\hat{\lambda}) = (0.014,5.96)$. Using Monte Carlo simulation for the random oocyst concentration the average yearly risk is now 0.49×10^{-4}, and the 95% level 0.9×10^{-4}.

That risk estimate may seem to indicate that the water quality is compliant with the 1 in 10 000 limit, but note that we have still assumed that treatment always reduces the oocyst concentration by 5 $\log_{10}$ units. As the upper 95% level of our risk estimate remains close to 1 in 10 000 there is still reason for concern.

4 Hazard characterization

4.1 General

A generic hazard characterization for various exposure pathways will be described. Quantitative risk assessment methods exploit the conditional relations between exposure, infection and illness. To become ill from an infectious pathogen, a person must have been infected (colonized) by that pathogen. And to become infected, a person must have been exposed, i.e. have ingested at least one infectious pathogenic organism. In terms of conditional probabilities:

$$P(ill) = P(ill \mid inf) \times P(inf) = P(ill \mid inf) \times P(inf \mid exp) \times P(exp) \quad (16)$$

where P(ill|inf) is the conditional probability of illness given infection: the probability that an infected host develops symptoms of acute illness. This may be a dose dependent probability. Assuming exposure to n pathogenic organisms, P(ill|inf; n) is a (conditional) illness dose response relation. P(inf|exp) is the conditional probability of infection given exposure: the probability that an exposed host becomes colonized by the infectious pathogen. This probability is thought to always depend on the numbers n of pathogens that were ingested, and P(inf|exp;n) is a (conditional) infection dose response relation. P(exp) is the probability that a host is exposed (to, say, n pathogenic organisms) when in contact (ingestion, inhalation, . . .) with a contaminated environment. Thus P(exp) = P(n|dose), as determined by exposure assessment, where dose represents some characterization of the distribution of n (often its mean). Thus, there are two main contributions to the risk that must be characterized: exposure and health effects resulting from that exposure. Exposure assessment is specific to any problem that is studied: often, direct enumeration of microbes in contaminated foods or water is not possible, because of low concentrations, or because contamination is a rare event. Therefore, indirect methods for exposure assessment have been developed, measuring pathogen concentrations in environmental sources (including e.g. raw foodstuffs) and the (log) changes caused by transport, storage, preparation, etc.

4.2 Data sources

Studies in risk assessment usually are based on proxy dose response information: a surrogate pathogen in humans, or a surrogate symptom in a surrogate (animal) host. As the objective is to determine health risk, the question remains how much uncertainty and bias such substitute data introduce.

Human challenge studies
For some pathogens human challenge studies have been published. Compared to quantitative risk assessments in toxicology, this is a great advantage. Health effects of genuine pathogens have been studied in the appropriate host (the one that the pathogen co–evolved with). However, as exposure of human volunteers to serious risk would be unethical, only relatively mild pathogens can be used, in immune competent hosts of good general health. As a consequence, the risk of infection may be underestimated, and it is likely that the (conditional) risk of illness given infection is underestimated in volunteer studies. Really harmful pathogens causing severe illness cannot be studied in human volunteers.

Natural experiment studies
Incident or outbreak reports deal with human patients linked to exposure to some microbial hazard and can be considered prime evidence for the infectivity and/or pathogenicity of the pathogen at hand. In outbreaks, it is likely to find severely infectious (causing many cases) and pathogenic strains (causing severe illness) in weakened hosts, as this increases the chance of detecting the outbreak. A pathogen that causes outbreaks of severe illness is highly likely to be relevant for public health, and of interest to quantitative risk assessment. Therefore, use of outbreak data for assessing infectivity and pathogenicity is important: this would include pathogens of interest (as they do cause outbreaks) and provide dose response information from the opposite end of the scale, compared to volunteer studies. The most infectious and pathogenic organisms in the most vulnerable (most susceptible) hosts. Exposure often is not known, but there are exceptions: sometimes food inspectors can obtain a sample of the implicated food and estimate the pathogen concentration. A single outbreak provides only limited information, a single exposure estimate, with corresponding numbers of exposed and affected subjects is the best that can be achieved, usually. Data from different outbreaks show considerable variation in observed attack rates, caused by many characteristics that may vary among outbreaks. The exposed population may differ, e.g. in age, or prior experience (acquired immunity to the pathogen); the vehicle may be different (different food, differences in preparation of the food); the pathogen may have a different history prior to its presence in the food. For such a meta–analysis combining data from different outbreaks, a hierarchical model is appropriate, to explicitly incorporate variation in infectivity between outbreaks. In Teunis et al. (2007), such a set of outbreaks is studied with a two–level dose response model. An additional complication is overdispersion in exposure. Unlike experimental studies, the pathogen may be distributed unevenly in the vehicle (e.g. food or water) and the quantity ingested may also vary. Such heterogeneous exposures affect the dose response relation.
Animal challenge studies depend on a proxy host, but may allow study of variation in dose response due to pathogen properties (typing). These studies are often difficult to translate to human dose response, due to differences in host response to infection and different pathogenesis.
In vitro experiments may be used to characterize a specific host barrier, and the effects of variation on its performance (e.g. gastric pH barrier; attachment/invasion of cultured intestinal mucosa cells).

4.3 Dose response models

The idea that exposure to large numbers of pathogens would be more likely to cause infection than exposure to only few organisms is appealing. The host defenses might be 'overwhelmed' by a massive number of viruses or bacteria, leaving the door to infection wide open. The idea of a 'threshold' dose, above which infection occurs seems to be as popular in microbial risk assessment as it is in toxicology. The threshold dose would be the lowest dose that causes a biological effect. There is a difference between exposure to a low dose of a chemical substance and exposure to a low dose of microbes however: low doses of a chemical substance consist of many molecules. For instance, an LD50 for Botulinum Toxin A in mice of 47pg (Krewski and van Ryzin, 1981) corresponds to 5.7×10^8 molecules. In contrast, a low dose of microbes often consists of only a few organisms. This difference is of the utmost importance in quantitative microbial risk assessment. A sample from a suspension with a low concentration may contain a small number of microbes, and there may be a nonzero probability that such a sample does not even contain a single organism. This is the basis for much of exposure assessment. And it is also the starting point for an important class of dose response models: hit theory models.
Some important classes of models relevant for microbial dose response assessment are:
Cooperation, inhibition models: Interactions between pathogens in causing infection (and/or illness), can result in a nonlinear dose response relation. Cooperative effects tend to steepen the slope of the dose response relation, inhibition has an opposite effect. The incorporation of heterogeneity also makes the any dose response relation less steep. Therefore, we have here a problem of identifiability: any observed slope could have occurred as a result of some balance between cooperative effects and inhibition or heterogeneity.
Extreme value models: the idea that only the "fittest" organisms in an inoculum elicit a response. Can be described as a specific form of heterogeneity.
Hazard models: The timing of events after challenge with a pathogen dose may be modelled as a hazard: the probability per unit time of an effect occurring. Useful when such temporal information is available, as is often the case (especially in experimental studies).
Time dose response models (Huang and Haas 2009): time post-inoculation may be used as a covariable to predict the probability of acute health effects. These models have been shown to be useful for (animal models) for bioterror agents.
Predator–prey models: Can be used for modelling the interaction between a (colonizing/invading) micro–organism and an adaptive defensive response of the host (like the immune response). Allows highly specific description of the mechanisms in action at the expense of many parameters and a high degree of complexity, especially when heterogeneity is incorporated.
Hit theory models: infection as the result of a random hit process: any (number of) organism(s) in the inoculum may succeed in breaching the host defenses and initiate infection. Many variants are possible: independent action or collective action; single hit or multiple hits; heterogeneity or not; exposure to aggregates or dispersed particles; . . . The familiar single hit model with independent action only consists of the probability of exposure (ingestion of an organism) and the probability that one or more of the ingested organisms elicit a response.
The Beta-Poisson model is the most commonly used model for microbial dose response. It is based on the assumptions that the dose is a Poisson sample, any ingested pathogen has probability to survive and colonize that is a sample from a Beta distribution, and independent

action: survival of any ingested pathogen is independent from any other pathogens in the same inoculum.
It should be noted that for some pathogens, more than one dose-response model is in use, *e.g. Campylobacter* (Chen et al., 2006; Teunis et al. 2005a), *Cryptosporidium* (Chappell et al., 2006; Messner and Berger, 2016; Teunis et al., 2002), *Salmonella* (Oscar, 2004; Teunis et al., 2010), and STEC (Haas et al., 2000); Teunis et al., 2004). An overview of dose response data and models can also be found at *http://qmrawiki.canr.msu.edu/index.php/Dose_Response*

4.4 Beyond infection

During infection, there is a probability of becoming ill: the longer infection is present, the higher the probability that symptomatic illness will occur. Conversely, the duration of infection may be related to the balance between the health status of the host (immunity) and the infectious potential of the pathogen. A host with strong defences against the pathogen is assumed to clear infection rapidly. Conversely, a highly virulent pathogen is assumed to be able to sustain growth for a long period. These three basic building blocks: presence of infection, illness hazard, and duration of infection can be translated into a model: the hazard model for illness dose–response (Teunis et al., 1999c). The hazard function can have arbitrary shape, as long as the illness risk is proportional to the duration of infection: when the duration of infection is twice as long, the integral of the hazard over this period becomes exactly twice as large. Arbitrarily, is has been assumed that the scale parameter (and not the shape parameter) of the (gamma) distribution of the duration of infection depends on the dose.
If the simplest possible relation, direct proportionality, is chosen, there are three possible alternatives: the duration of infection increases with dose, it decreases with dose, or it is independent of the dose. In Salmonella dose response, a higher dose not only causes an increased probability of infection (as seems to always be the case), but also, in an infected person, an increase of the probability of becoming ill (Teunis et al. 2010). A biological explanation may be that the higher initial dose may allow the pathogens to reach numbers that are damaging to the host more quickly, before defensive responses can slow down growth in order to prevent tissue damage.
Microbial infection may not only lead to acute illness, but there may also be long-term sequelae: systemic disorders that manifest months or years after infection, and may persist for long periods. Some sequelae result from chronic infections (e.g. chronic Q fever), others seem related to auto-immune disorders (Campylobacter and Guillain-Barré syndrome, Salmonella and reactive arthritis). As there is not a clear relation with exposure, such long-term health effects are usually described as a (fixed) conditional probability: the risk of long-term health impairment given infection. These risks tend to be small, but the health effects may be severe: disability and premature death. Therefore these sequelae cause the majority of the health burden for the pathogens where they are known to occur.
In a small fraction of the cases, acute symptoms may be severe enough to lead to mortality, in particular in infants or the elderly.

4.5 Immunity, to infection and/or illness

A basic approach to including immunity into dose response models is to include a covariable to describe the variation in susceptibility. For incorporation of covariables it is an advantage when parameters in the model are (biologically) meaningful. A candidate marker for susceptibility could be a serological parameter, like the baseline concentration of antibodies against the pathogen that is used for challenge. High baseline levels of serum antibodies could indicate recent infection or illness. If this would be associated with some protective effect, it should be possible to detect this protective effect with an appropriate model. The concentration of antigen–specific antibodies can be measured, and can be used as an indicator of the degree of protective immunity. For example, published experimental dose response studies indicate that there indeed may be an association between high IgG-levels and protection against infection by the protozoan parasite *Cryptosporidium parvum* (DuPont et al., 1995; Chappell et al., 1999). The dose response relation

$$P_{inf}(D \mid \mathrm{p}_m) = 1 - e^{-Pm} \tag{17}$$

can be modified by letting the single hit probability of infection depend on a covariable x

$$log\left(\frac{Pm}{1-Pm}\right) = \alpha + \beta \mathrm{x} \tag{18}$$

where the covariable x is the log of the baseline serum antibody concentration. This "wraps" the logistic regression model into the single hit dose response relation. The resulting model still is a single hit model (Teunis et al. 2002). The dose response relation now is a function of two variables, dose and antibody level, which can be visualized in a three–dimensional graph. A graph of pm against baseline IgG shows that increased baseline IgG is associated with protection against infection with *Cryptosporidium parvum*. Note that this does not imply that the serum antibodies are protective; rather, high baseline IgG indicates recent infection, and more recent infection appears more likely to retain protective immunity.
At the population level, transient (acquired) immunity to an infectious pathogen may cause a counter-intuitive relation between infection pressure and illness incidence. For *Campylobacter* a decrease in infection pressure, e.g. due to enhanced food hygiene, may lead to a temporary increase in the numbers of acute symptomatic cases, because of decreased boosting (Swart et al. 2012).

It should be attempted to restrict analysis of factors in quantitative risk assessment to human pathogens in human hosts. Exposure analysis assumes data on the occurrence of human pathogens in food or water sources; analysis of infectivity and pathogenicity assumes data from human volunteer studies or outbreaks. Such data are not abundant and often their availability is a major limiting factor in quantitative risk assessment. When crucial data cannot be found, surrogate data are used, challenging the "quantitative" part of quantitative risk assessment.

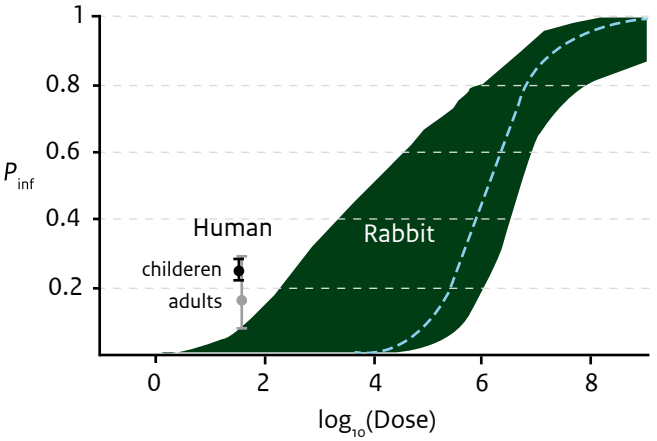
4.6 Surrogates for infectivity/pathogenicity

In dose response assessment for pathogenic microbes the availability of human challenge data is a great advantage, compared to toxicology. However, given the variability of microbial pathogens it is safe to say that any pathogen studied in a human challenge study is a surrogate. The single isolate, often propagated in lab conditions for a long time, is a surrogate for any study in microbial risk where a similar pathogen (the same species, genetic variant?) infects human hosts from a different vehicle, with a different history in the environment. Possibly with greatly increased heterogeneity, but it is not known how genetic or phenotypic heterogeneity influences infectivity and/or pathogenicity in humans. Sometimes, an outbreak happens to be studied in such detail that dose response information may be gleaned from such data, providing an opportunity to study infectivity and pathogenicity "in the wild". It may be expected to find more virulent strains of a pathogen in an outbreak, compared to an experimental study with a laboratory–adapted strain. Fortunately, limited data so far indicate no strong differences, so that it seems safe to use human challenge data in risk assessment.

4.7 Surrogate hosts

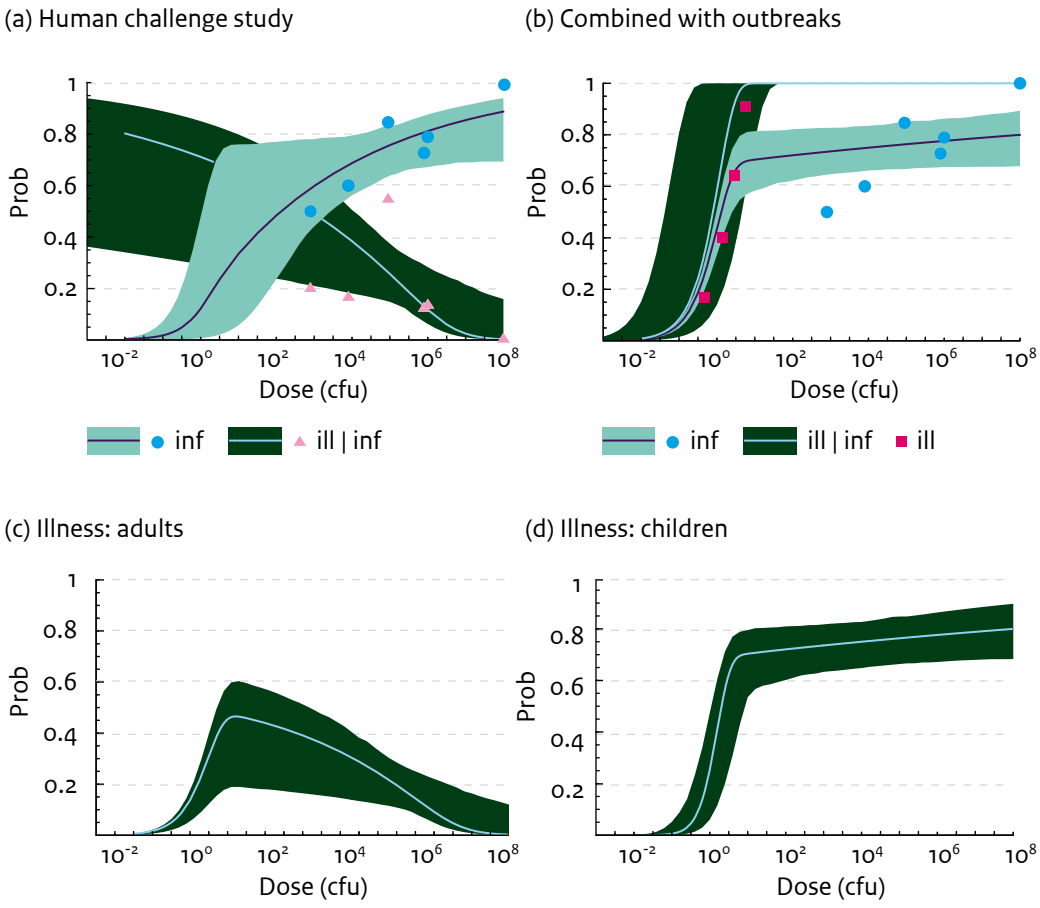
Studies with human subjects are expensive and require expert personnel, special facilities and safety-tested inocula. Relatively few studies exist and safety of clinical studies is scrutinized more than ever, making it unlikely that many more experimental human dose response studies can be done in the future, however valuable these studies are for understanding the processes of infection and pathogenesis. There are also many pathogens for which it is unlikely that there will ever be human experimental data, because they are too dangerous. Yet for many of these dangerous pathogens dose response information is wanted the most. An example is E. coli O157:H7, which causes, in addition to diarrhea, renal failure, and many infections, especially in infants, end in death. Animal studies have been proposed as a solution to this problem. For instance, a study exists where rabbits have been challenged with E. coli O157:H7 (Haas et al., 2000). An outbreak of E. coli O157:H7 enteritis in a boarding school in Morioka city, Japan, in 1996 was studied in unusual detail, even providing data on faecal excretion of the pathogen in all pupils and teachers, as well as a dose, that appeared to be low (31 cfu) (Teunis et al., 2004). This outbreak gives us the opportunity to compare infection probabilities of rabbits and humans, and shows that it is highly unlikely that these are the same (Figure 2).

Figure 2: Dose response relation for *E. coli* O157:H7 in rabbits (Haas et al., 2000) and fraction infected in outbreak in Morioka, Japan in 1996 (Teunis et al., 2004). Both pupils (208/828) and teachers (7/43) are shown, with 95% confidence intervals. Note that it is highly unlikely that infectivity of this pathogen is the same in humans and rabbits.



This example highlights a fundamental problem with animal data for human dose resonse assessment: not just the pathogenesis should be (more or less) comparable with that of humans, but also the magnitude of the probabilities of health effects should be comparable. We cannot know if that is the case when there are no observations in humans, and may only guess whether a pathogen is more or less virulent in an animal than in its human host.
A pathogenic microorganism is often specialized in infecting a particular host species, after a period of co-evolution with that host species. Such co-evolution may have optimized the interaction into various directions, making the pathogen more virulent or, less virulent. Available evidence seems to indicate that infectivity tends to be high for pathogens infecting their proper host, and lower in case of a "mismatch" (a host species barrier). Returning to *E. coli* O157:H7, this is an important pathogen causing many outbreaks, and for some of these outbreaks dose response data could be obtained, allowing construction of a dose response relation, based entirely on outbreak data (Teunis et al., 2008b). In rare circumstances it may even happen that appropriate outbreak data are available for a pathogen that has been studied in a human challenge study. Two outbreaks of campylobacteriosis among children who visited a cheese farm and drank raw milk appeared to produce a fine dose response relation for the attack rate of diarrhea and amount of milk consumed. This dose response relation appeared to not only be consistent with challenge study dose response data for infection, but provided additional information on its shape (Figure 3).

Figure 3: Infection and conditional illness dose response (illness among infected subjects): posterior mode curves and predictive intervals. (a) Human feeding study (Black et al., 1988); (b) Combined model based on human feeding study and the two milk outbreaks (Teunis et al., 2005); (c) Unconditional probability of illness (probability of becoming infected and ill) in adults (volunteer study of Black et al. (1988)) and (d) children (milk outbreaks, Teunis et al. (2005)).



Interestingly, contrary to the dose response relation for infection, the conditional probability for illness among infected subjects also appeared to depend on dose, but quite differently in the milkborne outbreak and the experimental study (compare Figures 3c and 3d). It is tempting to attribute that difference to differential susceptibility: the volunteers in the challenge study had been exposed before (some even still had serum antibodies) whereas the children were young enough to not be immune.
Use of animal dose response data in human risk assessment is common practice for chemical and radiation hazards but it is not quite as common in risk assessment for infectious diseases. It is a fortunate circumstance that human challenge data as well as an – as yet still mostly unexplored – reservoir of outbreak data is available, so that perhaps we may not need to resort to animal studies very often.

4.8 Safety factors

Use of safety factors is a common practice used in toxicology to account for unknown or uncertain steps in chemical risk assessment. For instance, when dose response is based on animal data, or when a benchmark dose for adults must be translated to infants, the dose is lowered by a certain factor, thus making the result safer by presumably erring on the safe side. As long as the uncertainty in the estimated risk is not known, the effect produced by a safety factor also remains unknown. Although originally designed on the basis of distributions of uncertainty in translating dose response relations between species, the procedure has deteriorated into allowing factors of 10 for any uncertain step in risk assessment. As such an approach replaces a proper study of uncertainty (statistical and/or conceptual) it should be avoided and instead, efforts should be made to describe and quantify uncertainty.

5 Risk characterization

5.1 General

Once all components of risk have been identified and quantified, they can be used to characterize the risk. This may involve nothing more than calculating a point estimate of exposure and using a dose response model to transform exposure estimates into estimates of infection risk. Usually, however, risk characterization is more complicated: variability and uncertainty must be evaluated, and endpoints may be integrated metrics of risk, estimating loss at the population level. Starting with the simplest scenario: estimating infection risk for repeated exposures, for instance the accumulated risk over a one-year period, we proceed to methods for weighting risk and methods for calculating risk, including Monte Carlo simulation. Then we finish with a few remarks on generalizing risk, and how to verify the outcomes of a risk assessment.

The probability of a specific health effect results from a chain of events, each with their probabilities of occurrence. The probability of exposure to a certain dose results from the probabilities of events determining food or water contamination, and behavioural events determining treatment (e.g. preparation of food) and uptake (consumption) of contaminated media. The resulting probabilities of infection and illness depend on the dose response relations, possibly influenced by additional variables determining virulence of the pathogen and host susceptibility.

Dose response relations are special because they are generic: most risk assessments use dose response information from published literature. Care must be taken that the used numbers are appropriate for the risk study. Various models exist for the dose response relation for infection and the dose response relation for illness when infected (the conditional relation for illness given infection). Most are based on hit theory, evaluating the probability of infection resulting from any single inoculated (ingested) pathogen particle acting independently to cause infection (Haas, 1983; Teunis and Havelaar, 2000). Note that there are two different forms of the Beta–Poisson model, an exact and an approximate function. Care should be taken that when parameter estimates for the exact model are not valid for the approximate model (rule of thumb: $\beta < 10$ and $\alpha > \beta/10$) these parameter estimates should NOT be used for predicting risks with the approximate model because they will result in serious errors in the calculated risk.

In addition to these dose response models for infection there are also models for predicting the probability of acute illness in infected subjects. Illness dose response relations are not always available because most dose response data are from human challenge studies with severe study size limitations.

As ususally, only a fraction of those infected becomes ill, while only part of the exposed subjects become infected, the numbers involved are often quite small. In case there are data on the occurrence of symptoms of acute illness in challenged volunteers, the probability of becoming ill when infected may be dose dependent or not, in other words, there may be a dose response relation for the conditional probability of becoming ill when infected, or not (Teunis et al., 1999b). In the latter case, there may be an estimate of a (fixed) probability of infected subjects becoming ill, for instance based on outbreak studies. It is therefore often possible to proceed from infection risk towards risk of acute illness. In terms of probabilities

$$P(ill, dose) = P(ill \mid inf, dose) \times P(inf, dose) \tag{19}$$

In case the (conditional) illness probability is dose-dependent (Teunis et al., 1999b, 2005, 2010), this has consequences for epidemiology. The reason is that when a population is exposed to low doses of an enteric pathogen some people may become infected but few of these infected subjects will become ill, so that cases are sporadic. When the dose is high there is not only a higher probability of infection, but also those that are infected have a higher probability of becoming ill, so that a cluster of enteric illness may be detected.

5.2 Long term illness and sequelae

It is sometimes possible (and necessary!) to proceed beyond acute illness and estimate probabilities of developing severe long–term sequelae: chronic illnesses caused by infection with a pathogenic micro–organism. Examples are symptoms of auto–immune disease following bacterial infection (*Salmonella* and arthritis; *Campylobacter* and Guillain–Barré syndrome). The probabilities of such long– term health effects may not be dose dependent (a moot point because it is hard to imagine a study establishing such a relationship) but is possibly associated with other factors like genetic makeup of the pathogen and the host. Even though the probabilities of illness may be quite small (in particular for long– term sequelae) including such endpoints may be relevant because the associated health burden is considerable, for two reasons: (1) the symptoms may be quite severe, and (2) they may be present, at least in part, for many years, possibly as long as the affected subject lives.

5.3 Scale

Usually, we are not interested in the infection risk for a single subject following a single exposure event, but rather in the accumulated risk over a specific period. For instance, the yearly risk associated with use of a particular foodstuff. Often, it is necessary to calculate risks resulting from repeated exposure events. When the probability of infection or illness, associated with a single intake event is known, calculation of the probability for a series of such events seems straightforward. It should be noted that, given the heterogeneity that is usually present in risk models, such calculations (as in text box 6) can be done using Monte Carlo simulation, generating a new probability for each intake event.

Imagine that these events are a day apart: every day one drinks a glass of water, for instance. For such a consumption pattern it seems plausible that exposure events are independent. By the time a glass of water is consumed, the previously consumed portion has passed the intestinal tract completely. That is not always so clear, however: in case of two or three meals a day, each containing a contaminated ingredient, it is not so plausible that the previous meal has been digested completely. And even when passage has completed between intakes, any pathogens succeeding in attaching to the intestinal mucosa may remain there for a longer period than one day and it is conceivable that interaction spans more than a day. In case two or more intake events are only a short period apart, doses may be assumed additive and the resulting risk can be calculated. In real world situations the intake pattern may be intermediate. Many risk models however, deal with low doses, for instance as encountered in drinking water risk assessments. Low dose approximations for both models of repeated exposure are identical.

Box 5 Long-term risk from repeated exposure, assuming independence: multiplicative model.

Estimates of daily risk may be extrapolated to yearly risk. When p_1 is the probability of a positive effect (infection, illness) on day 1, and p_2 the same on day 2, and on day i the probability is p_i, then the probability of becoming positive (once, or more times) within a period of n days is

$$P_n = 1 - (1 - p_1)(1 - p_2)...(1-p_n) = 1 - \Pi_{i=1}^{n} (1 - p_i) \quad (20)$$

If all p_i are very small so that $P_n << 1$ then

$$P_n \approx p_1 + p_2 + ... + p_n = \Sigma_{i=1}^{n} p_i \quad (21)$$

Box 6 Long-term risk from repeated exposure with short intervals between consumption events: additive model.

When the time intervals between exposure (ingestion of contaminated food or drinks) is short, exposures may not be considered independent and the cumulated risk must be calculated differently When several portions of food/drink are ingested shortly after one another, doses are additive. The risk may be calculated by first adding all ingested doses

$$V_{tot} = V_1 + V_2 + \ldots + V_n = \Sigma_{i=1}^{n} V_i \quad (22)$$

and then using the dose response model to calculate the probability of infection or illness

$$P_n = f(cV_{tot} \mid \theta) \quad (23)$$

if all V_i are very small so that $cV_n << 1$ then the single hit model is may be approximated by a linear relation

$$P_n \approx E(p_m) cV_{tot} \quad (24)$$

so that in the latter case the risk is equal to the result we got with independent exposure.

More sophisticated dose-response models consider timing between exposure events, and take into account immune system interactions with the inoculated pathogens (Pujol et al., 2009), potentially leading to decreased risks after repeated exposure.

5.4 Calculations

The simplest possible procedure of calculating risk is to take a sample of the estimates for all of the involved factors (or parameters) and use these to calculate probabilities of infection or illness, and/or other risk outcomes. For an assessment of microbial risk in food or water such a calculation could be as follows:

First, calculate exposure See also equation (7) in section 4.1. Next, translate exposure estimates into infection.

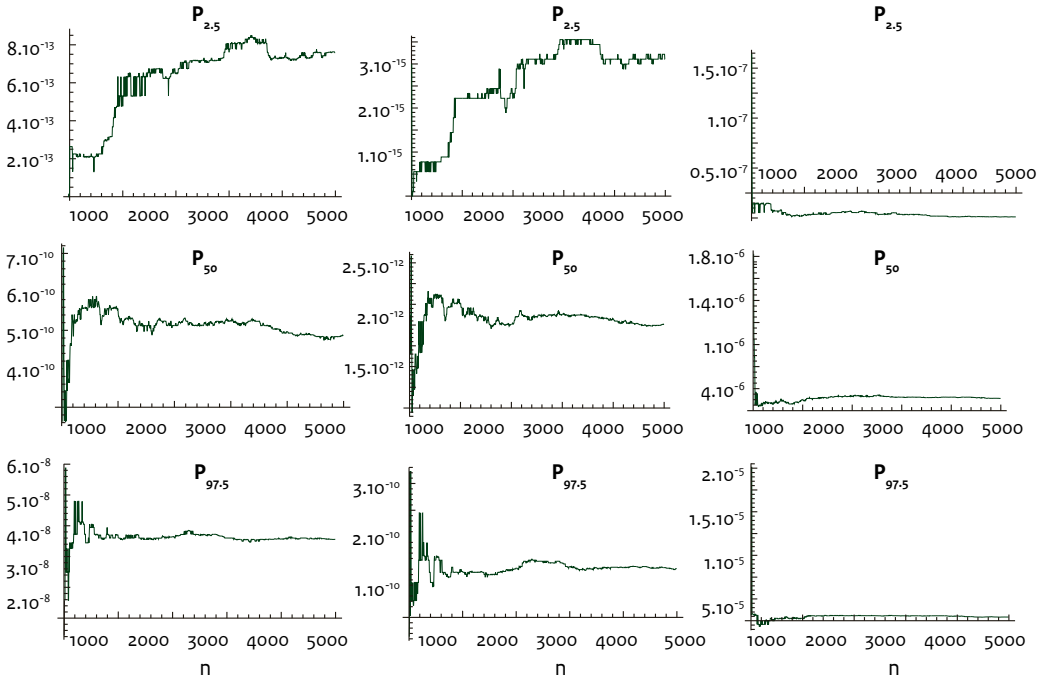
$$P(inf) = P_{inf}(D|\theta) \quad (26)$$

Using the dose response relation, the infection risk may be calculated. Other dose response models for acute illness may be inserted. This calculation may be repeated, each time repeating the sampling for all parameters, resulting in a Monte Carlo sample of the risk. When any (or all) parameters are defined as distributions, representing variability, these calculations can be repeated many times to obtain a (marginal) distribution of the infection (or illness) risk. In case the uncertainty is also available, a nested Monte Carlo simulation can be performed (Vicari et al. 2007, Pouillot and Delignette-Muller 2010).

5.5 Propagation of uncertainty

The above algorithm may be repeated for assessing variability and uncertainty. The source concentration C, all reductions by treatment (expressed as fractions), ingested volumes and the dose response parameters θ, may be sampled from probability distributions characterizing their variability and/or uncertainty. These distributions should have been found by statistical analysis of the available data, during exposure assessment and hazard characterization. In case of uncorrelated factors, a single random value can be generated for each variable, and a risk (say, a probability of infection) can be calculated. This can be repeated as often as necessary, generating a distribution of risks, of arbitrary size. How many iterations are needed? Figure 4 shows how percentiles of Monte Carlo samples on various risk estimates change with the number of iterations, for a risk assessment of Cryptosporidium in drinking water, for a water supply with infrequent failure in treatment (Teunis and Havelaar, 2001). It is clear that several thousands of iterations are needed before the higher and lower percentiles stabilize.
A common default is ten thousand iterations.

Figure 4. Stability of percentiles of dose and daily and yearly risk, with increasing number of (Monte Carlo) iterations n. From (Teunis et al., 1999b).



a) Daily dose (b) Daily risk (c) Yearly risk

Usually, the distributions of the various factors of risk are uncorrelated. If not, for instance when a multilevel analysis has been done to characterize exposure, the result is a joint multivariate distribution, and Monte Carlo samples can be constructed from such distributions by means of Markov chain Monte Carlo techniques (Gilks et al., 1996). In fact, such samples often directly result from the statistical analysis, if prediction of the outcome (the source concentration, or the combined removal or probability of passage) is done. Calculations of risk often involve a chain of events, resulting in dose as a product of several factors. When the dose is not too high the infection risk is again the product of dose and the (single hit) infectivity of the pathogens, as most infection dose response models are linear at low doses. Conditional probabilities of acute illness are often given as a fixed probability, that is multiplied by the infection risk to produce illness risks. And finally, accumulated risks (over a year, for instance) again involve products of many terms, for each individual exposure event. For those reasons, risk calculations usually are done in a multiplicative realm. A consequence of this is that risk distributions tend to assume a lognormal shape.

5.6 Transmission cycle

To translate infectious disease risk to the population level, the transmission cycle must be considered. Variable doses among the members of an exposed population may lead to infection in some of the exposed subjects. Once these infected (colonised) subjects start shedding pathogens, they may infect others, with whom they share contacts. This is called secondary infection (Diekmann et al. 2012). Thus, whenever an exposed subject is infected, they may cause a cluster of cases: an outbreak. The potential for secondary transmission is summarised in the reproduction number: the number of secondary cases any infectious subject causes. This may depend on immunity in the exposed population: the basic reproduction number Ro represents the number of secondary cases produced by an infectious subject released into a completely susceptible population (Diekmann et al. 2012). Thus, the risk caused by environmental exposure of the index (primary) case is amplified, by the numbers of secondary cases caused by this index case. Some food- or waterborne pathogens have epidemic potential, e.g. norovirus. In a real-world outbreak of enteric illness, the numbers of cases infected by any infectious vary strongly, as may be seen e.g. in healthcare associated norovirus outbreaks (Sukhrie et al. 2012, Teunis et al. 2013).
Sometimes it may be advantageous to model a closed transmission cycle, for instance when the infected population sheds pathogens into the environment, and these environmental pathogens are the source of exposure. This may occur e.g. in coastal areas where human shellfish consumption causes outbreaks of enteric viruses, while the shedded viruses are transported in sewage to the harvesting grounds of e.g. oysters, where they can contaminate these same oysters. Also, for foodborne zoonotic pathogens it may be important to consider transmission in the animal population.

5.7 Case: Why point estimates are a bad idea

A quick way to assess the risk might seem to take best (e.g. maximum likelihood) estimates for all the factors in the risk chain and calculate a point estimate of the risk. While it is often argued that this gives an impression of the magnitude of risk, it is a potentially dangerous approach. It is known from experience that distributions of risk usually are strongly (left–) skewed: while low levels of risk may be likely, there is a small probability of very high risks. Often, it is exactly the very small probability of a very severe outcome that we need to know. For instance, consider a drinking water treatment plant that removes pathogens with about 4 $\log_{10}$ units, nominally, thus rendering the finished water safe for consumption. If this treatment plant operates within design limits 364 days in a year, but at one day every year (on average) performance is degraded, so that only 1 $\log_{10}$ unit is removed, there is a small probability (about 0.27%) that the water contains a 1 000–fold higher pathogen concentration. In case the daily infection risk under nominal operation is 10^{-7}, it is 10^{-4} when performance is degraded. Without days of degraded operation, the yearly risk is approximately 0.37×10^{-4}, with degraded operation it becomes approximately 1.37×10^{-4}. The single degradation event dominates the yearly risk. Under such strongly skewed exposure conditions the arithmetic average of daily risk (and, in case it is low enough, the arithmetic average of daily dose) is still

the unbiased estimator of yearly risk. Such an average of a strongly skewed distribution can only be estimated with precision when the tails are well described, requiring many samples, or a method of capturing degradation events and sampling during such events.
When the goal of risk assessment is decision support (it usually is) then it is desirable to know quantiles of the estimated risk. The often debated infection risk level of 1 in 10 000 is only enforceable when there is a confidence level associated with it. In order to use a risk assessment to determine whether a water utility complies with the legal limit, the probability of exceeding that set risk limit must be below a certain level, say 95%. In the QMRA for drinking water in the Netherlands, the 95-percentile of the annual infection risk is the accepted level to test for compliance (Schijven et al., 2011). Therefore, a point estimate is useless. The confidence range associated with microbial estimates of risk may easily exceed a factor 100 in either direction from the average risk. Whether the average risk is below 1 in 10 000 therefore hardly provides any confidence in compliance.

6 Integrated risk metrics

6.1 General

Risk metrics combine probability estimates and measures of severity for a disease into a single number. This is nothing different from the expected loss calculations in the introduction. At the individual level, when the costs involved with sickness (its cost, or "severity") are c and the probability of becoming ill is p, the risk is

$$R = cp \tag{29}$$

Instead of a single outcome, the disease might progress into several categories of severity$\{c_1,c_2,\ldots,c_m\}$ with corresponding probabilities$\{p_1,p_2,\ldots,p_m\}$. The risk now is the weighted sum

$$R = \sum_{i=1}^{m} c_i p_i \tag{30}$$

In public health, risk usually is not defined on the level of the single individual, but on the population level. We have seen that the population risk is not simply N (the population size) times the individual risk, because of heterogeneity – variability in risk among individuals – and because often risks of different magnitudes are weighted differently. Before dealing with those problems we need to know the size of the affected population.

Population affected
In a city with a million inhabitants the drinking water may carry an individual yearly risk of infection with rotavirus of 10^{-5}. Does that mean that every year 10 people will be infected? On average, that may be true. But even when the infection risk is fixed at 10^{-5} the numbers infected will show random fluctuations around this average of 10. A better description would be that the numbers of inhabitants infected yearly is binomially distributed with $n = 10^6$ and $p = 10^{-5}$, and a good approximation is a Poisson distribution with parameter $\mu = np = 10$ (Nicas, 1996). In case the probability of infection is high and the exposed population is big so that the expected numbers are high, the variance in numbers infected due to random fluctuations is small compared to the expected numbers and can often be neglected. It is rare, in quantitative risk assessment, to have a fixed probability of infection or illness. There are always sources of heterogeneity: in exposure, and/or susceptibility of the host or virulence of the pathogen and these cause the probability of infection or illness to have a distribution itself, which is often highly skewed. The distribution of numbers affected then is a mixture, of the binomial (or its approximation, the Poisson) distribution and a distribution for p.

This distribution of the probability of a health effect p can have any shape, depending on the characteristics of exposure and dose response, so that in general it may not be easy (or even possible) to find the mixture distribution for numbers affected in closed form. It is however usually not difficult to simulate such a mixture distribution with Monte Carlo sampling. Some microbial infections may cause secondary transmission: primary infection in subjects exposed to pathogens in food or water (or aerosol) causes them to start excreting pathogens, rendering them infectious to other susceptible subjects. Given the right circumstances, there is considerable potential for rapid growth of the numbers infected, multiplying the numbers of primary cases by a large factor and causing an outbreak: a cluster of cases. Instead of the binomial distribution of numbers of cases the numbers affected then must be estimated from a transmission model.

6.2 Risk endpoint

The impact on society of outbreaks of large size can be disproportionally severe compared to small outbreaks. This is an aspect that has received very little attention in QMRA. Public transportation, logistics for goods and supplies, and public health services will suffer increasingly when too many citizens are incapable of performing their public functions.
The increase in severity with increasing size of the outbreak can be expressed in a (nonlinear) weighting function, or (dis-)utility function (Cox and Hinkley, 1974). This is a simple and statistically sound method to account for nonlinear effects in the severity of a large outbreak, compared to that of a small one. Suppose we have a probability distribution for the numbers of subjects affected by some infectious disease, p(n). If the costs associated with any infection are fixed, say c (in arbitrary units), the risk may be calculated

$$R = E(cn) = \sum_{n=0}^{N} c\, np(n) = c \sum_{n=0}^{N} np(n) = cE(n) \tag{31}$$

where N is the size of the exposed population. Here the losses have been expressed as

$$u(n) = cn \tag{32}$$

is a very simple utility function returning the costs of an (adverse) health effect as a function of the numbers affected. If we would like to give additional weight to large outbreaks, we could use instead the expectation of a function u(n) as

$$R = E\,(u(n)) = \sum u\,(n)p(n) \tag{33}$$

where u(n) > cn for large n, for instance

$$u(n) = cn^a,\ a > 1 \tag{34}$$

Different functions for weighting risk can be used for different problems, and weighting functions can also be employed to various categories of risk, for instance different endpoints of disease caused by the same pathogen, or different diseases (health burdens) caused by different agents in the same food or water. For a set of health effects with costs {u1(n), u2(n), . . . , um(n)} with numbers affected n with probabilities $\{p_1(n), p_2(n), \ldots, p_m(n)\}$ the risk is calculated as

$$R = \sum_{k=1}^{m} E\left(u_k(n)\right) = \sum_{k=1}^{m} \sum_{n=1}^{N} u_k(n) p_k(n) \tag{35}$$

Assuming that each individual can have any combination of health effects simultaneously. For mutually exclusive endpoints the numbers in each category are not independent. The joint distribution of these numbers $n = \{n_1, n_2, \ldots, n_m\}$ is then multinomial with probability p(n) and the risk becomes

$$R = E\left(p(n) \sum_{k=1}^{m} u_k\left(n_k\right)\right) \tag{36}$$

Such mixed utility functions can be useful in transforming risks to a common scale.

6.3 A specific mixed utility function: the DALY

When intervention measures have two or more disjunct, potentially competing effects, an integrated risk metric may be necessary. For instance, drinking water disinfection reduces the risk of infection and gastro-intestinal illness, but at the same time may increase the risk of long term illness caused by the toxicity of disinfection by–products (Havelaar et al., 2000). When both end-points – diarrhea and toxicity – can be measured on a common scale, quantitative weighting is possible. One such common metric that has gained popularity is the DALY (Disability Adjusted Life Year) originally proposed by Murray (1996) in the Global Burden of Disease study. Disease may result in reduction of survival time (quantity of life), in reduction of the quality of life or both. The loss of healthy life years in a population, measured in DALYs, is calculated as

$$\mathrm{DALY} = \mathrm{YLL} + \mathrm{YLD} \tag{37}$$

where YLL is the number of life years lost due to mortality and YLD is the number of years lived with a disability, weighted with a factor between 0 and 1 for the severity of the disability. For further details, see Murray (1996). If we think of severity weights as utility functions this is an example of the methods for calculating risks introduced in the previous section. The Life Years Lost have severity 1 (they are lost completely) and the Years Lived with Disability are a combination of any possible categories of illness with associated severities. For many enteric illnesses, a specific severity weight is not available. The Global Burden of Disease study (Murray, 1996) specifies a mean (median) weight of 0.066 (0.054) for illnesses characterized as “watery diarrhea”. Similar measures (Quality Adjusted Life Years) are commonly used in Medical Technology Assessment and Medical Decision Making to quantify the utility of

different options in a decision model (Murray, 1996). DALYs fulfill a similar role in decisions about the cost–effectiveness of intervention strategies in the drinking water or food chain, or to balance risks of interventions that reduce one kind of risk but increase another (Havelaar et al., 2000). Likewise, DALYs can be used as a rational, integrated measure of public health impact when defining Drinking Water or Food Safety Objectives. For health effects with impact over a very long period discounting of DALYs may be used, reducing the calculated numbers of DALYs by a certain yearly fraction, thereby reducing the influence of effects that are remote in time. For instance, future life years may be weighted less when they are further in the future.

7 Computational QMRA tools

A computational tool with a user-friendly interface to conduct QMRA may provide decision makers and other stakeholders, including experts and students, an open and science-based tool to predict infection and/or illness risks from exposure to pathogens in food or water routinely.

Based on experience with QMRAspot (Schijven et al., 2011), ideally, the tool has the following characteristics:

- The tool contains a probabilistic risk assessment, multiplicative model with functions for estimating factors from user provided surveillance data.
- For a set of index pathogens, the model computes infection and/or illness risks including variability and uncertainty in any contributing factor.
- In case of user-provided data, like from monitoring programs, maximum likelihood estimation assesses distribution parameter values.
- In the case of absent data or when data provide insufficient information, there is a well-described procedure for choosing default parameters, based on experience in deployment of the model and from scientific literature.
- The risk outcome can be tested against a standard or legal level, like a health based target.
- The focus of the tool is on robustness and stability so that it analyses many data sets well.
- The tool requires no extensive prior knowledge about QMRA by the user, because the tool provides guidance to the user on the quantity, type and format of raw data and performs a complete and automated analysis of the raw data.
- The uniform approach promotes proper collection and usage of raw data and, warrants quality of the risk assessment as well as enhances efficiency, i.e., less time is required.
- The tool must be transparent and science-based, and, therefore rely on publications in peer-reviewed journals, as well as have a (online) manual including the executable code, its use and detailed explanation of the risk assessment model.
- Workshops should be organised to train users. The interactive tool itself is educational.
- Version control should follow quality control standards. Corrections to operational versions are guided by feedback from users and by analysing the expanding database of tested surveillance data.
- The tool and manual provide directions for feedback, to record comments, recommendations, and to catch errors.
- Results are displayed as graphs (time series, histograms, pdfs) and tables (mean, quantiles).
- The tool produces a QMRA report including data, distribution parameter values, and risk outcomes for documentation.
- The tool contains some “tornado” chart of correlation of calculated risk with contributing

factors for sensitivity analysis.
- Functionality is in place to have users compare different scenarios, e.g. adding nonzero observations to a set of occurrence data without any positives; adding or removing treatment stages, or modifying existing stages.
- The studies providing the basic information for the submodels on occurrence, detection efficiency, treatment effects, and dose response have dealt with uncertainty (in conjuction with variability) and compared different models to address model uncertainty (cited).
- The tool is freely available
- Continual updating is warranted by internal and external feedback communicated by users (in routine use, but also in courses and individual use. The review and update activities are coordinated by the main author (and contact).
- A user-friendly interface, the guidance in data collection, the provision of default values, a user-manual, workshops and courses make the risk assessment tool highly educative.

Below a non-exhaustive list of existing QMRA tools is given:

- *www.foodrisk.org*
- *https://foodrisklabs.bfr.bund.de/foodrisk-labs/*
- ILSI report (Basset et al., 2012)
- QMRAspot (Schijven et al., 2011) covers most of the abovementioned characteristics. QMRAspot has been developed to calculate infection risks by consumption of waterborne pathogens in drinking water. It does not include an uncertainty analysis.
- FDA-iRISK is an online tool from the Food and Drug Administration (FDA) for constructing risk models and estimating health burden. (Chen et al., 2013)
- The Interactive online catalogue on risk assessment (ICRA) is an open repository for risk assessment models (*http://www.icra-edu.org*). It currently includes several models regarding bacteria in meat and eggs.
- MicroHibro is an online tool for microbial risk assessment in vegetables and meat. A draft manual is also available (*www.microhibro.com*).
- The World Health Organization (WHO) and Food and Agricultural Organization (FAO) link to a few tools regarding: Cronobacter in infant formula; Campylobacter and Salmonella in chicken; and assessment of sampling plans..
- R Packages for Risk Assessment (Pouillot and Delignette-Muller). Two R1 packages specifically developed to help risk assessors in their projects are now available: “fitdistrplus”, gathers graphical and statistical tools for choosing and fitting distributions. “mc2d”, helps to build and study two-dimensional (or second-order) Monte-Carlo simulations.
- The QMRA package in R (Brecht, 2016)provides mMaximum-likelihood and Bayesian parametric methods for exposure and dose-response assessment.

The list non-exhaustive because it is meant as a starting point for a “rolling revision” type of updates and as examples.

8 Acknowledgements

This document was written in collaboration with WHO and FAO. We like to thank Rei Nakagawa, Angelika Maria Tritscher (WHO, Geneva) and Sarah Cahill (FAO, Rome) for initiating and helpful comments.

The authors highly acknowledge the reviewing of this report by two international experts: The report has been reviewed by two international experts, namely Graham McBride, Principal Scientist – Water Quality,The National Institute of Water and Atmospheric Research (NIWA), New Zealand and Maarten Nauta, Senior Researcher, Research Group for Risk-Benefit, National Food Institute, Technical University of Denmark, Kongens Lyngby.

The reviewing by RIVM-colleagues Joke van der Giessen, Harold van de Berg and Ana Maria de Roda Husman is highly acknowledged as well.

This is a product of the WHO Collaborating Centre Risk Assessment of Pathogens in Food and Water, National Institute for Public Health and the Environment, RIVM. Contact: P.O. Box 1, 3720 BA Bilthoven, The Netherlands. E-mail: whoccmicro@rivm.nl

9 References

Bassett, J., Nauta, M., Lindqvist, R. and Zwietering, M. 2012 Tools for Microbiological Risk Assessment. ILSI Europe Report, Brussels.

Benford D. Principles of Risk Assessment of food and drinking water related to human health. ILSI Europe concise monographs. ILSI Press, One Thomas Circle, NW, 9th floor, Washington, D.C. 20005-5802, USA: ILSI Europe, 2001.

Black RE, Levine MM, Clements ML, Hughes TP, Blaser MJ (1988). Experimental Campylobacter jejuni infection in humans. Journal of Infectious Diseases 157(3):472-479.

Burmaster DE, Anderson PD (1994). Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk asessments. Risk Analysis 14(4):477–481.

CDC. FoodNet surveillance – Burden of illness pyramid. 2005. *http://www.cdc.gov/foodnet/surveillance_pages/burden_pyramid.htm*.

Chappell CL, Okhuysen PC, Sterling CR, Wang C, Jakubowski W, DuPont HL (1999). Infectivity of Cryptosporidium parvum in healthy adults with pre-existing anti-C. parvum serum immunoglobulin G. American Journal of Tropical Medicine and Hygiene 60(1):157-164.

Chappell, C. L., Okhuysen, P. C., Langer-Curry, R., Widmer, G., Akiyoshi, D. E., Tanriverdi, S., & Tzipori, S. (2006). Cryptosporidium hominis: experimental challenge of healthy adults. The American journal of tropical medicine and hygiene, 75(5), 851-857.

Chen, L., Geys, H., Cawthraw, S., Havelaar, A., & Teunis, P. (2006). Dose response for infectivity of several strains of Campylobacter jejuni in chickens. Risk Analysis, 26(6), 1613-1621.

Chen, Y., Dennis, S. B., Hartnett, E., Paoli, G., Pouillot, R., Ruthman, T., & Wilson, M. (2013). FDA-iRISK – a comparative risk assessment system for evaluating and ranking food-hazard pairs: case studies on microbial hazards. Journal of food protection, 76(3), 376-385.

Clough HE, Clancy D, French NP (2009). Quantifying exposure to Vero-toxigenic Escherichia coli O157 in milk sold as pasteurized: a model-based approach. International Journal of Food Microbiology 131(2-3):95-105. DOI: 10.1016/j.ijfoodmicro.2008.12.036.

Clough HE, Clancy D, O'Neill PD, Robinson SE, French NP (2005). Quantifying uncertainty associated with microbial count data: a Bayesian approach. Biometrics 61(2):610-616. DOI: 10.1111/j.1541-0420.2005.030903.x

Codex Alimentarius Commission. Principles and guidelines for the conduct of microbiological risk assessment. Report CAC/GL-30, Food and Agriculture Organization of the United Nations and World Health Organization, 1999.

Codex Alimentarius Commission. Risk assessment procedures used by the Codex Alimentarius Commission, and its subsidiary and advisory bodies. Technical report, Commission du Codex Alimentarius, joint FAO/WHO food standards programme, Geneva, 1993. ALINORM 93/37, 20th session, 28 june - 7 july.

Codex Committee on Food Hygiene. Proposed draft process by which the Codex Committee on Food Hygiene could undertake its work in microbiological risk assessment/risk management. Report CX/FH 03/6, Food and Agriculture Organization of the United Nations and World Health Organization, Orlando, Florida, 2003. Joint FAO/WHO Food Standards Programme.

Cox DR, Hinkley DV (1974). Theoretical Statistics. Chapman and Hall, London. isbn 0 412 16160 5

CRAN: Comprehensive R Archive Network (CRAN), 2015. *http://cran.r-project.org/*.

Curtis V, Biran A, Deverell K, Hughes C, Bellamy K, Drasar B (2003). Hygiene in the Home: Relating Bugs and Behaviour. Social Science & Medicine 57(4): 657–672. doi:10.1016/S0277-9536(02)00409-4.

Diekmann O, Heesterbeek H, Britton T (2012). Mathematical tools for understanding infectious disease dynamics. Princeton University Press. isbn 9780691155395.

DuPont HL, Chappell CL, Sterling CR, Okhuysen PC, Rose JB, Jakubowski W (1995). The infectivity of Cryptosporidium parvum in healthy volunteers. New England Journal of Medicine 332(13):855-859.

EHC 2009: Principles and Methods for the Risk Assessment of Chemicals in Food. Food and Agriculture Organization of the United Nations/World Health Organization. 2009. Environmental Health Criteria 240. ISBN 978 92 4 157240 8

FAO/WHO 2003: Hazard Characterization for Pathogens in Food and Water. Food and Agriculture Organization of the United Nations/ World Health Organization. 2003. Microbiological Risk Assessment Series Volume 3.

FAO/WHO 2008: Exposure assessment of Microbiological Hazards in Food. World Health Organization/Food and Agriculture Organization of the United Nations. 2008. Microbiological Risk Assessment Series Volume 7.

FAO/WHO 2009: Risk Characterization of Microbiological Hazards in Food. World Health Organization/Food and Agriculture Organization of the United Nations. 2009. Microbiological Risk Assessment Series Volume 17.

FAO/WHO. Principles and guidelines for incorporating microbiological risk assessment in the development of food safety standards, guidelines and related texts. Report isbn: 92-5-104845-2, World Health Organization, Kiel, Germany, 2002. Joint meeting of WHO, FAO and Institute for Hygiene and Food Safety of the Federal Dairy Research Center, Germany.

FAO/WHO. Report of a joint FAO/WHO consultation on the role of government agencies in assessing HACCP. Report, Food and Agriculture Organization of the United Nations and World Health Organization, Geneva, 1998. *http://www.who.int/foodsafety/publications/*.

FAO/WHO. The interaction between assessors and managers of microbiological hazards in food. Report WHO/SDE/PHE/FOS/007, World Health Organization, Kiel, Germany, 2000. Joint meeting of WHO, FAO and Institute for Hygiene and Food Safety of the Federal Dairy Research Center, Germany.

FAO/WHO. The use of microbiological risk assessment outputs to develop practical risk management strategies. Metrics to improve food safety. Report, Food and Agriculture Organization of the United Nations and World Health Organization, Kiel, Germany, 2006. Joint meeting of WHO, FAO and The Federal Ministry of Food, Agriculture and Consumer Protection, Germany.

Galassi M, Davies J, Theiler J, Gough B, Jungman G, Alken P, Booth M, Rossi F (2011). GNU Scientific Library Reference Manual. Free Software Foundation, 3rd edition. *http://www.gnu.org/software/gsl/*.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian Data Analysis. 3rd edition. Chapman & Hall/CRC texts in Statistical Science. CRC press, Boca Raton FL.

Gilks WR, Richardson S, Spiegelhalter DJ (1996). Markov Chain Monte Carlo in Practice. Chapman and Hall, London.

Haas CN (1983). Estimation of risk due to low doses of microorganisms: a comparison of alternative methodologies. American Journal of Epidemiology 118(4):573–582.

Haas CN, Rose JB, Gerba CP. Quantitative microbial risk assessment. John Wiley & Sons, 1999.

Haas CN, Thayyar-Madabusi A, Rose JB, Gerba CP (2000). Developement of a dose-response relationship for Escherichia coli O157:H7. International Journal for Food Microbiology 57:153-159.

Havelaar AH, de Wit MA, van Koningsveld R, van Kempen E (2000). Health burden in the Netherlands due to infection with thermophilic Campylobacter spp. Epidemiology and Infection 125(3):505-522.

Huang Y, Haas CN (2009). Time-Dose-Response Models for Microbial Risk Assessment. Risk Analysis 29(5): 648–661. doi:10.1111/j.1539-6924.2008.01195.x.

ILSI/EPA. Revised framework for microbial risk assessment. Workshop report, ILSI Risk Science Institute, 1126 Sixteenth Street, N.W., Washington, D.C. 20036-4810 USA, 2000.

Krewski D and van Ryzin J (1981). Dose response models for quantal response toxicity data. pp 201-231 in: Proceedings of the International Symposium on Statistics and related topics. eds. Csörgö M, Dawson DA, Rao JNK, Saleh AKMdE. North Holland Publishing Company, Ottawa, Ontario.

LeChevallier MW, Di Giovanni G, Clancy JL, Bukhari Z, Bukhari S, Rosen JS, Sobrinho J, Frey MM (2001). Variability of Giardia and Cryptosporidium concentrations in surface water. American Water Works Service Company, Inc. Voorhees, NJ.

Mataragas, M., Drosinos, E. H., Siana, P., Skandamis, P., & Metaxopoulos, I. (2006). Determination of the growth limits and kinetic behavior of Listeria monocytogenes in a sliced cooked cured meat product: validation of the predictive growth model under constant and dynamic temperature storage conditions. Journal of food protection, 69(6), 1312-1321.

Medrano-Félix A, Martínez C, Castro-del Campo N, León-Félix J, Peraza-Garay F, Gerba CP, Chaidez C (2011). Impact of Prescribed Cleaning and Disinfectant Use on Microbial Contamination in the Home. Journal of Applied Microbiology 110(2): 463–471. doi:10.1111/j.1365-2672.2010.04901.x.

Messner, M. J., & Berger, P. (2016). Cryptosporidium infection risk: results of new dose-response modeling. Risk Analysis, 36(10), 1969-1982.

Murray CJL, Lopez AD (1996). The global burden of disease: a comprehensive assessment of mortality and disability from disease, injury, and risk factors in 1990 and projected to 2020. Global burden of disease and inhjury series. WHO/Harvard University Press.

Mylius SD, Nauta MJ, Havelaar AH (2007). Cross-Contamination During Food Preparation: a Mechanistic Model Applied to Chicken-Borne Campylobacter. Risk Analysis 27(4): 803–813. doi:10.1111/j.1539-6924.2006.00872.x.

National Academy of Sciences. Drinking water and health. NAS, Washington DC, 1977.
National Research Council. Risk assessment in the federal government: managing the process. Washington, D.C.: National Academy Press, 1983.
Nauta MJ, Jacobs-Reitsma WF, Evers EG, van Pelt W, Havelaar AH. Risk assessment of Campylobacter in the Netherlands via broiler meat and other routes. Technical Report 250 911 006, RIVM, Bilthoven, 2005. *http://www.rivm.nl/bibliotheek/rapporten/250911006.html*.
Nauta, M. J. (2002). Modelling bacterial growth in quantitative microbiological risk assessment: is it possible?. International Journal of Food Microbiology, 73(2), 297-304.
Nauta, M. J., Fischer, A. R., Van Asselt, E. D., De Jong, A. E., Frewer, L. J., & De Jonge, R. (2008). Food safety in the domestic environment: the effect of consumer risk information on human disease risks. Risk Analysis, 28(1), 179-192.
Nauta, M.J., (2008). The Modular Process Risk Model (MPRM): a Structured Approach for Food Chain Exposure Assessment. In: Schaffner, D. W (ed.) Microbial Risk Analysis of Foods. pp. 99-136 ASM Press, Washington, D.C
Oscar, T. (2004). Dose-Response Model for 13 Strains of Salmonella. Risk Analysis, 24(1), 41-49.
Plummer M (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), pages 1–10, Vienna, Austria, March 20-22 2003.
Plummer M (2013). rjags: Bayesian graphical models using MCMC, 2013. R package.
Pouillot, R., and Delignette-Muller, M. L. (2010). Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. International Journal of Food Microbiology, 142(3), 330–340. *http://doi.org/10.1016/j.ijfoodmicro.2010.07.011*
R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for StatisticalComputing, Vienna, Austria, 2015.
Rai, S. N. and Krewski, D. (1998), Uncertainty and Variability Analysis in Multiplicative Risk Models. Risk Analysis, 18: 37–45. doi:10.1111/j.1539-6924.1998.tb00914.x)
Regli S, Rose JB, Haas CN and Gerba CP (1991). Modeling the risk from Giardia and viruses in drinking water. Journal of the American Water Works Association 83: 76-84.
Rusin P, Orosz-Coughlin P, Gerba CP (1998). Reduction of Faecal Coliform, Coliform and Heterotrophic Plate Count Bacteria in the Household Kitchen and Bathroom by Disinfection with Hypochlorite Cleaners. Journal of Applied Microbiology 85(5): 819–828.
Schijven JF, Teunis PFM, Rutjes SA, Bouwknegt M, de Roda Husman AM (2011). QMRAspot: a tool for Quantitative Microbial Risk Assessment from surface water to potable water. Water Research, 45(17), 5564–5576. *http://doi.org/10.1016/j.watres.2011.08.024*
Slob W (1994). Uncertainty analysis in multiplicative models. Risk Analysis 14(4):571-576.
Stan Development Team. 2016. Stan Modeling Language Users Guide and Reference Manual. Version 2.15.0. *http://mc-stan.org*
Sukhrie FHA, Teunis PFM, Vennema H, Copra C, Beersma MFC, Bogerman J, Koopmans M. (2012). Nosocomial Transmission of Norovirus Is Mainly Caused by Symptomatic Cases. Clinical Infectious Diseases 54(7):931-937. *http://doi.org/10.1093/cid/cir971*

Tenenhaus-Aziza F, Ellouze M. 2015 Software for predictive microbiology and risk assessment: A description and comparison of tools presented at the icpmf8 software fair. Food Microbiol 45(Pt B):290-299

Teunis PFM and Havelaar AH (2001). Evaluation of the ILSI/EPA risk assessment framework in assessing risks of waterborne pathogens. pp 165-178 in: Microbial pathogens and disinfection by-products in drinking water: Health effects and management of risks. Ed: Craun GF, Hauchman FS, Robinson DE. ILSI press, Washington D.C.

Teunis PFM, Chappell CL, Okhuysen PC (2002). Cryptosporidium Dose-Response Studies: Variation Between Hosts. Risk Analysis 22(3): 475–485.

Teunis PFM, Evers EG, Slob W (1999b). Analysis of variable fractions resulting from microbial counts. Quantitative Microbiology 1: 63-88.

Teunis PFM, Figueras MJ (2016b). Reassessment of the Enteropathogenicity of Mesophilic Aeromonas Species. Frontiers in Microbiology 7(1395):1–12. doi:10.3389/fmicb.2016.01395.

Teunis PFM, Havelaar AH (1999a). Cryptosporidium in drinking water: Evaluation of the ILSI/RSI quantitative risk assessment framework. Technical Report 284 550 006, RIVM, Bilthoven, The Netherlands.

Teunis PFM, Havelaar AH (2000). The Beta Poisson Dose-Response Model Is Not a Single-Hit Model. Risk Analysis 20(4): 513–520.

Teunis PFM, Heijne JCM, Sukhrie F, van Eijkeren J, Koopmans MPG, Kretzschmar MEE (2013). Infectious disease transmission as a forensic problem: who infected whom? Journal of the Royal Society, Interface 10(81), 20120955–20120955. *http://doi.org/10.1098/rsif.2012.0955*

Teunis PFM, Kasuga F, Fazil A, Ogden ID, Rotariu O, Strachan NJC (2010). Dose-Response Modeling of Salmonella Using Outbreak Data. International Journal of Food Microbiology 144(2): 243–249. doi:10.1016/j.ijfoodmicro.2010.09.026.

Teunis PFM, Koningstein M, Takumi K, van der Giessen JWB (2012). Human Beings Are Highly Susceptible to Low Doses of Trichinella Spp. Epidemiology and Infection 140(2): 210–218. doi:10.1017/S0950268811000380.

Teunis PFM, Lodder WJ, Heisterkamp SH, de Roda Husman AM (2005b). Mixed plaques: statistical evidence how plaque assays may underestimate virus concentrations. Water Research 39(17):4240-4250.

Teunis PFM, Medema GJ, Kruidenier L, Havelaar AH (1997). Assessment of the risk of infection by Cryptosporidium or Giardia in drinking water from a surface water source. Water Research 31(6):1333-1346.

Teunis PFM, Nagelkerke NJD, Haas CN (1999c). Dose Response Models for Infectious Gastroenteritis. Risk Analysis 19(6): 1251–1260.

Teunis PFM, Ogden ID, Strachan NJC (2008). Hierarchical Dose Response of E. Coli O157:H7 From Human Outbreaks Incorporating Heterogeneity in Exposure. Epidemiology and Infection 136(6): 761–770. doi:10.1017/S0950268807008771.

Teunis PFM, Reese HE, Null C, Yakubu H, Moe CL (2016a). Quantifying Contact with the Environment: Behaviors of Young Children in Accra, Ghana. Journal of Tropical Medicine and Hygiene 94(4): 920–931. doi:10.4269/ajtmh.15-0417.

Teunis PFM, Rutjes SA, Westrell T, de Roda Husman AM (2009). Characterization of drinking water treatment for virus risk assessment. Water Research 43(2):395-404.

Teunis PFM, Takumi K, Shinagawa K (2004). Dose response for infection by Escherichia coli O157:H7 from outbreak data. Risk Analysis 24(2):401-408.
Teunis PFM, van den Brandhof W, Nauta M, Wagenaar J, van den Kerkhof H, van Pelt W (2005a). A reconsideration of the Campylobacter dose response relation. Epidemiology and Infection 133(4):583-592.
Van Boekel, M. A. (2002). On the use of the Weibull model to describe thermal inactivation of microbial vegetative cells. International journal of food microbiology, 74(1), 139-159.
Vicari, A. S., Mokhtari, A., Morales, R. A., Jaykus, L.-A., Frey, H. C., Slenning, B. D., and Cowen, P. (2007). Second-order modeling of variability and uncertainty in microbial hazard characterization. Journal of Food Protection, 70(2), 363–372.
Vose D. Quantitative Risk Analysis. A guide to Monte Carlo modelling. Chichester: John Wiley & Sons, 1996.
Wang Y, Moe CL, Null C, Raj SJ, Baker KK, Robb K, Yakubu H, Ampofo JA, Wellington N, Freeman M, Armah G, Reese HE, Peprah D, Teunis PFM (2017). Quantitative Assessment of Exposure to Fecal Contamination for Young Children in a Crowded Low-Income Urban Environment in Accra, Ghana. Revision in review in: American Journal of Tropical Medicine and Hygiene.
WHO 2009: Risk Assessment of Cryptosporidium in Drinking Water. World Health Organization. 2009
WHO 2016: Quantitative Microbial Risk Assessment: application for water safety management. World Health Organization. 2016. ISBN 978 92 4 156537 0
WHO. Application of risk analysis to food standards issues. Report WHO/FNU/FOS/95.3, World Health Organization, Geneva, 1995. Joint FAO/WHO Expert Consultation.
WHO. Guidelines for drinking water quality, volume 1. Geneva: World Health Organization, 3rd edition, 2004. isbn: 9789241546386.
WHO. The risk to health of microbes in sewage sludge applied to land. Technical Report 1981, 92 890 1220 X, Zoon 52, V35, World Health Organization, Regional Office for Europe, Stevenage U.K., Copenhagen, 1981. prepared by a WHO Working Group.
WHO. Training considerations for the Hazard Analysis Critical Control point System. Technical Report WHO/HPP/FOS 93.3, World Health Organization, Geneva, 1993.
Wolfram Research Inc. Mathematica. Wolfram Research, Inc., Champaign, Illinois, 2015. version 10.2.

P. Teunis and J.F. Schijven

RIVM report 2017-0188

Published by:

National Institute for Public Health and the Environment, RIVM
P.O. Box 1 | 3720 BA Bilthoven
The Netherlands
www.rivm.nl/en

September 2018

Committed to *health and sustainability*

010555