6. Quality assurance

Risk characterization not only synthesizes the results of the previous parts of the risk assessment but also summarizes the overall findings and presents the strengths and limitations of the analysis to risk managers. The validity of the risk assessment is based on the soundness of the model structure, its input, the underlying assumptions and the interpretation of results. Therefore, quality assurance is a crucial element of risk characterization. Quality assurance can be achieved through a variety of methods. Data quality assurance is discussed in Section 6.1. Assessing the weight of evidence is discussed in Section 6.2. The sensitivity analysis is described in Section 6.3, while uncertainty analysis is addressed in Section 6.4. Model verification, anchoring and validation are addressed in Sections 6.5, 6.6 and 6.7, respectively. A specific method for model validation, involving comparison to epidemiological data, is discussed in Section 6.8. Model robustness and issues pertaining to model extrapolation are addressed in Section 6.9. The criteria for risk assessment credibility discussed in Section 6.10 include proper documentation of the analysis and peer review of the assessment. Public review is discussed in Section 8.5.

6.1 Data quality assurance

The results of sensitivity or uncertainty analysis are conditional on the data and other information used to develop the risk assessment model. Because it serves as the primary vehicle for communicating the risk assessment findings to risk managers, a risk characterization should briefly summarize the primary strengths and limitations of the data, methods, and analyses identified in the hazard identification, exposure assessment, and hazard characterization. Typically, these analyses require risk assessors to synthesize and draw inferences from disparate data sources not specifically or originally intended for use in risk assessment. In some cases, this requires the use of unconventional or non-routine methods that might be highlighted for particularly close scrutiny to ensure that they are reasonable and correctly applied. For relevant details, see the FAO/WHO hazard characterization and exposure assessment guidelines (FAO/WHO, 2003, 2008).

6.1.1 Data collection

Usually the suitable data for a microbiological risk assessment are sparse. In practice, assessors should initially collect all reasonably obtainable data consistent with the assessment objective, and subsequently investigate the quality of the different data sources. When collecting data for input distributions, several issues should be considered in order to evaluate data quality. The following considerations apply to empirical data and information elicited from experts.

Ideally, risk assessors would have access to raw, un-summarized data. With raw data (if consisting of sufficient observations), statistical methods such as Goodness-of-Fit tests are available to define a suitable parametric distribution describing the data. Alternatively, empirical distributions or non-parametric simulation methods can be used to characterize input distributions. Raw data, however, are frequently inaccessible. Often results are reported as aggregated summary statistics, such as the estimated mean, standard deviation or standard error of the mean. In order to develop a distribution from data summary statistics, it is necessary to obtain information on the assumed distribution of the underlying data, together with the sample size.

It is useful to collect as much background information on the data sources as possible, such as the year of completion, country of origin, the type of sample, possible transformation of the data, methods of analysis, microbiological strain and population demographics. This information could be important with regard to treatment or use of the data or to support the decision on whether or not to include these data in the model. An example is given in Box 6.1.

Box 6.1 Example of a Danish risk assessment of *Campylobacter jejuni* in chicken.

For the risk assessment, quantitative data were needed to describe the relative change in pathogen concentration over a given step in a poultry slaughterhouse (e.g. over the washing and chilling step, Figure 6.1). Because Danish data were unavailable, data from foreign studies were applied to assess the efficacy of the wash and chiller process in reducing the pathogen levels on chicken carcasses. Data for the microorganism of interest were available, but the data were obtained from different sample units (neck skin samples, whole carcass wash, and swab samples). This mix of sample types all reflected surface contamination of chicken carcasses. In synthesizing the data, it was assumed that the relative reduction in pathogen concentration over the process was independent of the type of surface measure. In Figure 6.2, the slopes reflect differences in log-concentration over the process. Since all the slopes appear to be similar, all data sets were used in describing the reduction over the 'wash + chiller' process. (Christensen et al., 2001).

Data for the specific microorganism under study may not always be available or of suitable quantity and quality (e.g. due to rare occurrence or imprecise collection methods). In that case, data from a surrogate microorganism can be used, provided that the surrogate behaves similarly under the process of interest (e.g. generic *E. coli* to estimate cross-contamination during slaughter procedures). In practice, data from different surrogate organisms could be used to model different steps in the same model, based on their availability and suitability. In some cases, sampled data with different units (e.g. absolute concentration or change in concentration) can be used in describing the same process, as the example below illustrates. Depending on how the data are used in the model (e.g. describing a change in concentration over a step or describing the concentration level, Figure 6.2), different parameters may be evaluated in a sensitivity analysis to ensure data quality objectives are satisfied.

Sensitivity analysis is a useful data quality assurance tool. Specific data sources and model inputs identified to have an important influence on model outputs warrant careful assessment. The available data may understate the true range of variability in a model input. In the example described above, the available data only covers two countries, and the variability may be greater than the empirical data alone suggest. Hence, techniques such as nominal range sensitivity analysis can be employed to evaluate the sensitivity of the model output to varying the model input across its entire range of plausible values. In other cases, the available data may not be considered representative of the population of interest. In such cases, the data may be excluded from the analysis or incorporated with appropriate adjustment. The bases for decisions regarding the treatment of non-representative data are context specific and need to be clearly articulated. For example, data from a particular source may be considered non-representative for the purposes of providing an estimate of central tendency (e.g. the mean) but useful for the purposes of characterizing the spread of an input distribution (e.g. plus or minus an order of magnitude).

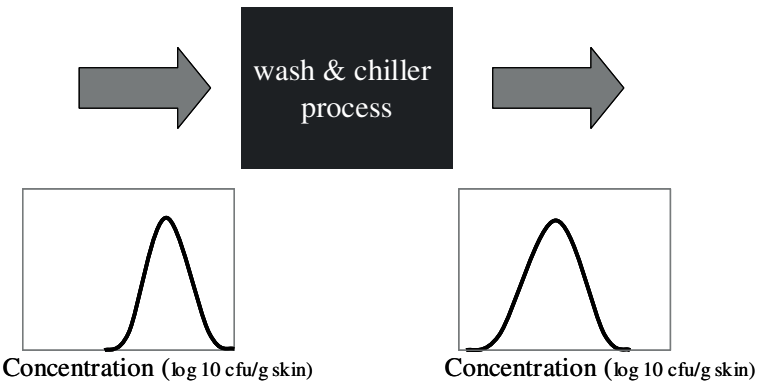

Figure 6.1 Illustration of a 'black-box' sub-model connecting two observed data sets (i.e. 'anchor points') over a process. The relative reduction of the *Campylobacter* load on chicken carcasses was assumed to be independent of where on the carcass the sample was taken. When data are given as log CFU values, this means that the relative change in concentration over the process (wash + chiller) is obtained by subtracting the concentrations before and after the process.

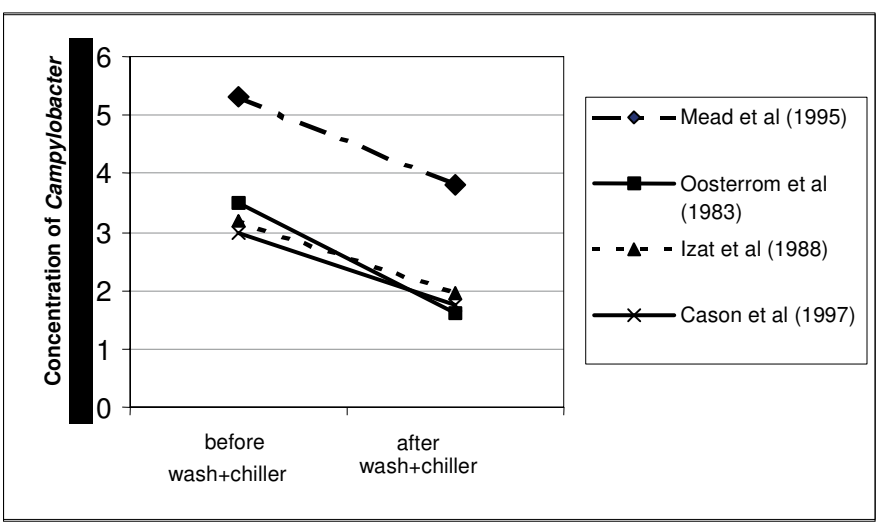

Figure 6.2 The influence of a selected slaughterhouse process on the *Campylobacter* concentration on chicken carcasses. The change in pathogen concentrations before and after the process is represented by a line connecting data points originating from the same study.

6.1.2 Sorting and selecting data sources

After collecting potentially suitable data sets, one should evaluate each of them critically and select the data that will provide the best possible model input for a specific purpose, such as describing the level of contamination, prevalence or changes over a process. Plotting the parameter of interest with the 95% confidence intervals provides a useful overview (see Figure 6.3).

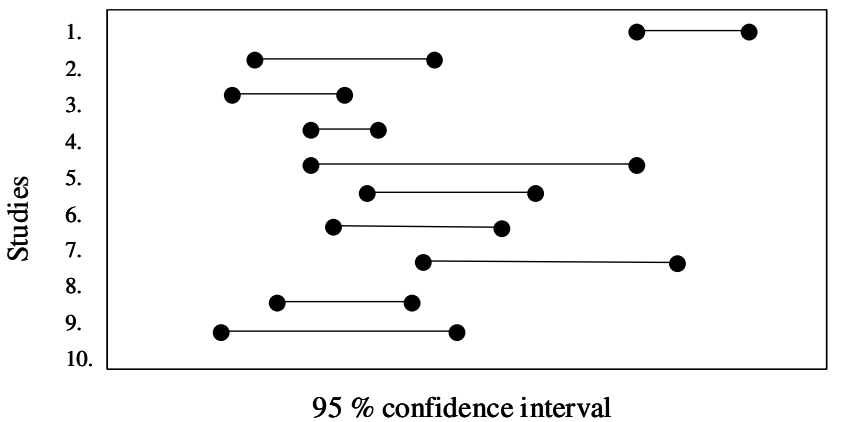

Figure 6.3 Example of an overview of data from different studies, with their 95% confidence intervals.

In selecting the suitable data sets for incorporation into the risk assessment, both subjective and analytical criteria may be applied. Subjective evaluation criteria may include the representativeness of the geographical and temporal properties of the candidate study. For example, if study no. 1 in Figure 6.3 is the only foreign study and it is significantly different from the rest (based on analytical criteria), this data set could be excluded. In contrast, if the 10 studies all originate from the same country, same year, etc., but are reported by different laboratories, the differences may be due to variability among the laboratories and the assessor might decide to incorporate all of the studies in the model.

6.2 Progression and weight of evidence

Whether an assessment is quantitative or qualitative, the public health risk posed by a micro-organism can be conceived at a basic level as the product of hazard, exposure and susceptible consumers (Figure 6.4).

Figure 6.4 Epidemiological Triangle.

If any one of the three elements of the epidemiological triangle equals zero, then there is no risk. A preliminary quality assurance step, therefore, is to evaluate whether a risk assessment

reflects this logical progression of threshold questions, to which the risk assessor could respond yes or no (perhaps with a qualifying level of confidence). If the response to a threshold question is 'no', then the analysis proceeds no further. At each threshold, the weight of evidence should be evaluated according to clearly specified, scientific criteria. As more criteria are satisfied, the weight of evidence indicates a more credible risk. Although there is a prima facie public health risk posed by several pathogens commonly associated with acute foodborne illnesses, in the future, risk assessors are likely to confront more cryptic and increasingly complex risk management questions, such as the risk posed by antibiotic-resistant microorganisms, the burden of chronic sequelae, the effect of specific growth-inhibiting food product formulations, and the susceptibility of individuals with underlying health problems. Some preliminary quality assurance guidance is provided here, therefore, in the anticipation that weight-of-evidence determinations will become increasingly prominent in risk assessments of microbiological pathogens in food.

6.3 Sensitivity analysis

Complex risk assessments may have many input and output variables that are linked by a system of equations or other model structures. Sensitivity analysis is a broad set of tools that can provide insights to risk assessors and risk managers about the relative importance of the components of a risk assessment to the risk management question. The plausibility of important components is essential to the overall quality of the risk assessment. Changes in important components also can be expressed in terms of the influence that these inputs have on the answers to risk-management questions.

A key criterion for sensitivity analysis is that it must be relevant to a decision. Sensitivity analysis evaluates the effect of changes in model input values and assumptions on the model output, and thus on decisions that would be based on the model output. It can be used during model development to evaluate and refine model performance and can play an important role in model verification and validation throughout the course of model development and refinement. Sensitivity analysis can also be used to provide insight into the robustness of model results when making decisions.

Sensitivity analysis can also be used as an aid in identifying important uncertainties for purposes of prioritizing additional data collection or research. For these purposes, value of information (VOI) analysis can complement sensitivity analysis methods, because the return to risk management decision-making on research and data collection expenditures depends on a variety of additional considerations (e.g. cost and time).

Microbiological risk assessment models typically have the following characteristics, which can pose substantial challenges to the application of sensitivity analysis methods:

- non-linearities;
- thresholds (e.g. below which there is no growth of a microbiological pathogen);
- discrete inputs (e.g. integer numbers of animals or herds; yes or no indicators of contamination);
- incorporation of measurement error;
- variation in the scale (units and range) and shape of distributions of model inputs; and
- temporal and spatial dimensions, including dynamics, seasonality or inter-annual variability.

Ideally, a sensitivity analysis method should provide not just a rank ordering of key inputs, but also some discriminatory quantitative measure of sensitivity, such that it is possible to clearly distinguish the relative importance of different inputs. For example, are there groups of inputs among which several inputs are of comparable importance, and is there clearly a difference in importance between such groups? Statistical-based methods such as regression analysis or analysis of variance (ANOVA) produce quantitative indicators of the relative importance of different inputs. Moreover, techniques such as regression analysis also provide an indication of the statistical significance of differences in sensitivity among inputs, based upon confidence intervals for regression coefficients.

This section emphasizes sensitivity analysis in quantitative risk assessment models, although some of the techniques (e.g. exploratory methods) may apply to both quantitative and qualitative assessments.

6.3.1 Sensitivity analysis in qualitative risk assessment

In examining an association between an agent and a putative adverse health effect, widely accepted criteria (e.g. Hill's Criteria) have been established for determining whether the evidence is weak, moderate or compelling (e.g. Tomatis, 1990). Narrative criteria may be inherently subjective, and therefore difficult to reproduce. To the extent that the criteria can be evaluated objectively, however, different assessors using the same information should be able to independently reproduce a determination of whether the criteria have been satisfied. For example, the weight of evidence for causality is stronger if detection of the association has been independently reported from multiple sources, if the strength of association is correlated with the level of exposure to the agent, or changes in the putative causative agent precede changes in the observed effect. Determining whether such criteria are satisfied is evidence-based. If the results of a qualitative assessment are invariant to an accumulation of evidence regarding an association or, alternatively, to contradictory evidence, then the assessment is insensitive to the established criteria for evaluating causality. In a qualitative hazard characterization, an assessment based solely on the criteria of acute health outcomes could be insensitive to information regarding known chronic sequelae. Alternatively, a qualitative hazard characterization may be highly sensitive to weak evidence regarding chronic sequelae associated with an opportunistic pathogen that rarely causes acute illness. If a qualitative assessment finds that a pathogen poses a negligible risk based on the assumption that the pathogen does not grow under certain environmental conditions, and new information indicates that the pathogen is capable of growing under these conditions, then the sensitivity of the findings of the risk assessment to this new information may depend on prespecified criteria, e.g. Have the results been independently reproduced? Have the methods been exposed to peer review? At a minimum, the scientific basis and criteria for characterization of a qualitative risk assessment needs to be sufficiently transparent to permit assessment of the impact of new information or plausible alternative assumptions on the findings.

6.3.2 Sensitivity analysis in quantitative risk assessment

There are several approaches to sensitivity analysis. Saltelli, Chan and Scott (2000) provide a thorough exploration of the topic, summarized below.

Exploratory methods

Exploratory methods for sensitivity analysis are typically applied in an *ad hoc* manner, but can be of central importance to the assessment of key sources of uncertainty in an analysis. Some

key sources of uncertainty in an assessment include qualitative features, such as the conceptual representation of the system under study, structure of the model, level of detail of the model, validation, extrapolation, resolution, boundaries and scenarios. It is not uncommon, for example, for the uncertainty about the true model form to be of much greater importance than the uncertainty associated with any model input for a given statistical model. An assessment of sensitivity of an analysis to changes in assumptions would not be complete unless consideration was given as to whether the scenario underlying the analysis is well specified. Methods for dealing with uncertainty regarding qualitative features of the analysis typically involve comparison of results under different structural assumptions. For example, a method for assessing the importance of different exposure pathways is to estimate the exposure associated with each pathway and to determine whether total exposures are dominated by only a few critical pathways. Similarly, if there is uncertainty regarding model structure, a common approach is to compare predictions based upon different models, each of which may have a different theoretical and mathematical formulation.

Statistical methods

Examples of statistical sensitivity analysis methods (also referred to as variance-based methods) include regression analysis, ANOVA, response surface methods, Fourier amplitude sensitivity test (FAST), mutual information index (MII), and classification and regression trees (CART) (Frey and Patil, 2002). Most of these methods are applied in conjunction with or after a Monte Carlo analysis. Regression analysis, ANOVA, FAST and MII provide quantitative measures of the sensitivity for each input. Regression analysis requires the assumption of a model form.

Graphical methods

Graphical methods represent sensitivity typically in the form of graphs, such as scatter plots and spider plots. The results of other sensitivity analysis methods (e.g. rank order correlation) also may be summarized graphically (e.g. by tornado charts). These methods can be used as a screening method before further analysis of a model, or to represent complex dependencies between inputs and outputs (For example, see McCamly and Rudel, 1995). For example, complex dependencies could include thresholds or non-linearities that might not be appropriately captured by other techniques.

Evaluation of sensitivity analysis methods

Each sensitivity analysis method provides different information regarding sensitivities of the inputs such as the joint effect of inputs versus individual effects, small perturbations of inputs versus the effect of a range of variation, or apportionment of variance versus mutual information. Because agreement among multiple methods implies robust findings, two or more different types of sensitivity methods might be applied where practicable, in order to compare the results of each method and draw conclusions about the robustness of rank ordering of key inputs. Non-parametric methods (e.g. Spearman's rank correlation) are applicable to monotonic, non-linear models. Vose (2000) recommends the use of spider plots to illustrate the effect of individual input variables on the uncertainty of the model output.

6.4 Uncertainty analysis

Uncertainty analysis evaluates the range and likelihood of model predictions. In the context of quality assurance, uncertainty analysis is a useful tool for characterizing the precision of model predictions.

In combination with sensitivity analysis, uncertainty analysis can also be used to evaluate the importance of model input uncertainties in terms of their relative contributions to uncertainty in the model outputs (Morgan and Henrion, 1990). There are a variety of methods for estimating uncertainty in a model output based upon uncertainty in model inputs. The choice of method depends on what information is of most interest, the functional form of the model, and, to some extent, the number of inputs for which uncertainty is characterized.

Methods typically applied include Monte Carlo simulation for generating samples from distributions assigned to each input. Sensitivity analysis methods such as regression and ANOVA can be used in combination with Monte Carlo simulation to identify model inputs that contribute most to uncertainty in model predictions. Helton and Davis (2002) provide an extensive literature review of methods for sensitivity analysis used in combination with sampling methods.

6.5 Model verification

Model verification is achieved by auditing the model to ensure that it operates as intended by the developer(s). Model verification should precede model validation. This process includes validation of the software code used to implement the model. Verification requires thorough documentation and transparency in the data, methods, assumptions and tools used, so that the model is independently reproducible. A well organized model structure facilitates the verification audit.

There are several major elements in model verification:

- Assess the correctness of the model formulation. For example, are the analytical equations correctly derived and free of error?
- Is the computerized version of the analytical model correctly implemented?
- Are the inputs correctly specified?
- Do the units of measurement propagate correctly through the model?
- Is the model internally consistent? For example, if an assumption is made in one part of the model, is it consistently applied throughout the model? Is there consistency within the model between the intermediate outputs and inputs?

It may be difficult in some cases to quantitatively verify computer code, especially for large models that are developed in a short time. However, the verification of computer code will be facilitated if good software engineering practices are followed, including clear specification of databases, development of a software structure design prior to coding, version control, clear specification of interfaces between components of a model, and good communication among project teams when different individuals are developing different components of a model. Model documentation and peer review are critical aspects of the verification process.

6.6 Model anchoring

Anchoring is a technique in which the model is adjusted or calibrated to be more compatible with observed data. For example, model parameters may be adjusted to achieve agreement between model predictions and observed data. Anchoring is a generally accepted practice in health risk assessment and environmental modelling, and has been employed in one fashion or another in risk assessments in the United States of America on *Salmonella* Enteritidis in eggs, *Listeria monocytogenes* in ready-to-eat foods, *Escherichia coli* O157:H7 in ground beef, and for an international risk assessment on *Vibrio vulnificus* in oysters (FAO/WHO, 2005). Data from outbreaks could be considered as the ultimate 'anchor' for dose-response models and are also an important way to validate risk assessments. There is a trade-off, however, because anchoring compromises the ability to validate the model output through comparison with the observed data in situations without sufficient data to support both. In general, anchoring approaches that weight model inputs in proportion to their likelihood in light of the observed data are superior to using simple adjustment factors or censoring input values that are incompatible with the observed data (National Academy of Sciences, 2002).

Whatever the anchoring approach, considerable care must be taken to ensure that the adjustment procedure is well reasoned and transparent. If the model is to be both anchored and validated (using a withheld portion of the independent data), then anchoring should precede model validation.

6.7 Model validation

A judgement needs to be made as to whether the risk assessment model response is reasonable. Stated less formally, model validation procedures are aimed at answering the following types of questions: (1) does the model make sense?; (2) does the model respond in an appropriate manner to changes in input assumptions; and (3) do predictions respond in an appropriate manner to changes in the structure of the analysis. This is also referred to by some as a 'reality check', 'laugh test' or 'confidence building'.

Model validation is highly dependent on the risk-management question, and the degree of validation required should be proportionate to the stakes of the decision. FAO/WHO (2003) defines model validation as demonstrating the accuracy of the model for a specified use and refers to different aspects of model validation. Conceptual validation concerns the question of whether the model accurately represents the system under study. Validation of algorithms concerns the translation of model concepts into mathematical formulae. Validation of software code concerns the implementation of mathematical formulae in computer language (see Section 6.5 on model verification). Functional validation concerns checking the model with independently obtained observations. Even if independent data are unavailable, a portion of the data may be withheld during model development to permit assessment of the model using the withheld data. When few data are available, however, the loss of information for model development may outweigh the benefit of withholding data for model evaluation.

Close agreement between an initial risk-modelling effort and independent validation data would be fortuitous. Agreement between the model output and validation data may be coincidental, however, and would not necessarily indicate that all of the intermediate models components are accurate. Typically, model development and refinement is an iterative process. Whether model validation or anchoring is considered, the credibility of the model may be strengthened by having multiple points at which the model can be compared to observed data. In general, the scientific credibility of a model is strengthened if consistent results are derived

from different relevant data sources (labs, regions) or types (observational or experimental), or a combination. The required degree of relevance and consistency is a context-specific judgement. The tolerance for inconsistent answers depends on what constitutes an 'important' difference with respect to changes in model results. In the risk assessment context, an important difference in model results is one that would significantly modify the risk management decision under the relevant decisional criteria.

There are situations in which it may be difficult or practically impossible to completely validate a model. For example, because risk assessment models are often attempting to predict low probability events, it can be difficult to obtain an independent data set of sufficient sample size to make statistically significant comparisons of predictions versus observations. However, even in such situations, it may be possible to validate components of the model. For example, it may be possible to validate portions of the model that deal with a particular exposure pathway by making measurements of contaminant levels in specific foods.

In many cases, there may be insufficient or no independent data with which to compare model predictions. In these situations, alternatives to validation include:

- screening procedures to identify the most important model inputs and pathways;
- sensitivity analysis to identify the most important inputs or groups of inputs;
- uncertainty analysis to evaluate the effect of uncertainty in model inputs with respect to predictions;
- comparison among predictions of different models; and
- evaluation of sensitivity of results to different assumptions regarding scenarios, model boundaries, model resolution and level of detail.

While none of these techniques provides a direct validation of the model, each of these techniques provides insight into the sensitivity of the model predictions to key assumptions regarding the analysis. The response of the predictions to these procedures can be evaluated with respect to prior expectations, comparison with analogous systems, and theoretical justifications.

6.8 Comparison with epidemiological data

In order to make a valid comparison with a foodborne pathogen risk estimate, at least three factors need to be considered in deriving an epidemiological estimate from human surveillance data (Powell, Ebel and Schlosser, 2001).

- *Cluster-weighted rate of illness*
  If the risk assessment estimates the incidence of illness at the national level, the epidemiological estimate will need to extrapolate the rate of illness beyond the surveillance area to permit comparison at the national level. In this case, the raw reported rate in each surveillance area may be weighted by the population of the region represented by the area (e.g. state population size) to obtain a weighted average rate of illness (e.g. cases per 100 000 in the national population). If multiple years of surveillance data are available, then the data can be used to characterize year-to-year variability in the rate of illness.
- *Adjust surveillance data to account for under-reporting*
  Estimating the actual incidence of illness requires adjustment for recognized sources of under-

reporting in human surveillance data. For example, some ill persons do not seek medical care, physicians do not obtain stool specimens from all patients, laboratories do not culture all stool samples for the pathogen of concern, and some proportion of the lab results are false negatives. If estimates are available on the proportion of cases at each step in the reporting process, the negative binomial distribution can be used in sequential fashion to estimate the number of cases missed at each step. In some cases, the proportions may be dependent on the nature or severity of symptoms. For example, a person with bloody diarrhoea may be more likely to seek medical care than one with non-bloody diarrhoea. In this case, the proportion of cases with different levels of symptoms must be estimated prior to accounting for the number of cases missed at each step, and the adjusted symptom-specific estimates are summed to estimate the total number of cases. In general, the degree of under-reporting tends to be substantial. The degree of under-reporting also varies among countries and between regions within countries.

- *Etiological fraction attributable to food product(s)*
  The etiological fraction refers to the proportion of cases attributable to an exposure pathway or a specific food product. If the scope of the risk assessment is limited to a particular food product, then the proportion of cases due to other exposure pathways (e.g. other foods, drinking water) needs to be subtracted from the overall estimate of illness obtained from the human surveillance data. In general, empirical data on the etiological fraction are scarce. It may be possible, however, to specify a range of uncertainty on the basis of expert judgement.

If observed epidemiological data are used to generate the dose-response model or to anchor the model, then these data are unavailable for independent model validation. If sufficient epidemiological data are available, however, a portion of the data may be withheld for the purposes of model validation.

6.9 Extrapolation and robustness

Model robustness refers to the performance of the model when its assumptions are violated. In this context, assumptions include model form and model inputs. Extrapolating model results to other settings may involve many forms of extrapolation: from the present to the future, from one geographical region to another, from one microorganism to another, from animals to humans, from human clinical trial subjects to the general population, from one human population to another, from the available data to values beyond the observed range of the data, from controlled experimental settings to operational environments, and so on. Some extrapolations can be made with relative confidence, while others require a leap of faith. Some degree of extrapolation is inevitable if risk assessment attempts to inform risk-management decisions, since the demands of risk management tend to outstrip the supply of relevant science. The importance of various forms of extrapolation made in risk assessment needs to be considered and, to the extent feasible and relevant to the decision at hand, characterized in a clear manner, either quantitatively or qualitatively.

Extrapolation is explicit when the selected values of model inputs are outside the range of values used to calibrate or validate the model, or both. However, there can also be hidden extrapolation. A hidden extrapolation occurs for a combination of values of each model input such that these values are enclosed by ranges used for calibration and validation, but for which that specific combination was not included or approximated during calibration or validation. Thus, simple range checks on each input will not guarantee that a hidden extrapolation cannot occur. Hidden extrapolation would typically be more of a problem for a system in which there are highly sensitive interactions among inputs.

A model that is calibrated to a narrow range of values for each input may not be robust when applied to sensitivity or uncertainty analysis. The use of ranges or distributions rather than point estimates could lead to hidden or explicit extrapolations of the model. In addition, situations can arise in which a joint set of model inputs are sampled in a Monte Carlo analysis for singularity points of a model, leading to problems such as division by zero or unbounded results. Such problems can often be traced to simplifying assumptions in model development, mis-specification of distributions for model inputs, or computer software limitations. Problems such as these can arise in practice, particularly when working with a model or computer code that someone else developed and for which documentation may be inadequate.

A model is considered to be robust if it responds in a reasonable manner to variation in input values, while at the same time not being easily subject to singularity points or other structural issues that lead to substantial magnification of errors in input values, whether because of uncertainty or user error. Moreover, a model that is based on sound theory might be used with more confidence compared with a purely empirical model that is essentially a curve fit to a calibration database. There is a distinction between the robustness of a risk assessment model and the robustness of a risk management decision. From an analytical perspective, a risk management decision is robust if the decision is beneficial over a reasonably wide range of possible future outcomes regarding uncertainties associated with the many factors that influence the decision. One such source of uncertainty would typically include the risk assessment model itself.

6.10 Credibility of the risk assessment

Documentation, validation, and review are necessary criteria for the credibility of a risk assessment. None of these criteria is sufficient by itself, however, as credibility depends on all three criteria being satisfied in a manner that is proportionate to the stakes of the decision.

6.10.1 Risk assessment documentation

At a minimum, risk assessment documentation must enable the analysis to be independently reproduced. The principle of transparency also requires that the source or basis for model inputs and assumptions be clearly stated (e.g. by references to scientific literature, evaluation criteria or expert judgement). The expectation for risk assessment documentation should be reasonable, however, because in some cases, assumptions may be based on common knowledge or generally accepted practices in the field. For example, the lognormal distribution is commonly assumed for modelling variables that are the product of several other variables. Because risk assessments are difficult to fully validate, and because such assessment are used to inform public decision-making at various levels, including local, national, and international, pertaining to public health, it is critically important that the information used for the assessment, including the model, be accessible for review by experts and the lay public. Ideally, subject to resource constraints, the following information should be included in documentation of a risk assessment:

- data or references to data sources;
- scenario, including the temporal and spatial aspects of the exposure scenarios, the specific hazards addressed, the specified pathogens included, exposed populations and exposure pathways;
- analytical model used for analysis, including the theoretical or empirical basis;

- discussion and comparison of alternative model formulations, and justification for choices made regarding model structure;
- assumptions regarding values assigned to model inputs, including point-estimates, ranges and distributions;
- model verification, including assessment of results from sensitivity and uncertainty analysis;
- model anchoring (calibration);
- model validation; and
- computer implementation of the analytical model, including software design.

6.10.2 Peer review

FAO/WHO (2003) notes that the credibility of risk assessment results can be improved by the process used to develop the results. Peer and public review of risk assessment results are an essential part of the process, but each type of review generates distinct and sometimes conflicting demands that should be addressed on their own terms. There is also a distinction between the scientific credibility of a risk assessment and the credibility of risk management decisions. Public review is addressed in Section 8.5.

Morgan and Henrion (1990) identify exposure to peer review as a basic tenet of good policy analysis. The focus of a scientific peer review is highly dependent on the risk management question that the risk assessment is intended to inform. Without reference to a well-defined and specific risk management question, peer review of a risk assessment may fail to focus on the particular uncertainties that are most likely to influence the risk management decision. For example, if the risk management question is "What is the likelihood that a specific pathogen occurs in a particular food production process?" then data gaps and other uncertainties regarding post-production processes are irrelevant to the decision. Peer review comments regarding the scope of the risk assessment, while potentially useful for future risk assessments, are not relevant to the adequacy of the risk assessment under review to inform the risk management decision for which it was intended. If a risk assessment has multiple objectives, peer review may help to identify which objectives an assessment satisfies, since an assessment that is adequate to inform one decision may be insufficient to support another. For a complex risk assessment, a thorough review can be difficult and time consuming, even if the documentation is adequate. In the case of large, complex risk assessments, a thorough review may require a multidisciplinary team and a significant budget. Therefore, the substantive and procedural benefits of peer review should be balanced by time and resource considerations. The level and extent of review should be proportionate to the stakes of the decision, taking into consideration the need for immediate action in the event of bona fide public health emergencies.