ACADEMIE UNIVERSITAIRE WALLONIE-EUROPE
UNIVERSITE DE LIEGE
FACULTE DE MEDECINE VETERINAIRE
DEPARTEMENT DES SCIENCES DES DENREES ALIMENTAIRES
SERVICE DE MICROBIOLOGIE

Quality Assurance in Quantitative Microbial Risk Assessment:

Application of methods to a model for *Salmonella* in pork

L'Assurance Qualité appliquée à l'appréciation quantitative des risques microbiologiques :

Etude d'un modèle de contamination par *Salmonella* dans la filière porcine

IDESBALD BOONE

**THESE PRESENTEE EN VUE DE L'OBTENTION DU GRADE DE
DOCTEUR EN SCIENCES VETERINAIRES
ORIENTATION SANTE ET PRODUCTION ANIMALES**

ANNEE ACADEMIQUE 2010-2011

ACKNOWLEDGEMENTS

This study has been carried out thanks to the financial support provided by the Belgian Federal Public Service of Health, Food Chain Safety, and Environment research programme (R-04/003-METZOON-project) ‘Development of a Methodology for Quantitative Assessment of Zoonotic Risks in Belgium Applied to the “*Salmonella* in Pork” Model’.

There are many people without whom this thesis would not have been possible. Many thanks to Dr. Koen Mintiens, who as the METZOON project leader gave me the opportunity to work out this doctoral thesis at the Coordination Centre for Veterinary Diagnostics of the Veterinary and Agrochemical Research Centre (CODA-CERVA). I also thank him and the general director dr. Pierre Kerkhofs for allowing me to follow courses in applied statistics at the University of Hasselt. I thank my co-promoter prof. Georges Daube, who accepted me as his Ph.D. student at the Department of Food Microbiology of the Veterinary Faculty at Liège University, and who supported me all along this project.

I am deeply indebted to my promoter dr. Yves Van der Stede, the current head of the Unit Coordination for Veterinary Diagnostics – Epidemiology & Risk Assessment (CVD-ERA), for supervising me and for his patience, encouragements and positive criticism.

I am very grateful to dr. Jeroen van der Sluijs (Utrecht University), whose work has greatly inspired this thesis and who always very effectively answered my questions. He also tipped me on various symposia and workshops related to uncertainty management and knowledge assessment methodologies, which led to meeting dr. Silvio Funtowicz and dr. Matthieu Craye (JRC, Ispra). I also thank prof. Cooke and dr. Dorota Kurowicka (TU Delft) for their help and remarks on the elicitation and analysis of structured expert judgement.

I would like to thank prof. Nicolas Korsak, dr. Martine Laitat, prof. A. Linden, prof. Jacques Mainil, prof. Claude Saegerman and especially dr. Lis Alban and dr. Jeroen van der Sluijs for reading the manuscript and being part of my jury.

I thank the colleagues from the METZOON-project for their collaboration in the NUSAP workshops as well as all the experts who participated in the expert judgement study, my co-authors (dr. Kaatje Bollaerts, dr. Winy Messens, prof. Marc Aerts, prof. Jeroen Dewulf, prof. Dominiek Maes, David Vose), all the colleagues at the Coordination Centre for Veterinary Diagnostics for the good working atmosphere: François Dekeyser, Cécile Herr, dr. Estelle Méroc and dr. Sarah Welby. My special thanks are due to dr. Marc Dispas for his critical remarks and his practical advice in preparing presentations, Conny Rogiers and Chantal Deknop from the library for their help with my documentation, and Michelle Dehantschutter for her support in the administration and during preparations of workshops.

I thank my family and friends for their continuous support, especially Timothy Vanagt, Hélène Van Ginderachter and Myriam Boone for reviewing the final version.

Last but not least, I would like to thank my parents who always encouraged me.

TABLE OF CONTENTS

Acknowledgements i

List of tables vii

List of figures ix

List of abbreviations and concepts xi

Chapter 1. General introduction 1

1.1. The METZOON project 4
1.2. Structure of the thesis 6

Chapter 2. Quality assurance in quantitative microbial risk assessment: a review 9

2.1. Principles of food safety risk analysis 10
2.2. Quantitative microbial risk assessment 12
2.3. Guidelines for QMRA 12
2.4. Examples of QMRAs 15
2.5. What is quality assurance? 19
    2.5.1. Definitions 19
    2.5.2. Quality Assurance in QMRA: Why? 19
2.6. Methods for the implementation of quality assurance in QMRA 22
    2.6.1. Uncertainty assessment 22
    2.6.2. Systematic review 24
    2.6.3. NUSAP/Pedigree approach for the evaluation of data quality 24
    2.6.4. Expert elicitation 26
    2.6.5. Critical evaluation of assumptions 27
    2.6.6. Sensitivity analysis 31
    2.6.7. What-if scenario analysis 32
    2.6.8. Checklist approach 33
    2.6.9. Peer review 34
    2.6.10. Model verification 36
    2.6.11. Model validation 36

2.6.12. Multiple model comparison 36
2.6.13. Quality of documentation and risk communication 37
2.7. The METZOON model 39

Chapter 3. Aims of the thesis 43

Chapter 4. Structured expert judgement in quantitative microbial risk assessment: evaluation of weighting schemes to combine expert opinions 45

4.1. Introduction 46
4.2. Materials and methods 47
4.2.1. Elicitation protocol 47
4.2.2. Calibration and information scores 52
4.2.3. Decision makers 52
4.3. Results 54
4.3.1. Participation of experts 54
4.3.2. Performance of experts 58
4.3.3. Robustness analysis 59
4.3.4. Combined distributions 60
4.3.5. Experts' self-rating of expertise 63
4.4. Discussion 64
4.5. Concluding remarks 66

Chapter 5. The NUSAP system: A tool for evaluating the data quality in quantitative microbial risk assessment 67

5.1. Introduction 68
5.2. Material and methods 69
5.2.1. Parameters for the Belgian QMRA model 69
5.2.2. NUSAP pedigree matrix for the evaluation of data quality 71
5.2.3. Analysis of pedigree scores 74
5.3. Results 77
5.3.1. Parameter ID Cards 77
5.3.2. Analysis of pedigree scores 78
5.3.3. Communication of pedigree scores 81
5.4. Discussion 82

5.5. Concluding remarks 86

Chapter 6. The NUSAP system: a method to evaluate the quality of assumptions in quantitative microbial risk assessment........ 87

6.1. Introduction 88
6.2. Methodology 89
6.2.1. NUSAP pedigree matrix for the evaluation of assumptions 89
6.2.2. Assumptions in the Belgian QMRA model........ 92
6.2.3. Analysis and reporting of the pedigree scores........ 94
6.2.4. Statistical analysis 95
6.3. Results 95
6.3.1. Ranking of key assumptions 95
6.3.2. Pedigree scores 96
6.4. Discussion 101
6.4.1. Ranking of key assumptions 101
6.4.2. Evaluation of the NUSAP/Pedigree methodology 102
6.4.3. Extending the workshop 104

Chapter 7. Checklists as a quality assurance tool for the evaluation of QMRA models: application to the METZOON model 105

7.1. Introduction 106
7.2. Material and methods 106
7.3. Results: quality audit of the METZOON model by means of the Paisley (2007) checklist 108
7.3.1. General questions (Purpose, Identification of actors, Output) 108
7.3.2. Uncertainty assessment 109
7.3.3. Knowledge base 109
7.3.4. Modelling structure, modelling practices and quality assurance methods..... 110
7.3.5. Risk communication........ 111
7.3.6. Conclusion and overall evaluation of the model 115

Chapter 8. General discussion 117

8.1. Expert opinion in QMRA 118
8.2. NUSAP/Pedigree assessment for QMRA 120

8.3. Checklist approach 123
8.4. Risk communication 124

Chapter 9. Recommendations for an integrated quality assurance for QMRA 125

References 129

Summary 147

Résumé 151

LIST OF TABLES

Table 1: General principles of microbial risk assessment according to the Codex Alimentarius Commission 13
Table 2: Overview of available (non exhaustive) farm-to-fork quantitative microbial risk assessments, with indication of applied quality assurance methods 17
Table 3: Criteria determining the quality of a QMRA, with indication of methods to address the validity and utility criteria 21
Table 4: Pedigree matrix for the evaluation of data quality 25
Table 5: Pedigree matrix for reviewing the quality of assumptions 30
Table 6: Examples of checklists useful for the evaluation of quantitative microbial risk assessments 34
Table 7: Variables of interest elicited as part of an expert opinion workshop in 2007 for a quantitative microbial risk assessment on *Salmonella* in the pig production chain… 49
Table 8: Seed variables elicited as part of an expert opinion workshop in 2007 for a quantitative microbial risk assessment on *Salmonella* in the pig production chain ……..50
Table 9: Background of selected experts participating in an expert opinion workshop in 2007. 55
Table 10: Results of the performance assessment for 11 experts participating in an expert elicitation workshop in 2007. 56
Table 11: Experts' robustness analysis for the performance-based weighting schemes: the global weight and the item weight decision maker (DM) 57
Table 12: Inventory and characterization of the variables in the questionnaire for sources of information in QMRA 70
Table 13: Characterization of the input parameters as foreseen in the ID-card 70
Table 14: List of potential input parameters for the QMRA model 72
Table 15: NUSAP pedigree matrix used to score the parameter strength 74
Table 16: Summary statistics for the strength of each criterion and overall strengths within the exposure pathways 80
Table 17: Pedigree matrix for reviewing the quality of assumptions 91

Table 18: List of assumptions scored on quality by nine experts from the METZOON consortium, during a NUSAP workshop. 93
Table 19: Average pedigree scores (Average) and standard deviations (SD) of the assessed criteria: situational limitations, plausibility, choice space and agreement among peers for the 13 selected key-assumptions.. 97
Table 20: Model quality checklist... 107
Table 21: Quality audit of the METZOON model.. 113

LIST OF FIGURES

Figure 1: Number of identified human salmonellosis cases in Belgium from 1989 to 2009… 4

Figure 2: Framework of the risk analysis process adopted by the Codex Alimentarius Commission 11

Figure 3: Components of a microbial risk assessment according to the Codex Alimentarius Commision 14

Figure 4: Diagnostic diagram for the representation of the quality of data. It combines scores for input parameters obtained in the pedigree assessment with their sensitivity 26

Figure 5: Graphical representation of pedigree scores. a) Example of a kite diagram ; b) Example of a pedigree chart for the evaluation of assumptions 29

Figure 6: Schematic representation of the METZOON model 41

Figure 7: Self-rating of expertise-Box 51

Figure 8: Distributions for the 11 seed variables based on 11 experts participating in an expert elicitation workshop in 2007 61

Figure 9: The decision maker's (item weight DM, global weight DM, equal weight DM, user weight DM) and the best expert's distributions estimating the *Salmonella* concentration increase/decrease in $\log_{10}$ CFU at abattoir processing stages. 62

Figure 10: The decision makers' distributions (item weight DM, global weight DM, equal weight DM, user weight DM) and the best expert's distributions estimating the *Salmonella* prevalence at abattoir production stages. 62

Figure 11: The combined distributions of four variables of interest expressed by the 5th, 50th and 95th percentiles derived from the item weight (1), the global weight (2), the equal weight (3), the user weight (3) decision maker (DM) and comparison of the best expert's (5) uncertainty distribution. 63

Figure 12: Overall parameter strength (bottom diagram) and pedigree criterion strength (upper 4 diagrams) of 101 parameters distributed according to the production pathways 79

Figure 13: Illustration of pedigree scores of four parameters by means of kite diagrams). ... 82

Figure 14: Diagnostic diagram for the identification of weak links in the calculation chain of the METZOON model. 98

Figure 15: Kite diagrams with pedigree score results indicating the subjective nature (or value-ladenness) of the key assumptions in the METZOON model. 100
Figure 16: Pedigree charts of the key assumptions in the METZOON model. 101
Figure 17: Integrated quality assurance approach for QMRA 127

LIST OF ABBREVIATIONS AND CONCEPTS

DM	Decision-maker
EFSA	European Food Safety Authority
FASFC	Federal Agency for the Safety of the Food Chain
ILVO	Institute for Agricultural and Fisheries Research
METZOON	Methodology for Quantitative Assessment of Zoonotic Risks in Belgium
MS	Member State
NUSAP	Numeral, Unit, Spread, Assessment, Pedigree notational system for the evaluation of uncertainty in policy-relevant science
PDF	Probability density function
QMRA	Quantitative microbial risk assessment
QA	Quality Assurance
Risk assessor	Person/consortium carrying out the risk assessment
SA	Sensitivity Analysis
Stakeholder	Body/organisation which is involved/can be involved in the risk assessment or who can be affected by the policy process resulting from the implementation of the model (e.g. universities, research institutes, public bodies, consumer’s organisation, meat processing industry, etc.)

CHAPTER 1:

GENERAL INTRODUCTION

CHAPTER 1. GENERAL INTRODUCTION

Foodborne diseases present a major threat to public health, and cause serious (socio)-economic losses worldwide. In developing countries, an estimated 2.2 million people, mainly children die each year as a result of diarrhoea, mostly caused by contaminated food or water (2006). In the USA, each year around 76 million cases of foodborne illnesses occur, among which more than 300 000 persons were hospitalized and 5000 person died. The health-related costs from foodborne illnesses are estimated at US$152 billion per year (Anonymous, 2010). In Europe, campylobacteriosis and salmonellosis were the two most common reported foodborne illnesses in humans, accounting for 131 468 and 190 566 confirmed cases, respectively, as reported by the Member States in 2008 (EFSA, 2009a) and it is estimated that thousands people die each year in the EU as a result of salmonellosis (EFSA, 2010). However, due to underreporting the actual incidences are likely to be much higher.

In its White Paper on Food Safety (Anonymous, 2000a), the EU has developed its policy towards food safety and emphasised on the fundamental role of risk analysis with its interrelated components risk assessment, risk management and risk communication. The general principles and requirements of the EU general food law were laid down in the EC Regulation N° 178/2002 (Anonymous, 2002). It stresses on adopting an integrated farm-to-fork approach, encompassing all stages of the food production chain: primary production, processing, transport, distribution through the end consumer. This regulation also established the European Food Safety Authority (EFSA), as a scientific point of reference which undertakes risk assessment in an independent, objective and transparent manner. Further, EFSA takes the role of independent source of advice and risk communication in order to improve consumer confidence, which had been greatly damaged as a result of the (mis)management of the BSE crisis.

Quantitative microbial risk assessments (QMRA) is a method that can be used to estimate the risk of pathogens along the food chain and aims to support management decisions for the reduction of food-safety risks. The degree of credibility that can be attached to risk assessment results depends largely on the quality and quantity of the data, the model structure and the assumptions taken. Adopting a quality assurance framework (QA) aims to improve the transparency in the QMRA process and is helpful to evaluate if a QMRA achieved its objectives. This QA should enhance the confidence of decision-makers in the results of the QMRA. Possibilities to apply quality assurance methods in QMRA were explored within a Belgian research project (the METZOON project) aiming to assess the risk of human salmonellosis through the consumption of fresh minced pork meat.

1.1. THE METZOON PROJECT

In response to the EU legislation on the monitoring and control of zoonoses and zoonotic agents (Directive 2003/99/EC and Regulation (EC) N°2160/2003), Members States have to take measures to detect and control zoonoses and zoonotic agents along the food chain (Anonymous, 2003a, b). According to Mead et al. (1999), 95% of the salmonellosis cases have a foodborne origin. In Belgium, non-typhoidal salmonellosis and campylobacteriosis are the most frequently reported foodborne illnesses. In 2009, 3208 cases of human salmonellosis were diagnosed and reported to the National Reference Centre for *Salmonella* and *Shigella* (SIPH, 2010). The evolution of the human salmonellosis cases between 1989 and 2009, as well as of the two most common serotypes *Salmonella enterica* serovar Enteritidis and *Salmonella enterica* serovar Typhimurium is shown in Figure 1. After a peak in 1999, a decrease was observed in the number of reported cases. This decline is mainly attributed to the decrease in *S.* Enteritidis cases, while the number of *S.* Typhimurium cases remained almost stable since 2000, and became the predominant serovar since 2006.

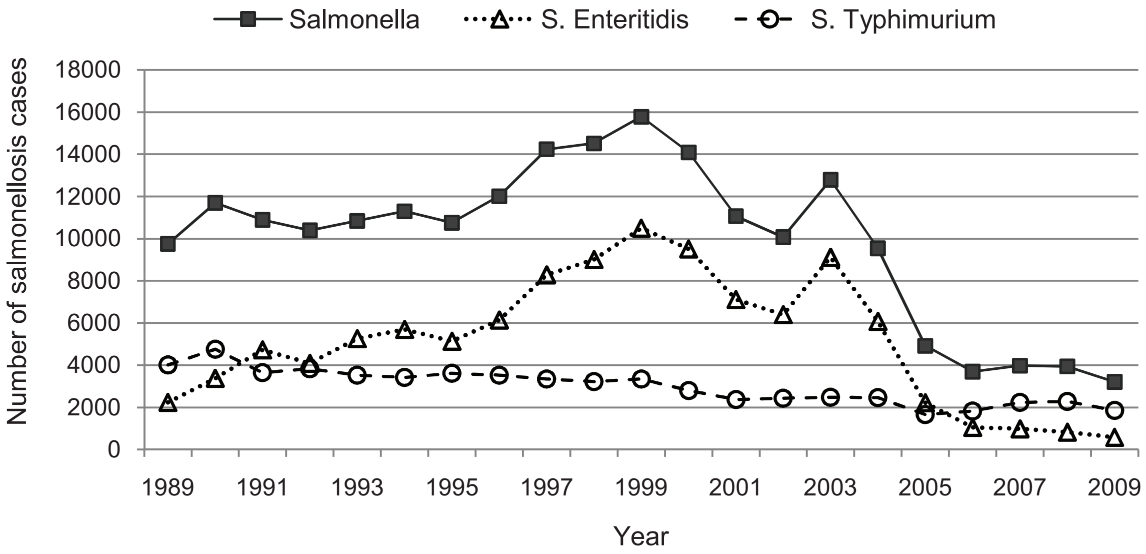

Figure 1: Number of identified human salmonellosis cases in Belgium from 1989 to 2009 (Source: SIPH, 2010).

The most common sources of foodborne salmonellosis are eggs, poultry and pork. Since *S.* Typhimurium is predominantly isolated in slaughter pigs and in pork meat, as well as in human isolates, it is suggested that around 10 to 20% of human salmonellosis cases can be attributed to the pig reservoir (EFSA, 2010). According to the baseline study in slaughter pigs carried out in 2006-2007 in 25 Member States, the EU weighted prevalence of *Salmonella* in lymph nodes of slaughter pigs was 10.3% (EFSA, 2009b). In this study, the prevalence of Belgian slaughter pigs was with 13.9% above the EU average, and compared to the other MSs located in the group with an intermediate prevalence. At processing and cutting plants, Belgium reported 5.7% of *Salmonella* infected samples which was the highest proportion of reported positive samples in the EU in 2008 (EFSA, 2009a).

The Belgian Federal Public Service of Health, Food Chain Safety, and Environment (FPS) therefore decided to fund a research project aiming to assess the risk of human salmonellosis due to the consumption of fresh minced pork meat.

This project denoted as the METZOON-project was carried out from April 2005 to March 2008 by a consortium of six partners. The METZOON-partners were the Veterinary and Agrochemical Research Centre (VAR), the Schools of Veterinary Medicine of both Liège and Ghent Universities, the Institute for Agricultural and Fisheries Research (ILVO), the Federal Scientific Institute for Public Health and the Center for Statistics of Hasselt University.
The deliverables of the METZOON project were:

1. The development of a quantitative microbial risk assessment (QMRA) to assess the risk of human salmonellosis through the consumption of fresh minced pork meat. This included the development of a modular risk model (the METZOON-model, see section 2.7.) which covered the pork production from farm to fork, and allowing to test mitigation strategies by means of what-if scenarios;
2. The identification of data gaps and collecting new data to be used in the QMRA;
3. The evaluation and development or refinement of statistical and mathematical methodology;
4. The development of a methodology to describe and assess the quality of the information and assumptions used to build the QMRA model.

The METZOON project resulted in several peer-reviewed publications and doctoral theses. For the results of the Deliverables 1, the reader is referred to Bollaerts et al. (2009; 2010) and Delhalle et al. (2009b) for Deliverable 2, Delhalle et al. (2008; 2009a) and for Deliverable 3, Bollaerts et al. (2008; 2009). An overview of information related to the METZOON project can be found at http://www.metzoon.be.

The topics in the present thesis are related to Deliverable 4 “Development of a methodology to describe and assess the quality of the information and assumptions used to build the QMRA model”.

1.2. STRUCTURE OF THE THESIS

In chapter 2, the role of quantitative microbial risk assessment in food safety is outlined and an overview is given of quality assurance methods and tools of interest in QMRA.

QMRA integrates different types of knowledge, such as survey data, research data, experimental data, modelled data and expert opinion. In chapters 4 and 5, QA methods were used to evaluate the quality of input knowledge of the METZOON model. More specifically, chapter 4 deals with the evaluation of expert judgement during QMRA. Hereto, a structured expert judgement approach using Cooke’s classical model (Cooke, 1991) was chosen to obtain missing input parameters for the METZOON model. To quantify uncertain input parameters, expert judgements were elicited as subjective probability density functions (PDF). The performance of experts as probability assessors was measured by the experts’ ability to correctly and precisely provide estimates for a set of seed variables (variables from the experts’ area of expertise for which the true values were known to the analyst). Subsequently different weighting schemes were applied using Cooke’s classical model in order to obtain combined PDFs as a weighted linear combination of the expert’s individual PDFs, which can be used as input for the QMRA. The importance of the quality of expert judgement in QMRA is critically discussed.

Chapter 5 was related to an application of the Numeral Unit Spread Assessment Pedigree (NUSAP) system for the evaluation of the quality of potential input parameters in the METZOON model. This experimental study corresponds to Boone et al. (2009b). The NUSAP system was used to provide an objective basis in the selection of input parameters and to promote a structured debate about the relevance of the data in the QMRA model.

QMRAs inevitably include assumptions, and some can be determinant for the quality of a QMRA. Chapter 6 is dedicated to the assessment of the value-ladenness of assumptions. In particular, the protocol described by Kloprogge et al. (2010) and the NUSAP system were used to evaluate assumptions in the METZOON model. Model assumptions were first identified, the most important assumptions selected by a panel of experts, and the value-ladenness of assumptions of these key assumptions assessed by a set of criteria. The impact of assumptions on the output of the QMRA was also evaluated.

In chapter 7, a quality audit checklist of QMRA models was applied to evaluate the quality of the METZOON model. The checklist allowed to give an overview of the status of the model and to identify its strengths and possible pitfalls. This information can help to fine-tune the model and thereby improve the decision-making process. A general discussion is provided in chapter 8, followed by recommendations for an integrated quality assurance in QMRA in chapter 9.

CHAPTER 2:

QUALITY ASSURANCE

IN QUANTITATIVE MICROBIAL RISK ASSESSMENT:

A REVIEW

Modified from: Boone, I., Van der Stede, Y., Aerts, M., Mintiens, K., Daube, G. (2010). Quantitative microbial risk assessment: methods and quality assurance. Vlaams Diergeneeskundig Tijdschrift, 76: 367-380.

2.1. PRINCIPLES OF FOOD SAFETY RISK ANALYSIS

In food safety, risk is defined as the probability that an adverse health effect will occur, and the severity of that effect resulting from one or more biological, chemical or physical hazards present in food. Food safety risk analysis is mostly carried out by international, national or regional food safety authorities. The Codex Alimentarius has been designed as one of the official organizations to develop risk analysis as a framework to deal with risks in food safety (Codex Alimentarius Commission, 2010). Risk analysis is defined by the Codex Alimentarius Commission (1999) as the process consisting of three interrelated components: risk assessment, risk management and risk communication (Figure 2).

Risk assessment is considered as the scientific component of risk analysis and can be defined as the characterization of the potential adverse health effects associated with exposure to hazards over a specific time period (FAO/WHO, 2006). Beside hard scientific facts, risk assessment may also involve value judgements and choices (assumptions) that are not entirely scientific (Haas et al., 1999; FAO/WHO, 2006). **Risk management** is the process of weighing policy alternatives, in consultation with all interested parties, taking into account risk assessment and other factors relevant for the protection of consumers including values, social, economic, environmental, cultural, ethical and legal issues. **Risk communication** is the interactive exchange of information concerning risk among risk assessors, risk managers and stakeholders (industry, consumers etc). It includes the explanation of risk assessment findings and the basis for risk management decisions.

Quantitative microbial risk assessment is able to provide policy makers with a scientific basis for selecting from among the various appropriate intervention options, for determining the risk-based food safety targets, and for establishing the levels of protection between countries. It thereby plays an important role in international trade (Havelaar et al., 2004; Nauta and Havelaar, 2008).

According to the Codex Alimentarius Commission (2010) risk assessment can further be divided in i) hazard identification, ii) exposure assessment, iii) hazard characterization, and iv) risk characterization (Figure 3).

The **hazard identification** identifies whether known or potential health effects are associated with biological, chemical or physical agents when present in food.

The **exposure assessment** is the estimation of the likely intake of biological, chemical or physical agents via food and along the food-chain with the potential to cause an adverse health effect.

The **hazard characterization** is the description of the adverse health effects as a result of the intake of biological, chemical or physical agents in food. An important part of the hazard characterization - if data are available - is the elaboration of a dose-response model to estimate the amount of pathogens causing illness.

Finally, **risk characterization** is the process of integrating hazard identification, hazard characterization and exposure assessment to give an overall estimate including associated uncertainties, of the probability and severity of adverse health outcomes in a given population.

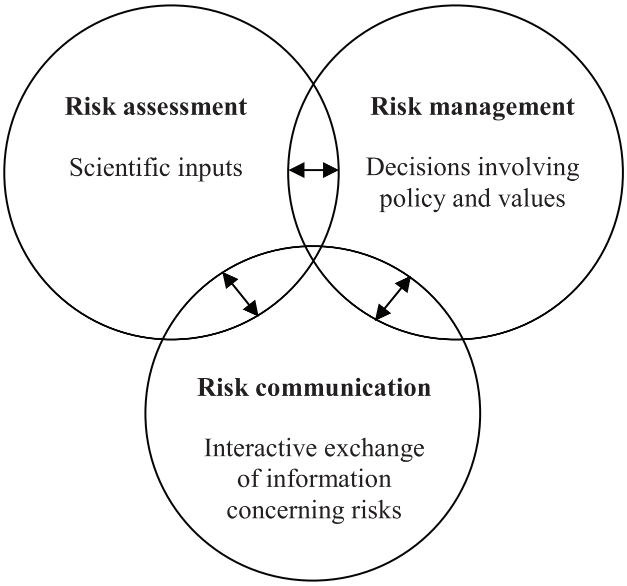

Figure 2: Framework of the risk analysis process adopted by the Codex Alimentarius Commission (Source: FAO/WHO, 2006).

A risk assessment is likely to be commissioned if the exposure pathways are complex, data on hazards and health impacts are incomplete, and the risk issue is of concern (FAO/WHO, 2006).

The discipline that deals with the risk of biological agents including microbes, viruses, parasites, toxins of biological origin, prions, etc. is denoted as microbial risk assessment. Historically, the methodology for microbial risk assessment (MRA) was derived from chemical risk assessment, but there is an essential difference between the two. QMRAs have to take into account that microbial agents - unlike chemical agents - can multiply and/or be inactivated or die within or on food products during the consecutive phases of the entire food chain and beyond. The highly variable dose-response relation between the biological agent and the adverse health effects, which is due to the specific characteristics of the micro-organisms involved and the existence of susceptible sub-populations within the target human and/or animal populations, makes the implementation of QMRAs a very challenging undertaking (Voysey and Brown, 2000).

Microbial risk assessment integrates information from various sources, including published and unpublished scientific studies, monitoring data, surveillance data and laboratory diagnostic data. The data can originate from disease outbreak investigations, food consumption surveys, national and international risk assessments, and so on. When data are lacking, assumptions are made and expert opinion can be used as a complementary source of information (FAO/WHO, 2006).

2.2. QUANTITATIVE MICROBIAL RISK ASSESSMENT

Microbial risk assessment includes both qualitative and quantitative microbial risk assessment. Quantitative microbial risk assessment (QMRA) is the approach that is most advanced in terms of complexity and resource requirements. QMRA is defined as a science-based process to *quantitatively* estimate the probability and severity of adverse health effects resulting from the exposure to micro-organisms in a given foodstuff, together with an indication of the attending uncertainties (Codex Alimentarius Commission, 1999; Haas et al., 1999; Lindqvist et al., 2002; Nauta et al., 2003; Walls, 2006).

As opposed to qualitative microbial risk assessments, which express outputs in descriptive terms, QMRAs require quantitative data in order to provide numerical expressions of risk, which allow for the quantification of uncertainty and variability. QMRAs can be classified as either deterministic or stochastic. In the deterministic approach the variables are represented by single-point estimates, whereas in the stochastic (or probabilistic) approach, probability distributions are used to describe variables. The stochastic approach is generally regarded to be the most optimal to adequately represent or mimic the real world, even though it is often complex and difficult to generate (FAO/WHO, 2006).

2.3. GUIDELINES FOR QMRA

Very often the QMRA process is subject to constraints due to poor data quality, the limited amount of time and resources, the assumptions made, the deficiencies in the model structure, and the interpretation of the results. In this regard, the Codex Alimentarius (1999), which stated general principles for microbial risk assessment (Table 1), can serve as a useful basis for a quality assurance framework.

Table 1: General principles of microbial risk assessment according to the Codex Alimentarius Commission (1999)

1	Scientific basis
2	Separation between risk assessment and risk management
3	Includes hazard identification, hazard characterization, exposure assessment and risk characterization
4	Purpose and the format of the output (risk estimate) clearly stated
5	Transparency
6	Description of constraints (costs, resources, time, assumptions) on the risk assessment, and description of consequences
7	Description of uncertainty, and location of the uncertainties in the risk assessment process
8	Data should allow determination of the uncertainty in the estimate. Data and data collection should be of sufficient quality and precision that uncertainty in the risk estimate is minimized
9	Consideration of the dynamics of microbial growth, survival, and death in foods and the complexity of the interaction between human and agent following consumption as well as the potential for further spread
10	Reassessment over time by comparison with independent data
11	Re-evaluation as new relevant information becomes available

A formal QMRA must be subdivided into four phases, namely hazard identification, exposure assessment, hazard characterization and risk characterization (Principle 3) (Figure 3). The FAO/WHO has developed guidelines for carrying out three segments of the microbial risk assessment process: the hazard characterization (FAO/WHO, 2003), the exposure assessment (FAO/WHO, 2008), and the risk characterization (FAO/WHO, 2009).

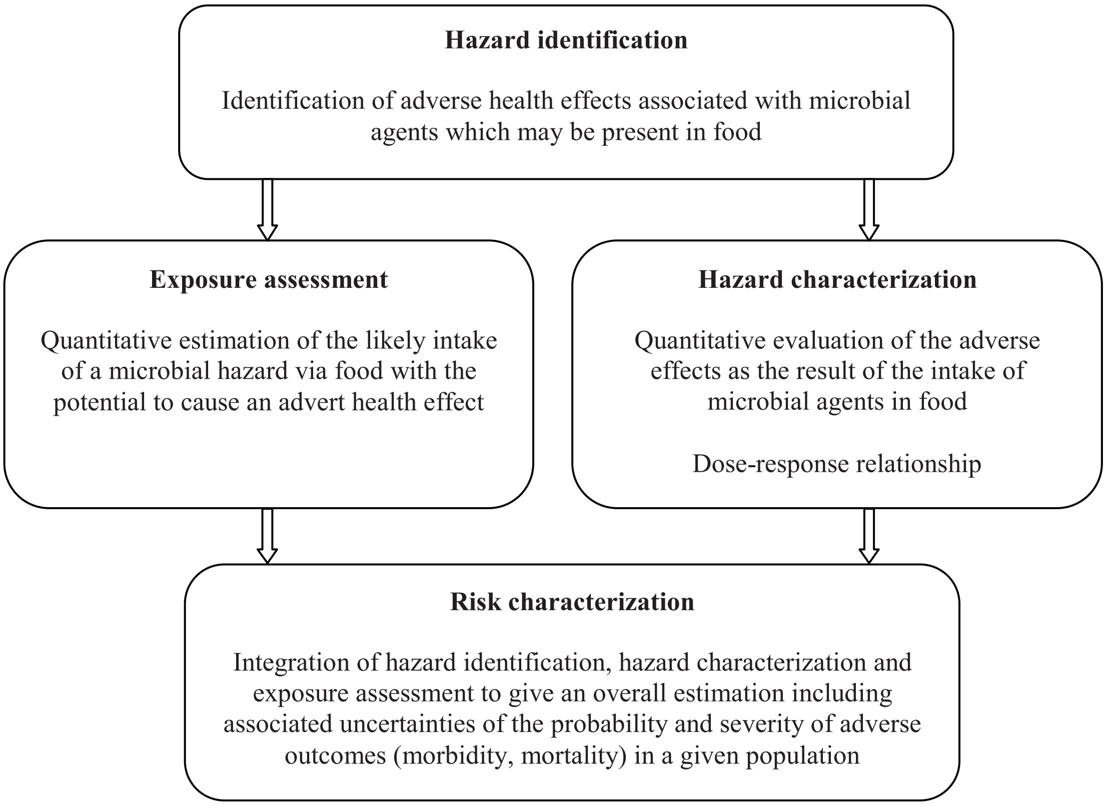

Figure 3: Components of a microbial risk assessment according to the Codex Alimentarius Commision (Source: Voysey and Brown, 2000).

Although QMRA must have a scientific basis (Principle 1, Table 1), value judgements, choices and assumptions are often unavoidable. A functional separation between risk assessment and risk management is necessary (Principle 2). Transparency should play a major role throughout the risk assessment process (Principle 5). Constraints, assumptions and value judgements should be documented systematically (Principle 6), and the risk estimates should contain a full description of the uncertainties, including their location within the risk assessment process (Principle 7). The data needs to be of sufficient quality (Principle 8) and the influence of the estimates and assumptions used in the risk assessment on the final outcome should be evaluated (Principle 6). Dynamics of microbial growth, survival, death in different matrices, and interactions between human, agent and environment should be taken into account (Principle 9). Principles 10 and 11 deal with the issue of validation of data, and improving the risk assessment with updated information. In addition, there must be a clear communication of the purpose and output of the risk assessment, which requires the interaction between risk assessors, risk managers, and stakeholders (Principle 4).

2.4. EXAMPLES OF QMRAS

QMRAs are carried out by developing a model which is a schematic or mathematical representation describing a food safety problem in as great a detail as possible. Every QMRA model aims to give an answer to one or several risk management questions, based on the available knowledge.

Up to now, few QMRAs cover the entire food production chain encompassing the primary production, processing, distribution, food preparation and consumption stages, which are typically modelled as modules, with the results of one module being exploited as inputs for the following module. Such large-scale risk assessments, termed farm-to-fork risk assessments, are usually commissioned by environmental, veterinary, public health or food safety authorities, and carried out by multidisciplinary consortia. The choice of the modelling techniques applied should be according to the problem which has to be modelled, and it depends also on the available data and expertise. A variety of modelling techniques for (pre-harvest) microbial risk assessment were reviewed by Lo Fo Wong et al. (2006b) on usefulness, strengths and weaknesses. Infectious diseases modelling can be time-consuming and mathematically complex, but can be easily updated and are useful to assess the impact of control strategies. Predictive microbiology is often used to predict the behaviour of growth, survival, inactivation of bacteria and is an essential tool in the hazard identification and exposure assessment. Predictive models yield point estimates instead of stochastic ones. Reliable predictive microbiology models require datasets of good quality, and are most useful in combination with advanced probabilistic modelling techniques.

The modular process risk modelling (MPRM) method (Nauta et al., 2001; Bollaerts et al., 2009) is designed to model the transmission of micro-organisms along the food pathway, by breaking down the pathway into consecutive modules and then modelling the basic microbial processes that take place in each module, such as growth, inactivation (pathogen-related) and the production processes (mixing, partitioning, removal and cross-contamination) (cf. Table 1: Principle 9).

An inventory of representative farm-to-fork QMRAs with an indication of country, agents and quality assurance methods applied is shown in Table 2. These QMRAs were principally related to three food-borne bacteria: *Salmonella*, *Escherichia coli* and *Campylobacter* in a variety of animals and food products, and they were carried out by North American and European consortia between 1997 and 2010.

In a review of early QMRAs, Schlundt (2000) commented that few formal QMRAs had been carried out in accordance with to the Codex Alimentarius guidelines. From the QMRAs reviewed it was not clear whether a critical evaluation of input data had taken place. In addition, the variability and uncertainty of the data were often not described in sufficient detail and assumptions having an impact on the final result were not often clearly presented or critically evaluated. Unfortunately, even in more recent QMRAs the same drawbacks relating to the lack of a coherent quality assurance system are frequently encountered. The role of the different quality assurance methods and their usefulness in QMRA are further explained in the next section.

Table 2: Overview of available (non exhaustive) farm-to-fork quantitative microbial risk assessments, with indication of applied quality assurance methods

Pathogen	Food product/ animal species	Country	Quality assurance methods[a]										Remarks	Reference
			CEA	EE	EPR	MQC	MCA	MVA	NUSAP	PR	SA	WI		
Bacillus cereus	Pasteurised milk	NL											No uncertainty analysis described	Notermans et al. (1997)
Campylobacter spp.	Broiler chicken	NL		X			X	X			X	X	Uncertainty not quantified, assessed by scenario analysis.	Nauta et al. (2007) Havelaar et al. (2007)
	Broiler chicken	DK					X					X		Rosenquist et al. (2003)[b]
	Broiler chicken	IT					X				X		Discussion of assumptions	Calistri and Giovannini (2008)[b]
Escherichia coli O157:H7	Ground beef hamburger	CA					X				X	X	No uncertainty included in the final risk estimate	Cassin et al. (1998)
	Beef	IE					X	X			X	X		Duffy et al. (2006)[b]
	Ground minced beef	US					X	X			X	X	No clear risk management question No full distinction between uncertainty and variability	FSIS (2001)
	Steak tartare	NL		X			X	X			X	X	No quantitative uncertainty assessment (only variability)	Nauta et al. (2001)

[a] CEA: Critical Evaluation of Assumptions, EE: Structured Expert Elicitation, EPR: Extended Peer Review / Public Review, MQC: Model Quality Checklist, MCA: Monte Carlo Analysis (Tier 3), NUSAP: NUSAP/Pedigree for data quality assessment, MVA: Model Validation, PR: Peer Review, SA: Sensitivity Analysis, WI: What-if Scenario Analysis. Multiple Model Comparison is not included in the table; [b] no complete farm-to-fork QMRA.

(Continued)

Table 2 (continued)

Pathogen	Food product/ animal species	Country	Quality assurance methods[a]										Remarks	Reference
			CEA	EE	EPR	MQC	MCA	MVA	NUSAP	PR	SA	WI		
Salmonella	Eggs and broiler chicken	Not country-specific		X	X		X			X			Discussion of assumptions No complete quantitative uncertainty analysis	FAO/WHO (2002b)
S. Typhimurium	Pork, bacon, mixed meat products	UK					X	X			X	X	Only variability modelled, no uncertainty Clear presentation of assumptions	Hill et al. (2003; 2008)
S. Typhimurium DT 104	Danish dry-cured pork sausages	DK					X					X		Alban et al. (2002)
Salmonella	All pork-derived foods	FI					X						Bayesian model	Ranta et al. (2004)[b]
	Fresh minced pork meat	BE					X	X			X	X		Delhalle et al. (2009b)[b]
	Fresh minced pork meat (pure and mixed)	BE	X	X		X	X	X	X		X	X		Bollaerts et al. (2009), Boone et al. (2009a; 2009b; 2010)
	Slaughter and breeding pigs	EU					X	X		X	X	X	Farm-to-fork risk assessment at EU level	VLA-DTU-RIVM (2010), EFSA (2010)

[a] CEA: Critical Evaluation of Assumptions, EE: Structured Expert Elicitation, EPR: Extended Peer Review / Public Review, MQC: Model Quality Checklist, MCA: Monte Carlo Analysis (Tier 3), NUSAP: NUSAP/Pedigree for data quality assessment, MVA: Model Validation, PR: Peer Review, SA: Sensitivity Analysis, WI: What-if Scenario Analysis. Multiple Model Comparison is not included in the table; [b] no complete farm-to-fork QMRA.

2.5. WHAT IS QUALITY ASSURANCE?

2.5.1. Definitions

According to the scientific discipline numerous definitions for quality and quality assurance have been formulated. As a general definition, the US Office of Management and Budget defined “quality” as a multidimensional term including objectivity, utility and integrity (OMB, 2002). Objectivity refers to how accurate, clear, complete and unbiased the information is presented. Utility indicates how useful the information is while integrity is related to the protection of information from unauthorised access or revision. Essential for quality in a QMRA is the “fitness for purpose” principle, which means that quality is a relative and context depending concept.

The assurance of quality (Quality Assurance (QA)) is the framework that is provided to ensure that all tasks included in the risk assessment are executed in a technically and scientifically correct manner, and that all model-based analyses are reproducible. The aim of the QA process is to enhance the credibility of the model results: (i) by means of a proper interaction between risk assessor and risk managers and clearly defining the purpose of the risk assessment, (ii) by means of rigorous validation tests against independent data, (iii) by means of uncertainty assessment and (iv) by means of independent peer review of the various stages of the risk assessment (Refsgaard et al., 2005). The criteria that can be used to determine the validity and utility of a QMRA, and that are relevant for the QA process, have been summarised by Lammerding et al. (2007) (Table 3).

2.5.2. Quality Assurance in QMRA: Why?

As QMRA is a decision-support tool used by risk managers that can have an impact on a variety of stakeholders (policy-makers, funding organisations, farmers, meat processing industry, consumers, etc.), it is essential to know whether the results provided by the risk assessment process are sufficiently relevant, robust, credible and accurate to provide an answer to the risk problem.

In order to facilitate decision making, risk assessors need to clearly explain and communicate to decision makers the level of confidence associated with the results and to report the relevant uncertainties (where are the uncertainties; how large are they?) and assumptions (where are the assumptions in the model; what impact do they have?).

The importance of QA and a harmonised approach to microbial risk assessment is addressed in the FAO/WHO guidelines (FAO/WHO, 2003; 2008; 2009) and by Havelaar et al. (2007a). The need for transparency, which is a key element in QA has recently been highlighted in a guidance document released by EFSA (2009c).

The purpose of this review is to present a summary of methods allowing to meet the general guidelines set up by the Codex Alimentarius (1999), and thereby to contribute to the quality assurance of the QMRA process.

Table 3: Criteria determining the quality of a QMRA, with indication of methods to address the validity and utility criteria (adapted from: Lammerding et al., 2007)

Criteria relevant to validity	Definition	Methods
Quality and treatment of data	Relevant and timely data, criteria for inclusion/exclusion data	MCA, NUSAP, PR, UA
Inference of probability	Appropriate choice of distributions, adequate number of iterations	MQC, MVA, PR
Internal consistency	Sound logic and inference	MQC, MVA, PR
Appropriateness of assumptions, expert opinion	Soundness of assumptions	CEA, EE, MQC, PR
Epidemiological and biological credibility	Outcomes should be within plausible limits	MQC, MVA, PR
Transparency	Systematic development of the QMRA steps, indication of data used, data gaps (use of expert opinion, assumptions). Identification and communication of the uncertainty in the models, data, assumptions, what-if scenarios.	CEA, EE, NUSAP, MQC, SA, UA, WI
Peer-review	Independent review of data, models, analysis	PR, MQC
Stakeholder involvement	As appropriate for data input, QMRA should reflect its scope to the segment of stakeholders (farmers, industry, public)	EPR
Criteria relevant to utility		
Addresses the risk management question	Clear definition of the problem formulation, application of the results of the QMRA.	MQC, PR
Clarity for different audiences	Tiered series of reports for different groups ranging from very detailed to summary reports for non-technical audiences. Progressive disclosure of uncertainties	EPR, RC
Explicit statement of limitations	Description of constraints of the model (time, money, application of results)	CEA, MQC, NUSAP, PR, SA, WI
Identification of risk-determining steps, knowledge gaps, conflicting evidence	Helps decision-makers to focus on important steps. Clear statement of uncertainties, in data, assumptions. Identification of data needs	CEA, MQC, NUSAP, PR, SA UA,
Inclusion of what-if scenarios, evaluation of potential risk reduction strategies	Requires defining scenarios in interaction with risk managers	EPR, WI, RC,
Applicable to stakeholders	The QMRA enhances knowledge on the food production processes and can inform stakeholders	EPR, MQC, RC

CEA: Critical Evaluation of Assumptions, EE: Structured Expert Elicitation, EPR: Extended Peer Review / Public Review, MQC: Model Quality Checklist, MCA: Monte Carlo Analysis (Tier 3), MVA: Model Validation, NUSAP: NUSAP/Pedigree for data quality assessment, PR: Peer Review, RC: Risk communication, SA: Sensitivity Analysis, UA: Tiered Uncertainty Analysis, WI: What-if Scenario Analysis

2.6. METHODS FOR THE IMPLEMENTATION OF QUALITY ASSURANCE IN QMRA

2.6.1. Uncertainty assessment

Uncertainty is defined as the lack of knowledge concerning input data, models and assumptions (EPA, 2003). Refsgaard et al. (2007) state that both subjective and objective aspects are important in assessing the degree of uncertainty, which they define as the degree of the lack of confidence that one has about the validity of the information obtained. There is no straightforward relationship between the degree of uncertainty and the level of quality, as a risk assessment involving a high degree of uncertainty might be of good quality if the degree of precision is in accordance to the risk management questions (fit-for-purpose) (Krayer von Krauss, 2005).

A document outlining general guidelines (not specifically for QMRA) for characterizing and communicating uncertainty in exposure assessment was released by the World Health Organisation (WHO, 2008). The most detailed guidelines for uncertainty assessment and uncertainty communication have been developed for environmental risk assessment by the Dutch Environmental Agency (RIVM-MNP) (Janssen et al., 2003; Petersen et al., 2003; van der Sluijs et al., 2003; 2004). These uncertainty guidelines are increasingly used within the Dutch Environmental Agency to deal with uncertainty in performing and reporting environmental assessments (Petersen et al., in press).

2.6.1.1. Uncertainty terminology

The use of a coherent typology of uncertainties, such as that proposed by Walker et al. (2003b) and Janssen et al. (2005), and recently applied to environmental burden of disease assessment by Knol et al. (2009) is essential for a thorough uncertainty assessment. In this typology, uncertainty is interpreted as a multidimensional concept and distinctions are made between the location of the uncertainty, its nature, its level (range), the qualification of the knowledge base and the value-ladenness of assumptions resulting from subjective choices.

The **location** indicates where the uncertainty manifests itself in the assessment and includes the description of the risk problem, the model structure, the data and the model output. For the **nature** of uncertainty a distinction between epistemic uncertainty and stochastic uncertainty should be made. The first type results from incomplete knowledge that can be reduced by research, while the latter can be considered as variability, inherent to the system and that cannot be reduced by additional empirical research.

The **level** of uncertainty indicates where the uncertainty manifests itself in the scale ranging from statistical uncertainty to scenario uncertainty to ignorance. Statistical uncertainty is appropriate when the uncertainty can be described in statistical terms (e.g. central estimate and confidence intervals), such as with measurement uncertainty due to sampling errors. In scenario uncertainty a range of outcomes is possible due to underlying assumptions. Ignorance refers to relationships of processes that are poorly understood, so that statistical properties cannot be estimated.

Subsequently, the fourth dimension of uncertainty is the **qualification of the knowledge base** within the different parts of the assessment, as determined by quality criteria (proxy, empirical basis, methodology and validity, see section 2.6.3.). The fifth dimension is the evaluation of the **potential value-ladenness of assumptions** as a result of subjective choices in the assessment (see section 2.6.5.).

The results of the uncertainty assessment should be clearly communicated in terms useful and understandable by the different target groups (risk assessors, risk managers or stakeholders) (see section 2.6.13).

2.6.1.2. Tiered uncertainty analysis

Depending on the scope and the desired level of uncertainty assessment in the QMRA process, a tiered approach (Tier 1, 2 and 3) is recommended both by EFSA (2006b) and by WHO (2008). Tier 1 analysis starts with a qualitative estimate of all the uncertainties and provides a description of the most significant uncertainties and the relative magnitude of their influence on the assessment output. Tier 2 and Tier 3 are quantitative uncertainty assessment approaches. Tier 2 consists of the deterministic analysis of uncertainties. Different alternative point estimates are filled in for uncertain inputs in the assessment and their impact on the assessment outcome is calculated. The most detailed level and resource intensive type of uncertainty analysis is obtained via a probabilistic analysis of uncertainties (Tier 3). Compared to the Tier 1 and Tier 2 approaches, Tier 3 produces probability distributions as outputs. What is essential in a Tier 3 approach is the specification of probability distributions for the model inputs. Hereafter, computations will identify how the variability and uncertainty propagate through the model, resulting in the quantification of the variability and uncertainty in the output. In addition, a sensitivity analysis can be performed to assess how the variation of the output is affected by changes in the model inputs. The most common approach for performing Tier 3 Uncertainty assessment includes Monte Carlo Analysis (MCA), Bootstrapping, and Bayesian analysis (FAO/WHO, 2008).

By identifying uncertainties qualitatively, deterministically and/or probabilistically, information on data gaps can be obtained. In order to take decisions, risk managers can ask for additional data collection to reduce the uncertainties.

2.6.2. Systematic review

QMRAs generally require a diversity of data sources to build a model. It is therefore good practice to make an inventory of what is known in the literature on a specific risk problem. A systematic review approach can be utilised to obtain quality data to be used as input in a QMRA. Systematic review is a rigorous and replicable method for the identification, evaluation and synthesis of scientific evidence for the purposes of addressing a specific topic (Sargeant et al., 2005). The steps in a systematic review include (i) the development of a focused study, (ii) the identification of relevant types of research using a structured strategy, (iii) the screening of abstracts for relevance to the study question, (iv) the quality assessment of the relevant literature using pre-determined criteria, (v) the extraction of data of sufficient quality, and (vi) the synthesis of data. In meta-analysis, a statistical technique is used (e.g. meta-regression) to combine results to provide a single estimate, whereby higher weights can be attributed to studies according to their study characteristics (study population, study method, sample size, sampling plan, etc.).

The absence of published literature on a specific topic can serve as a motivation for initiating additional research, to contact database owners for the exchange of (unpublished) data, and/or to set up new experiments.

2.6.3. NUSAP/Pedigree approach for the evaluation of data quality

Good quality data is data that is complete, relevant, fit-for-purpose and valid. A prerequisite for the evaluation of the data quality is that data should be sufficiently documented, with respect to its references, sampling characteristics (e.g. sample size, sample methods, temporal/geographical representativeness, etc.) and validation status. A systematic review approach (see section 2.6.2.) can be helpful in this documentation process.

The NUSAP/Pedigree approach is a method that provides a basis for the structured evaluation of data quality. The purpose of the NUSAP (Numeral, Unit, Spread, Assessment and Pedigree) system aims is to analyse the uncertainty in scientific procedures used to support decision-making (Funtowicz and Ravetz, 1990). NUSAP uniquely integrates quantitative uncertainty information (Numeral, Unit and Spread) and qualitative uncertainty information by using expert judgement (Assessment) and a multi-criteria assessment

(Pedigree) of the scientific knowledge basis of a risk assessment. NUSAP is able to address uncertainties that are hard to quantify, such as methodological and epistemological uncertainties and which are not systematically taken into account in scientific studies (van der Sluijs et al., 2005a).

The pedigree assessment is the most innovative aspect of NUSAP. It introduces a set of criteria brought together in a pedigree matrix, that capture the essential characteristics of the data, such as the proxy representation, the empirical basis, the methodological rigour and the degree of validation (Table 4).

Table 4: Pedigree matrix for the evaluation of data quality (Source: Risbey et al., 2001a)

	Pedigree criteria			
Score	Proxy	Empirical	Method	Validation
4	Exact measure of the desired quantity	Large sample Direct measurements, Controlled experiments	Best available practice in well-established discipline	Compared with independent measurements of the same variable over long domain
3	Good fit or measure	Historical/field data uncontrolled experiments small sample direct measurements	Reliable method common within established discipline, best available practice in immature discipline	Compared with independent measurements of closely related variable over shorter period
2	Well correlated but not measuring the same thing	Modelled/derived data / indirect measurements	Acceptable method but limited consensus on reliability	Measurements not independent proxy variable, limited domain
1	Weak correlation but commonalities in measure	Educated guesses indirect approximation rule of thumb estimate	Preliminary methods with unknown reliability	Weak and very indirect validation
0	Not correlated and not clearly related	Crude speculation	No discernible rigour	No validation performed

The proxy criterion evaluates the closeness of resemblance between the input parameter available from the data source and the actual variable that would be required in the model. The empirical basis criterion is used to evaluate the degree to which direct observations were used to estimate the input parameter. A higher pedigree score for the empirical basis was attributed to input parameters obtained from the field data compared with indirect, modelled data or data obtained by expert judgement. The methodological rigour refers to the norms used in the collection and checking of the data and the degree of acceptance of these norms by the peer community in the relevant discipline. Lastly, the validation criterion evaluates the degree to which one was able to cross-check the data against independent sources.

This pedigree matrix is an instrument used by risk assessors in discussing and evaluating data. The matrix can be used to attribute scores to each criterion on a discrete numeral scale from 0 (weak) to 4 (strong). By aggregating scores over the different criteria, overall pedigree strengths are obtained. Pedigree strengths can be graphically represented within a diagnostic diagram (Figure 4) representing the overall strengths of input parameters on the x-axis and the sensitivity of the input parameters (obtained by sensitivity analysis) on the y-axis (van der Sluijs et al., 2004; 2005a). The two metrics taken together - strength and sensitivity - are a measure of the quality of a parameter. The position of the input parameters within the diagnostic diagram is a helpful tool for obtaining an overview of the weak and strong links within the model and can thereby lead to model improvement.

How the NUSAP/Pedigree method was applied to assess the quality of potential input parameters for a Belgian QMRA on *Salmonella* in the pork production process is discussed in chapter 5.

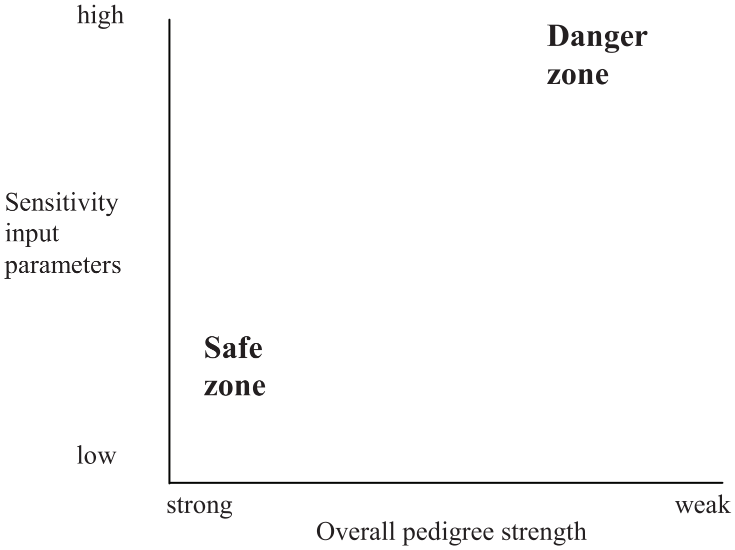

Figure 4: Diagnostic diagram for the representation of the quality of data. It combines scores for input parameters obtained in the pedigree assessment with their sensitivity (Source: van der Sluijs et al., 2004).

2.6.4. Expert elicitation

Expert elicitation is the process of eliciting subjective judgements from experts. It is used to provide input for QMRA when empirical data are either lacking, or of poor quality or difficult to obtain (van der Fels-Klerx et al., 2005). Since the elicitation of expert judgement involves subjectivity, it is prone to bias from the expert providing his/her judgement, as well as from the elicitor (person collecting the expert judgement) and from the elicitation protocol used, all of which may ultimately have an impact on the validity of the decisions based on a QMRA. The aim of a structured elicitation procedure is to reduce this bias as much as possible and this requires thorough preparation.

Elicitation of expert judgement can be described by subjective probability density functions reflecting the expert's degree of belief (Cooke, 1991). A key issue when eliciting expert opinion is related to the method used to aggregate multiple expert opinions. There are two ways to combine expert distributions: the behavioural approach aiming to elicit a consensus distribution (such as the Delphi method), and the mathematical approach aiming to combine individual expert's distribution (e.g. linear opinion pools or Cooke's Classical model) (Cooke, 1991; Clemen and Winkler, 1999). A structured expert elicitation involves the selection of the experts, explanation to the experts of the problem and the elicitation procedure, a clear definition of the quantity to be assessed, a discussion of the gaps in the knowledge, specification of the experts' belief in a distribution, and the decision whether or not to aggregate the distributions of the different experts (van der Sluijs et al., 2004). A successfully structured expert elicitation also implies solid training in elicitation techniques. For an overview of general guidelines to carry out expert elicitation the reader is referred to Morgan and Henrion (1990), Cooke (1991) and Slottje et al. (2008). In veterinary science, the most common structured expert elicitation methods include the Delphi method, and Cooke's Classical model (van der Fels-Klerx et al., 2002; 2005).

In particular, structured expert opinion following Cooke's Classical model was used to provide input in a QMRA for *Campylobacter* (van der Fels-Klerx et al., 2005). In addition, Cooke's model has been applied for the elicitation of expert judgement study to fill in data gaps of the METZOON model (Boone et al., 2009a). This is discussed in chapter 4.

2.6.5. Critical evaluation of assumptions

The quality of a QMRA also depends largely on the assumptions made in constructing the model. It is therefore necessary to identify these assumptions and to screen the model for hidden or implicit assumptions. In QMRA publications, assumptions are often systematically listed (e.g. Calistri and Giovannini (2008), FAO/WHO (2002a) and Hill et al. (2003)) but not evaluated on their quality.

A novel method for the critical evaluation of a model's assumptions was developed by Kloprogge et al. (2005; 2010). This method consists of three sections, the analysis of the assumptions, the revision section, and the section related to the communication of the analysis. These three sections are made up of subsequent methodological steps: first, assumptions (and hidden assumptions) are identified in the model; after this, the model's most important assumptions (or key-assumptions) are identified and prioritized. Hereafter the potential value-ladenness (degree of subjectiveness) of the key assumptions is assessed. Subsequently "weak" links in the model are identified. The next methodological steps include the further analysis of the potential value-ladenness of the key assumptions.

The revision of the assessment includes an evaluation of the sensitivity of the assumptions, and of the effect of diversifications or different choices made with respect to the assumptions. The last methodological step deals with what should be communicated on the basis of the assumptions analysis (key-assumptions, alternatives and underpinning of choices, impact of key-assumptions on results, and robustness).

To promote a structured discussion about the assumptions, Kloprogge et al. (2010) incorporated the NUSAP/Pedigree approach (see section 2.6.3.) and proposed a pedigree matrix (Table 5) containing six pedigree criteria: (i) the influence of situational limitations, (ii) the plausibility, (iii) the choice space and (iv) the agreement among peers, (v) the agreement among stakeholders, (vi) the sensitivity to the analyst's views and interests. The pedigree matrix contained an additional column, the "influence on results".

The influence of situational limitations refers to the degree to which the choice for an assumption is influenced by the limited amount of data, time, software, hardware and human resources. The plausibility criterion designates the degree to which an assumption is in accordance with the "reality" while the choice space indicates the degree to which alternatives were available to choose from at the moment of making the assumption. Agreement among peers addresses the degree to which the choice of peers is likely to coincide with the analyst's choice. Agreement among stakeholders addresses the degree to which the analyst's choice is likely to agree with the stakeholders' views. The influence of the analyst's views, background and interests are taken into account in the criterion 'sensitivity to view and interests of the analyst'.

The "influence on results" does not evaluate the value-ladenness of the assumptions, but rather provides a rough indication of the influence of an assumption on the end result of the risk assessment. The pedigree matrix is used as a tool to score the assumptions for the different pedigree criteria. As for the evaluation of the quality of data, a diagnostic diagram can be used to identify weak and strong links within a risk model (Figure 4). Individual scores for the different pedigree criteria can be represented graphically either by kite diagrams

(Kloprogge et al., 2005), or by pedigree charts (Figure 5) (Wardekker et al., 2008; Kloprogge et al., 2010) (for further details, see section 6.2.3.). The critical evaluation of the assumptions can be applied after the risk assessment has been carried out. It is, however, preferable also to apply it iteratively during the development of the risk assessment so that the insights gained from the assumptions analysis can be used for the improvement of the risk assessment. For an application of the critical evaluation of assumptions-method for the Belgian QMRA model for *Salmonella* in minced pork meat the reader is referred to chapter 5.

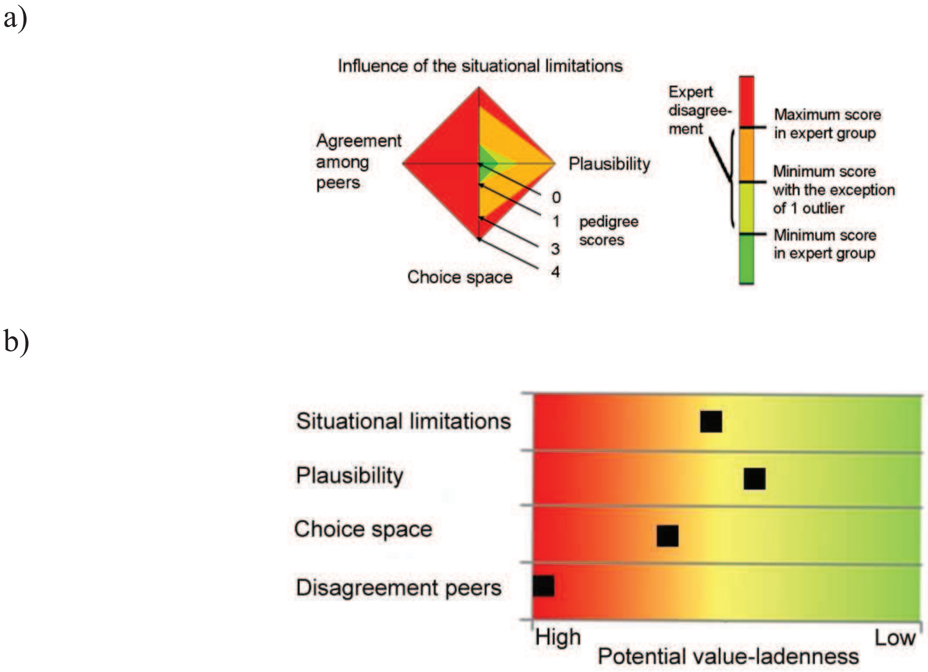

Figure 5: Graphical representation of pedigree scores. a) Example of a kite diagram (based on van der Sluijs et al., 2005b); b) Example of a pedigree chart for the evaluation of assumptions (based on Wardekker et al., 2008). The ■ indicates the average pedigree score.

Evaluation of assumptions is of the utmost importance in a QMRA with high policy relevance (target settings, for example for the whole EU), and when uncertain outcomes may have a large effect on policy making. The proposed method inevitably depends on expert judgement and on the composition of the expert groups making the evaluations. Since QMRA mostly involve a wide range of disciplines, it may sometimes be difficult to ask experts to evaluate assumptions that are outside their expertise field. For an optimal reflexive debate on the quality of assumptions, it may then be preferable to form subgroups including only experts with sufficient expertise on specific topics (e.g. primary production, processing, preparation and consumption).

Table 5: Pedigree matrix for reviewing the quality of assumptions (Source: Kloprogge et al., 2005)

	Criteria						Influence on results
Score	Influence of situational limitations	Plausibility	Choice space	(Dis)agreement among peers	(Dis)agreement among stakeholders	Sensitivity to view and interests of the analyst	
2	Choice assumption hardly influenced	The assumption is plausible	Hardly any alternative assumption available	Many would have made the same assumption	Many would have made the same assumption	Choice assumption hardly sensitive	The assumption has only local influence
1	Choice assumption moderately influenced	The assumption is acceptable	Limited choice from alternative assumptions	Several would have made the same assumption	Several would have made the same assumption	Choice assumption moderately sensitive	The assumption greatly determines the results of the step
0	Totally different assumption had there not been limitations	The assumptions is fictive or speculative	Ample choice from alternative assumptions	Few would have made the same assumption	Few would have made the same assumption	Choice assumption sensitive	The assumption greatly determines the results of the indicator

2.6.6. Sensitivity analysis

Sensitivity analysis (SA) aims to assess how the variation in the output of a model can be apportioned to the different sources of variation in the model's input parameters (Saltelli et al., 2000). SA can be used as a quality assurance method for the purpose of obtaining a better insight into the model. It is helpful in the selection process of appropriate risk management strategies and is recommended when the aim is: (i) to prioritize potential critical control points in the model, (ii) to identify key sources of uncertainty and variability, (iii) to refine, verify and/or validate the model, (iv) to prioritize additional data collection or research, and (v) to develop what-if scenarios (Frey et al., 2004).

As a preparation for SA, it is essential that the QMRA model be clearly structured and documented, and that a clear distinction be made between inputs and outputs (Frey et al., 2003). The choice of the output variable is very important in SA. In modular farm-to-fork QMRAs, it can be more straightforward to perform a SA on the output variables for the different modules separately (e.g. primary production, transport and lairage, slaughter and processing, and preparation and consumption), than to perform a SA on the model as a whole. In this modular approach, a clear one-to-one relationship between output and inputs may be more easily identified, whereas this relationship is often very hard to observe in the end output of a complete model (VLA-DTU-RIVM, 2010). Secondly, SA can be particularly difficult across modules, where units of interest are variable (e.g. the random selection of individual pigs in the primary production stage, the transport of a batch of pigs to the slaughterhouse, the half-carcasses and meat-cuts at the processing stage, the meat portions etc.).

Guidance to select and apply SA methods in food safety risk assessment as well as to the interpretation and presentation of its results is provided by Frey et al. (2004). The choice of a SA method depends on its scope, applicability and the characteristics of the model. Frey et al. (2003) carried out an evaluation of nine SA methods (including nominal range sensitivity analysis, differential sensitivity analysis, standardized linear regression analysis, rank regression, correlation coefficients, ANOVA, Classification and regression trees (CART), scatter plots and conditional sensitivity analysis) and applied it to two QMRAs, *Listeria monocytogenes* in Ready-to-Eat foods, and *Escherichia coli* in ground beef. In this review, ANOVA and CART were considered to deal best with the simultaneous variations in all inputs, both the qualitative and the quantitative inputs, the non–linearity and the interactions. On the other hand, sample correlation coefficients (Pearson coefficients) and linear regression was judged to be the weakest with respect to application to nonlinear QMRA models. Spearman rank coefficients were found to be inappropriate for non-monotonic models. However, the most commonly used SA methods in QMRA are precisely the Pearson sample and the Spearman rank correlation coefficients, both using commercial software such as @Risk® (Palisade, NY, USA) and Cristal Ball® (Decisioneering Inc., Denver, USA). Although, the aforementioned software packages are easy to use, they are often neither very flexible, nor model-independent, and they may be of limited use when there are many interactions between inputs and huge numbers of correlation coefficients need to be calculated. There are other powerful and promising variance-based SA methods, such as the Fourier amplitude sensitivity analysis test (FAST) and Sobol's method which may also be suitable to the characteristics of QMRA models (Frey and Patil, 2002), but these have not been applied yet to food safety. Application of these methods using commercial software such as SAS, S-PLUS, Matlab, R, SIMLAB,... may require recoding of the model and higher skills in mathematics.

Various graphical SA methods such as scatter plots and conditional sensitivity analysis plots can be helpful to complement and interpret the results of SA methods (Frey et al., 2003).

2.6.7. What-if scenario analysis

What-if scenario analysis is a conditional analysis in which specific goals for risk mitigation can be established and evaluated. A prerequisite is that the structure of the model should allow for what-if scenario analysis. In scenario analysis, different alternative scenarios (compared to the baseline risk model) can be explored, along with their associated uncertainties. The best case and worst case scenarios can be interesting for decision makers, as they show those scenarios that explore the relevant extremes of input variables as compared to the baseline model (van der Sluijs et al., 2004). Scenario analysis inevitably involves scenario uncertainty which is associated with the quality of input data and assumptions. While what-if scenarios provide a basis for risk management, it is also a necessary quality assurance tool, since it makes it possible to explore the possibilities and usefulness of the QMRA model.

Before doing a scenario analysis, the scope and objectives of the analysis should be clearly defined, through interaction between the risk analysts, the risk managers and the stakeholders, and each scenario should be transparently documented (van der Sluijs et al., 2004). Most published QMRA studies include what-if scenarios for the purpose of exploring mitigation strategies (see Table 2).

2.6.8. Checklist approach

Checklists offer a structured tool to help modellers during the model building and quality control process of risk models (van der Sluijs et al., 2004) and are intended for internal use by risk assessors or external use by peer reviewers for the purpose of identifying (i) pitfalls in complex models (ii) details in the model that are critical to policy choices, and (iii) value-laden choices. A comparison of available checklists for model evaluation is represented in Table 6. A checklist for quality assistance in environmental modelling developed by Risbey et al. (2005) is also helpful for the evaluation of QMRA models. The checklist contains questions related to the description of the objectives of the model and what role it can play in policy making. Other questions focus on the internal strength and quality aspects of the model inputs and parameters, the treatment of uncertainties, assumptions and robustness of the model, and whether the model output matches the requirements of the users. Finally, there are questions that focus on how the model results are communicated to and used by the risk managers, and how the stakeholders have been involved in the risk assessment process. The filled in checklist is used to analyze the main pitfalls in the risk assessment process and to draw conclusions concerning how fit-for-purpose the model is.

A web-based checklist (Petersen et al., 2003), used by risk assessors at the Dutch Environmental Assessment Agency (PBL) offers guidance for the uncertainty assessment of environmental risk assessment (available at http://leidraad.pbl.nl).

Macgill et al. (2001) proposed a checklist to assess the quality of waterborne risk assessments. The questions in the checklist are divided into five parts: (i) the observations or input data used in the risk assessment, (ii) the methodology used, (iii) the output of the risk assessment, (iv) the peer review process, and (iv) the validity of the model. On the basis of the answers given to each of the questions in the checklist, scores are attributed, which are added up to provide a total score. Total scores, as well as the scores of the different parts of the checklist can be used to evaluate and compare the quality of risk assessments as a whole or of components thereof. Both the scores and the *rationale* behind the scores are used to improve, if necessary, the quality of the risk assessment.

As an aid in the evaluation and peer review of veterinary import risk assessment (e.g. Classical Swine Fever, Foot and Mouth Disease), Paisley (2007) developed a quality audit checklist. Answers to the questions in the checklist are scored on a scale from 0 to 5, and subsequently aggregated to provide the total score. The checklist is used to audit the risk assessments in terms of the risk question and the purpose of the risk assessment, the uncertainty assessment, the methods used, adherence to international guidelines, the data used, the description and plausibility of the assumptions and scenarios, the risk communication and the reporting. The checklist was used by de Vos et al. (2009) to compare the quality of 16 import risk assessments (IRA). Based on the answers to the checklist, she concluded that IRAs were relatively heterogeneous with respect to quality. Although this checklist is still in the development phase, its compactness presents clear advantages to be used as an auditing tool.

The checklists by Paisley (2007) was used to review the quality of the METZOON model (see chapter 7).

Table 6: Examples of checklists useful for the evaluation of quantitative microbial risk assessments

Checklist name	Risk assessment type	Characteristics	Reference
Risbey	Environmental modelling	Quality assistance for internal use No scoring Identification of pitfalls in the model Identification if the model is fit for its purpose Identification of value-laden assumptions Long but complete checklist Generic checklist	Risbey et al. (2001b; 2005)
Dutch Environmental Agency	Environmental modelling	Easy-to-use web-based application Very flexible: quickscan checklist, with elaboration if necessary Focus on policy relevance Identification of uncertainties and pitfalls No scoring Increasingly used for quality assurance of research projects of the Dutch Environmental Agency (PBL)	Petersen et al. (2003)
Macgill	Waterborne risk assessment	Scoring Short checklist	Macgill et al. (2001)
Paisley	Import risk assessment	Comprehensive, not too detailed Worksheet-based Scoring Generic, applicable to QMRA	De Vos et al. (2009), Paisley (2007)

2.6.9. Peer review

Peer review is the independent review of data, logic, scientific interpretation, models, assumptions and analysis of all steps in the QMRA process, to ensure that it meets the standards of the scientific community (Lammerding, 2007). In the absence of specific guidance for the peer review process in QMRA, general guidelines have been set out by the Office of Management and Budget (OMB, 2004). Comments by peer reviewers can be helpful in terms of identifying biases and ignored uncertainties, reconsidering assumptions and/or modifying and/or improving the design of data collection and (statistical) analysis. The main objective of the peer review process is to improve the credibility and transparency of a QMRA. In determining the appropriate type and format of the peer review, the following aspects should be considered: individual versus panel review, timing and resources, scope of the review, selections of the reviewers, public participation, and the processing of the reviewer comments. The OMB recommends peer review already in the early stages of the risk assessment process, such as when determining which input data and the model to use. Selection of peer reviewers is a challenging task, as most QMRAs are carried out by a multidisciplinary team. Therefore, experts from different disciplines should be involved in the peer-review process, such as statisticians, veterinarians, microbiologists, epidemiologists and medical doctors.... When necessary, economists and social scientists can be involved too. To allow for peer review, QMRAs should be transparently documented, and the reviewers should have access to all data and the models. Internal peer-review by colleagues is recommended to precede proper peer review. Checklists can offer a standardised format as a support tool for the review process (Risbey et al., 2001b; Paisley, 2007). In an inventory of 18 pre-harvest microbial risk assessments, Lo Fo Wong (2006a) indicated that 15 QMRAs had undergone a form of review process. QMRAs were considered as externally peer reviewed if it had succeeded a peer review publication process, but is not stated if this peer review process also included a formal review of the model code, assumptions etc. Three QMRAs presented in Table 2 mentioned that external peer review had been carried out. These included two QMRAs for *Salmonella* on eggs and broiler chicken, and a QMRA for *Salmonella* in slaughter and breeding pigs, commissioned by the USDA-FSIS (1998), FAO/WHO (2002b) and the EFSA (2010), respectively. The greatest limiting factor of peer review is the time and resources one is willing to allocate, especially when quick decisions are required for high-stakes decision problems.

Peer review should be distinguished from extended peer review, in which the review process is carried out by stakeholders (industry, public,...). These are defined as the actors involved in the risk assessment, who can be directly or indirectly affected by decisions based on a risk assessment. Consultation of stakeholders in the review process can contribute to the quality of a risk assessment, as they may have different views on the problem formulation than scientists, and may be critical towards assumptions taken by the risk assessors. Guidance on how to involve the extended peer community is provided in the guidance reports of the Dutch Environmental Agency (MNP/RU, 2008).

2.6.10. Model verification

Model verification is defined as the process of verifying that the mathematical expressions, the definitions of the data inputs, and the logic of the model are correct and correctly implemented. It involves checking the correctness of the model formulation, the inputs, and the internal consistency of the model, and it should precede model validation (see section 2.6.11.). Model verification is facilitated when the data, model structure, methods, tools and assumptions are clearly documented (FAO/WHO, 2009). There are however no standard guidelines for model verification.

2.6.11. Model validation

Model validation consists in verifying whether a model corresponds with the reality and is fit for its purpose. Model validation includes conceptual validation (the model represents accurately the system under study), the validation of algorithms (the model concepts have been translated adequately into mathematical formulas), the software code validation (the mathematical formulas have been correctly implemented in computer language), and the functional validation (checking the model with independent observations). A model is said to be validated when there is a close match between the model output and independent validation data. Model validation is highly dependent on the risk management questions and should be proportionate to the stakes of the decisions. In many QMRAs, validation or even partial validation is difficult to achieve due to the lack of data or comparable independent data. As an alternative to model validation when independent validation data is scarce or lacking, screening procedures and sensitivity analysis can be applied to identify the most important inputs, uncertainty assessment, and multiple model comparisons (FAO/WHO, 2009).

2.6.12. Multiple model comparison

A model is always a simplification of the reality. The mismatch between the modelled system and the reality inevitable causes model structure uncertainty. As an example, in a Danish environmental risk assessment study, five alternative conceptual models were developed by five independent consultants, who used the same raw data as input for their models.

The five consultants all used different approaches to answer the risk management question, which resulted in substantially different model predictions (Refsgaard et al., 2006). Large differences between alternative models may cause confusion in the results of a QMRA and delay or hinder management decisions. On the other hand, alternative models yielding similar conclusions can support and facilitate decision-making. When time and resources are limited, it is usually better to develop a single detailed QMRA model, instead of several alternative (less detailed) QMRAs. The quality of alternative models can be assessed and compared by means of previously discussed methods, such as the checklist approach, NUSAP/Pedigree, critical evaluation of assumptions, uncertainty analysis, sensitivity analysis and scenario analysis. Obviously good modelling skills are necessary to use the multiple modelling comparison technique.

2.6.13. Quality of documentation and risk communication

A clear documentation of all stages of the QMRA is essential. This should include a clear representation of strengths and limitations of the model (data quality, critical assumptions, model structure, uncertainties), and information on how the quality assurance has been dealt with. In turn, the implementation of the different quality assurance methods (e.g. peer review, NUSAP, etc.) also depends on the clarity of the documentation of the risk assessment process, the description of the data and assumptions etc.

The way in which the results of a QMRA is documented should be adapted to different target audiences (analysts, stakeholders, decision-makers), using the progressive disclosure of information approach (PDI) (Kloprogge et al., 2007). This implies that a full technical document with all model details for risk assessors should be complemented with a less technical report that is comprehensible for decision-makers and stakeholders. Special attention should be focussed on the documentation of the uncertainties and assumptions. For guidelines on the contents, style and degree of reported uncertainty information at different PDI layers, see Kloprogge et al. (2007). The clarity of the information can be improved by including tables and charts. For example, the quality of the data and the assumptions can be represented with kite diagrams, pedigree charts and diagnostic diagrams (as in Figures 4 and 5).

The uncertainty information should be tailored to the target groups. In a workshop on the communication of uncertainty information, it appeared that decision-makers are primarily interested in policy relevant uncertainties, and that these should be placed in the main text and summary instead of in the appendices, which remains often unread. Jargon should be avoided for risk managers (Kloprogge et al., 2007) and emphasis should be put on the implications of uncertainties for policy advice while uncertainties should be documented in detail (probability density functions, nature, extent and sources of uncertainty) only for risk assessors.

2.7. THE METZOON MODEL

The METZOON model was developed in Belgium to evaluate the health risks associated with the household consumption of fresh minced pork meat contaminated with *Salmonella*. The development of this model was done in two phases. At the beginning of the METZOON project, the model was mainly based on the model of Hill et al. (2003) as was described in Grijspeerdt et al. (2007). Subsequently, this model was considerably modified and published in its definitive form by Bollaerts et al. (2009). The different quality assurance methods that were developed and tested in this thesis (see chapter 3) were implemented in the earlier model described by Grijspeerdt et al. (2007).

In accordance with the Codex Alimentarius Commission (1999) this model included the hazard identification, the exposure assessment, the hazard characterisation and the risk characterisation. For an extensive overview of publications related to the METZOON model, the reader is referred to http://www.metzoon.be.

The **exposure assessment** was carried out by building a modular risk model covering the minced pork meat production from farm to fork. The food production pathway was split up in six consecutive modules: primary production, transport and lairage, slaughterhouse, post-processing, distribution and storage, and preparation and consumption (Bollaerts et al., 2009). The model is schematically represented in Figure 6. Matlab and Monte Carlo simulation was used to obtain stochastic estimates of the output variables. The input variables were expressed as distributions to reflect natural variability and/or stochastic uncertainty.

In the primary production module, the *Salmonella* serological status of pigs of which meat cuts will end up in the same meat mix are modelled. The number of seropositive animals in a batch was simulated using a density estimate of the within-herd seroprevalence. The latter was obtained using data for 2006 from the *Salmonella* surveillance programme organized by the Belgian Federal Agency for the Safety of the Food Chain (FASFC).

In the transport and lairage module, the pigs serological status was converted first into internal contaminated status (mesenteric lymph nodes and/or colon content positive) and then into external contaminated status using conditional probabilities.

From the slaughterhouse onwards, changes in prevalence, bacteriological concentration and unit size within each module are modelled by means of six basic processes, two of which are microbial processes (e.g. growth and inactivation) and the remaining four are food handling processes (e.g. removal, cross-contamination, mixing and partitioning). At the slaughterhouse, changes in the external contamination status of the carcasses are modelled at 5 stages assuming *Salmonella* increasing stages (killing, evisceration and polishing) and *Salmonella* decreasing stages (singeing and chilling) and probability calculations. From the status after chilling onwards, the *Salmonella* concentrations were modelled using Belgian data.

In the post-processing module, cutting of the carcass, mixing of the meat cuts and partitioning of the minced meat were distinguished. In the distribution and storage module, microbial growth was modelled at retail, during transport from retail to home and during storage at home.

In the preparation and consumption module, the process of preparing a meal, partially consisting of minced pork meat and another food item, is simulated. It is assumed that the meat is cooked while the other food item was consumed raw. Microbial inactivation due to cooking as well as cross-contamination from the meat to another food item were modelled. Baseline results showed that after chilling, 4.3% of the carcasses were contaminated with *Salmonella*. The results for distribution and storage suggested that growth, although limited, primarily happened during storage at home.

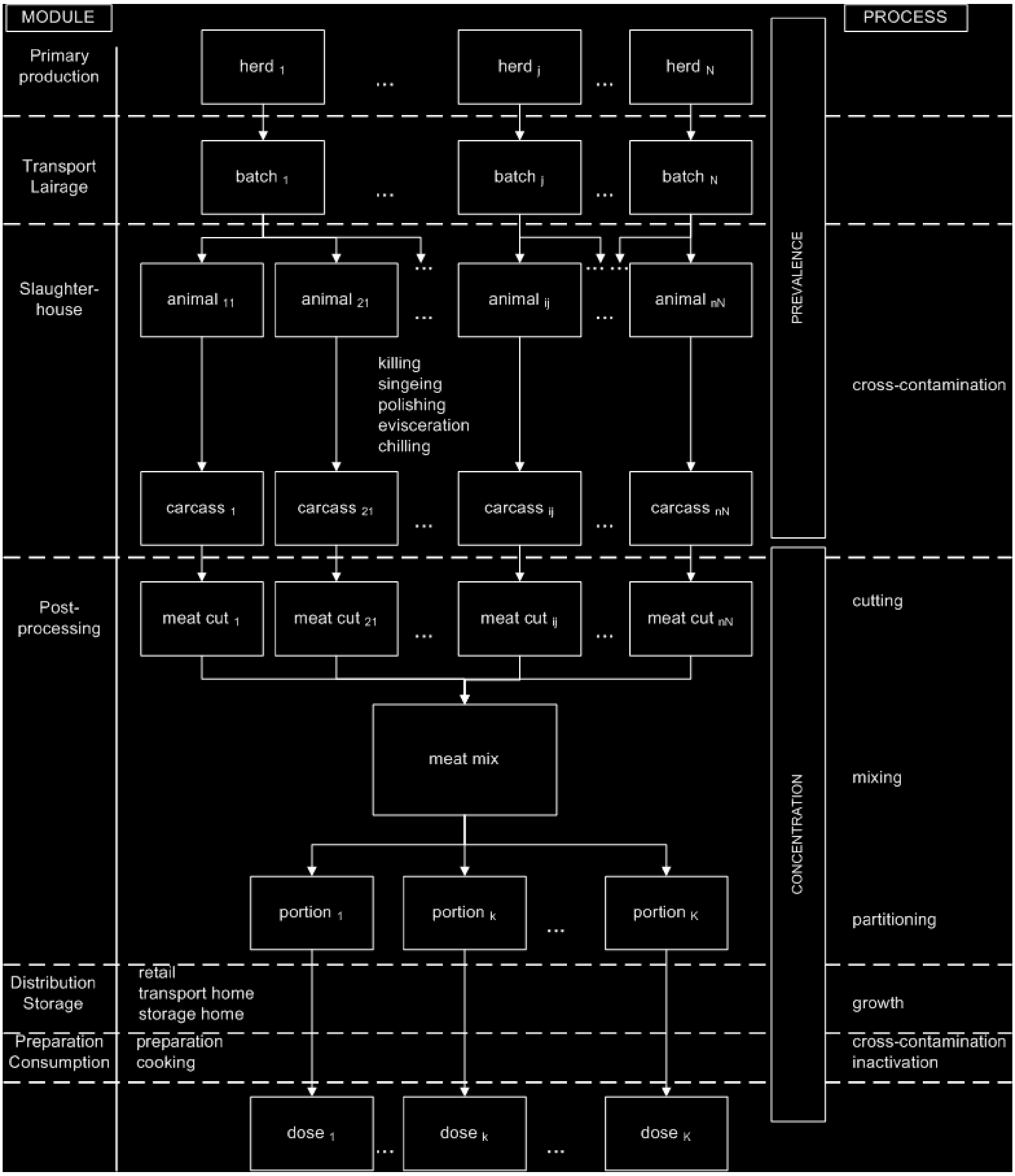

Figure 6: Schematic representation of the METZOON model (Bollaerts et al., 2009).
The left side illustrates the six consecutive modules, the middle part is a flowchart and the right part displays the basic process modelled.

In the **hazard characterization**, the dose-response relationship developed by Bollaerts et al. (2008) based on outbreak data was used.

In the **risk characterization**, the susceptible fraction or YOPI (young, old, pregnant, immune-compromised) group of the population in Belgium was estimated to be 24%. Considering the yearly number of servings of fresh mixed minced pork meat and the total population size in Belgium, the yearly number of salmonellosis cases for susceptible and normal population could be estimated. The yearly number of human salmonellosis cases in Belgium within the susceptible population was estimated as 13517 [90%: 7887 – 21691] and within the normal population as 6996 [90%CI 2045– 16555]. The total number of annual cases attributed to fresh minced pork meat was estimated as 20513 and the corresponding 90% percentile interval as [90%CI: 9932-38246] and was mainly due to undercooking and for a smaller extent to cross contamination in the kitchen via cook's hands. The average risk of illness following consumption of minced pork meat was higher for the susceptible (estimated as 4.713×10^{-5}; 90%CI: $2.750\ 10^{-5} – 7.563 \times 10^{-5}$) compared to the normal population (estimated as $7.704\ 10^{-6}$; 90%CI $2.251 \times 10^{-6} – 1.822 \times 10^{-5}$).

The results of the METZOON model were partly validated using external data originating from Belgian monitoring and surveillance programmes within and at the end of the production chain. Through “what-if” scenario analysis the efficacy of the model to identify risk management options was explored (Bollaerts et al., 2010).

CHAPTER 3: AIMS OF THE THESIS

The general objective was to use, develop and evaluate quality assurance methods for quantitative microbial risk assessment. These methods were applied to a Belgian QMRA model aiming to evaluate the risk of human illness due to *Salmonella* spp. associated with the consumption of fresh minced pork meat (the METZOON model, see section 2.7.). Quality assurance methods were applied in order to increase the transparency and confidence in the results of the QMRA.

The specific aims of the study were related to the application and evaluation of four quality assurance methods:

1. to carry out a **structured expert judgement** study to assess if and how the opinions from a heterogeneous panel of experts can be combined to provide input distributions for uncertain parameters in the QMRA (the METZOON model) (chapter 4);

2. to screen and evaluate the **quality of potential input parameters** for the METZOON model, by applying the **Numeral Unit Spread Assessment Pedigree (NUSAP) system** in order to provide an objective basis in the selection of input parameters. In addition, the NUSAP method as well as new graphical risk communication tools were evaluated for its relevance in promoting a structured debate about the quality of the data in the QMRA model (chapter 5);

3. to identify **assumptions** in the METZOON model and assess their potential value-ladenness as well as their impact on the output of the QMRA (chapter 6);

4. to evaluate the usefulness of **quality audit checklists** applied to evaluate the METZOON model (chapter 7).

CHAPTER 4:

STRUCTURED EXPERT JUDGEMENT IN QUANTITATIVE MICROBIAL RISK ASSESSMENT: EVALUATION OF WEIGHTING SCHEMES TO COMBINE EXPERT OPINIONS

Modified from: Boone, I., Van der Stede, Y., Bollaerts, K., Messens, W., Vose, D., Daube, G., Aerts, M., Mintiens, K. (2009). Expert judgement in a risk assessment model for *Salmonella* spp. in pork: the performance of different weighting schemes. Preventive Veterinary Medicine, 92: 224-234.

4.1. INTRODUCTION

Expert judgement is frequently used to provide input for a quantitative microbial risk assessment (QMRA) when empirical data are missing, difficult to obtain or of poor quality (Nauta et al., 2001; Stärk et al., 2002; Alban and Stärk, 2005; van der Fels-Klerx et al., 2005). As with empirical data, the input obtained through expert opinion may be biased and imprecise and this can have an impact on the outcome of the risk assessment (Walker et al., 2003a). These authors defined the quality of an expert judgement as how accurate the judgement estimates the true but unknown value and how well it is related to what the expert knows about the subject. Since the quality of expert judgement may ultimately determine the validity of decisions based on a QMRA, it is important that the expert judgements are elicited and treated through a structured approach with transparent and objective methodological rules (see section 2.6.4).

Many expert elicitation methods and protocols have been described. A useful method for eliciting expert opinion is known as the Delphi technique. This method includes a high level of interaction among the experts and aims to reach a consensus among the experts (van der Fels-Klerx et al., 2002). Another method, ELI (elicitation), is a graphically oriented computer interface that facilitates the quantification of expert knowledge on uncertain quantities (Horst et al., 1998; van der Fels-Klerx et al., 2002). This ELI method helps the experts to elicit unbiased subjective probability density functions (PDFs). The peak of a PDF represents the experts' guess for the uncertain variable, whereas the upper and lower bound of the PDF refers to the expert's uncertainty related to this best guess (van der Fels-Klerx et al., 2002).

An important issue in expert judgement elicitation is related to the aggregation method for combining the subjective PDFs of multiple experts. These combination methods can be classified in behavioural approaches (such as the Delphi method) and in mathematical approaches (e.g. linear or logarithmic opinion pools, Bayesian approaches) (Cooke, 1991; Clemen and Winkler, 1999; Scholz and Hansmann, 2007).

In this chapter, the mathematical combination approach known as Cooke's classical model was used to aggregate the individual experts' PDFs over model parameters into a single combined distribution (Cooke and Goossens, 2000). This model was chosen because it aims to produce a consensus on the combined distributions by complying with four necessary conditions (Cooke, 1991):

1. the whole process should be open for peer review;
2. expert assessment must allow for empirical quality control;

3. the elicitation procedure should encourage the experts to state their true opinion and not bias the results, and;
4. the experts should not be pre-judged before processing the results.

The classical model is a performance-based weighted averaging model allowing to aggregate individual experts' PDFs into a single combined PDF. The weights are derived from the experts' performance measured on seed variables (Cooke, 1991). Seed variables are variables whose true values are known to the analyst but unknown to the experts at the moment they are expressing their opinions, or which will become known post hoc. The performance of the experts to the seed variables is taken as indicative of their performance for the specific variables of interest. Seed variables are used to serve three objectives: (i) measure expert performance, (ii) enable performance based weighted combination of distributions elicited by experts and (iii) evaluate the resulting combined expert assessment.

The classical model contains four different weighting schemes also called "decision makers" (DMs), namely the equal weight DM, the global weight DM, the item weight DM and the user weight DM. The DMs are a combination of expert assessments characterised by different weights assigned to each expert's individual assessment in order to form combined distributions (Cooke and Goossens, 2008).

The purpose of this study was to carry out a structured expert judgement study —using the classical model— to obtain PDFs for missing input parameters in order to complete the QMRA-model (METZOON model) for human salmonellosis associated with the consumption of fresh minced pork meat (Grijspeerdt et al., 2007). More specifically, it aimed to evaluate the quality of the aggregated PDFs elicited by means of different weighting schemes contained in the classical model. The correlation between the weights derived from the experts' self-rating of expertise and performance-based weights was analysed to assess for possible over and/or under confidence of the experts.

4.2. MATERIALS AND METHODS

4.2.1. Elicitation protocol

The elicitation protocol consisted of three parts according to Cooke and Goossens (2000) and van der Fels-Klerx et al. (2002; 2005): (i) the preparation for elicitation (definition of the case structure, identification of variables of interest and seed variables, identification of experts, design of the quantitative elicitation session and dry-run session), (ii) the elicitation itself, and

(iii) the postelicitation processing (combination of the experts' assessments, robustness analysis and documentation of the results). These parts are described below.

4.2.1.1. Preparation of the expert elicitation process

The elicitation documents consisted of two questionnaires with two parts[1]. The first part was related to the background of the experts, the second part to specific questions on knowledge gaps relevant for the QMRA-model pertaining to the epidemiology of *Salmonella* spp. in the pork production chain. In total, 6 and 18 questions were related to the variables of interest (Table 7) and the seed variables (Table 8) respectively. The variables of interest were related to (i) the bacterial contamination load and prevalence of *Salmonella* contamination of pig carcasses during selected processing steps in the slaughterhouse, (ii) the agreement between bacteriological and serological test results related to *Salmonella* in pigs, (iii) the effect of transport and lairage on the bacteriological prevalence of *Salmonella* in pigs, and (iv) the impact of improper cleaning at the cutting plant on the *Salmonella* bacteriological prevalence in contaminated pork carcasses. In order to calibrate the experts, seed variables were chosen related to the same topics on *Salmonella* prevalence as well as to different pork production processes. The true values for the seed variables were obtained from published and unpublished data (Table 8).

An expert was considered to be a professional involved in the pork meat supply chain with an advanced knowledge of the epidemiology and/or microbiology of *Salmonella* in pigs (pork). This condition was assumed to be fulfilled by a selection of delegates of the 7th International Symposium on the Epidemiology and Control of Foodborne Pathogens in Pork (Safepork, 9-11 May, 2007, Verona, Italy). A list of the presenting authors (either poster or oral, n = 123) was obtained from the conference organisers. Those delegates having submitted an abstract on a *Salmonella* topic were contacted (n = 61). Nine additional experts having submitted an abstract on a different topic were also contacted as a result of a search on PubMed which revealed that they had at least one publication on a *Salmonella* topic as a first or second author. The questionnaire was pre-tested by two Belgian experts, not involved in the expert elicitation workshop.

---

[1] Questionnaires available at : http://www.metzoon.be

Table 7: Variables of interest elicited as part of an expert opinion workshop in 2007 for a quantitative microbial risk assessment on *Salmonella* in the pig production chain

Variable ID	Name
Question 1	*Salmonella* concentration during different steps in slaughterhouse :
V1	Unloading pigs from truck at the slaughterhouse
V2	Lairage
V3	Stunning and killing
V4	Scalding
V5	De-hairing
V6	Singeing
V7	Polishing
V8	Evisceration
V9	Splitting
V10	Meat Inspection
V11	Chilling
Question 2	*Salmonella* bacteriological prevalence (%) in subsequent steps at slaughterhouse (starting with a 7% prevalence when pigs are leaving the farm):
V12	Unloading at the abattoir
V13	When euthanized
V14	After singeing
V15	After polishing
V16	After evisceration
V17	After meat inspection
V18	Chilled carcass
Question 3 (V19)	Percentage of bacteriological positive pigs given these pigs are serologically positive (positive test agreement bacteriology and serology given serologically positive)
Question 4 (V20)	Percentage of pigs excreting *Salmonella* spp. after transport and lairage in the slaughterhouse (starting with a *Salmonella* excretion percentage of 5% when pigs are leaving the farm)
Question 5 (V21)	Increase of *Salmonella* spp. bacteriological prevalence among pig carcasses due to improper cleaning of the conveyor belt and work surface at the cutting plant
Question 6 (V22)	Relative contribution of *Salmonella* Typhimurium human salmonellosis cases due to the consumption of minced pork (versus non-minced pork)

Table 8: Seed variables elicited as part of an expert opinion workshop in 2007 for a quantitative microbial risk assessment on *Salmonella* in the pig production chain

Variable ID	Name	True value
S1[b]	Within-herd apparent sero-prevalence (in %) for *Salmonella* in Belgian pig herds in 2006.	36.8
S2[b]	Percentage of serological samples found positive for *Salmonella* spp. in Belgian pigs sampled in 2005 (used cut off 0.25 SP ratio).	39.25
S3[a]	Incubation time for the *Salmonella* detection test ISO 6579: 2002 during pre-enrichment in Buffered Peptone Water	18
S4[a]	Ideal pH of water in pigs' drinking water at the farm in order to reduce *Salmonella* spp. after adding acids to the water	3.9
S5[b]	Average duration for fasting pigs before slaughtering (in hours) in Belgium	17
S6[b]	Average number of slaughtered pigs per hour in 10 biggest slaughterhouses in Belgium	398.5
S7[b]	Average duration of pigs kept at the lairage in the slaughterhouse in Belgium (minutes)	126
S8	Duration of the singeing process (in seconds) in the slaughterhouse	10.6
S9[a,b]	Average pig's carcass weight in Belgium (kg)	82.5
S10[b]	*Salmonella* bacteriological prevalence through swabbing of 600 cm² of a pig's carcass in 5 large slaughterhouses: (a) after polishing	11.1
S11[b]	(b) after splitting	13.7
S12[b]	(c) of chilled carcasses	2.2
S13[a]	Minimum growth temperature (°C) for *S.* Typhimurium in pig meat	9
S14[b]	Average temperature in the working hall of the 11 largest Belgian cutting plants (°C)	9.6
S15[a,b]	Duration (minutes) for manipulation of pig carcasses in the working hall in Belgian cutting plants (half carcasses are cut into shoulder, back, belly, ham, and into smaller pieces)	38
S16[b]	Average number of pig meat servings per person per year in Belgium	11
S17[a]	Number of salmonellosis cases in EU in 2005 (EFSA, 2006)	176 395
S18[a,b]	Number of salmonellosis cases in EU in 2006	NA

The realisations of the seed variables retained in the study originated from unpublished Belgian data, mostly collected during the METZOON research project (METZOON, 2006). NA: true value not available at the time of elicitation.

[a] Seed variables discarded from the analysis.

[b] Seed variables included in the second mailed questionnaire.

4.2.1.2. Elicitation

A workshop was organised at the end of the first day of the Safepork conference. The Belgian QMRA model was orally presented in a plenary session earlier during the conference (Grijspeerdt et al., 2007). During the workshop, a short summary of the model was presented and clear instructions on how the questionnaire should be filled in were given to the workshop participants. To represent the uncertainties associated with the variables using PDFs, the experts were asked to provide a most likely value, a minimum and a maximum value for both the seed variables and the variables of interest (Vose, 2000; van der Gaag and Huirne, 2002).

For each question, the expert's self-rating of his/her competence was asked by means of a coloured scaled 5-point scoring box (Figure 7). All experts had to complete the questionnaire individually.

Figure 7: Self-rating of expertise-Box

Due to missing answers, a second questionnaire, containing questions on 13 seed variables was sent later by e-mail to a selection of experts having filled in at least the question on the *Salmonella* spp. prevalence on pig carcasses at various stages at the slaughterhouse (Question 2, V12-V18, Table 7). The answers to the questions in the second questionnaire were individually filled in by the experts and returned by e-mail.

4.2.1.3. Post-elicitation

Before the onset of the analysis, the answers were screened for inconsistencies and verified for misunderstanding of questions. For the aggregation of the experts' subjective PDFs, weights were computed using the classical model which has been implemented in the Excalibur software (Pro-version v 1.0, developed by TU Delft, R.M. Cooke; EXCALIBUR light version available at http://risk2.ewi.tudelft.nl/oursoftware/6-excalibur) in order to obtain one combined PDF.

A triangular distribution was chosen to model the expert opinions for both the seed and variables of interest, since this distribution is appropriate when little is known outside the experts' estimates for minimum, the most likely and the maximum value. The EXCALIBUR software was used to obtain the according 5%, 50% and 95% quantiles from the triangular distributions. These quantiles are necessary to allow the classical model to produce and evaluate a combined PDF for each variable as a weighted linear combination from the individual experts' PDFs. The experts' weights are derived from two quantitative performance measures, namely calibration and information, measured on seed variables as explained below (Cooke, 1991; van Rooij, 2005).

4.2.2. Calibration and information scores

The expert's assessments are statistically accurate, if for *N* assessed seed variables, 90% of the true values fall within this 90% probability interval. Calibration measures the statistical likelihood that the true values of the seed variables correspond with the experts' assessments. That is, the true values of the seed variables are sampled independently from distributions corresponding with the experts' PDFs. The calibration is measured per expert by a calibration score which is the *p*-value of a standard Chi-square goodness of fit test. Low calibration scores (close to zero) indicate that it is likely that the expert's probabilities are not statistically supported by the set of seed variables (Cooke and Goossens, 2008).

The information of a distribution refers to the degree to which an expert's PDF is concentrated or peaked. It requires to fit a density function to the expert's quantiles which has a minimal Shannon's information relative to a background measure. This background measure was taken as the uniform distribution over an intrinsic range for each variable. The intrinsic range for each variable corresponds to the smallest interval containing all the assessed quantiles and the realizations (true values for the seed variables) if available, of all experts, increased or decreased with 10%.

For each expert, an average relative information score for all variables is obtained by summing the information scores for each variable and dividing by the number of variables. Larger information scores are obtained when the experts elicit quantiles that are located closely together (Cooke and Goossens, 2008).

The classical model combines the calibration and information scores into a single overall combined weight, in which the calibration score dominates over the information score to determine the DM.

4.2.3. Decision makers

The decision makers used in this study are described below. For a more detailed description of the scoring rules related to these DMs, we refer to Cooke (1991) and Cooke and Goossens (2000).

The equal weight DM results from the simple arithmetic average of the experts' individual distributions, by assigning equal weight to each expert's density. When there are N experts, the weights for each PDF equals 1/N. If N experts have assessed a given set of variables, the equal weight DM's distribution is given by:

$$f_{eqdm,i} = \left(\frac{1}{N}\right) \sum\nolimits_{j=1}^{N} f_{j,i}$$

where $f_{j,i}$ is the density associated with expert j's assessment for variable i.

Additionally, two performance-based DMs are available in EXCALIBUR: the "global weight DM" and the "item weight DM". Both weighting schemes are subject to a proper scoring rule constraint. This means that the expert achieves his/her maximal expected weight if and only if his/her stated assessment corresponds to his/her true opinion (Cooke, 1991).

In the global weight DM, the weights per expert are defined by the normalized product of the calibration and the overall information scores on the seed variables, i.e. for expert j the same weight (w_j) is used for all variables. For variable i, the global weight decision maker's density is:

$$f_{gwdm,i} = \frac{\sum_{j=1}^{N} w_j f_{j,i}}{\sum_{j=1}^{N} w_j}$$

where $f_{j,i}$ is the density of expert j for variable i and w_j is the normalized weight of the j th expert.

In the item weight DM, the weights are determined per expert and per variable (item). Whereas in the global weight DM an overall measure of information is used, the item weight DM is sensitive to the experts' information score for each variable itself. The item weight decision maker's density for variable i is given by:

$$f_{iwdm,i} = \frac{\sum_{j=1}^{N} w_{j,i} f_{j,i}}{\sum_{j=1}^{N} w_{j,i}}$$

where $f_{j,i}$ is the density of expert j for variable i as above and $w_{j,i}$ is the normalized weight of the j th expert for variable (item) i.

The performance-based DMs are optimised by finding a set of weights such that its weight maximises the product of calibration and information. The optimisation is achieved by the following procedure. If a calibration score of an expert falls below a certain minimum or cut off level (α), the expert should be unweighted (i.e. receives zeroweight). The remaining experts are then pooled according to their normalized weights to obtain the DM. This is done in an iterative process and subsequently for each value of α, there will be a new DM (DM α) composed as a weighted linear combination of the experts with a calibration score above the cut-off value α. These DMαs are scored with respect to calibration and information. In the optimisation procedure of the DM, the cut-off level is chosen in a way that the unnormalized weight of the resulting DMα is maximal. This subset of experts is then used to provide the combined distributions for the variables of interest (van der Fels-Klerx et al., 2002).

Alongside the equal weight DM and the performance-based DMs, the user-weight DM assigned weights to each expert based on the expert's self-rating scores for the seed variables. Experts used a 5-point scale from no knowledge to high expertise to express their competence with respect to all the variables assessed. The user weights per expert were computed using the sum of the experts' self-rating scores for the set of seed variables. To obtain the normalized user weights per expert, the sum of the experts' self-rating scores was divided by the sum of the self-rating scores of all experts. The obtained weights were introduced in EXCALIBUR and the user weight DM was evaluated with respect to its calibration and information performance on the seed variables.

The different DMs in this study were evaluated by comparing the performance measures (calibration, information and combined weight) of the different DMs.

A robustness analysis on experts and seed variables was performed to check how the results would change by loss of a single expert or seed variable. Hereto, the experts and the seed variables, respectively were excluded from the analysis one at the time and the calibration and the relative information scores of the new DM were computed. If the difference between the new DM and the original DM is small, relative to the differences among the experts themselves, then the results are considered to be robust against the choice of experts (Cooke and Slijkhuis, 2003).

4.3. RESULTS

4.3.1. Participation of experts

Out of 70 participants at the Safepork symposium who were considered as *Salmonella* experts and who were invited at the expert elicitation workshop, 27 experts, originating from nine European countries (Belgium, Czech Republic, Denmark, Germany, Ireland, the Netherlands, Norway, Spain, UK), the USA and Canada, agreed to participate in the study and completed the first questionnaire. Among these 27 experts, 14 of them attended the workshop during the symposium. Ten experts who could not attend the workshop agreed to submit the questionnaire later on during the symposium, and three experts sent their completed questionnaire by mail or by e-mail shortly after the symposium.

Based on the questionnaire completion rate, a second questionnaire containing the majority of the seed variables was sent to 21 experts (on 27 experts). These 21 experts had all filled in at least question 2 (Table 7: V12-V18) which was considered to contain key variables necessary to complete the QMRA model.

This second questionnaire was returned by 11 experts (on 21 experts), originating from seven European countries (Denmark, Germany, Ireland, the Netherlands, Norway, Spain, UK), the USA and Canada. The 11 retained experts covered all the areas in the pork production chain (Table 9), but experts active in the animal production domain (n = 6) and in the consumer and public health part (n = 5) were predominant (several listings possible). Seven experts peer reviewed publications related to *Salmonella* in the pork production chain.

Table 9: Background of selected experts participating in an expert opinion workshop in 2007. Multiple listings possible with a total of 11 experts

		Number of experts	%
Activity in the pork production chain			
	Animal production	6	55
	Animal transport and slaughter	4	36
	Distribution and retail	2	18
	Consumer and public health	5	45
Organisation			
	Research institute/university	7	64
	Industry	2	18
	Government/policy	4	36
	Laboratory	2	18
Field of interest			
Laboratory	Diagnostics	6	55
	Detection	5	45
	Typing	4	36
	Antibiotic resistance	9	82
Field	Epidemiology	10	91
	Veterinarian	4	36
	Risk assessment	6	55
Statistician		2	18
Other		5	45

Table 10: Results of the performance assessment for 11 experts participating in an expert elicitation workshop in 2007. Performance was assessed using calibration, information and weights. Four different decision makers (DMs) were compared: the equal weight DM, the user weight DM, the global weight DM, and item weight DM.

Expert ID	Calibration[a]	Relative information[b]	Weight[c]	Information relative to EDM[d]	Normalized weight			Unnormalized weight
					Equal weighting[e]	User weighting[f]	Global weighting[g]	Item weighting[h]
Exp. 1	<0.0001	1.74	<0.0001	1.64	0.09	0.05	0	0
Exp. 2	<0.0001	1.44	<0.0001	1.34	0.09	0.11	0	0
Exp. 3	<0.0001	1.47	<0.0001	1.06	0.09	0.12	0	<0.0001
Exp. 4	0.018	0.99	0.018	0.71	0.09	0.05	0.10	0.018
Exp. 5	<0.0001	1.75	<0.0001	1.53	0.09	0.08	0	0
Exp. 6	<0.0001	1.58	<0.0001	1.35	0.09	0.09	0	0
Exp. 7	0.0008	0.77	0.0006	0.67	0.09	0.08	0	0.0006
Exp. 8	<0.0001	2.30	<0.0001	1.75	0.09	0.05	0	0
Exp. 9	0.019	1.34	0.025	0.91	0.09	0.09	0.13	0.025
Exp. 10	<0.0001	1.81	<0.0001	1.66	0.09	0.14	0	0
Exp. 11	0.21	0.67	0.14	0.56	0.09	0.14	0.77	0.14
Equal weight DM	0.49	0.24	0.12					
Global weight DM	0.39	0.44	0.17					
Item weight DM	0.62	0.54	0.33					
User weight DM	0.37	0.24	0.09					

[a] Calibration: score for the statistical likelihood to which the true values of the 11 seed variables correspond with an expert's assessments.

[b] Average relative information on the 11 seed variables.

[c] Product of calibration and information scores.

[d] Relative information of the experts with respect to the equal weight DM.

[e] Normalized weight used to obtain the equal weight DM by assigning equal weight to each expert.

[f] User weights for the user weight DM resulting from assigning normalized weights to the experts corresponding to experts' self-rating of expertise.

[g] Global weights used in the global weight DM based on the experts' calibration and overall information performance on the seed variables. Zero when the calibration score is below the cut-off level ($\alpha = 0.018$).

[h] Item weights used for the item weigh DM based on the experts calibration and information scores, depending on the expert and on the variables, in a way which is sensitive to the experts' information score to each variable. Zero when the calibration score is below the cut-off ($\alpha = 2.11 \times 10^{-5}$).

Table 11: Experts' robustness analysis for the performance-based weighting schemes: the global weight and the item weight decision maker (DM)

Expert excluded	Global Weight decision maker[a]			Item Weight decision maker[b]		
	Calibration[c]	Relative information[d]	Relative information to original DM[e]	Calibration	Rel. Information	Relative information to original DM
None	0.39	0.44	0	0.62	0.54	0
Exp.1	0.39	0.41	0.0012	0.62	0.52	0.0019
Exp.2	0.39	0.44	0.0003	0.62	0.53	0.0002
Exp.3	0.39	0.44	<0.0001	0.62	0.54	<0.0001
Exp.4	0.22	0.66	0.19	0.22	0.66	0.24
Exp.5	0.39	0.40	0.002	0.62	0.51	0.026
Exp.6	0.39	0.44	<0.0001	0.62	0.54	0.0002
Exp.7	0.39	0.43	0.0006	0.62	0.53	0.0008
Exp.8	0.39	0.43	0.0008	0.62	0.53	0.0016
Exp.9	0.39	0.44	0.063	0.39	0.48	0.10
Exp.10	0.39	0.44	0.0002	0.62	0.55	0.011
Exp.11	0.33	0.85	0.57	0.05	0.91	0.53

[a] DM based on the experts' calibration and overall information performance on 11 seed variables.

[b] DM based on the experts' calibration and information scores, depending on the expert and on the variables, in a way which is sensitive to the experts' information score to each variable.

[c] Calibration: score for the statistical likelihood to which the true values of the 11 seed variables correspond with the experts' individual assessments.

[d] Altered DM's average relative information on the seed variables.

[e] The DM's average relative information with respect to the original DM (all experts are included in the original DM).

4.3.2. Performance of experts

In total, 11 out of 18 seed variables were used to weight the experts' opinions since that was the number of seed variables assessed by all 11 experts. From the initial 18 seed variables, seven seed variables were dropped (Table 8): two seed variables were ambiguously defined (S4 and S13). For seed S17, pre-knowledge of the realisation by one of the experts was obvious while for S18 the true value was not available within the timeframe of the study. The experts' elicitations for three seed variables (S3, S9 and S15) were not provided by one of the 11 experts.

Table 10 shows the calibration, information scores and the experts' weights obtained after applying the different DMs (the equal weight DM, global weight DM, item weight DM and user weight DM) as well as the performance of the resulting DMs.

The best calibrated experts were experts 4, 9 and 11 (Table 10). The remaining experts had significantly lower calibration scores (< 0.01). The highest information score to the seed variables (2.30) was for expert 8 while expert 11 obtained the highest calibration score (0.21). However, the latter had the lowest information score of all experts (0.67). Despite this, the combined weight (unnormalized weight) of expert 11 was the highest of all experts resulting in the highest contribution to the performance-based DMs (global weight DM and item weight DM). In general, the information score for the variables of interest were higher than for the seed variables (results not shown). This was most pronounced with the best performing expert (expert 11) for whom the information score for all variables (1.24) was twice the information score on the seed variables only. The experts' relative information with respect to the equal weight DM indicates how well the experts agreed among themselves. These values ranged from 0.56 to 1.75, with an average value over all experts of 1.20. The experts were only slightly more informative relative to the uniform distribution over the intrinsic range (average over all experts = 1.44) than with respect to the equal weight DM. The fact that the experts are not more informative with respect to the uniform distribution over the intrinsic range that to the equal weight DM indicates there is little overlap between the experts 90% uncertainty ranges.

Among the DMs applied, the item weight DM obtained the highest calibration and information scores as compared to the global, equal and user weight DMs. All DMs obtained a higher calibration score than the best calibrated expert (expert 11). From all the evaluated DMs, it appears that the item weight DM can provide the best estimates of the combined PDFs for the variables of interest. In the optimisation procedure, a weight was allocated to three experts (4, 9 and 11) in the global weight DM, and five experts were retained in the item weight DM (3, 4, 7, 9 and 11), while all other experts received weight "0" (below cut off level). Experts 3 and 7 contributed little to the item weight DM.

The user weight DM obtained the lowest calibration and information scores. The equal weight and user weight DM obtained substantially lower information scores than the performance-based DMs.

4.3.3. Robustness analysis

Robustness analysis was used to identify the importance of each expert in relation to the DM. One expert at a time was excluded and then the relative information and calibration scores of the DMs containing the remaining experts were computed (Table 11).

The calibration scores of the item weight DM were lower if experts 4, 9 and 11 were removed (altered DM's calibration score of respectively 0.05, 0.22 and 0.39) (Table 10). This indicates that these experts contributed significantly to the results. Expert 11 was the expert who obtained the highest weight; removing this expert from the pool of experts resulted in a considerable loss in the calibration score of the item weight DM (from 0.62 to 0.05).

Perturbing the model by removing experts resulted in changes, but the information of the altered DMs (Table 11, columns 4 and 7) was still smaller than the inter-expert differences (ranging 0.56-1.75; Table 10, column 5). Robustness against the choice of experts was higher in the global weight DM than in the item weight DM, since only removal of experts 4 and 11 resulted in a small loss of the global weight DM calibration score.

A robustness analysis carried out on the choice of the seed variables, indicated that none of the seed variables was influential on the calibration score of the item weight DM or the global weight DM (data not shown).

4.3.4. Combined distributions

Using the weights resulting from the application of the DMs, combined distributions were constructed for both the seed variables and the variables of interest. These combined distributions are denoted as decision maker's distributions (DM distributions).

4.3.4.1. Seed variables

The 5th, 50th and 95th percentiles of the DM distributions for the 11 seed variables are shown in Figure 8. The true values of ten seed variables fell within the 90% uncertainty interval of the item weight DM distribution, whereas the realisation of seed S16 fell below the 5th percentile value. The uncertainty interval between the 5th and the 95th percentiles of the item weight DM were narrower in 4 out of 11 seed variables as compared to the global weight DM.

4.3.4.2. Variables of interest

One variable of interest (V22) was discarded from the analysis since this question was not well formulated and misinterpreted by a number of experts. Figure 9 represents the DMs distributions as expressed by their 5th, 50th and 95th percentiles for the first question concerning the $\log_{10}$ increase or decrease in colony-forming units (CFU) on a pig's carcass in subsequent processing steps at the slaughterhouse (V1-V11, see Table 7). The uncertainty intervals (5th to 95th percentile) were narrower for the item weight DM, than for the other DMs.

Figure 10 (V12-V18, see Table 7) shows the 5th, 50th and 95th percentiles of the DM distributions estimating the *Salmonella* spp. prevalence in subsequent processing stages at the abattoir. The uncertainty intervals in the combined distribution obtained by the item weight DM were similar to those using the global weight DM, but were smaller than those of the equal weight and user weight DM. This was observed in most of the production processes at the slaughterhouse (from singeing until chilling) indicating that combined distributions based on the performance based DMs were more informative. Figure 11a (V19, Table 7) shows that the 90% uncertainty intervals for the four DMs were almost identical. The estimates of the 50th percentile were slightly higher in the performance based DMs. The performance-based DMs produced higher estimates for the medians, than the equal weight and user weight DM (Figure 11b; with reference to V20, Table 7). For question of interest 5 (V21, Table 7), almost no difference in the uncertainty range was observed with respect to the combined PDF obtained by the item weight DM, the user and equal weight DM (Figure 11c).

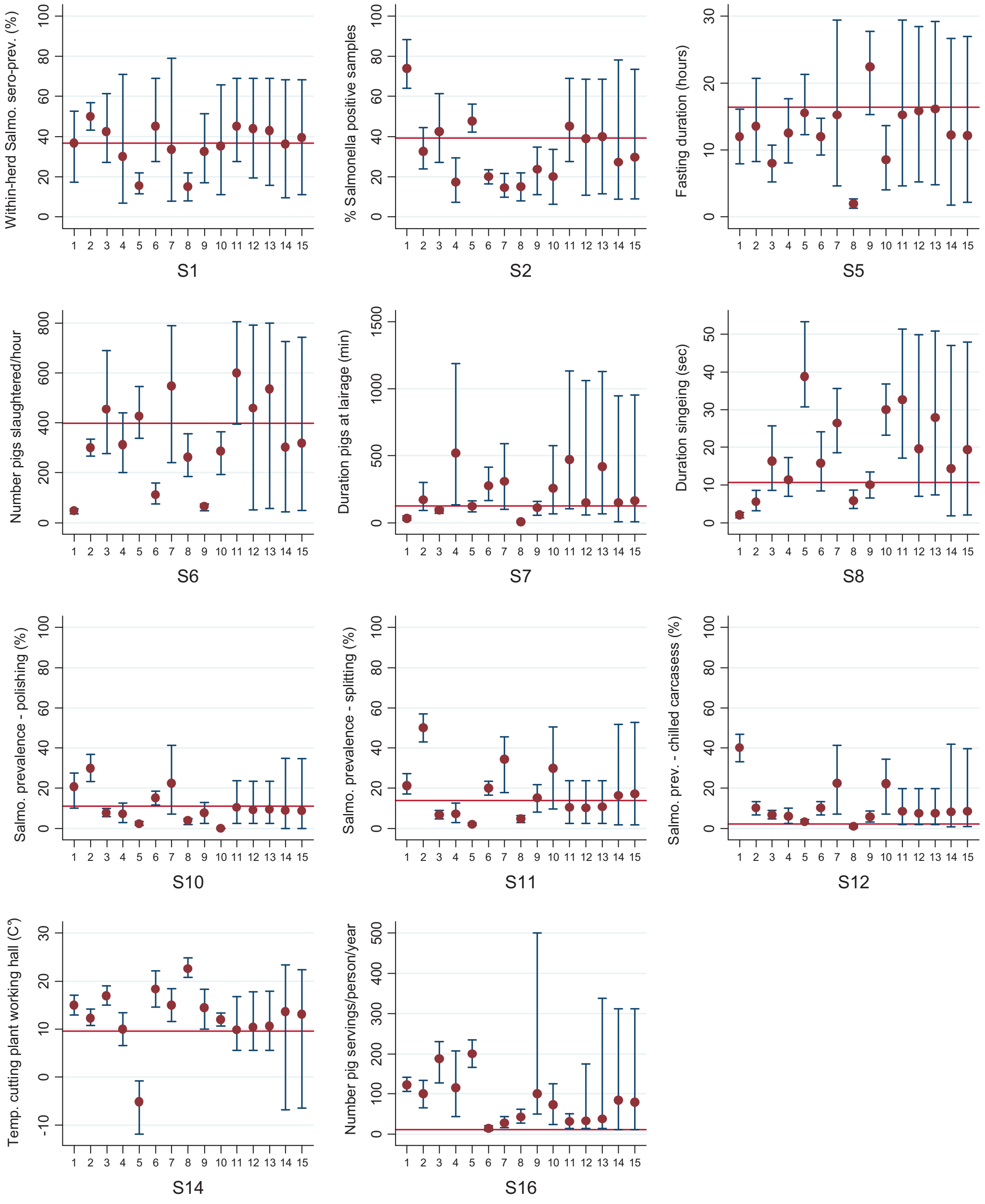

Figure 8: Distributions for the 11 seed variables based on 11 experts participating in an expert elicitation workshop in 2007.

Expert distributions (1, ..., 11) represented by their 5th percentile (lower cap), 50th (dot) and 95th percentiles (upper cap). Combined distributions derived from the item weight DM (12), the global weight DM (13), the equal weight DM (14) and the user weight DM (15). See Table 8 for an overview of the seed variables. The horizontal lines show the true values for the seed variables.

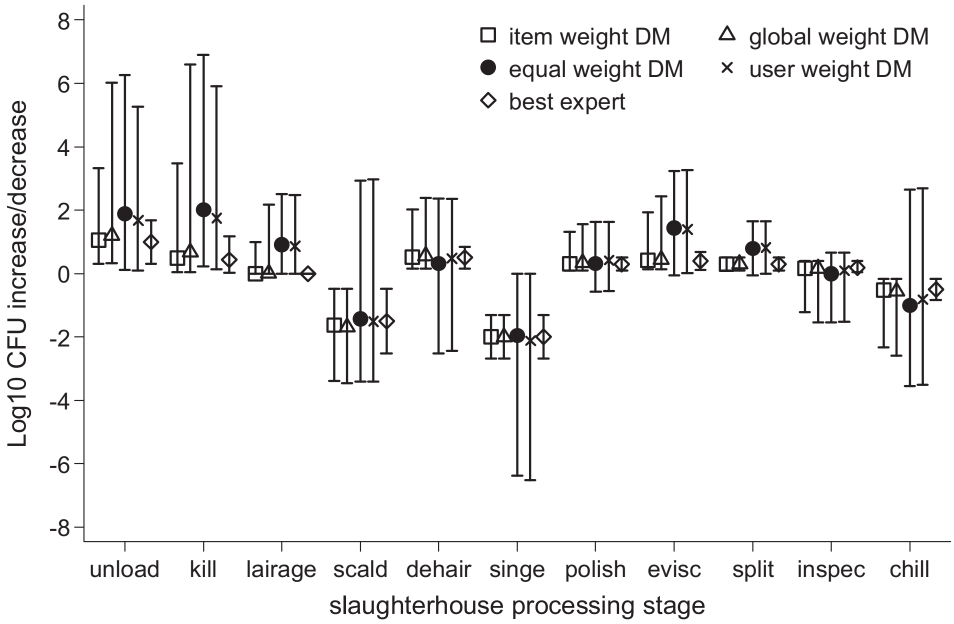

Figure 9: The decision maker's (item weight DM, global weight DM, equal weight DM, user weight DM) and the best expert's distributions estimating the *Salmonella* concentration increase/decrease in $\log_{10}$ CFU at abattoir processing stages.

Distributions expressed by the 5th, 50th and 95th percentiles. Unload = unloading pigs from truck slaughterhouse, kill = stunning and killing, singe = singeing, polish = polishing, evisc = evisceration, inspect = meat inspection, chill = chilling

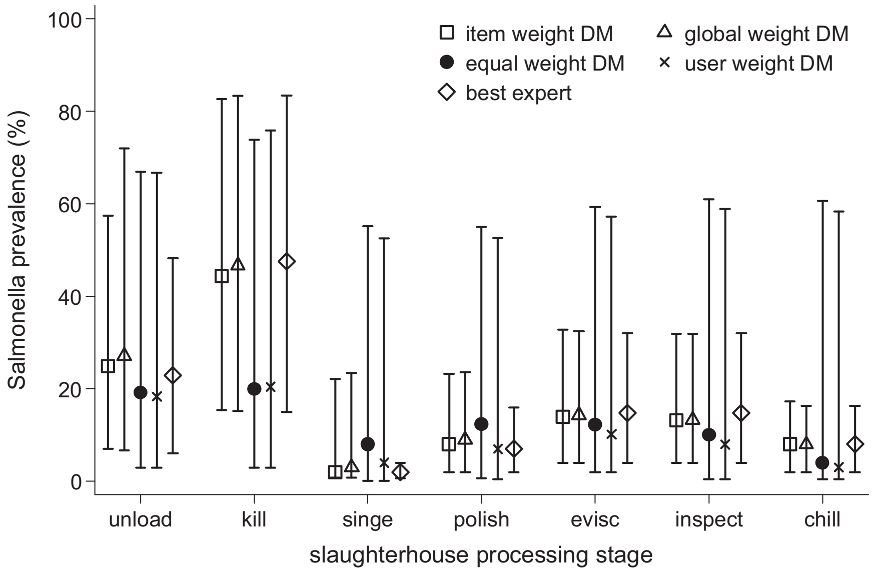

Figure 10: The decision makers' distributions (item weight DM, global weight DM, equal weight DM, user weight DM) and the best expert's distributions estimating the *Salmonella* prevalence at abattoir production stages. Distributions expressed by the 5th, 50th and 95th percentiles. A starting bacteriological prevalence of 7% (90 CI 6-8%) was assumed when pigs were leaving the farm.

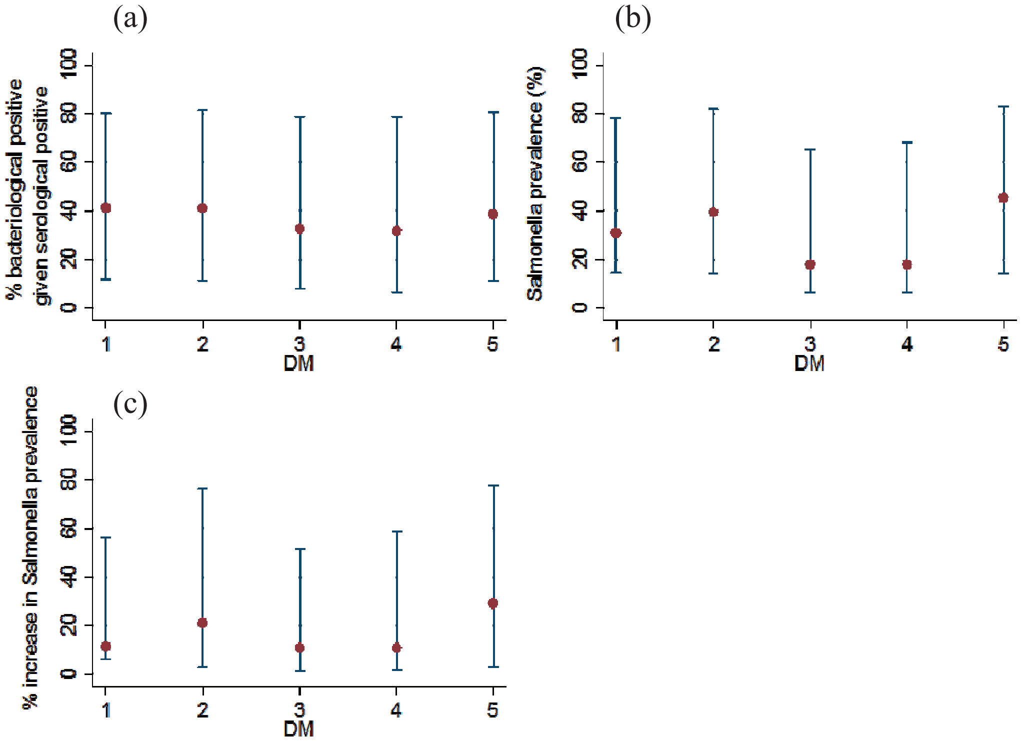

Figure 11: The combined distributions of four variables of interest expressed by the 5th, 50th and 95th percentiles derived from the item weight (1), the global weight (2), the equal weight (3), the user weight (3) decision maker (DM) and comparison of the best expert's (5) uncertainty distribution.
(a) The DM's distributions estimating the percentage of bacteriological *Salmonella* prevalence of a pig on the farm given it is also serologically positive (Table 7, question 3, V19); (b) the increase in bacteriological *Salmonella* prevalence (%) after transport and lairage (Table 7, question 4, V20); (c) the increase in bacteriological *Salmonella* prevalence (%) due to improper cleaning of the conveyor belt and work surface at the cutting plant (Table 7, question V21).

4.3.5. Experts' self-rating of expertise

By attributing the experts' self-rating of expertise as weights in the user weight DM, a DM was obtained with lower performance measures (calibration and information scores) as compared to the equal, global and item weight DM (Table 10). This suggests that in our study, the experts' self-expertise was on average not indicative for their performance to the seed variables as measured by the information and calibration scores (Spearman r = 0.37, P = 0.13). This means that some experts were over- and some under-confident. Experts 11, 9 and 4 were the experts with the largest performance-based weights. Agreement was found for expert 11, since this expert also obtained the highest weights related to his/her self-expertise. Expert 10 was an example of an over-confident expert who was unweighted in the performance-based DMs, but who had a self-expertise weight as high as that of expert 11. Expert 4 could be considered as an under-confident expert, with the third largest unnormalized performance weight (0.018, Table 10, column 4) in combination with low weight for self-assessment of expertise (0.05, Table 10, column 7).

4.4. DISCUSSION

In this study, a structured judgement approach was chosen in order to aggregate distributions provided by a panel of 11 experts into a one distribution using different weighting schemes (called decision makers, DMs). The aim was to enhance a rational consensus on the quality of combined distributions for variables necessary to complete a QMRA model. Different DMs for combining expert subjective probability distributions were compared.

The protocol used in this study was slightly different from the one described by Cooke and Goosens (2000) and van der Fels-Klerx et al. (2005), where experts were asked to present their PDFs by providing quantiles. The reason for eliciting expert judgements through minimum, most likely and maximum values in this study was that it is more straightforward (Vose, 2000) than elicitation by means of quantiles.

Several questions of our initial questionnaire could not be exploited due to missing answers, or misinterpretation of some questions. The missing answers were most likely caused by lack of knowledge. The misinterpretation of some questions was due to wording which might be have been interpreted differently by experts from different EU countries and/or continents. Large differences were observed between the calibration scores of the 11 experts retained in the analysis. The majority of the experts in the panel (8/11) were unweighted from the performance based DM because they obtained a calibration score below the cut-off level. The fact that not all experts performed well with respect to their calibration and information scores might be due to the selection of the seed variables, which might have been too country-specific for some of the experts. Secondly, some experts may have performed better if they were trained in quantifying their uncertainty in terms of subjective PDFs (Cooke, 1991).

Subsequently, experts originated from different backgrounds with large differences in the pork supply chain can misunderstand some questions. Stark et al. (2002) and van der Gaag and Huirne (2002) highlighted that with a QMRA model that is country specific, the judgements of local experts should be elicited in order to avoid bias. Despite the lack of local experts on the variables of interest, Cooke's classical model was used in the present study because it is able to adequately deal with a heterogeneous panel of experts (van der Fels-Klerx et al., 2002).

According to the experts' performance on the seed variables, the item weight DM obtained the highest calibration and information score, as compared to the global weight, equal weight and user weight DM. Both performance-based DMs outperformed the best expert in terms of unnormalized weight. Although there is no mathematical theorem that the performance-based DMs should outperform the equal weight DM, in practice the performance-based DM is usually better than the equal weight DM in most expert judgement studies using the classical model (Cooke and Slijkhuis, 2003; Cooke and Goossens, 2008).

The combined PDFs for the variables of interest obtained under the item weight DM can readily be used to provide the input parameters for the QMRA model, since this DM obtained the highest performance. The performance-based DMs were more informative than the equal weight and user weight DM.

The success of the implementation of the classical model depends to a large extend on finding the adequate seed variables. Indeed, the performance of the experts to the seed variables is judged as indicative for their performance on the variable of interest. The number of effective seed variables (11) in the present study was judged successful since Goossens and Cooke (2006) stated that 10 seed variables is sufficient. However, the more seed variables the better but the choice for finding the ideal set of seed variables was difficult in this study. Seed variables must resemble as much as possible the variables of interest, and the true values of the seed variables must be readily accessible to the analyst during the time of the study. This was obtained for those seed variables related to *Salmonella* prevalence, but could not be achieved for seed variables on *Salmonella* concentration data. In this study seven seed variables were discarded due to e.g. prior knowledge of an expert and/or missing values. Many experts argued that providing the estimates for the seed variables and variables of interest was difficult.

The user weight DM which was given weight according to the experts' self-rating of expertise, was judged unsatisfactory. A high or low self-rating of expertise did not automatically result in a high or low weight. Although the user weight DM was better calibrated than the best expert, its information score was the lowest of all the evaluated DMs. The user weight DM's weight was even lower than that of the equal weight DM. We conclude that the use of the self-rating of experts did not provide a rational objective basis in this study for weighting the experts. The performance on well chosen seed variables offers a more objective basis for weighting the experts.

The performance-based weights and the expert uncertainty distributions related to the bacteriological prevalence at the slaughterhouse obtained in this study (Question 2, Table 7) were used to fill in data gaps in the METZOON model (Bollaerts et al., 2009).

4.5. CONCLUDING REMARKS

The aggregated PDFs for the variables of interest obtained under the item weight decision maker can be used in the QMRA model for *Salmonella* in pigs and can be considered fit for purpose. This performance-based DM yielded the highest calibration score of all the DMs and generally produced more informative distributions than by using the other DMs.

The proposed protocol was judged useful to evaluate weighting schemes and to combine PDFs to provide input in future QMRA models, and is likely to enhance the quality of the QMRA process.

CHAPTER 5:

THE NUSAP SYSTEM: A TOOL FOR EVALUATING THE DATA QUALITY IN QUANTITATIVE MICROBIAL RISK ASSESSMENT

Modified from: Boone, I., Van der Stede, Y., Bollaerts, K., Vose, D., Maes, D., Dewulf, J., Messens, W., Daube, G., Aerts, M., Mintiens, K. (2009). NUSAP method for evaluating the data quality in a quantitative microbial risk assessment model for *Salmonella* in the pork production chain. Risk Analysis, 29: 502-517.

5.1. INTRODUCTION

A QMRA relies on the availability of numerical data in order to model food production and consumption pathways, to provide decision makers with science-based risk management options. The reliability of a QMRA depends on the quality of the model input parameters, the assumptions taken and the validity of the model. Very often, high quality data are not available due to diversity of biases, or due to incompleteness and missing data. Also, conflicting data may be found due to temporal and/or geographical variability (Lammerding and Fazil, 2000; Gardner, 2004; Havelaar, 2005; EFSA, 2007).

In general, a QMRA starts with the design of the different pathways process followed by making an inventory of candidate information sources that can provide the input parameters. This first step is essential but often in the QMRA there is no critical evaluation of the model input parameters. Doing this helps in identifying uncertainties pertaining to these parameters. In addition, it can allow risk assessors to judge parameters on their usefulness for the QMRA. Providing information on the strengths and weaknesses of the data is helpful for making decisions even when the highest quality data are not available. In addition, a documented quality evaluation avoids possible overconfidence in the QMRA model.

Several approaches were described in literature to analyze the quality of data (Funtowicz and Ravetz, 1990; EIIP, 1996; Tielemans et al., 2002; Sanchez et al., 2007). The novel approach chosen in the present study is known as the Numeral Unit Spread Assessment Pedigree (NUSAP) notational system (Funtowicz and Ravetz, 1990) (see also section 2.6.3.). NUSAP is a tool that helps decision makers in taking rational decisions based on adequate information on the uncertainties in the available scientific information provided by a risk assessment. Whereas mainstream uncertainty methods (e.g. Monte Carlo analysis, Bayesian updating) concentrate on quantifiable uncertainties, they do not address the unquantifiable uncertainties (related to problem framing, model structure, assumptions, value ladenness,…) (van der Sluijs et al., 2005a). The NUSAP system is able to address aspects of data quality resulting from uncertainties that are hard to quantify, such as methodological and epistemological uncertainties and which are not systematically taken into account in scientific studies. Moreover, it aims to promote a reflexive dialogue on the quality of information between risk analysts, stakeholders and decision makers (Craye et al., 2005). Within the NUSAP acronym, the pedigree qualifier is what is most innovative for assessing the quality of data. Crucial for a pedigree evaluation is the use of a pedigree matrix, which is expressed by scores on a discrete numeric scale from weak (score "0") to strong (score "4") for a set of criteria that determine the quality of data. To each score in the matrix, a label is attached giving a description of the scale (van der Sluijs et al., 2005a). The pedigree scores to assess the quality of data are obtained by qualitative expert judgement. Although the NUSAP pedigree method has been successfully applied to various complex models, mostly in environmental risk assessment (Krayer von Krauss et al., 2004; van der Sluijs et al., 2005a), it has not been applied before to a QMRA.

In the present study, the NUSAP pedigree method was applied for screening and evaluating the data quality of potential input parameters in a Belgian QMRA model on human salmonellosis due to the consumption of fresh minced pork meat (the METZOON model). For a description of this QMRA model the reader is referred to Grijspeerdt et al. (2007).

5.2.MATERIAL AND METHODS

5.2.1. Parameters for the Belgian QMRA model

The Belgian *Salmonella* QMRA model, called the METZOON model included the following successive modules as a backbone in the exposure pathway: (1) primary production, (2) transport, holding and slaughterhouse, (3) post processing, distribution and storage and (4) preparation and consumption. These modules have been supplied with available input parameters originating from various information sources such as (inter)national scientific literature, data from official monitoring and surveillance programs, and expert opinions.

The information sources were identified through a questionnaire survey (Table 12) which was sent to potential database owners on *Salmonella* in the Belgian (pig) pork production chain and/or on human consumption behaviour in Belgium. These database owners included the METZOON-consortium partners, governmental authorities, semi-official organizations, and industrial companies. Based on this survey, a total of 101 potential parameters were further identified (Table 14) and registered in an Access database (Microsoft Office Professional Edition 2003). The number of parameters introduced into the database corresponded to those available during the first build-up stage of the risk assessment model. The parameters were expressed by means of an identity-card (ID-card), containing the parameter characteristics and their value information (Table 13).

Table 12: Inventory and characterization of the variables in the questionnaire for sources of information in QMRA

Questionnaire part	Specifications
Contact details	Reporter of the database / Affiliation Owner of the database Administrator of the database
General database information	Name of the database Format (MS Excel, MS Access, Oracle, My SQL, etc.) Recommended format: XML Purpose of the database Type of study: cross-sectional, longitudinal, case-control, cohort, experimental, other Users of the database Accessibility
Database coverage information	Compulsory or voluntary program Timeframe Geographical coverage of the data Data related to which module in the model
Database contents	Presence of links to other databases via relational identifiers Size of the database (number of variables and records) Information on zoonotic agents contained in the database Sampling method (simple random sample, stratified, systematic,...) Format cell consistency (date, number,…) Error checking of the database References of reports or journal articles for which the database has been used

Electronic questionnaire format is available at: http://www.metzoon.be

Table 13: Characterization of the input parameters as foreseen in the ID-card

General	Sampling	Value*
Reference	Target population	Central tendency
Name	Study population	Spread
Parameter identification number	Sampling year	Distribution
	Sampling unit	Unit
	Sampling frame	
	Sample frequency	Information on validation
	Sampling method	
	Sample size	
	Sample type	
	Test method	
	Diagnostic test sensitivity – specificity	
	Missingness, non-response rate	

Electronic questionnaire format is available at : http://www.metzoon.be
*The value part corresponds to Numeral, Unit and Spread in the NUSAP system.

5.2.2. NUSAP pedigree matrix for the evaluation of data quality

A NUSAP pedigree matrix, adapted from van der Sluijs et al. (2005c) was used (Table 15). The pedigree criteria were proxy, empirical basis, methodological rigour, and validation. The proxy criterion evaluates the closeness of resemblance between the input parameter available from the data source and the actual variable that would be required in the model. The empirical criterion evaluates the degree to which direct observations were used to estimate the input parameter. A higher pedigree score for empirical basis was attributed to input parameters obtained from field data compared with indirect, modelled data or data obtained by expert judgement. The methodological rigour refers to the norms used in the collection and checking of data and the degree of acceptance of these norms by the peer community of the relevant discipline. Lastly, the validation criterion evaluates the degree to which one was able to cross-check the data against independent sources.

The 101 input parameters were assessed by individual experts of the METZOON consortium by using the NUSAP pedigree matrix and attributing pedigree scores (from “0” to “4”) to each of the parameters. The experts were allowed to motivate and/or comment the rationale for their scores which is of use for further selecting or improving the quality of input parameters. In addition, the experts could indicate their degree of expertise pertaining to each parameter. All the scores and additional comments were directly completed in MS Access query forms which were compiled in an MS Access database. The expert panel from the METZOON consortium was composed of 12 persons involved in the primary production, transport and slaughter, distribution and retail, consumption and public health phase, as well as statisticians. The methodology was pretested on 12 parameters randomly selected along the production pathways, during a workshop with 10 partners from the METZOON consortium. The definitions of some criteria—proxy representation and empirical basis in particular—were clarified and fine-tuned during the pre-test workshop. Some essential keywords, according to the experts, were added in the pedigree matrix such as aspects of geographic representativeness within the proxy criterion and temporal representativeness within the empirical basis criterion. To capture the absence of scoring, two categories were distinguished (1) due to insufficient information provided in the parameter ID-card and (2) due to an insufficient level of expertise. The revised pedigree matrix is also shown in Table 15 (underlined).

Table 14: List of potential input parameters for the QMRA model

Parameter identification number	Parameter	Reference
	Primary production	
1	Apparent prevalence of *Salmonella* in pig feed (Walloon Region, Belgium)	Korsak et al. (2003)
2-4	Apparent prevalence of *Salmonella* in pig feed: Federation for Belgian Feed Manufacturers (2); auto control data from members (3)	http://www.bemefa.be
	Database from the Federal Agency for Safety of the Food Chain (FASFC) (4)	EFSA (2006a), FAVV (2006)
5	Time period from weaner to finisher pig	Belgian expert opinion
6, 8	Number of pigs per pen in a pig farm (6); number of pens per pig farm (8): UK data	Hill et al. (2003)
7, 9	Number of pigs per pen (7); number of pens per pig farm in Belgium (9)	Belgian expert opinion
10	Number of sources to replace stock on a pig farm	METZOON (2006)
11, 12	Shedding duration of *Salmonella* in infected pigs (11); duration of carrier status in carrier pigs (12)	Hill et al. (2003)
13	Apparent prevalence of *Salmonella* in Belgian sows herds (Walloon Region)	Nollet et al. (2005)
14, 16, 17, 19	Apparent prevalence of *Salmonella* in Belgian sows (14), weaners (16), growers (17) fattening pigs (19) (Flemish Region, herd-level)	Animal Health Care (2005)
15, 18	Apparent prevalence of *Salmonella* in Belgian sows herds (15) and fattening herds (18) (national level)	Van Vlaenderen et al. (1999)
24-30	Apparent prevalence of *Salmonella* in closed pig production system (animal-level)	Korsak et al. (2003)
36	Apparent prevalence of *Salmonella* in Flemish farms (Belgium)	Animal Health Care (2005)
37-39	Apparent prevalence of *Salmonella* in pig herds at herd-level and animal-level (Flanders, Belgium)	Botteldoorn et al. (2003)
31	Within-herd apparent prevalence *Salmonella* in French sow herds	Beloeil et al. (2003)
32	Within-herd apparent prevalence of *Salmonella* in Belgian sow herds	Nollet et al. (2005)
35	Within-herd apparent prevalence of *Salmonella* in Belgian pig farms (mesenterial lymph nodes)	Huysmans et al. (2003)
33	Within-herd apparent sero-prevalence of *Salmonella* in Belgian sow herds	Nollet et al. (2005)
34	Within-herd apparent sero-prevalence of *Salmonella* in Belgian pig farms	Huysmans et al. (2003)
20	Apparent sero-prevalence of *Salmonella* in COVAVEE-farms (Pig Production Cooperation) (herd-level)	Huysmans et al. (2003)
21-23	Apparent sero-prevalence of *Salmonella* in Belgian pig herds (21, 22: herd level) (23: animal-level)	Van der Stede et al. (2007)
	Transport, holding, slaughterhouse	
40	Proportion of pigs that were fasted previous to slaughter	De Sadeleer (2008)
41	Transport time of pigs from farm to lairage	De Sadeleer (2008)
42	Time of pigs in lairage	De Sadeleer (2008)
43-49	Average number of colony forming units (CFU) per 100 cm² carcass swab after killing (43), scalding (44), dehairing (45), singeing (46), polishing (47), evisceration (48) and after splitting (49)	Hill et al. (2003)
50-53	Proportion of positive *Salmonella* samples in lairage (overshoes: 50), scalding tank (sterile cotton tube, 51), splitting machine (swabs of knife blade: 52), chilling room (overshoes, 53)	Botteldoorn et al. (2003)
54	Frequency of disinfection of the carcass splitter between two carcasses	Delhalle et al. (2008)
55	Temperature of the water in the scalding tank	Delhalle et al. (2008)
56	Time period for a carcass remaining in the scalding tank	Delhalle et al. (2008)
57	Frequency of disinfection of the dehairing machine	METZOON (2006)
58	Time period of singeing	Delhalle et al. (2008)
59-62	Prevalence of *Salmonella* in pig carcass swabs (600 cm²) at end of slaughtering process (59, Walloon region; Flemish region, 62); in cooling room (Flemish region: 60; Walloon region: 61)	Korsak et al. (2003) Huysmans et al. (2003) Botteldoorn et al. (2003)
63	Prevalence of *Salmonella* in pig carcasses in Belgian slaughterhouses	FAVV (2006), Ghafir et al. (2005)
64	Semi-quantitative enumeration of *Salmonella* on a pig's carcass	Ghafir et al. (2005)

(Continued)

Table 14: List of potential input parameters for the QMRA model (*continued*)

Parameter identification number	Parameter	Reference
	Postprocessing, distribution, and storage	
65	Minimum growth temperature *Salmonella* Typhimurium in pork	Hill et al. (2003)
66	Temperature of carcass during transport from slaughterhouse to storage room in cutting plant	METZOON (2006)
67	Carcass storage in Belgian cutting plant (time)	METZOON (2006)
68, 69	Temperature in storage room (68), ambient temperature in working hall (69) at cutting plant	De Sadeleer (2008)
70	Duration of manipulation carcasses in the working hall of cutting plant	De Sadeleer (2008)
71, 76	Semi-quantitative enumeration of *Salmonella* in cut meat parts at cutting plant (71), in minced meat (76)	Ghafir et al. (2005)
72	Prevalence of *Salmonella* in cut meat in the work area where the meat is being cut in cutting plants	Ghafir et al. (2005)
73-75	Prevalence of *Salmonella* of minced meat processed in processing plant (73), at retail (74), at processing plant and distribution centre (75)	EFSA (2006a)
77	Temperature in the second storage room of the cutting plant	METZOON (2006)
78	Duration of carcass storage in the second storage room of the cutting plant	METZOON (2006)
79	Temperature in truck during transport meat from cutting plant to distribution centres	METZOON (2006)
80	Transport time from working hall until display pork for customer in distribution centres	METZOON (2006)
81, 82	Temperature in storage room (81), in working hall (82) at distribution centres	METZOON (2006)
83	Time lapse between shopping and storing of raw meat in the fridge/freezer (Ireland)	Kennedy et al. (2005)
84	External temperature	RMI
85	Percentage of people not respecting the use-by date of food product	Devriese et al. (2006)
86	Percentage of people unfreezing pork in fridge (Belgium)	Devriese et al. (2006)
87	Percentage of people unfreezing pork at room temperature	Kennedy et al. (2005)
88; 89	Temperature minced meat (88) and ham (89) in fridge in Sweden	Marklinder et al. (2004)
90	Temperature in fridge (lowest drawer) in Belgium	Devriese et al. (2006)
91	Hourly log growth of *Salmonella* Typhimurium in sterile ground chicken breast (Oscar growth model)	Hill et al. (2003)
	Preparation and consumption	
92	Total weight of pig meat per person/year bought for home consumption	VLAM
93	Average daily consumption of meat in Belgium	Devriese et al. (2006)
94	Total net weight pork in kg/person for Belgian market	NIS
95	Average weight of consumed vegetables in Belgium	Devriese et al. (2006)
96	Probability washing hands after handling poultry meat	FAO/WHO (2002b)
97	Number of person washing correctly hands or rinsing after manipulating raw meat or poultry in Belgium	Devriese et al. (2006)
98	Frequency of eating pork meals per year per person in the UK	Hill et al. (2003)
99, 101	Percentage of people nearly (99) and never (101) eating meat in Belgium	Devriese et al. (2006)
100	People eating pork in Belgium	NIS + crude estimations

Parameters that have been underlined in the first column have been selected for inclusion in a preliminary version of the QMRA model (Grijspeerdt et al., 2007). NIS: National Institute for Statistics, RMI: Belgian Royal Meteorological Institute, VLAM: Flanders Agricultural Marketing Board, based on GfK Panel Services Benelux (unpublished).

Table 15: NUSAP pedigree matrix used to score the parameter strength

	Pedigree criteria			
Score	Proxy	Empirical basis	Methodological rigour	Validation
4	Exact measure of the desired quantity (e.g. measurements from the same geographically representative area as that being investigated)	Large sample direct measurements, recent data, controlled experiments	Best available practice in well-established discipline (accredited method for sampling / diagnostic test)	Compared with independent measurements of the same variable over long domain, rigorous correction of errors
3	Good fit or measure (e.g. measurements used from another geographical area but representative)	Small sample, direct measurements, less recent data, uncontrolled experiments, low non-response rate	Reliable method common within established discipline, best available practice in immature discipline (sampling / diagnostic test)	Compared with independent measurements of closely related variable over shorter period
2	Well correlated but not measuring the same thing (e.g. large geographical differences, less representative)	Very small sample Modelled/derived data / indirect measurements, structured expert opinion	Acceptable method but limited consensus on reliability	Compared with measurements not independent, proxy variable, limited domain
1	Weak correlation (e.g. very large geographical differences, low representativeness)	One expert opinion, rule of thumb estimate	Preliminary methods with unknown reliability	Weak very indirect validation
0	Not clearly correlated	Crude speculation	No discernible rigour	No validation
A				
B				

Underlined keywords are changes made in comparison to the pedigree matrix by Risbey et al. (2001a). Rows A and B were used to register missingness in two categories: A = no score due to insufficient information, B = no score due to insufficient expertise.

5.2.3. Analysis of pedigree scores

The scores attributed by the experts were analyzed in two ways: (1) by assessing criteria and overall pedigree strength, (2) by graphical presentations.

5.2.3.1. Analysis of scores through overall pedigree strength

The strength of all the 101 potential input parameters was calculated taking into account: (1) the expertise of the experts, (2) the consistency in scoring between experts and (3) the number of experts attributing scores to a specific pedigree criterion of a parameter. Each candidate input parameter *i* (*i* = 1,…,101) was scored with respect to *j* pedigree criteria (*j* = 1,…,4) by *k* experts (*k* = 1,…, N_{ij}). The scores for each pedigree criterion x_{ijk} can take integer values ranging from 0 to 4.

The participating experts indicated their degree of expertise concerning each parameter *i* on a 3-point scale (low, average or high, corresponding to weights with a numerical value of 1, 2 and 3, respectively). To account for the differences in expertise between experts, a weighted mean of the scores x_{ijk} is calculated for each criterion *j* of each parameter *i* with the weights reflecting the expert's expertise or:

$$\bar{X}_{ij}^{*} = \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} \frac{w_{ik}}{w_{max}} x_{ijk}$$

for each *i* = 1,…,101; *j* = 1,…,4, where w_{ik} is the expertise of expert *k* concerning parameter *i*, and $w_{\max}$ is the theoretical maximum level of expertise (= 3).

The consistency in rating was taken into account by multiplying the weighted mean $\bar{X}_{ij}^{*}$ with C_{ij} defined as:

$$C_{ij} = 1 - \frac{s_{ij}}{E_{ij}}$$

for each *i* = 1,…,101; *j* = 1,…,4, where s_{ij} is the standard deviation (*SD*) of the scores, or

$$\sqrt{\frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} \left(x_{ijk} - \bar{x}_{ij}\right)^{2}}$$

with $\bar{x}_{ij}$ being the ordinary mean of the scores and where E_{ij} is the maximal *SD* possible when N_{ij} experts attribute scores to a pedigree criterion of a parameter. Denote the minimum score that can be given to pedigree criteria *j* of input parameter *i* as n_{ij} and the maximum score as M_{ij}. In case N_{ij} is an even number, the maximal *SD* E_{ij} is obtained when $N_{ij}/2$ of the experts give the minimum score m_{ij} and the remaining experts the maximum score M_{ij}. In case of an odd number of experts, the maximal standard deviation E_{ij} is obtained when $(N_{ij}+1)/2$ of the experts give the minimum score m_{ij} and the remaining experts give the maximum score M_{ij} (or the inverse). Then, the maximal *SD* E_{ij} is given as:

$$E_{ij} = \begin{cases} \dfrac{M_{ij} - m_{ij}}{2} & \text{in case } N_{ij} \text{ is even number} \\ \sqrt{\dfrac{(N_{ij}+1)(N_{ij}-1)}{N_{ij}}} \left(\dfrac{M_{ij} - m_{ij}}{2}\right) & \text{in case } N_{ij} \text{is odd number} \end{cases}$$

When all experts provide the same score, the *SD* s_{ij} is 0, and C_{ij} equals 1. When the scoring experts maximally disagree, the *SD* s_{ij} equals the theoretical maximum E_{ij}, and C_{ij} is 0. An alternative and frequently used approach to account for (rater) consistency is to divide the mean score by the *SD* of the scores. However, this approach suffers from a major drawback as it yields infinite results when the standard deviation equals 0.

Next to the differences in expertise and consistency in scoring, the number of experts attributing a score to a pedigree criterion of a parameter was accounted for by multiplying the weighted mean $\bar{X}_{ij}^{*}$ with:

$$R_{ij} = \frac{N_{ij}}{N}$$

for each *i* = 1,…,101; *j* = 1,…,4, where *N* is the number of experts involved in the study (here *N* = 10, i.e. 10/12 experts returned the parameter ID-cards). In case all 10 experts involved in the study give scores to pedigree criterion *j* of parameter i, R_{ij} equals 1 whereas R_{ij} equals 0 in case no scores are given.

The strength of each criterion *j* of each parameter *i* was calculated as:

$$S_{ij} = \frac{1}{\max(x_{ij})} R_{ij} C_{ij} \bar{X}_{ij}^{*}$$

for each i = 1,…,101; j = 1,…,4, where $\max(x_{ij})$ is the highest score possible that can be given for each criterion implying that S_{ij} equals 1 if all experts involved in the study attribute the highest score to criterion j of parameter i ($\max(x_{ij})$ = 4). Finally, the overall strength or the quality of a parameter was calculated as the mean of the strengths of its four pedigree criteria or:

$$S_i = \frac{1}{4}\sum_{j=1}^{4} S_{ij}$$

for each *i* = 1,…,101.

In the calculation of the overall strength, equal weight was attributed to the strengths of all four pedigree criteria, because we did not have a scientific justification to apply a weighted average. Clearly, the overall parameter strength S_i equals 1 if all experts involved in the study attribute the highest score possible to all pedigree criteria of parameter i.

The overall pedigree scores were computed using SAS 9.1. (SAS Institute, NC, USA). Wilcoxon Mann-Whitney test (Lehmann, 1975) using StatXact 4 (Cytel Software

Corporation, Cambridge, MA) with the Bonferroni (Shaffer, 1995) corrected alpha values was used to test pairwise comparisons in strengths among the exposure pathways and among the pedigree criteria, and to compare the average expertise score of experts scoring a parameter between the exposure pathways. The family-wise significance level was set at alpha = 0.05, with six pair-wise comparisons, the Bonferroni corrected significance level was set at 0.05/6 = 0.0083.

5.2.3.2. Graphical analysis of pedigree scores

For each parameter the resulting NUSAP pedigree scores were represented by a kite diagram (Risbey et al., 2001a). This type of diagram aims to quickly identifying low versus high pedigree scores. This gives opportunities to look for quality improvement in a fast and comprehensive way. Per parameter, each of the four pedigree criteria was placed on one of the axes of the diagram, with the lowest score (0) in the centre and the highest score (4) on the corner point of the polygon. Four colours were used to visualize the scores elicited by the experts (Figure 13).

The minimum scores among the group of scoring experts were depicted as a green area to reflect the minimal consensus of each parameter. A light green area was used to indicate the minimum score given by the expert panel if the expert who attributed the lowest score for a specific criterion would have been excluded. Scores from the minimum up to the maximum given by the expert panel were represented by an amber surface in order to reflect expert disagreement for one or more pedigree criteria. The remaining area up to the borders of the kite diagram was coloured red. The larger red area in the kite diagram represents lower pedigree strength for one or more criteria (Kloprogge et al., 2005; van der Sluijs et al., 2005c). The kite diagrams were constructed using the interactive “kite diagram maker” (van der Sluijs, 2007).

As pointed out by Wardekker et al. (2008), the conformation of the area sizes is dependent on the arrangements of the criteria in kite diagrams.

5.3. RESULTS

5.3.1. Parameter ID Cards

The completed ID-cards of the 101 input parameters containing the quality assessment were returned by 10 experts (response rate = 83%) three weeks after they were sent out. The expert disagreement on elicited scores was large in most parameters for the proxy, the empirical and the method criteria, while it was small for the validation criterion. For the latter criterion the lowest pedigree score was given unanimously in 31.7% of the scored parameters. Low nonresponse rates were observed for the proxy (0.4%), empirical (8%), and the method (16%) criterion, whereas for the validation criterion the nonresponse rate was substantially higher (51%). Within the transport, holding and slaughterhouse module, 57% of scores related to validation were missing among the experts. As a reason for nonresponse in the proxy and the empirical criterion, the experts indicated that not enough information was provided in the ID-cards to allow them to attribute a score. In addition, a second reason for nonresponse was that the experts did not have enough expertise to score for the method (36% of the missing scores) and the validation criterion (23% of the missing scores).

5.3.2. Analysis of pedigree scores

Figure 12 shows an overview of the criterion strengths and the overall strengths of the assessed parameters. The overall parameter strength varied between 0.04 and 0.39. Fifteen parameters obtained overall strength scores below 0.1, whereas 52 parameters were given scores ranging between 0.1 and 0.2, and in 34 parameters the scores were equal or higher than 0.2. A large variability was observed in the strengths for proxy, empirical and method indicate, whereas this was not observed with respect to the validation criterion, except for two parameters with strengths higher than 0.29.

Significantly lower strengths were obtained for the validation criterion compared to the other criteria ($p < 0.001$: Table 16). Moreover, significant differences were observed by comparing each of the exposure pathways within each criterion. Indeed, the strengths for the proxy and empirical criterion were higher in the primary production module as compared to the transport, holding and slaughterhouse module, the post-processing, distribution and storage module and the preparation and consumption module ($p < 0.001$). The methodological rigour was significantly higher in the primary production pathway compared to the transport, holding and slaughterhouse module ($p < 0.001$), which was in turn characterized by parameters with a lower score than those in the postprocessing, distribution and storage module.

The parameters obtained lower scores for validation in the transport, holding and slaughterhouse pathway compared to the primary production ($p < 0.001$), the postprocessing, distribution and storage module ($p = 0.007$) and the preparation and consumption module ($p = 0.001$). By comparing the overall strength among the different exposure pathways, significantly higher strengths were observed in the primary production module compared with the other pathways (Table 16). The same is true for the average self-assessment of expertise associated with the parameters.

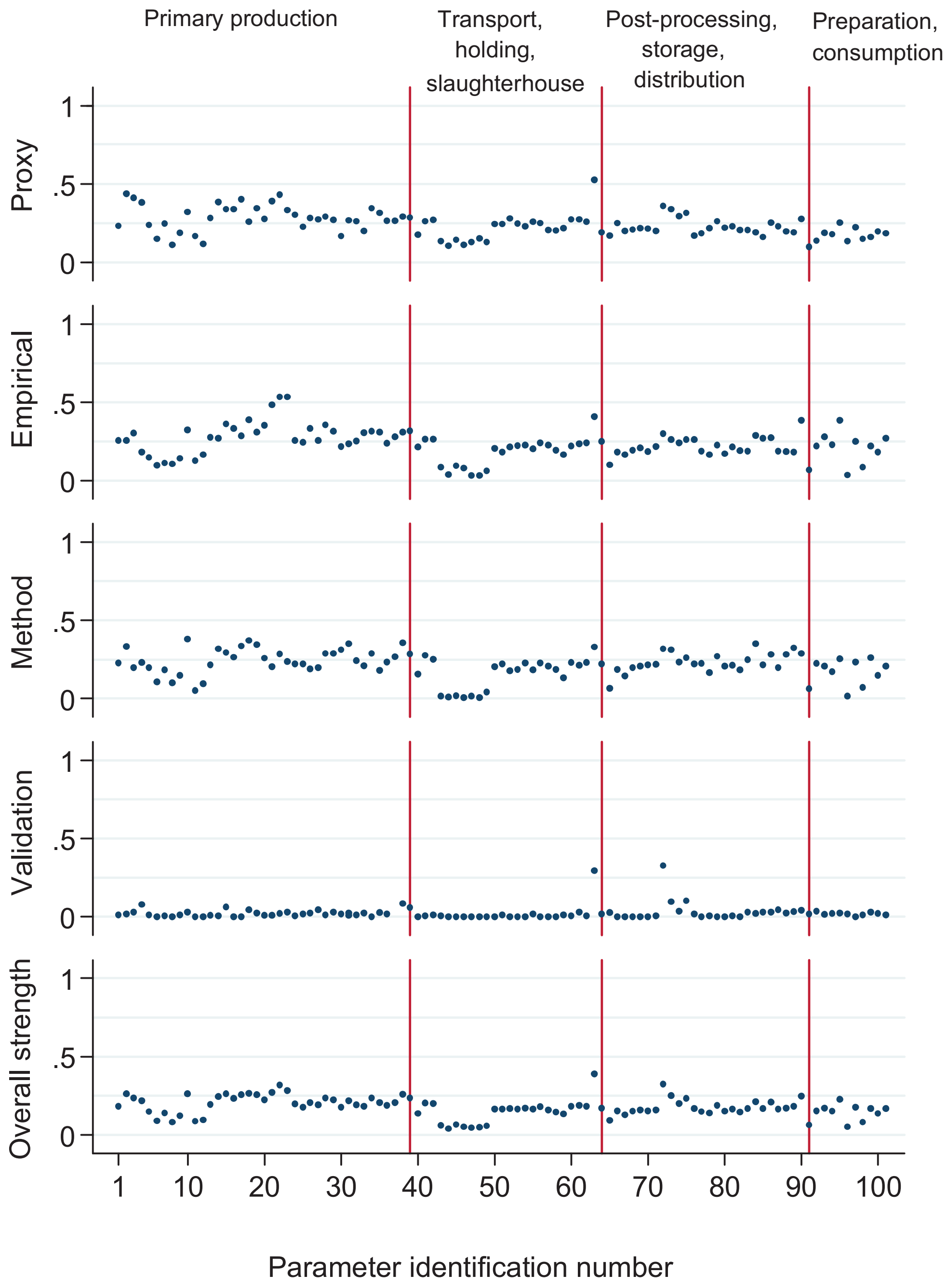

Figure 12: Overall parameter strength (bottom diagram) and pedigree criterion strength (upper 4 diagrams) of 101 parameters distributed according to the production pathways (1) primary production (parameters 1–39), (2) transport, holding, and slaughterhouse (parameters 40–64), (3) postprocessing, distribution, and storage (parameters 65–91), and (4) preparation and consumption (parameters 92–101) (Table 14). Calculation of the strength was done as described in Section 5.2.3.1.

Table 16: Summary statistics for the strength of each criterion and overall strengths within the exposure pathways

	Criterion									
	Proxy		Empirical		Method		Validation		Overall strength/pathway	
Exposure pathway	Median	Range	Median	Range	Median	Range	Median	Range	Median	Range
Primary production (n=39)	0.28^{a}	(0.11-0.44)	0.28^{a}	(0.10-0.53)	0.24^{a}	(0.05-0.38)	0.02^{a}	(0-0.08)	0.22^{a}	(0.08-0.32)
Transport, holding, and slaughterhouse (n=25)	0.23^{bc}	(0.11-0.53)	0.21^{b}	(0.03-0.41)	0.19^{b}	(0.08-0.33)	0.00^{b}	(0-0.29)	0.16^{b}	(0.04-0.39)
Post-processing, distribution, and storage (n=27)	0.21^{b}	(0.10-0.36)	0.19^{b}	(0.07-0.39)	0.22^{ac}	(0.06-0.35)	0.02^{a}	(0-0.33)	0.17^{b}	(0.06-0.33)
Preparation, and consumption (n=10)	0.18^{c}	(0.13-0.25)	0.22^{ab}	(0.03-0.39)	0.21^{ab}	(0.02-0.26)	0.02^{a}	(0-0.04)	0.16^{b}	(0.05-0.23)
Overall criterion score*	0.24^{A}	(0.10-0.53)	0.23^{A}	(0.03-0.53)	0.22^{A}	(0.01-0.38)	0.01^{B}	(0-0.33)	0.17	(0.04-0.39)
		N=101		*N*=101		*N*=101		*N*=101		*N*=101

The strengths (median and its range [min-max]) in the pathways were pairwise compared using the Wilcoxon-Mann-Whitney test with a Bonferroni correction. The medians in the same column within the four exposure pathways, with different lowercase superscripts, indicate statistically significant differences. The family-wise significance level was alpha = 0.05. The medians of the overall criterion score in the last row (*) with different uppercase superscripts, indicate groups that are statistically significantly different. *N* = total number of parameters; *n* = number of parameters within an exposure pathway

5.3.3. Communication of pedigree scores

As an example four different parameters randomly chosen along the different exposure pathways, are presented in the kite diagrams (Figure 13a-13d). The kite diagrams of the remaining 97 parameters can be obtained on request. The kite diagram for parameter 23 (Figure 13a) refers to the seroprevalence of *Salmonella* at the animal level obtained through the surveillance program of the Belgian Federal Agency for the Safety of the Food Chain (FASFC). This parameter is characterized by a high overall strength, which is due to a high score for the empirical criterion. A high level of expert disagreement was observed for the validation criterion. Some experts argued that the seroprevalence is not a good indicator for *Salmonella* prevalence at animal level as explained by the amber colour. The kite in Figure 13b illustrates parameter 55 “Temperature of the water in the scalding tank of the slaughterhouse”, with intermediate to high scores for proxy, and a large expert disagreement for empirical basis and method. As the value for this parameter was obtained by interviewing the quality manager of the slaughterhouse, some experts suggested that the pedigree scores could have been higher if the temperature of the scalding water had also been verified by the interviewer (low validation score) The kite diagram for parameter 77, “temperature in the second storage room of the cutting plant” (Figure 13c), shows full agreement upon the validation criterion (= no validation). Finally, the kite diagram of parameter 98 “frequency of eating pork per year per person in the UK” (Figure 13d), shows a large red area indicating that the scientific value of that particular parameter is weak.

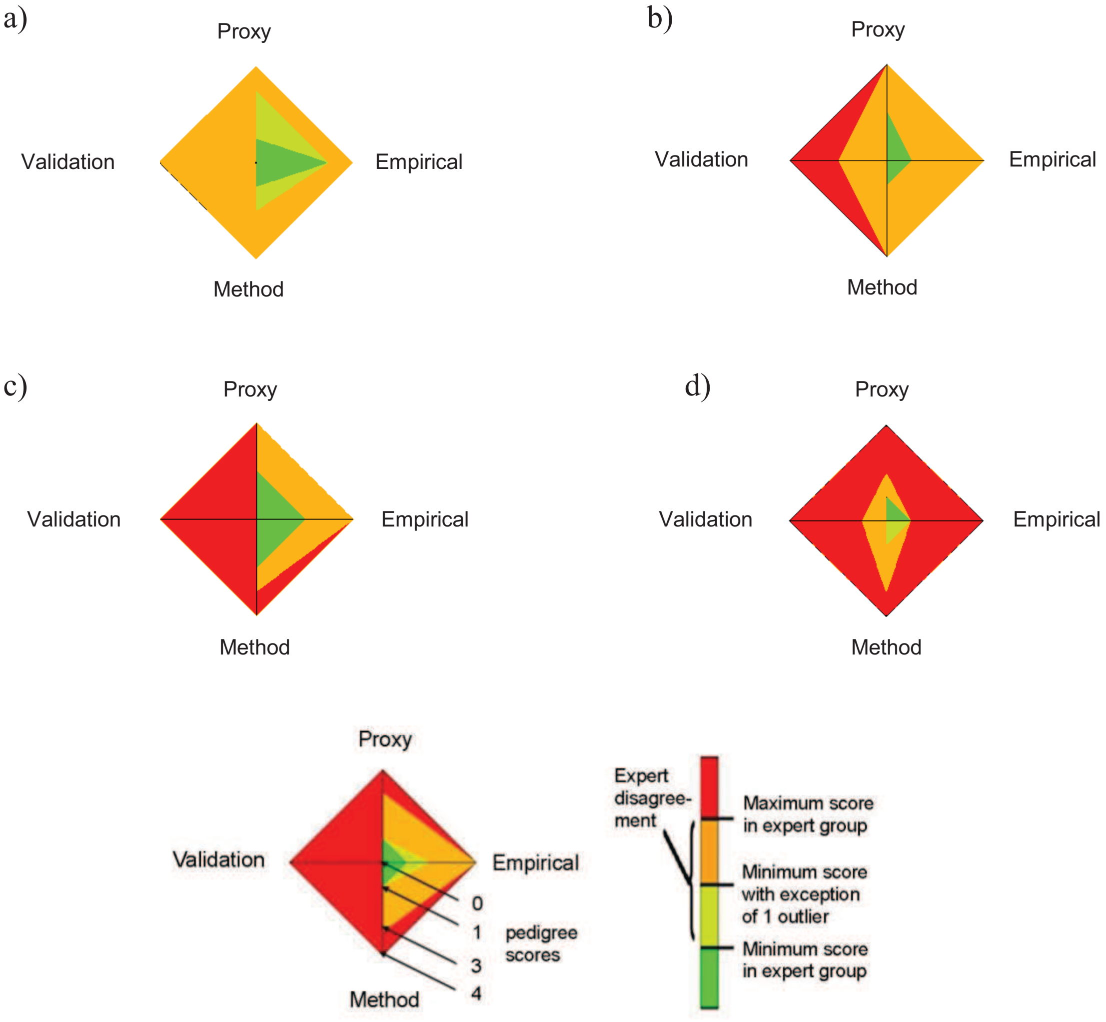

Figure 13: Illustration of pedigree scores of four parameters by means of kite diagrams. Kite diagrams are polygons with one axis for each pedigree criterion (proxy, empirical basis, method, and validation), with the lowest pedigree score (0) in the centre and the highest score (4) on the corner point of the polygon. The legend for the colour codes used in a kite diagram is shown at the bottom figure. (a) Parameter 23: seroprevalence of *Salmonella* at the animal level obtained through the surveillance programme of the Belgian Federal Agency for the Safety of the Food Chain. (b) Parameter 55: temperature of the water in the scalding tank. (c) Parameter 77: temperature in the second storage room of the cutting plant. (d) Parameter 98: frequency of eating pork meals per year per person in the UK. Results and interpretation of the kite diagrams are described in Section 4.3.3.

5.4. Discussion

In this study, the NUSAP methodology was applied to systematically explore the parameter quality in a QMRA model. Quality scoring of input parameters by experts has been greatly facilitated by building up a structured data format template that combined a detailed description of the sources of information from which the parameters originated (Table 12), as well as of different data characteristics described in the parameter ID-cards (Table 13). The characteristics are comparable to those presented in a systematic review by Sanchez et al. (2007). These authors identified the study design and sampling design, the diagnostic procedures, and the source of data to be associated with apparent *Salmonella* prevalence. Systematic review and the NUSAP/pedigree method can be seen as two methods to evaluate the quality of information. Although the two methods use objective quality or eligibility criteria for their quality assessment, a degree of subjectivity may still be encountered due to differences in interpretation of the criteria between reviewers performing a systematic review and among experts attributing pedigree scores using a pedigree matrix. The main advantage of the NUSAP methodology was its ability to assess the quality of parameters in a structured way using a simple setting (the pedigree assessment) that did not require too much training. The training was realized during a NUSAP preworkshop in order to clearly explain the pedigree criteria. Although the scoring procedure itself was judged easy, some participants commented that the criteria proxy and empirical were perceived to somewhat overlap. It is advisable to better instruct the respondents on how to distinguish between the two criteria during the scoring. The empirical basis should be understood as the empirical basis for the proxy variable, not for the quantity for which the proxy is a stand in. Another advantage of the NUSAP method is that it provides conditions for an open debate on the quality of input used in decision making, which is not yet standard in the QMRA. As important as the scores itself, were the comments that experts gave to motivate their scores (Craye et al., 2005; Janssen et al., 2005; van der Sluijs et al., 2005a; 2008).

The same scoring scale was used for the pedigree matrix in the present study (from 0 to 4 score), as in the one used by van der Sluijs et al. (2002). A matrix with, for example, only three scales (from 0 to 2) could not cover enough detail to score the input parameters for the QMRA, whereas a scale with a larger number of classes needs to have slightly different definitions within the matrix, which is not easy and perhaps more confusing for the experts to fill in. Although it might be argued that in a 5-point scale, experts are inclined to attribute a middle score “2,” this was not supported by our data. Although it accounted for 22% of all the elicited scores, it was not the most frequent one in any of the criteria. Other scores as well as the extreme score “0” or “4” (respectively, 8.6% and 16.9% of all elicited scores) were attributed by the experts to score the parameters.

To calculate the strength of each parameter a weighted average was used which took into account the scoring design (the experts’ self-assessment of competence for a parameter, the consistency in scoring, and the number of experts scoring each parameter). The advantage of the weighted average is a better management of the calculated strengths as it is not necessary to communicate *SD*s. Furthermore, this weighted average is bounded between 0 and 1, which facilitates the interpretation of a parameter's strength: a parameter strength of 0 indicates extremely poor quality, whereas a parameter strength of 1 indicates that the parameter is of the best possible quality. This procedure was different as compared with van der Sluijs et al. (2002), Krayer von Krauss (2005) and Krayer von Kraus et al. (2008). In these studies, an unweighted average was calculated of the scores of the different pedigree criteria for each expert. Afterwards, these pedigree strengths were averaged over all experts. Tielemans et al. (2002) observed that the evaluation of the quality of data largely depends on the quality of the judgement of the individual assessor or expert group. Ideally, the expert panel could have been supplemented with external peers from outside the research consortium. Indeed, with experts composed of the core members of the METZOON consortium only, one may argue that the scoring of the parameters could be subject to motivational bias, as several experts were asked to score parameters originating from studies in which they were personally involved. On the other hand, because the experts were familiar with the parameters related to the Belgian situation, it allowed them to have a better picture of how the data were collected and estimated.

Although the expert panel was composed of individuals with backgrounds distributed over all the exposure pathways, the parameters within the primary production module were associated with experts who attributed themselves a higher score for self-expertise, as compared with the other modules. Although the experts' self-expertise level was related to each of the parameters, it may indicate that more qualified experts were available with expertise related to the primary production modules as compared with the other modules. To avoid potential selection bias, it is necessary to select a number of qualified experts for each of the different modules. Obviously, there will always be modules in which it is more easy or difficult to find these experts.

Surprisingly, most assessed parameters obtained low scores for the validation criterion or were not scored by the experts due to a lack of information for this criterion. This can be explained by the origin of the data. Indeed, most data originated from various epidemiologic studies. It is well-known that validation of these studies is often difficult to obtain although it is an essential aspect of data quality (Anonymous, 2000b; Klapwijk et al., 2000; Lammerding and Fazil, 2000). However, a few parameters (i.e. parameters 63 and 72 estimating *Salmonella* prevalence at the end of the slaughterline and in the cutting plants, obtained through the zoonoses monitoring plan of the FASFC) did receive relatively high scores for validation, since they were validated with corresponding data from the private industry (Delhalle, 2006). It would be beneficial to focus on validation issues when setting up new studies or collecting new data, as parameters with increased validation scores will improve the overall strength of model parameters. In addition, future research should evaluate the need for developing minimum quality requirements (Tielemans et al., 2002) with respect to the criterion and overall strengths in order to in- or exclude a parameter into a model. These minimal requirements (cut-off level) will have to be adjusted according to the objectives of the QMRA and the precision level requested by the decision makers.

Within the different exposure pathways, the parameters related to the transport, holding and slaughterhouse module obtained lower scores with respect to the empirical basis, methodological rigour and validation criteria. Management strategies within this module were identified to be very effective in the QMRA models to decrease *Salmonella* prevalence at the end of the slaughterline (Alban and Stärk, 2005). However, the quality assessment indicated that at the time of evaluation the scientific basis for the candidate input parameters within the transport, holding and slaughterhouse-module with respect to the Belgian QMRA was rather weak. Therefore, taking decisions based on a module supplemented with such uncertain parameters should be done carefully.

A way to communicate the quality of parameters was to use different ways to evaluate as much as possible the range of the scores elicited by the experts. The kite diagrams can enhance the risk communication by visualizing the contribution of the pedigree scores related to a parameter as well as to expert disagreement in a comprehensive way (van der Sluijs et al., 2002; Kloprogge et al., 2005). High versus low scoring disagreement are important elements in the total evaluation of a parameter. It should be noted here that low criterion or low overall strengths do not necessarily indicate bad research, but can also refer to the limits of knowledge for a specific parameter (van der Sluijs et al., 2004).

After the assessment of all parameters it was agreed at a METZOON project meeting that 20 of them could be selected as input into the QMRA model. On the other hand, ten parameters yielded relatively high (≥ 0.25) overall strengths but were not retained in the model. These parameters included those obtained through the official Belgian monitoring program for *Salmonella* in pig farms, slaughterhouses and cutting plants. These parameters were not included into the model, but were retained for model validation purposes. Using the NUSAP methodology, important data deficiencies were discovered. This allowed, where possible, replacement of data or imputation of new data by collection or expert opinion. Examples discussed here with respect to the transport, holding and slaughterhouse-module were (1) a structured expert elicitation on *Salmonella* prevalence and concentration data from primary production until the slaughterhouse (Boone et al., 2009a), (2) a study on the *Salmonella* prevalence variation within different stages in five commercial slaughterhouses (De Busser et al., 2008).

A useful application of pedigree strengths with respect to the evaluation of the quality of the parameters within the QMRA-model, lies in the production of a diagnostic diagram (van der Sluijs et al., 2002; 2005a; 2005b; 2005c). This type of diagram consists of two axes: the overall strengths of the parameters included in the QMRA model form the x-axis, whereas the sensitivity of the input parameters (relative contribution of a parameter to the variance in a given model output) forms the y-axis (van der Sluijs et al., 2004). A diagnostic diagram is a helpful tool for decision-makers to have an overview of the real quality of the parameters within the model. Highly sensitive parameters together with a high overall parameter strength can focus decision-makers to take intervention measures directed towards these parameters, because they are not surrounded by much uncertainty. In large-scale farm-to-fork QMRAs the sensitivity analysis is often applied at the end of a risk assessment task. It is however strongly advised to perform a crude sensitivity analysis each time the model is modified and to build up each time a new diagnostic diagram. Successive diagnostic diagrams of the various versions of the model can be considered as part of a quality assurance of the model building process. The NUSAP system is only one of the decision support tools. A multicriteria evaluation (incorporating cost-benefit analysis, stakeholder concerns etc.) will be equally important in the regulatory decisions to be taken along the food pathways to reduce the risk for human salmonellosis cases.

5.5. CONCLUDING REMARKS

The proposed pedigree methodology for assessing data quality permits their use in a QMRA. It is recommended by the authors to implement the method in the risk assessment process. In addition, a standardised parameter ID-card may help to build up an (inter)national database and would allow that the parameters could be shared by various risk assessment teams. The NUSAP/pedigree process should be regarded as an essential step for quality assurance in the QMRA and can contribute to a better risk communication. We argue that the proposed method should be applied in combination with the framework for the collection of data used in microbial risk assessment (Walls, 2007).

CHAPTER 6:

THE NUSAP SYSTEM:

A METHOD TO EVALUATE THE QUALITY OF ASSUMPTIONS IN QUANTITATIVE MICROBIAL RISK ASSESSMENT

Modified from: Boone, I., Van der Stede, Y., Dewulf, J., Messens, W., Aerts, M., Daube, G. Mintiens, K. 2010. NUSAP: a tool to evaluate the quality of assumptions in a quantitative microbial risk assessment model for *Salmonella* in the pork production chain. Journal of Risk Research, 13(3): 337-352.

6.1. INTRODUCTION

Quantitative microbial risk assessment relies on the availability of numerical data in order to model food production and consumption pathways, and aims at providing decision makers with science-based risk management options. The overall quality of a QMRA depends largely on the quality of the data, expert judgement as well as the assumptions made.

According to Beard (2004) one should be cautious in decision making based on risk assessment for three reasons. Firstly, one cannot make the assumption to know all possible consequences of future events. Secondly, some possible consequences may be effectively ignored or underestimated when they are considered to be unlikely for the decision-making problem. Thirdly, model assumptions may be made which are not realistic. As a consequence, decision-making solely based on such (incomplete) models may result in wrong decisions. The danger for decision-making, resulting from the lack of transparency and consideration of all assumptions is well illustrated by the case of bovine spongiform encephalopathy (BSE), a disease that transfers from animals to humans, for which until 1996 the UK denied the assumption of the potential harmful role of BSE to human health. When it finally was accepted that the transfer had taken place, it turned into a major public scandal, which undermined the consumer's confidence in the EU regulatory institutions (Krapohl and Zurek, 2006).

A second example of the impact of not considering assumptions on decision-making was demonstrated in the "de Kwaadtsteniet affair". In 1999, de Kwaadsteniet, a senior statistician of The Netherlands Institute of Public Health and the Environment, accused the institute of lying to the public by using computer models in assessment studies that were poorly validated and based on (unrealistic) assumptions rather than on data. This led to a heated debate about the credibility of assessment studies in decision-making in the Netherlands (van der Sluijs, 2002; Kloprogge et al., 2005).

To prevent similar crises it is important to be as open and cautious as possible and question the assumptions made, as well as to be creative in finding alternative assumptions. Carrying out a risk assessment inevitably involves (many) subjective choices and value-laden assumptions (van der Sluijs, 2006). Value-ladenness refers to the fact that making an assumption involves a choice process. Although the Codex Alimentarius Commission (2010) recommends that in QMRA assumptions should be considered at each step of the risk assessment and documented in a transparent manner, it does not focus on the value-ladenness or the subjective component of assumptions.

Due to the impact of taking wrong decisions as a result of inadequate or subjective assumptions an appropriate evaluation of assumptions is necessary.

The NUSAP/Pedigree method was applied to assess the value-ladenness of assumptions within the Belgian QMRA for human salmonellosis through the consumption of minced pork meat (the METZOON model). This QMRA model included six successive exposure pathway modules: (1) primary production, (2) transport and lairage, (3) slaughterhouse, (4) post-processing, (5) distribution and storage, and (6) preparation and consumption. Since the exposure pathway modules are based on data of various quality levels (see chapter 5 for the evaluation of data quality in the METZOON model), and several value-laden assumptions may have been taken, a method for a clear communication of these uncertainties is very important as this information can contribute in taking more rational decisions based on the results of a QMRA (see also section 2.6.5).

The aim was to prioritize the key assumptions in the Belgian QMRA model and to evaluate their subjective nature, using a structured and transparent approach. This approach aims to help policy makers by providing them with objective criteria to weigh up the pros and cons of various policy options.

6.2. METHODOLOGY

The methodology developed by Kloprogge *et al.* (2005), using the concept of "pedigree of knowledge" was applied to analyse the assumptions in the METZOON model. The methodology starts with the identification of the assumptions in the risk model, followed by the identification and prioritization of the model's key assumptions. Subsequently, the potential value-ladenness of the key assumptions is evaluated by a NUSAP/pedigree assessment. Hereafter "weak" links in the model are identified. The next methodological steps include the further analysis of the potential value-ladenness of the key assumptions, the revision of the assessment and finally its communication.

6.2.1. NUSAP pedigree matrix for the evaluation of assumptions

A NUSAP pedigree matrix (adapted from Craye et al. (2009) and Kloprogge et al. (2005)), containing four pedigree criteria (Table 17) was used as a tool to discuss the value-ladenness of assumptions: (1) the influence of situational limitations, (2) the plausibility, (3) the choice space and (4) the agreement among peers.

The influence of situational limitations refers to the degree to which the choice for an assumption is influenced by limited amount of data, time, soft- and hardware and human resources. The plausibility criterion designates the degree, mostly based on intuitive assessment, to which an assumption is in accordance with the “reality”. The choice space indicates the degree to which alternatives were available to choose from at the moment of making the assumption. Agreement among peers addresses the degree to which the choice of peers is likely to coincide with the analyst’s choice.

The pedigree matrix contains an additional column, to estimate the “influence on the results” of an assumption. The “influence on results” does not evaluate the value-ladenness of assumptions, but aims to provide a rough indication of the influence of an assumption on the end result of the risk assessment.

Although Kloprogge et al. (2005) used a three-point scale (scores 0, 1 and 2) in their pedigree matrix for the evaluation of assumptions, the extended matrix with a five-point scale (from score 0 to 4) as in Craye et al. (2009) and in Honingh (2004) was judged more useful in the present study to allow the experts to provide more subtle scores. The lower the scores in the pedigree matrix, the higher the potential value-ladenness of the assumption.

Table 17: Pedigree matrix for reviewing the quality of assumptions (Adapted from: Kloprogge et al. (2005) and Craye et al. (2009))

	Criteria				
Score	Influence situational limitations	Plausibility	Choice space	Agreement among peers	Influence on results
4	Choice assumption hardly influenced	The assumption is very plausible (based on established theory, verified through peer review)	Hardly any alternative assumption available	A large majority (90-100%) among peers would have made the same assumption	The assumption has little or no impact on the results
3	Limited influence in choice assumption	Plausible (based on model with theoretical basis, empirically verified data)	Very limited number of alternatives	Many experts (75%) would have made the same assumption	The assumption has only a local impact
2	Choice assumption moderately influenced	The assumption is acceptable (based on a simple model, extrapolated data)	Limited choice from alternative assumptions	Several experts (50%) would have made the same assumption	The assumption greatly determines the results in a major step in the calculation
1	Important influence in choice assumption	Assumption is doubtful (based on not verified empirical data)	Average number of alternatives	Few experts (25%) would have made the same assumption.	The assumption has a moderate impact on the end result
0	Totally different assumption had there not been limitations	The assumption is fictive or speculative	Ample choice from alternative assumptions	Controversial assumption, hardly any expert (1%) would have made the same assumption	The assumption greatly determines the end result

6.2.2. Assumptions in the Belgian QMRA model

In total, a list of 39 assumptions was drafted by reviewing the QMRA model in collaboration with the modelling experts of the METZOON project. This list of assumptions was sent to ten project participants for review and these experts were asked to complete the list with extra assumptions they judged missing in the list.

The most important assumptions (key-assumptions) were prioritized by three experts: two of them were involved in model-building and one expert had an advanced knowledge of *Salmonella* spp. in the pork production chain. Importance was defined here as the expected influence of an assumption to the final outcome of the risk assessment. To prioritize the assumptions, each expert was asked to attribute seven points to the most important assumption, six points to the second most important assumption, until one point for the seventh most important assumption, in accordance with Honingh (2004). The remaining assumptions received zero points. The attributed ranking points of all scoring experts were subsequently summed per assumption. A final list of ten assumptions having obtained the highest ranking points was withheld. Due to developments during fine-tuning the METZOON model, three additional assumptions were identified and judged useful for scoring. These extra assumptions were considered as important although they had not been ranked.

The 13 selected assumptions (labelled from 1 to 13), which covered all the modules of the METZOON model were discussed and scored during a NUSAP/Pedigree workshop (Table 18). Prior to this workshop a try-out workshop was organized to familiarize the participants with the workshop protocol and to clarify the pedigree matrix. All the 13 assumptions were introduced by the workshop facilitator followed by a group discussion with nine experts, including the three experts who selected the key assumptions. Hereafter, the discussion was closed by scoring each of the criteria using the pedigree matrix. After the scoring, workshop participants were invited to clarify the reasoning behind their scores. The participants were allowed to review their scores for assumptions based on the group discussion.

Table 18: List of assumptions scored on quality by nine experts from the METZOON consortium, during a NUSAP workshop.

Number	Assumptions	Module[a]	Points[b]
1	The *Salmonella* concentrations, at the start of the slaughterhouse module, of 4 log CFU/0.1 m² (70% of the cases) and 7 log CFU/0.1 m² (30% of the cases) are meaningful and representative for the Belgian situation (Cfr. Hill et al. (2003)). All subsequent multiplication factors in the following abattoir steps are related to these starting concentrations.	3	17
2	The *Salmonella* seroprevalence in Belgian pigs represents the infected (excreting pigs) + carriers (infected but not excreting situation).	1	14
3	During cutting and mixing of the meat, the *Salmonella* cells are not mixed homogeneously within the meat mix, and the model for partitioning described by Nauta et al. (2001) for steak tartare is valid for minced pork meat.	4	9
4	In Belgium all people expose the *Salmonella* cells in the “protected areas” of the minced pork meat to temperatures of between 60°C and 70°C, for a period of between 0.5 to 1.5 minutes (cfr. Hill et al. (2003)).	6	8
5	The odds-ratio’s for external prevalence of *Salmonella* versus internal prevalence (intestinal carriage) is 1.16 (70% of the cases) and 0.6 (30% of the cases) (cfr. Hill et al. (2003)).	3	7
6	Adequate cooking of minced meat (at least 70 °C during 2 min) destroys *Salmonella* cells.	6	7
7	The concentration of *Salmonella* per 100 cm² is homogenous over the entire pig carcass.	3	5
8	At the end of transport and lairage, carrier pigs have the same likelihood as excreting pigs of being contaminated with *Salmonella* on their exterior (cfr. Hill et al. (2003)).	2	3
9	The number of pigs excreting *Salmonella* after 4 hours of transport and lairage will increase with a factor mean 3.8, 90% CI [1.6-12.4] (based on expert opinion).	2	2
10	No growth of *Salmonella* occurs below 10°C in minced pork meat.	4	2
11	The *Salmonella* serologically positive pigs represent the excreting pigs (infectious) + the carriers (infected but not excreting) + the immune.	1	Not done
12	The multiplication factors in the slaughterhouse are independent from the starting *Salmonella* bacteriological prevalence. The multiplication factors for prevalence have been obtained based on a hypothetic starting prevalence of 7% when pigs were leaving the farm.	3	Not done
13	The dose-illness model using WHO outbreak data (Bollaerts et al., 2008) is valid to estimate the *Salmonella* dose-illness relationship based on consumption of minced pork meat.	6	Not done

[a] Module: 1. primary production, 2. transport and lairage, 3. slaughterhouse, 4. Post-processing, 5. distribution and storage, 6. preparation and consumption.

[b] Points attributed by three experts in order to select the key-assumptions.

6.2.3. Analysis and reporting of the pedigree scores

An average score of the four pedigree criteria (proxy, empirical basis, method and validation) was calculated per scoring expert for each assumption. Subsequently, an overall pedigree strength was obtained by averaging these mean pedigree scores over the nine scoring group members, for each of the 13 assumptions assessed. Using the classification inspired by van der Sluijs et al. (2005a), scores below 1.4 were considered to reflect high subjectivity or a high degree of value-ladenness, scores between 1.4 and 2.6 moderately subjective or value-laden, and scores above 2.6 suggested a low degree of subjectivity. Results were communicated using a diagnostic diagram. This type of diagram consists of two axes: the assumptions' overall pedigree strengths are represented on the (inverted) x-axis, while the average scores for the criterion "estimated influence on the results of the QMRA" form the (inverted) y-axis. Assumptions with low overall pedigree strengths and having a strong estimated influence on the results of the QMRA can be considered as weak links in the model. These assumptions are plotted in the right upper quadrant of the diagnostic diagram. A diagnostic diagram is a helpful tool for the analysts, peer reviewers and decision makers to have a quick overview of the strengths and weaknesses in the model (van der Sluijs et al., 2002; 2005a).

In order to present all four criteria in one figure, each assumption was represented individually by using a kite diagram (Risbey et al., 2001a). This type of diagram aims at identifying the reasons for low versus high pedigree scores. Per assumption, each of the four pedigree criteria was placed on one of the axes of the diagram, with the lowest score (0) in the centre and the highest score (4) on the corner point of the polygon. Four colours were used to visualize the scores elicited by the experts. Minimum scores among the group of scoring experts were depicted as a green area in order to reflect the minimal consensus of each parameter. A light green area is used to indicate the minimum score given by the expert panel if the expert who attributed the lowest score for a specific criterion would have been excluded. Scores from the minimum up to the maximum scores given by the expert panel were represented by an amber surface in order to reflect expert disagreement for one or more pedigree criteria. The remaining area up to the borders of the kite diagram is coloured red. Larger red areas in the kite diagram represents low pedigree strength for one or more criteria (van der Sluijs et al., 2002; Kloprogge et al., 2005). The kite diagrams were constructed using the NUSAP Kite Diagram Maker tool (van der Sluijs, 2007). In addition, pedigree charts representing average criterion scores per assumption were constructed, based on Wardekker et al. (2008).

6.2.4. Statistical analysis

The inter-rater reliability, which is a measure for the degree of concordance between raters, was computed by using the generalized kappa of Fleiss (1971) between (1) three experts attributing points for ranking the key-assumptions, and (2) the experts attributing pedigree scores during the NUSAP workshop. For this purpose the SAS Macro (INTER_RATER.TXT) was used (Gwet, 2002). According to Landis and Koch (1977) a kappa value between 0 and 0.2 suggests a poor agreement; kappa values between 0.2 and 0.4 indicate a fair agreement, while values between 0.4 and 0.6 represent moderate agreement, between 0.6 and 0.8 a good agreement, and values between 0.8 and 1 suggest an excellent agreement between raters.

The spearman-rank correlation between the points given to rank the assumptions and the “influence on results” was computed (StatXact, 1999). Wilcoxon Mann-Whitney test (Lehmann, 1975) with Bonferroni (Shaffer, 1995) corrected alpha values were used to test pairwise comparisons in strengths between the pedigree criteria.

6.3. RESULTS

6.3.1. Ranking of key assumptions

The prioritized list of assumptions is presented in Table 18. From the initial list of 39 assumptions, two assumptions (1 and 2) were attributed points by all three ranking experts. Three assumptions (3, 4 and 5) obtained points from two experts, whereas assumptions (6, 7, 8, 9 and 10) were attributed points by only one expert.

In total, three assumptions not contained in Table 18 also received ranking points. Two assumptions related to thawing of pork meat receiving points from a single and two experts respectively, were left out as it was decided to focus the METZOON model on fresh meat only. A third assumption considered to be a proxy for assumption 4 obtained points from one expert.

The overall kappa between the three ranking experts was 0.29 (*P* = 0.07) indicating a fair inter-rater reliability.

6.3.2. Pedigree scores

Thirteen key-assumptions (Table 18) were evaluated and scored with respect to four pedigree criteria (influence of situational limitations, plausibility, choice space, agreement among peers) and the “influence on results”. The average scores and the respective standard deviations for the pedigree criteria for these 13 assumptions are shown in Table 19. The overall strength (average of the four pedigree criteria) of the assessed assumptions varied between 1.1 and 3.2. Assumption 1 obtained an overall strength less than 1.4 (highly subjective or value-laden), whereas assumptions 3 and 6 were given scores greater than 2.6 reflecting a low degree of value-ladenness. The remaining 10 assumptions were given scores ranging between 1.4 and 2.6.

Seven assumptions were given scores less than 1.4 for the “influence of situational limitation”. Experts judged that for these assumptions investing time and resources may have led in taking different assumptions in the model. Two assumptions (3 and 6) on the other hand were considered as hardly influenced by situational limitations (scores > 2.6). The pedigree scores for the plausibility criterion ranged from 0.6 to 3.9. Three assumptions (1, 7 and 9) were considered as fictive or unrealistic, whereas assumptions 3 and 6 were considered as very plausible; both highly plausible and fictive assumptions were characterized by a high degree of expert agreement (low standard deviations). Scoring experts attributed moderate to high average scores (> 1.3) for the choice space criterion. In the “agreement among peers” criterion, assumption 1 obtained a score below 1.4, reflecting an assumption for which most peers would have made a different choice, whereas in the case of assumptions 3, 6, 10 and 11, most peers would have made the same assumption as the one proposed by the METZOON consortium. No significant association was observed between the points obtained by ranking the assumptions and the expected “influence on results” criterion ($r = 0.38$, $P = 0.14$). No significant differences were observed in the pairwise comparisons between the average scores of the four assessed pedigree criteria.

Extreme scores (Score “0” in 8% and Score “4” in 12% of all scores) were less frequently attributed than scores 1, 2 and 3 (respectively, 30%, 27% and 22% of all scores).

Table 19: Average pedigree scores (Average) and standard deviations (SD) of the assessed criteria: situational limitations, plausibility, choice space and agreement among peers for the 13 selected key-assumptions.

	Pedigree criteria								Overall strength		Influence on results	
	Situational limitations		Plausibility		Choice space		Agreement peers					
Assumptions	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD
1	0.9	0.8	0.6	0.5	2.0	1.6	1.1	0.9	1.1	0.6	0.5	0.8
2	1.3	1.0	1.5	0.5	1.5	0.8	2.3	0.8	1.6	0.4	1.9	1.1
3	3.2	1.2	3.5	0.5	3.0	0.9	3.0	0.0	3.2	0.6	2.2	0.4
4	1.7	0.7	1.9	1.5	1.7	1.3	1.6	0.9	1.7	0.8	1.3	1.1
5	0.5	0.5	1.4	0.5	2.6	1.5	1.6	0.5	1.5	0.4	2.0	1.3
6	3.3	1.1	3.9	0.4	3.0	1.3	3.3	0.9	3.2	1.0	1.0	1.7
7	0.9	0.8	1.3	0.5	2.1	1.3	2.4	0.9	1.7	0.6	1.8	1.2
8	1.5	0.8	2.0	0.8	1.9	0.8	2.0	0.6	1.8	0.6	2.3	1.0
9	0.9	0.6	1.3	0.5	1.8	0.9	2.0	1.3	1.4	0.3	1.1	1.1
10	2.1	1.3	2.8	1.2	2.5	0.5	2.8	1.3	2.3	1.1	2.6	1.6
11	1.9	1.4	2.5	0.9	1.8	0.7	2.7	0.8	2.2	0.5	2.0	0.9
12	1.1	0.4	2.4	0.9	2.4	0.5	1.8	0.8	1.9	0.5	1.0	0.8
13	1.3	1.3	2.5	0.9	3.1	0.8	2.6	0.7	2.4	0.5	0.9	0.6

Scores < 1.4 were considered as highly value-laden, scores between 1.4 and 2.6 are moderately value-laden, and scores >2.6 represent a low degree of value-ladenness. See Table 18 for the description of the assumptions (1-13). The overall strength is the average over the four pedigree criteria. The overall strengths and the influence on results scores are used to build the diagnostic diagram (see Figure 14).

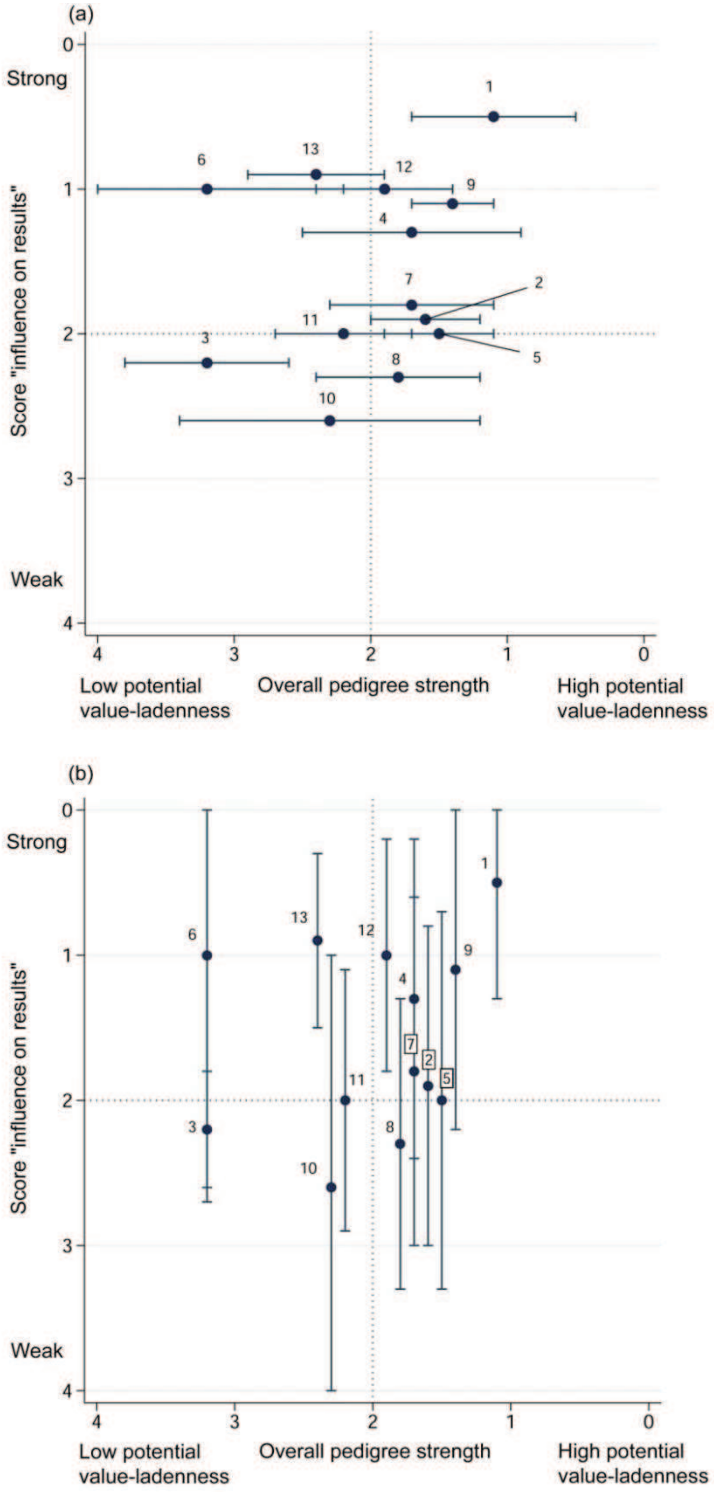

Figure 14: Diagnostic diagram for the identification of weak links in the calculation chain of the METZOON model.
(a) (•): average score; the range represents the average overall pedigree strength plus/minus one standard deviation, and (b) the average score for ‘influence on results’ plus/minus one standard deviation. Assumptions in the upper right quadrant are considered as weak (highly subjective in nature and a strong expected influence on the results). See Table 18 for the description of the assumptions (1–13).

The standard deviation for the plausibility scores was lower in seven assumptions (1, 2, 3, 5, 6, 7 and 9) in comparison with the other assessed assumptions. Assumption 10 was characterized by large standard deviations for the criteria "influence of situational limitations", "plausibility" and "agreement among peers". The scores for the "influence on results" criterion were characterized by larger standard deviations as compared to the other assessed criteria, such as in assumptions 6 and 10 (high expert disagreement). Only assumption 3 had a low standard deviation (< 0.5) suggesting a clear consensus among scoring experts for the "influence on results" criterion.

Overall kappa values were computed to determine the inter-rater reliability between scoring experts. None of the overall kappas for the pedigree criteria exceeded 0.2, indicating a poor concordance.

The diagnostic diagram (Figure 14) identified four assumptions located in the upper right quadrant (1, 4, 9 and 12), which were judged as problematic by the panel of experts. Assumptions 2, 5 and 7 were also located in the right upper quadrant having a low pedigree strength, but with a lower expected influence on results as compared to assumptions 1, 4, 9 and 12.

The pedigree scores are graphically represented in Figure 15 (kite diagrams) and as pedigree charts in figure 16. Assumptions 1, 2, 5, 7 and 9 were all characterized by a high degree of subjectivity, which can be deduced from the extensive red areas (Figure 15).

The low overall pedigree score of assumption 1 was due to the fact that situational limitations played a major role in the choice for this assumption, the assumption was largely implausible and there was a large disagreement among peers. Workshop participants disagreed on their score for the criterion "choice space" for this assumption.

Green areas in the Assumption 3, 6 and 13 corresponded to a low degree of subjectivity for the pedigree criteria assessed. Assumption 6 obtained the highest overall pedigree score, and experts agreed that the assumption was very plausible (Scores "3" or "4"). Experts disagreed, however, on attributing scores on the influence of situational limitations and the choice space (Scores ranging from Score "1" to "4").

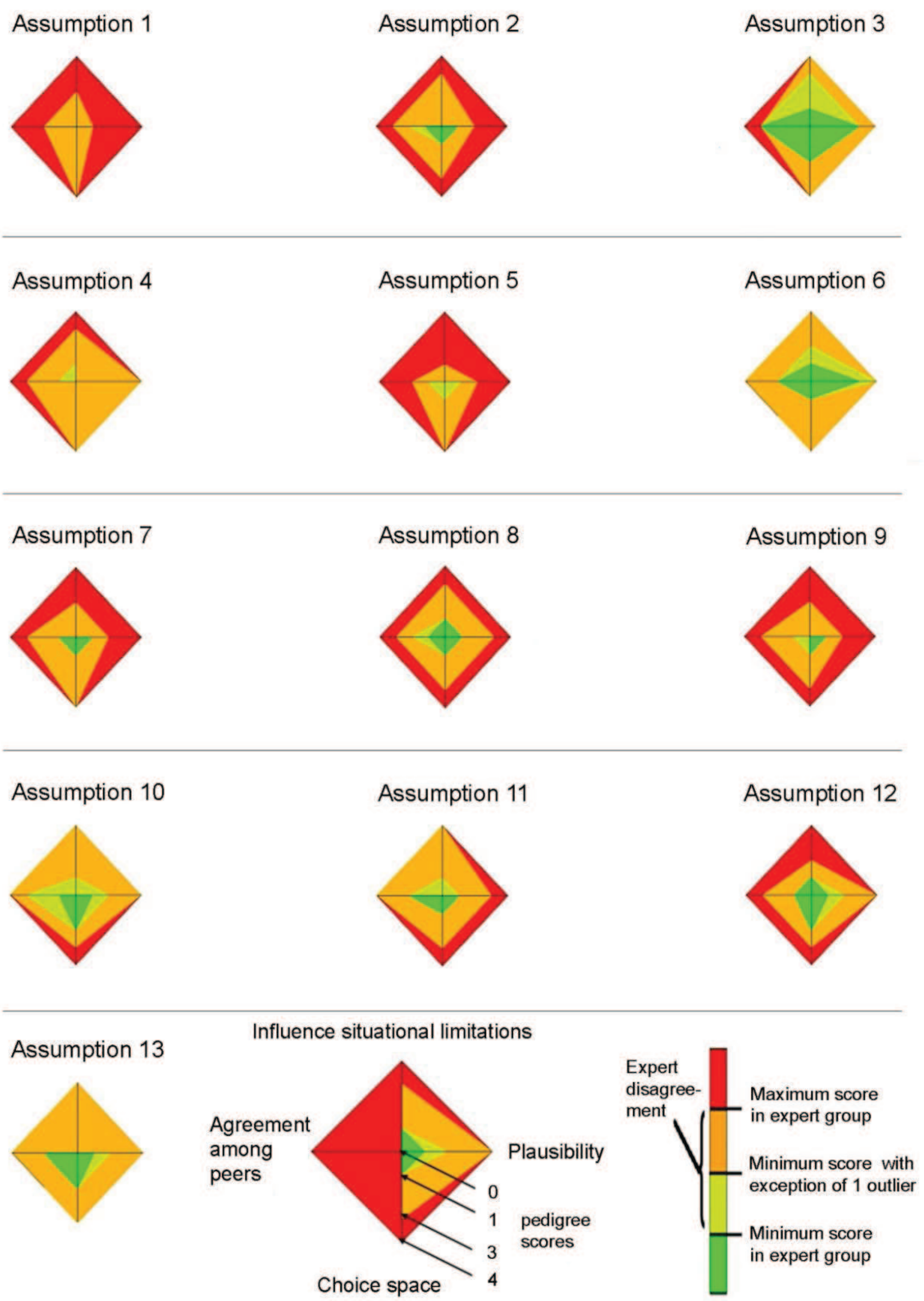

Figure 15: Kite diagrams with pedigree score results indicating the subjective nature (or value-ladenness) of the key assumptions in the METZOON model.
Legend for the colour codes used in a kite diagram is shown at the bottom right of the figure. A kite diagram is a polygon with one axis for each pedigree criterion, with the lowest pedigree score (0) in the centre and the highest score (4) on the corner point of the polygon. See Table 18 for the description of the assumptions (1–13).

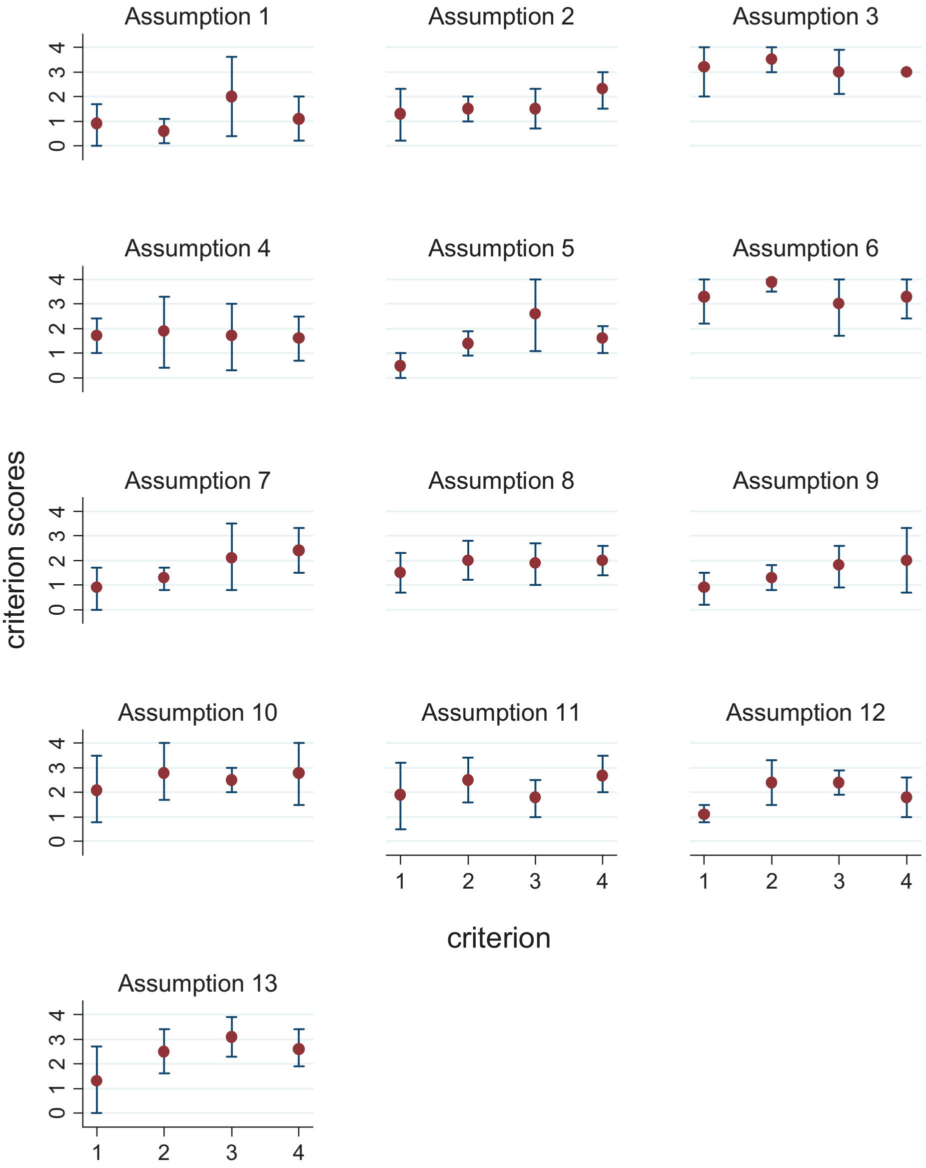

Figure 16: Pedigree charts of the key assumptions in the METZOON model.
Criterion 1 = influence of situational limitations; criterion 2: plausibility; criterion 3: choice space; criterion 4: agreement among peers. (•): average score; the range represents the average criterion strength plus/minus one standard deviation. See Table 18 for a description of the assumptions.

6.4. DISCUSSION

6.4.1. Ranking of key assumptions

The ranking exercise was carried out in order to obtain a list of the most important assumptions to be discussed during the NUSAP workshop. Starting with a list of 39 assumptions, a final list was reduced to 10 selected assumptions, and three additional assumptions. The reason to take 13 assumptions was motivated by the fact that the time allowed for the workshop was restricted to two hours. The study of the remaining assumptions not selected as key-assumptions may as well yield information on their importance in the risk model (Kloprogge et al., 2005).

A low agreement between the three experts who attributed points to rank the key-assumptions was reflected by a low overall kappa. A large variability in ranking points was expected, since a set of 39 assumptions needed to be ranked. In addition, it can be explained by the fact that the ranking experts, a veterinary microbiologist and two risk modellers, had different backgrounds and concerns. Only assumptions 1 and 2 were given points by all three experts ranking the assumptions. Kloprogge et al. (2005) described two ways to prioritize assumptions: one was obtained by a plenary group session, whereas in a second study the ranking occurred without a group discussion. The first one resulted in a low average standard deviation of the scores for the "influence on results". In contrast to this, the present study showed high variability and can be explained by the fact that the prioritization of the key assumptions was achieved without a group discussion. In addition, no significant correlation was observed between the average scores for the "influence on the results" and the ranking points of the key assumptions, whereas a correlation between these two components was seen in a study by Kloprogge et al. (2005).

Although it was initially planned that the key assumptions should be ranked by all project participants, most of them not directly involved in the model building declared not willing to select the key assumptions, because they had difficulties in defining the most important assumptions with respect to the model without the results of a sensitivity analysis. Formulating alternative assumptions and evaluating the effect of these on the final outcome is, however, not straightforward, since this may even require new models to be built (Kloprogge et al., 2005). In future work, comparison of the scores for the "influence on results" with the results of the sensitivity analysis can provide a partial refinement of the diagnostic diagram.

6.4.2. Evaluation of the NUSAP/Pedigree methodology

In order to train the participating experts in the NUSAP methodology, workshop participants had been previously familiarized with this methodology (Boone et al., 2009b). This training was important so that the workshop could focus on the actual discussion of the assumptions and was more time-efficient.

With respect to the pedigree matrix, workshop participants found it was hard to score the criterion “agreement among peers”. Participants argued it was difficult to imagine if peers would have made a different choice in assumptions as compared to the analysts’ choices in the assumptions of the METZOON model.

The NUSAP/Pedigree workshop aimed to provide a reflexive approach on the use of assumptions in the QMRA model, which is important to enhance an open debate on the assumptions and to communicate them towards the decision makers.

Unlike in other NUSAP studies where the evaluation of assumptions was carried out for finished models (van der Sluijs et al., 2005a; Craye et al., 2009), the present study was undertaken during the model building process. This resulted in the identification of weak assumptions in the model and was helpful in redesigning parts of the model structure before the end of the project. Initially the Belgian QMRA-model was derived from the model by Hill et al. (2003) and used the same assumptions. Subsequently and partly due to the NUSAP methodology for the evaluation of the quality of data (see chapter 5) and in the present study for assumptions, the model structure was modified to fit better the Belgian situation. As an example, the diagnostic diagram clearly identified assumption 1 as being highly subjective and inappropriate for the Belgian situation. With this information from the pedigree analysis, it was decided to use semi-quantitative data from a Belgian study (Ghafir et al., 2005; Delhalle et al., 2009a). Assumptions with low pedigree scores for the criterion “influence of situational limitations” are an indication that lifting these limitations (e.g. carrying out more experiments) can lead to a different assumption. If the assumption is at the same time highly implausible, carrying a high level of disagreement among peers, and has a large expected influence on the results (as in assumption 1), it needs reconsideration.

In the final METZOON model (Bollaerts et al., 2009), four assumptions (i.e. assumptions 3, 6, 8 and 10, see Table 18) were retained from the key assumption discussed in the present study. These were located in the safe zone of the diagnostic diagram (Figure 14) and were not considered as weak links in the final METZOON model. To obtain a clear overview of the quality of the assumptions in the METZOON model by Bollaerts et al. (2009) it is necessary to repeat the evaluation of assumptions method and build a new diagnostic diagram.

The used NUSAP approach in evaluating the quality of assumptions is innovative in QMRA. Hill et al. (2003) thoroughly described the assumptions taken in a farm-to-fork risk assessment model on *Salmonella* Typhimurium and in a paper on a *Salmonella* transmission model for pigs (Hill et al., 2008), but did not evaluate their potential value-ladenness. Likewise in a study by van der Gaag et al. (2002), a panel of experts was asked about their agreement with 12 presented assumptions related to *Salmonella* in the primary production and at the slaughterhouse, but no details were provided concerning the quality of these assumptions. The advantage of the NUSAP/Pedigree approach for the evaluation of assumptions consists in the fact that it offers a structured approach disentangling various components determining the value-ladenness of assumptions, which has not been dealt with before in QMRA.

6.4.3. Extending the workshop

In the present study, only experts from the METZOON consortium were involved in the selection and analysis of assumptions. As a part of an extended peer-review process, peers and stakeholders (veterinary services, farmers' organizations, slaughterhouse and processing industry, consumers' organizations) can also be included in the analysis of assumptions (Craye et al., 2005; 2009). Peers and stakeholders can be effectively involved to review the assumptions of the METZOON model using the same protocol as was used in this study. Including stakeholders in a NUSAP workshop may require technical reformulation of some assumptions. The involvement of stakeholders in the debate on assumptions can be beneficial in order to increase their confidence and acceptability in decisions based on QMRA models.

CHAPTER 7:

CHECKLISTS AS A QUALITY ASSURANCE TOOL FOR THE EVALUATION OF QMRA MODELS: APPLICATION TO THE METZOON MODEL

7.1. INTRODUCTION

Checklists are very often used as a practical tool for quality assessment and for ISO certification. As an example, the ISO/IEC 17025: 2005 is the standard used for testing and calibration laboratories, and auditors use checklists to implement a quality management system and to evaluate the competence of the laboratory (http://statbel.fgov.be/nl/binaries/6-101_NL_tcm325-56473.doc). To perform food inspections, the FASFC uses checklists, which aim to make controls as uniform and objective as possible (see http://www.favv.be/checklists-en). The answers to the checklist questions together with the comments and remarks allow the auditors to prepare the assessment report, which serves as a basis for improvements or corrective measures.

The most straightforward way to evaluate the quality of a model is by means of validation with independent data. However, such validation data are often lacking or scarce and provide generally only a partial understanding into the quality of a model. Checklists can be used as a tool to assess the purpose and the quality of complex models or as an aid in the modelling process (Risbey et al., 2005) and thereby improve the transparency and accountability of models. The aim of this study was to evaluate checklists as a tool to evaluate the quality of QMRA models and more specifically the METZOON model.

7.2. MATERIAL AND METHODS

Four checklists used for the evaluation of the quality of models were presented in section 2.6.8. In particular, the checklist by Paisley (2007) was designed as a review tool for the evaluation of import risk assessments. The answers to this checklist are translated in scores and comments can be added to justify the scores. It was used by de Vos et al. (2009) to compare the quality of published classical swine fever and food and mouth disease risk assessments.

The checklist by Risbey et al. (2005) was originally developed for the quality assistance of environmental risk assessment models in order to provide a diagnostic output and to identify pitfalls. It has been applied to assess the quality of an energy model to estimate the greenhouse gas emissions for the IPCC Special Report on Emission Scenario (SRES) B1 energy scenario. The questions related to whether the model is fit for its purpose, how it corresponds to the needs of the users and the stakeholders, and how it is used in decision-making are the principal topics in this checklist (see Table 20).

The checklist by Macgill et al. (2001) and the web-based checklist by Petersen et al. (2003) were not selected to evaluate the METZOON model, because of redundancy in questions with the previous checklists.

Table 20: Model quality checklist adapted from Risbey et al. (2005)

I.	**Problem context**
	Are the model's aims clearly described?
	Description of the model development
	Have the stakeholders been identified?
	What is intended use of the model? What role should the model play in policy-making?
II.	**Strength of the model**
	Has the strength of the input parameters been assessed?
	Was an uncertainty assessment carried out?
	What kind of model validation has been carried out?
	Has there been a systematic process for identifying and evaluating model assumptions, including their influence and possible pitfalls?
	Was a sensitivity analysis carried out?
	What is the accessibility level of the model? Is it sufficiently documented?
	Has the model been reviewed? How?
III.	**Interaction with the model users and the stakeholders**
	What level of expertise is required for a competent use of the model?
	What is the level of the stakeholder involvement?
	Are the model results used in the policy process? How are they used?
	Is the level of accuracy of the model results sufficient to be useful in the policy process?
	Is it clear to the users what the effects of the different assumptions are?
IV.	**Overall assessment**
	Provide your subjective overall assessment of the model and explain your judgement
	List the potential pitfalls encountered by reviewing the model

The Paisley (2007) checklist was applied to review the METZOON model, because it was considered most suitable for the evaluation of QMRA models. In addition, the obtained scores allow comparisons between different reviewers or different QMRAs. Compared to the original Paisley-checklist, questions were added related to the validation status, the quality of the assumptions and the accessibility of the model, while one question concerning the communication of the risk in relative terms (i.e. comparison with other risks) was removed. For a description of the METZOON model, see section 2.7., http://www.metzoon.be, or Bollaerts et al. (2009). The METZOON model was first evaluated by self-assessment using the checklist (Table 21) and hereafter by interviewing the main modeller of the METZOON model. The results should therefore be seen as a form of internal review of the QA process. In a later phase, this checklist could be used and scored by different risk assessors or stakeholders.

7.3. RESULTS: QUALITY AUDIT OF THE METZOON MODEL BY MEANS OF THE PAISLEY (2007)-CHECKLIST

The checklist is subdivided in general questions (purpose of the QMRA, identification of the actors, clarity of the output) and specific questions related to the uncertainty assessment, the methodology, the evaluation of the knowledge base, the model structure and the modelling practices, the validation and peer review status, and the risk communication. The following sections are a summary of the responses obtained by filling in the checklist in Table 21.

7.3.1. General questions (Purpose, Identification of actors, Output)

The purpose of the model and the different actors of the QMRA were clearly identified, i.e. the METZOON model was carried out as part of the METZOON project, which was funded by the Belgian Federal Public Service of Health, Food Chain Safety, and Environment (FPS). The model was developed by the METZOON consortium, which included experts from the primary production module until consumer's module (the Veterinary and Agrochemical Research Centre (VAR), the Faculties of Veterinary Medicine of the Universities of Liège and Ghent, and the Institute for Agricultural and Fisheries Research (ILVO), the Federal Scientific Institute for Public Health and the Interuniversity Institute of Biostatistics and Statistical Bioinformatics of Hasselt University). The main objective of the model was to evaluate the risk for human salmonellosis through household consumption of minced pork meat, and the development of what-if scenarios to reduce the salmonellosis risk level (Bollaerts et al., 2009; 2010). To achieve this goal, the QMRA involved the refinement of statistical and mathematical methods (Bollaerts, 2009), and methods for the evaluation of the quality of input data and assumptions used to build the model and impact of this quality on the final risk estimates.

Before the development of the METZOON model, an inventory of the available QMRA models for *Salmonella* in the pork production chain was made (METZOON, 2006). The QMRA model described by Hill et al. (2003) estimated the risk of salmonellosis due to *Salmonella* Typhimurium (STM) originating from pork, mixed meat products or bacon via cross-contamination or undercooking. This farm-to-fork QMRA was subdivided in six subsequent deterministic modules. Both the prevalence and the concentrations of STM were simulated along the food chain. A second model by Ranta et al. (2004) used a Bayesian approach in combination with simulation models to estimate the *Salmonella* prevalence and concentrations but was less detailed than the model of Hill et al. (2003). A third model by van der Gaag et al. (2004) focussed on economical aspects of intervention measures in the pork production chain. This model implemented a detailed infection model and the prevalence and concentrations were both modelled, using mostly expert opinion. It was not a farm-to-fork risk assessment as the human impact was not taken into account. The Alban and Stärk (2005) model simulated only prevalence through the pork production chain, and was mostly based on expert judgement. From the four reviewed models, the modular model structure by Hill et al. (2003) was chosen as a starting point to build the Belgian QMRA model. The preliminary version of the METZOON model was published by Grijspeerdt et al. (2007). It was subsequently considerably modified to fit better to the Belgian situation and published in its definitive form by Bollaerts et al. (2009).

7.3.2. Uncertainty assessment

Statistical uncertainty and variability were taken into account throughout the model. The uncertainty and variability was included by representing the input parameters as probability distributions and by using Monte Carlo Analysis for error propagation. Uncertainty and variability was explicitly separated in the dose-illness model by means of a two-stage Bootstrap (Bollaerts et al., 2008), but was not separated in other parts of the model due to the absence of relevant data and due the complexity of building second order models.

A source of uncertainty was due to the lack of adequate data. More specifically, the absence of bacteriological data from the primary production until the lairage stage resulted in modelling the primary production module as a black box. The temperatures at the post-processing module were obtained by interviewing the quality managers of processing plants, instead of own measurements which potentially caused motivational bias. Within the consumption module, there is uncertainty related to the accuracy of the Oscar growth model applied to the pork meat matrix. Structured expert opinion was used to quantify the uncertainty of missing parameters in the slaughterhouse module.

7.3.3. Knowledge base

Model input data included in the METZOON model and the assumptions taken were listed, referenced and discussed (Boone et al., 2009b; 2010). The question whether important publications or input data were overlooked can only be assessed by external peer review.

7.3.4. Modelling structure, modelling practices and quality assurance methods

The model was developed in accordance with the Codex Alimentarius guidelines (Codex Alimentarius Commission, 1999) which included the hazard identification, the hazard characterisation, the exposure assessment and the risk characterisation. In addition, new quality assurance methods were developed which had not or rarely been applied before to QMRA. These methods included the NUSAP/Pedigree assessment for the evaluation of the quality of data (chapter 5), the evaluation of the value-ladenness of the assumptions (chapter 6), and model quality checklists (chapter 7). As an example, the NUSAP workshop pinpointed that assumption 1, related to *Salmonella* concentration within the slaughterhouse module (chapter 6, Table 18) as particularly weak.

A structured expert elicitation method (chapter 4) has been used to obtain subjective probability distributions for missing data in the slaughterhouse module (Boone et al., 2009a). Structured expert elicitation methods are superior to informal expert elicitation as these methods aims to reduce bias and to reach a rational consensus of the obtained distributions.

The model was internally reviewed and then subjected to peer-review during the publication process of the manuscripts related to the METZOON model (see http://www.metzoon.be for an overview of peer-reviewed publications). For a thorough assessment of the biological/technical soundness of the METZOON model, an external peer review could be carried out in line with those carried out at EFSA, the USDA and the WHO.

The what-if scenarios developed by Bollaerts et al. (2010) indicated that the most effective scenarios were the ones taken at the end of the slaughterline, during post-processing as well as by improving consumer's awareness. To be used in decision-making, a cost-benefit analysis will have to be carried out for the different what-if scenarios. In addition, the effect of the multiple intervention scenarios in more than one module needs be investigated (Bollaerts et al., 2010).

A complete validation of the METZOON model was not possible due to the scarcity of the independent data that have not used as input data in the model. However, partial validation was obtained for the output at the end of the primary production, at end of the slaughterhouse and at the end of the post-processing stage by using independent data sets (Bollaerts et al., 2009). Whereas the METZOON model predicted a *Salmonella* prevalence of 20.6% in the colon content, Boudry et al. (2002) recorded in a Belgian study a prevalence of 45% (s.e. 11%). The percentage of contaminated chilled carcasses at the end of slaughter predicted by the METZOON model was 4.3%, which is in line with the 7.1% (s.e. 2.1%) based the survey data from the FASFC (FAVV, 2007). The METZOON model produced a prevalence of contaminated minced meat portions of 12.2% whereas a Belgian study obtained a prevalence of 13% (s.e. 4.5) (Korsak et al., 2002) and of 2.6% (s.e. 1.8) by the FASFC (FAVV, 2007). Although previous model results were in line with independent data, one should be cautious about the degree of validation of the METZOON model, as the model results were compared with data originating from small sampling survey studies from a different sample population and/or a different time period or geographical region. Although the model output was in line with the validation data, one should be careful in the interpretation of the validation status of the model results due to the fact that the input parameters varied in quality (see chapter 5, Figure 12). For example, the input distributions obtained via expert elicitation should be verified and/or replaced with empirical data once these become available.

Sensitivity analysis (SA) has not been carried out for the METZOON model. Due to the structure of the model, a SA using Spearman rank coefficients, or a one factor at a time SA method would have yielded erroneous results, due to the variation in the units as a result of portioning and mixing including pig herds, batches, animals, carcasses, meat cuts, meat mix and portions and due to large variability in input distributions. As an alternative to SA, what-if scenarios for the reduction of human salmonellosis were developed by changing one factor at a time. The most effective what-if scenarios were related to salmonellosis reductions at the end of the slaughterhouse module (Bollaerts et al., 2010), although we must take into account that this module is strongly based on (structured) expert judgement.

The model code of the METZOON model, written in the object-oriented MATLAB software, is well documented (Bollaerts et al., 2009), has been made publicly accessible (available at: http://www.metzoon.be) and can be run as a stand-alone. The results related to the METZOON model were published in peer reviewed journals (for an overview, see http://www.metzoon.be). Due to time and budget constraints, a formal external peer review of the METZOON model has not been carried out.

7.3.5. Risk communication

The METZOON risk assessment results were communicated with a combination of text, tables and graphics. Graphical tools were specifically used to communicate the quality of data (chapter 5) and the assumptions (chapter 6) of the METZOON model. These tools were essentially used internally to discuss the quality of the model. The results of the what-if scenario analysis were graphically represented (Bollaerts et al., 2010).

The stakeholders were identified as those involved in the pork production until consumption (e.g. farmers organisations, veterinary health care organisations (DGZ-Vlaanderen and ARSIA), the animal transport sector, slaughterhouse and meat producer organisations, the distribution and retail sector, consumer’s organisations, medical doctors etc.). The stakeholders were contacted for data availability at the beginning of the project. They were, however not involved in the development of the METZOON model, or in the discussion on the quality of the data and the assumptions. Information related to the METZOON model has been made accessible to the stakeholders (http://www.metzoon.be). In addition, the stakeholders have been invited at the end of the project at a stakeholders’ meeting day (for an overview of the presentations, see http://www.metzoon.be) in order to present and discuss the METZOON model. The stakeholder meeting was attended by representatives of Animal Health Care (DGZ-Vlaanderen, ARSIA), the Belgian Federal Public Service of Health, Food Chain Safety, and Environment, the FASFC, the ILVO, the VAR, Belgian universities, the Farmers organisations (Boerenbond and FWA). There were no representatives of slaughterhouse organisations. No major objections were made by the stakeholders against the METZOON model. The participants highlighted the need for a holistic approach from farm to fork to tackle salmonellosis. The FASFC will investigate the feasibility of the proposed what-if scenarios and implement the most adequate control measures for the reduction of human salmonellosis. It is however too early to report on how the model results will be used in policy-making, since socio-economic, and political considerations will have to be taken into account as well. Decision makers may also require further clarification, updating or collection of new data before what-if scenarios can be implemented.

Table 21: Quality audit of the METZOON model (Adapted from: Paisley, 2007). Scores are attributed on scale from 0 to 5. Maximum total score of the complete checklist: 185

	Answer	Score
Purpose of the QMRA		
Does the title adequately describe the report?	Yes	5
Does the report addresses a risk question?	Yes	5
Does the report state the purpose of the risk assessment?	Yes	5
Identification of actors		
Does the report identify the risk assessors?	Yes	5
Does the report identify the client(s)?	Yes	5
Does the report identify other stakeholders?	Yes	5
Output		
Does the analysis clearly communicate the expected risk including how this estimate was generated?	Yes	5
Uncertainty assessment		
Does the analysis adequately incorporate uncertainty and variability in the appropriate parameters in order to characterize the range of plausible scenarios (and their respective outcomes)?	Partly	3
Are significant sources of uncertainty clearly identified?	Partly	3
Knowledge base		
Does the document make clear what are data and what are assumptions?	Yes	5
Does the analysis consider the relevant peer-reviewed studies, including both those that support the risk estimation's conclusions, and those that do not?	Yes	5
Has the literature been cited accurately?	Yes	5
Have any important publications or other information been overlooked?	Unknown	3
Are the references cited appropriate? Are the critical epidemiological observations based on primary not secondary sources?	Yes	5
Does the analysis accurately characterize the cited literature?	Yes	5
Are the conclusions reached supported by the data and or the model?	Yes	5
Model structure, modelling practices and quality assurance methods		
Does the report adhere to international guidelines?	Codex Alimentarius	5
Is the approach biologically and technically sound?	Partly	4
Is the logic of the process clear?	Yes	5
Can the steps from hazard identification, hazard characterization, exposure assessment and risk characterization be easily followed?	Yes	5
Is it clear precisely what has been modelled?	Clear, complete	5
Have both the scenarios being modelled, and the modelling approach, been adequately described in the written text?	Clear, complete	5
Are the scenarios being modelled plausible, logical and appropriate?	Somewhat clear	4
Would every iteration of the model give a biologically plausible output?	Clear, complete	5

(*Continued*)

Table 21 (continued): Quality audit of the METZOON model (Adapted from: Paisley, 2007). Scores are attributed on scale from 0 to 5. Maximum total score of the complete checklist: 185

	Answer	Score
Is the structure of the model appropriate?	Clear, complete	5
Are appropriate data used (sufficient quality)?	Clear, complete	3
Were expert judgements elicited according to state-of-the-art guidelines/protocols?	Partly	4
Is the model mathematically sound and are the formulae used appropriate?	Clear, complete	5
Are the distributions used appropriate for the data or information being modelled?	Clear, complete	5
Are there any data or information that have been overlooked but which might be appropriate in the quantitative assessment?	Don't know	3
Was uncertainty analysis done?	Yes	5
Have assumptions been evaluated as well as their impact on the output results?	Partly	3
Was sensitivity analysis done?	No	0
Was the model validated?	Partly	3
Was the model peer reviewed?	Peer reviewed publications	4
Risk communication		
Was the risk(s) explained clearly and concisely in the report?	Clear; Complete	5
Was communication to stakeholders (other than the client) done?	Yes	4
Total score		161

7.3.6. Conclusion and overall evaluation of the model

Given the quality of the data and the assumptions, the different what-if scenarios for the reduction of salmonellosis can be used with confidence by decision-makers. To be used in policy making, the what-if scenarios need to be combined with a cost-benefit analysis. In addition, model results would improve if expert judgement was replaced by empirical data and if the data would be updated. Methods for the quality evaluation, such as the NUSAP/Pedigree assessment for data quality and the critical evaluation of assumptions were applied principally to check their usefulness as a QA method in QMRA. The quality evaluation of the input data was only carried out at the onset of the project to screen potential input parameters, but not for all parameters included in the final model. Therefore, it may be useful to organise a NUSAP/Pedigree workshop together with the stakeholders to evaluate the data and assumptions of the final model.

The scores linked to the checklist question provide an indication of the quality of a model. The maximal score obtainable in the modified checklist is 185 credits. By evaluating the METZOON model, a total score of 161 credits (87 %) was obtained. A higher score would have been obtained if external peer review, a sensitivity analysis, validation of the model with more recent data, updating/replacing expert judgement data with empirical data, repetition of the NUSAP/Pedigree assessment on the input data and assumptions of the final model would be carried out for the METZOON model. Checklist questions that did not obtain the maximum score should receive special attention. In some cases improvements can be made to improve the scores if this would be necessary. In other cases lower scores reflect the current state of the model. One of the most important questions with respect to the quality of a model is to evaluate if the model is fit for its purpose. The purpose of a QMRA model and the risk question should be clearly defined as well as how the model is used in the policy process. Questions related to how the model is used in policy-making are not contained in the Paisley checklist, but can be found in the Risbey et al. (2005) checklist.

CHAPTER 8: GENERAL DISCUSSION

The Bovine Spongiform Encephalopathy (BSE) crisis had a severe impact on the consumers' confidence in policy making (Krapohl and Zurek, 2006) and highlighted the importance of risk analysis in coping with food safety issues. In particular, QMRA is a framework used to assess the probability and severity of food safety risks and serves as the scientific basis for risk managers. Although a growing number of farm-to-fork QMRAs have been carried out in the last decade, its full capabilities as a decision-support tool have not yet been reached (Tuominen, 2009). For sound decision making, policy-makers need to be confident that the results of a QMRA are credible and defensible. This requires transparency in all phases of the QMRA which can be obtained by a sound quality assurance (QA) system before, during and after the risk assessment.

For policy problems characterized by political pressure, high decision stakes and large uncertainties, Funtowicz and Ravetz introduced the term Post-Normal Science (PNS) (Ravetz and Funtowicz, 1999). Whereas 'Normal' science seeks to find the ultimate truth for a scientific problem, PNS recognizes that this may be unachievable when there are large uncertainties and the stakes are high. It acknowledges that not all uncertainties can be quantified, and that unquantifiable uncertainties and assumptions may be more important than the quantifiable uncertainties. In this thesis, the Belgian QMRA model (the METZOON model) for human salmonellosis due to the consumption of fresh minced pork meat (Bollaerts et al., 2009) was treated as a PNS case where stakes are high and where uncertainties and assumptions to be taken are considered as challenging problems. PNS problems need to be addressed by installing a QA framework which requires appropriate QA methods. Although the Codex Alimentarius principles (1999) provide general guidance for QA, there is a need for more harmonized guidelines in order to put these general principles in practice. In the FAO/WHO guidelines for microbial risk assessment (FAO/WHO, 2003, 2008, 2009), QA is developed without going into details to novel QA methods. The METZOON model was used as a case study to such novel QA methods, including structured expert judgement elicitation (chapter 4), the evaluation of input data (chapter 5) and assumptions by means of a NUSAP/Pedigree approach (chapter 6) and model quality checklists (chapter 7). The lessons learnt from the implementation of these methods are important to characterise and/or enhance the quality of the QMRA process in general.

8.1. EXPERT OPINION IN QMRA

Due to the complexity of the risk questions asked by the decision makers, farm-to-fork QMRA models often require a large number of quantitative and/or qualitative data (Havelaar et al., 2008). Inevitably, data gaps occur in QMRAs and the available data, which have mostly not been collected for the purpose to feed QMRA models, are often of a 'worrying' quality. To cope with data gaps and to quantify the uncertainty of input variables, expert judgement is required in order to complete models. Another factor that motivates the use of expert judgement is that risk assessors need to provide an answer even before all the conclusive empirical evidence is available. Although numerous elicitation protocols have been developed these are until now not systematically applied to provide input in food safety and in QMRA.

In chapter 4, Cooke's classical model (Cooke, 1991) was chosen to provide a structured approach for the elicitation of missing variables in the METZOON model, by means of subjective probability distribution functions (PDFs). Cooke's classical model has been intensively used for the elicitation of expert judgement in a wide range of fields, such as in the nuclear industry, the chemical and gas industry, volcanology and hydrology (Cooke and Goossens, 2008). It was previously applied to obtain input data for a QMRA model on *Campylobacter* in broiler chicken (van der Fels-Klerx et al., 2005).

The method was chosen to fill in data gaps of the METZOON model because it allowed to aggregate expert PDFs to a single combined distribution, by comparing different weighting schemes and by the fact that it aims to achieve a rational consensus on the method used for combining the expert judgements.

The expert assessments were combined using four different weighting schemes (called decision makers (DMs)). The first weighting scheme is called the equal weight DM, which is defined by equal weights given to each expert. In two performance-based DMs, the weights are derived from the experts' performance on seed variables, which are variables whose true values are unknown to the experts, but will become known within the time frame of the study. The performance of the experts on the seed variables is assumed to be indicative for their performance on the variables of interest. Finally, the user-weight DM, which was based on the experts' self-perceived level of expertise, was evaluated.

There was a considerable variability in the PDFs of the individual experts. Based on the performance of the 11 experts on the 11 seed variables, only one expert performed very well, two experts obtained lower weights, and the remaining ones extremely low weights. The combined distributions of the variables of interest in the item weight DM were determined by five experts.

A major difficulty in the application of Cooke's model was to find enough appropriate seed variables, which resemble as much as possible the variables of interest. Although the elicitation questionnaire was pre-tested, a number of seed variables were discarded partly due to misinterpretation and missing estimates. This highlights the importance to test both seed variables and variables of interest for their clarity internally by members of the risk assessment team and if possible by external reviewers. A sufficient number of seed variables is important, since a reduction in seed variables reduces the power to calibrate the experts. The reason why some experts were lesser calibrated could be partly explained by the fact that they may be unfamiliar with the Belgian pork production chain or that they had difficulties in expressing uncertainty distributions. This could be improved by providing additional training to the experts previous to the elicitation.

No significant correlation between the weights based on the experts' self-rating scores (user DM) and those based on the experts' performance to the seed variables could be demonstrated which indicated that some experts were either over- or under confident. The performance based DMs were judged more optimal for the combination of the experts' assessments, compared to the equal weight DM and the user weight DM. One of the reasons for the high weight of a single expert (expert 11) can be explained by the fact that this expert originated from a neighbouring country with a similar pork production chain as in Belgium. This suggests that the composition of the expert panel is important and that one should include as much as possible local experts (Stärk et al., 2002). If possible, it is recommended to create subgroups of experts with expertise related to one or more food chain modules, but this will require extra time to carry out the expert elicitation as well as additional seed variables specific for these modules.

It is worthwhile to test other weighting schemes for the combination of expert judgement, including social network weighting (based on experts' citations) and likelihood weighting (the likelihood weights are proportional the expert's observed outcomes) (Cooke et al., 2008). According to a review by Cooke et al. (2008) these weighting schemes did not outperform the performance-based weighting schemes, although further research is necessary to confirm this statement.

In future expert judgement studies it is recommended to record the reasoning behind the elicited values from the experts as this may add to the transparency of the elicitation process (Cooke and Goossens, 2008). In the expert elicitation software (EXCALIBUR) (Cooke and Solomatine, 1992) data input has to be done manually and export requires an ad-hoc procedure. The Excalibur could therefore benefit from an upgrade to increase its user-friendliness. Recently Cooke's model was implemented in a R-package (the **expert** package) (Pigeon et al., 2009) and can be used as an alternative to Excalibur. The package offers enhanced graphical capabilities, as well as the implementation of another approach for expert elicitation, known as the Mendel-Sheridan model (Bayesian approach) (Goulet et al., 2009).

As pointed out by Knol et al. (2010), expert judgement is often not regarded as a reliable source of information compared to that obtained from empirical studies. This criticism can however be explained by a lack of knowledge about structured expert elicitation methods. Just as with expert judgement data, empirical data can also be criticised because their inclusion in a model also requires expert judgement. Empirical data may also contain implicit expert judgements and assumptions. It should be emphasized that although expert judgements remain subjective and are not necessarily correct or sufficiently precise, using a structured methodology increases the transparency in the risk assessment process and is of importance for a quality check of the data. To facilitate peer review, the expert elicitation procedure should be clearly described in the QMRA reports. Although the elicitation of structured expert judgement requires a thorough preparation, the data obtained through this approach can help in saving money and time to get a fast answer. Expert opinion should be replaced with empirical data once these become available. These newly obtained empirical data can also be used to check how accurate the experts were in there estimates.

8.2.NUSAP/PEDIGREE ASSESSMENT FOR QMRA

Although the NUSAP system was successfully applied in environmental risk assessments (Craye et al., 2005; Kloprogge et al., 2005; 2005b; van der Sluijs et al., 2005c; 2009; 2010), the method was introduced here for the first time in a QMRA model to evaluate the data quality and the quality of assumptions (Boone et al., 2009b; 2010). The NUSAP/Pedigree approach has the advantage to be relatively simple and comprehensive to apply, without too much training of the experts, and was accepted as a valuable method for the screening of input parameters by the METZOON project participants. A template was developed for the inventory and the description of potential parameters for the METZOON model (see chapter

4, Table 13), which could subsequently be used as background information for the experts evaluating the quality of these parameters. A weighted average of the scores for the pedigree criteria (proxy, empirical basis, methodological rigour, validation) was used to evaluate the input parameters taking into account the consistency in rating, the experts' self-perceived level expertise and the number of experts attributing scores for the different criteria. We believe that the used pedigree criteria captured well the basic features of data quality relevant for a QMRA.

Although the criteria were supposed to be independent, some overlap was perceived by the project members between the proxy and the empirical basis. In principle, other pedigree matrices can be used for the quality evaluation provided that the risk assessors agree on the criteria and the linguistic definitions for the scores. For example, a recent systematic review of prevalence data on welfare issues in broilers and broiler breeders in relation to genetic selection used an adapted matrix with the criteria proxy, empirical basis and methodological rigour for the selection/rejection of relevant references in order to provide input for a risk assessment (Lefebvre et al., 2010). Previous to the actual quality evaluation, it is important to present and discuss the pedigree criteria and the method to both the risk assessment team and the decision makers.

The NUSAP method provides a framework for a structured and reflexive dialogue on the quality of data. This debate on the quality could however not be fully exploited in this study, as the parameters had to be scored individually by experts instead of during a workshop, as was done for the evaluation of assumptions (chapter 6). As important as the scores were the experts' rationale for their scores, since this extra information can be used to motivate the selection of parameters into the model.

Although the NUSAP/pedigree assessment for the evaluation of data quality was applied only during pre-screening of parameters of an intermediate model, ideally, the procedure should be repeated after modifications in the model structure and after inclusion of new parameters in the model, and for the parameters in the finalised model. As such, the scores obtained over time can be used to monitor the evolution of the quality the QMRA model. The experts' average self-perceived knowledge with respect to the parameters was higher for the primary production module as compared to other modules. This suggested that it is recommended to have a balanced panel of experts with expertise pertaining to each of the different modules. Indeed, experience in a certain module may result in of over- or under-confidence and may have an influence on the scores given by the experts. At the same time, the results of the pedigree assessment indicated that the input parameters in the primary production obtained higher pedigree scores than those of the other modules. The quality assessment can thus motivate new research in specific modules in order to obtain better data or instigate expert elicitation to characterize the uncertainty in critical parameters.

Validation of data is often difficult to achieve, although it remains an important data quality feature. The majority of the parameters in the METZOON model obtained however very low scores for validation. This can be explained because most data in the QMRA originated from various studies and were not collected for the purpose to be used in a QMRA model. It would be beneficial to promote validation of results in all new research studies that might be used in future QMRA. The quality of the parameters and the impact this quality has on the model output should best be communicated by means of a diagnostic diagram (van der Sluijs et al., 2005a). This type of diagram allows the identification of the weakest links in a risk assessment model, i.e. the parameters characterised by a large impact on the output of a QMRA as well as low pedigree strengths.

In addition, the results of a pedigree assessment can also facilitate external peer review. In order to increase the acceptability in QMRA it should be tested in a large number of QMRA models. There is still more research necessary in how to communicate the results of the pedigree assessment using a combination of text, tables and diagrams to risk assessors, decision-makers and stakeholders. For example, the use of the colour codes, the scale and the interpretation of the surface in the kite diagrams were not straightforward to all experts of the METZOON consortium.

The applied formal evaluation of the assumptions in accordance with the protocol by Kloprogge et al. (2010) provides knowledge about the weak and strong parts of the QMRA and was helpful to describe the quality of the model and for the improvement of the model.

There was a low agreement between experts with respect to their ranking scores for the most important assumptions (key assumptions). Many experts did not want to provide a ranking of the key assumptions, judging that it was too difficult without a sensitivity analysis. In future work, it would be useful to ask for the experts' motivations for their ranking to discover if some assumptions received a higher ranking due to a scientific curiosity of the experts for specific assumptions, instead as for their importance in the METZOON model. To verify this, one could ask the experts to give a ranking of assumptions for which they would be willing to invest resources as in Krayer von Kraus et al. (2008).

Due to the limited available amount of time allocated to the NUSAP/pedigree workshop only a small number of assumptions from the original list of identified assumptions was taken for further assessment. It may be valuable to explore the quality of the other non-selected assumptions as well. The most difficult criterion for the experts to score in the used pedigree matrix was "the agreement among peers" as it was difficult for the experts to imagine how peers would have chosen assumptions. We therefore propose to down weight this criterion when a majority of experts do not feel to give a sensible answer. Two additional criteria in the matrix by Kloprogge et al. (2005) and Craye et al. (2009), namely the "agreement among stakeholders" and the "sensitivity to views of analysts" were not included in the evaluation of the assumptions of the METZOON model, because stakeholders were not involved during the model building phase and because it was assumed that the (political) views of the analysts did not play a role.

For QMRAs that are expected to play an important role in decision-making, it would be recommended to promote the use of NUSAP in international guidelines and include the approach in the Terms of Reference of QMRA projects. The NUSAP approach will most likely gain acceptability among food safety experts after it has been applied and tested successfully in a growing number of QMRAs and accepted by the risk managers as well.

8.3. CHECKLIST APPROACH

From four checklists (Macgill et al., 2001; Petersen et al., 2003; Risbey et al., 2005; Paisley, 2007), the checklist by Paisley (2007) was selected to evaluate the quality of the METZOON model, because this checklist allowed to screen the most important quality characteristics of the QMRA and the scoring system allowed to give a quick overview on the strengths and weaknesses of the model. The Paisley (2007) checklist can be used from the early phases of the development of the conceptual model, until the model is used in the decision-making process both for internal peer review purposes as for external peer review. The obtained checklist scores by reviewing the METZOON model were obtained by self-assessment. The checklist is most useful when filled in by more than one reviewer, as the scores and remarks can then be compared and used for model improvement. Special attention should be given to disagreement in scores. Whether a QMRA model is *in fine* of a sufficient quality in order to base sound decision on should be discussed together with the risk managers, and the checklist results can be helpful for that. The Paisley-checklist is in its current form still a very promising instrument for screening the quality of a model and to support model improvements. It should nevertheless be evaluated on a wide range of QMRAs in order to improve its acceptability as a quality assurance method.

The Risbey et al. (2005) checklist contained additional questions not contained in the Paisley (2007)-checklist (see Table 20, Section III.) related to the use of the model in policy-making and the involvement of stakeholders, but these questions could not be fully answered for the METZOON model as the model is not yet used for risk management. One of the key issues in the Risbey et al. (2005) checklist is to evaluate whether the model is fit for its purpose. This requires communication between all actors in the risk analysis process (risk assessors, decision-makers, stakeholders). It is recommended to improve the risk communication during the model-building phases with the decision makers and the stakeholders as this will improve the acceptability of the model results.

8.4. RISK COMMUNICATION

Risk communication is defined by the Codex Alimentarius Commission (1999) as the interactive exchange of information and opinions concerning risk and risk management among risk assessors, risk managers and stakeholders and is an essential part of the risk analysis process. Good practices in risk communication are therefore necessary in all stages of the risk assessment. In the METZOON project, the results were communicated to the scientific community by means of annual reports, peer reviewed publications and presentations at meetings and (inter)national conferences (for an overview see http://www.metzoon.be). The results of the NUSAP/Pedigree approach for the evaluation of the quality of data and assumptions were communicated by means of graphical tools such as kite diagrams, pedigree charts, diagnostic diagrams in order to facilitate discussions about the quality of the model. Although the usefulness of these methods within the risk assessors of METZOON team was acknowledged, there is still a need to test the relevance of these graphical risk communication tools on the decision makers and stakeholders.

Communication with the stakeholders was done in two phases. The identified stakeholders were contacted at the start of the METZOON project for the availability of data related to *Salmonella* in the pork production chain. At the end of the project, the stakeholders were invited at a stakeholders meeting in which the model results and the what-if scenarios where discussed. The website (http://www.metzoon.be) was made available for the stakeholders to summarise in a non-technical way the results, the methods, the what-if scenarios and publications resulting from the METZOON project.

CHAPTER 9: RECOMMENDATIONS FOR AN INTEGRATED QUALITY ASSURANCE FOR QMRA

QMRAs are increasingly used as a decision support tool in food safety. At the same time the actors in the risk analysis process need to have confidence in the results of these QMRAs. Especially in the absence of good quality data and facing assumptions, a transparent approach of all aspects in the QMRA process and a clear communication of the uncertainties, strengths and weaknesses of the risk assessment as well as their impact on the results is of paramount importance. In order to achieve the QMRA objectives, an integrated and pragmatic quality assurance is therefore advisable as we proposed in Figure 17. Optimal QA can be achieved by combining several QA methods, as the application of one QA method facilitates the application of another QA method.

An effective QA starts when the risk question is asked by the risk managers to the decision makers. At this stage a multidisciplinary and balanced consortium of experts needs to be composed (including epidemiologists, statisticians, veterinarians, general practitioners, economists,...) and a clear risk communication strategy needs to be prepared to exchange information between risk assessors, decision makers and the stakeholders so that the risk problem can be clearly framed. A first conceptual model can be evaluated using checklists and internal/external peer review. Subsequently, depending on the availability of data, time and resources or the degree of accuracy necessary to reply to the risk question, a model structure needs to be chosen. Either an existing (published) model is used or updated or a new one can be developed.

The data acquisition phase is crucial and requires a clear communication (based on trust, confidentiality, etc.) with the data-owners. Structured methods for data gathering and data quality evaluation include systematic review and the NUSAP/Pedigree system (see chapter 5), which combine the use of quality criteria with expert judgement in order to characterise the quality of data and to select or reject data as input in a model. Data gaps in the model can be filled in by planning structured expert elicitation following a clearly described protocol (e.g. Cooke’s classical model, see chapter 4).

Subsequently, the appropriateness of the model can be assessed by the combination of a variety of QA methods, including uncertainty assessment, model verification, sensitivity analysis, the evaluation of assumptions, scenario analysis and model validation. Model quality checklists can be used as a QA tool for model guidance or for peer review. The weaknesses and strengths identified through the implementation of the QA tools should guide the risk assessors in consultation with the risk managers and stakeholders to know if the model is fit for its purpose. QMRA model building is an iterative process and often improvements and modifications to the model structure, the data or the assumptions will have to be made before the model can be used for decision-making. Modified model versions will have to be reassessed on their quality using one or more of the QA methods. When a sufficient quality level has been achieved, good practices in risk communications (Pfeifle et al., 2006; Wardekker et al., 2008) are necessary for the dissemination of the model results to the risk managers, stakeholders and the general public, in the form of scientific publication, meetings, workshops and websites. In order to clarify and discuss the key assumptions related to a model or what-if scenarios, a NUSAP/Pedigree assessment can be organised before the further implementation of risk management measures. This will allow a further fine-tuning of the model and will increase the acceptability of the model results among the stakeholders.

Implementing an effective QA framework for QMRA can be improved:

- By further testing the novel QA methods proposed in this thesis (NUSAP/Pedigree assessment, checklist approach);
- By sensitizing the actors in the risk analysis process (risk assessors, risk managers, stakeholders) about the importance of QA for a transparent decision-making. This can be achieved by improving the risk communication (e.g. evaluation of the graphical representation of risk and the QA tools) and further training of the QA tools. Presenting in detail the uncertainties and quality evaluation of the QMRA, may give the impression of lack of certainty to the policy makers. Both the strengths and the limitations must thus be provided in a balanced manner;
- By the development of harmonised guidelines for QMRA containing novel state-of-the-art QA methods, including references to available software and case studies, at the EFSA and FAO/WHO levels;
- By developing a quality label or an ISO-norm for QMRA to which high stakes QMRAs need to comply with in order to be accepted as decision-support tool.

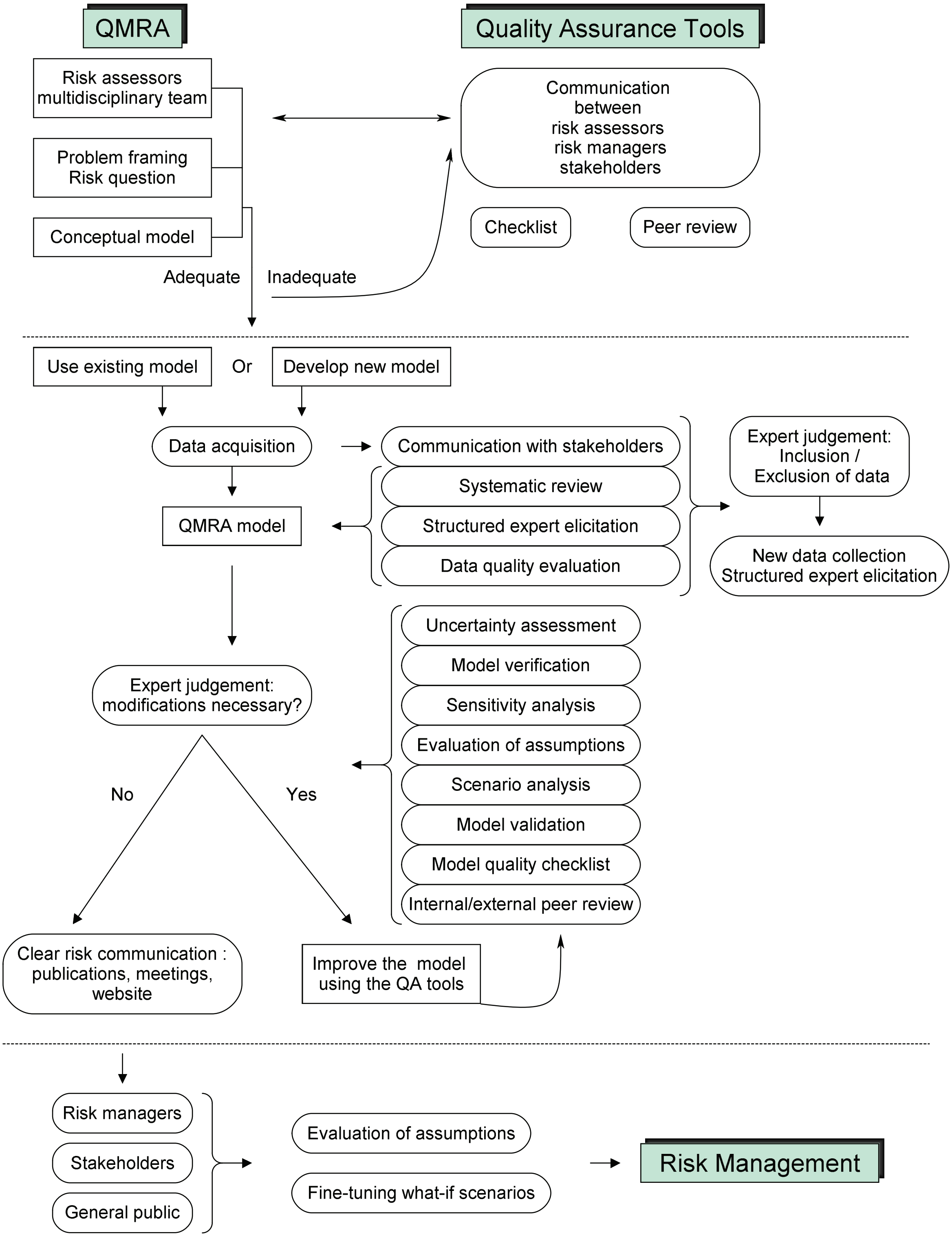

Figure 17: Integrated quality assurance approach for QMRA

REFERENCES

Alban, L., Olsen, A.M., Nielsen, B., Sorensen, R., Jessen, B. (2002). Qualitative and quantitative risk assessment for human salmonellosis due to multi-resistant *Salmonella* Typhimurium DT104 from consumption of Danish dry-cured pork sausages. *Preventive Veterinary Medicine 52*(3-4), 251-265.

Alban, L., Stärk, K.D. (2005). Where should the effort be put to reduce the *Salmonella* prevalence in the slaughtered swine carcass effectively? *Preventive Veterinary Medicine 68*(1), 63-79.

Animal Health Care (DGZ-Flanders) (2005). Report on monitoring Animal Health Care, Pigs. Available at: http://www.dierengezondheidszorg.be/ondersteuning/praktijk_advies_publikaties_varkens/rapport_va_2004.pdf (accessed 17-11-2010).

Anonymous (2000a). White paper on food safety. COM (1999) 719 final. Brussels, 12 January 2000, Commision of the European Communities. Available at: http://ec.europa.eu/dgs/health_consumer/library/pub/pub06_en.pdf (accessed 17-11-2010)

Anonymous (2000b). First report on the harmonisation of risk assessment procedures. The report of the Scientific Steering Committee's Working Group on harmonisation of risk assessment procedures in the scientific committees advising the European Commission in the area of human and environmental health. 26-27 October 2000. Part 2: Appendices. Available at: http://ec.europa.eu/food/fs/sc/ssc/out83_en.pdf (accessed 17-11-2010).

Anonymous (2002). Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. *Official Journal of the European Communities*, L 31.

Anonymous (2003a). Regulation (EC) No 2160/2003 of the European Parliament and of the Council of 17 November 2003 on the control of *Salmonella* and other specified food-borne zoonotic agents. *Official Journal of the European Union L 325*, 1-15.

Anonymous (2003b). Directive 2003/99/EC of the European parliament and the council of 17 November 2003 on the monitoring of zoonoses and zoonotic agents, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC. *Official Journal of the European Union L325*, 31-40.

Anonymous (2010). The price of foodborne illness in the USA. *Lancet 375*(9718), 866.

Beard, A.N. (2004). Risk assessment assumptions. *Civil Engineering and Environmental Systems 21*(1), 19-31.

Beloeil, P.A., Chauvin, C., Proux, K., Rose, N., Queguiner, S., Eveno, E., Houdayer, C., Rose, V., Fravalo, P., Madec, F. (2003). Longitudinal serological responses to *Salmonella enterica* of growing pigs in a subclinically infected herd. *Preventive Veterinary Medicine 60*(3), 207-226.

Bollaerts, K., Aerts, M., Faes, C., Grijspeerdt, K., Dewulf, J., Mintiens, K. (2008). Human salmonellosis: estimation of dose-illness from outbreak data. *Risk Analysis 28*(2), 427-440.

Bollaerts, K. (2009). Statistical models in epidemiology and quantitative microbial risk assessment applied to *Salmonella* in pork. Ph.D. Hasselt University, Belgium.

Bollaerts, K., Messens, W., Delhalle, L., Aerts, M., Van der Stede, Y., Dewulf, J., Quoilin, S., Maes, D., Mintiens, K., Grijspeerdt, K. (2009). Development of a quantitative microbial risk assessment for human salmonellosis through household consumption of fresh minced pork meat in Belgium. *Risk Analysis 29*(6), 820-840.

Bollaerts, K., Messens, W., Aerts, M., Dewulf, J., Maes, D., Grijspeerdt, K., Van der Stede, Y. (2010). Evaluation of scenarios for reducing human salmonellosis through household consumption of fresh minced pork meat. *Risk Analysis 30*(5), 853-865.

Boone, I., Van der Stede, Y., Bollaerts, K., Messens, W., Vose, D., Daube, G., Aerts, M., Mintiens, K. (2009a). Expert judgement in a risk assessment model for *Salmonella* spp. in pork: the performance of different weighting schemes. *Preventive Veterinary Medicine 92*(3), 224-234.

Boone, I., Van der Stede, Y., Bollaerts, K., Vose, D., Maes, D., Dewulf, J., Messens, W., Daube, G., Aerts, M., Mintiens, K. (2009b). NUSAP method for evaluating the data quality in a quantitative microbial risk assessment model for *Salmonella* in the pork production chain. *Risk Analysis 29*(4), 502-517.

Boone, I., Van der Stede, Y., Dewulf, J., Messens, W., Aerts, M., Daube, G., Mintiens, K. (2010). NUSAP: a method to evaluate the quality of assumptions in a quantitative microbial risk assessment. *Journal of Risk Research 13*(3), 337-352.

Botteldoorn, N., Heyndrickx, M., Rijpens, N., Grijspeerdt, K., Herman, L. (2003). *Salmonella* on pig carcasses: positive pigs and cross contamination in the slaughterhouse. *Journal of Applied Microbiology 95*, 891-903.

Boudry, C., Korsak, N., Jacob, B., Etienne, G., A., T., Daube, G. (2002). Ecologie de *Salmonella* dans le tube digestif du porc à l'abattage et étude de la contamination des carcasses. *Annales de Médecine Vétérinaire 146*, 353-360.

Calistri, P., Giovannini, A. (2008). Quantitative risk assessment of human campylobacteriosis related to the consumption of chicken meat in two Italian regions. *International Journal of Food Microbiology 128*(2), 274-287.

Cassin, M.H., Lammerding, A.M., Todd, E.C.D., Ross, W., McColl, R.S. (1998). Quantitative risk assessment for *Escherichia coli* O157:H7 in ground beef hamburgers. *International Journal of Food Microbiology 41*(1), 21-44.

Clemen, R.T., Winkler, R.L. (1999). Combining Probability Distributions From Experts in Risk Analysis. *Risk Analysis 19*(2), 187-203.

Codex Alimentarius Commission (1999). Principles and guidelines for the conduct of microbiological risk assessment. CAC/GL-30 Codex Alimentarius Commision (FAO/WHO). Available at: www.codexalimentarius.net/download/standards/357/CXG_030e.pdf (accessed 17-11-2010).

Codex Alimentarius Commission (2010). Procedural Manual 19th Edition. FAO, Rome. Available at: http://www.codexalimentarius.net/web/procedural_manual.jsp (accessed 17-11-2010).

Cooke, R.M. (1991). *Experts in Uncertainty. Opinion and Subjective Probability in Science*. Oxford University Press, New York.

Cooke, R.M., Solomatine, D. (1992). EXCALIBUR integrated system for processing expert judgments version 3.0, User's manual prepared under contract for Directorate-General XII. Delft, the Netherlands, Delft University of Technology and SoLogic Delft.

Cooke, R.M., Goossens, L.H.J. (2000). Procedures Guide for Structured Expert Judgement. Brussels, Belgium, European Commission, Directorate-General for Research, Technical report EUR 11820 EN

Cooke, R.M., Slijkhuis, K.A. (2003). Expert judgment in the uncertainty analysis of dike ring failure frequency, In: Blischke, W.R., Prabhakar Murthy, D.N. (editors). *Case Studies in Reliability and Maintenance*. pp. 331-350.

Cooke, R.M., ElSaadany, S., Huang, X. (2008). On the performance of social network and likelihood-based expert weighting schemes. *Reliability Engineering & System Safety 93*(5), 745-756.

Cooke, R.M., Goossens, L.L.H.J. (2008). TU Delft expert judgment data base. *Reliability Engineering & System Safety 93*(5), 657-674.

Craye, M., Funtowicz, S., van der Sluijs, J. (2005). A reflexive approach to dealing with uncertainties in environmental health risk science and policy. *International Journal of Risk Assessment and Management 5*(2/3/4), 216-236.

Craye, M., Laes, E., van der Sluijs, J. (2009). Re-negotiating the role of external cost calculations in the Belgian nuclear and sustainable energy debate, In: Pereira, A., Funtowicz, S. (editors). *Science for Policy*. Oxford University Press, Oxford.

De Busser, E., Maes, D., Dewulf, J., Houf, K., De Zutter, L. (2008). *Salmonella* contamination rate along the slaughter line in 5 different Belgian slaughterhouses. In: Proceedings 20th International Pig Veterinary Society Congress, 22-26 June 2008, Durban, South Africa, p. 180.

De Sadeleer, L., Dewulf, J., Van der Stede, Y., De Zutter, L., Houf, K., Ribbens, S., De Busser, E., Quoilin, S., Grijspeerdt, K., Maes, D. (2008). A qualitative risk assessment for human salmonellosis due to the consumption of fresh pig meat in Belgium. *Vlaams Diergeneeskundig Tijdschrift 78*(1), 34-43.

de Vos, C.J., Conraths, F., Adkin, A., Hallgren, G., Paisley, L. (2009). Quality of veterinary import risk analysis studies. In: Proceedings of the 3rd Annual Meeting of EPIZONE "Crossing Borders", Antalya, Turkey 12-15 May 2009, p. 52.

Delhalle, L. (2006). *Salmonella* et germes indicateurs dans la filière de production et de distribution de viande de porc en Belgique. DEA en Sciences vétérinaires, Service de microbiologie des denrées alimentaires d'origine animale, Faculté de Médecine vétérinaire, Université de Liège, Belgique.

Delhalle, L., De Sadeleer, L., Bollaerts, K., Farnir, F., Saegerman, C., Korsak, N., Dewulf, J., De Zutter, L., Daube, G. (2008). Risk factors for *Salmonella* and hygiene indicators in the 10 largest Belgian pig slaughterhouses. *Journal of Food Protection 71*(7), 1320-1329.

Delhalle, L., Saegerman, C., Farnir, F., Korsak, N., Maes, D., Messens, W., De Sadeleer, L., De Zutter, L., Daube, G. (2009a). Salmonella surveillance and control at post-harvest in the Belgian pork meat chain. *Food Microbiology 26*(3), 265-271.

Delhalle, L., Saegerman, C., Messens, W., Farnir, F., Korsak, N., Van der Stede, Y., Daube, G. (2009b). Assessing interventions by quantitative risk assessment tools to reduce the risk of human salmonellosis from fresh minced pork meat in Belgium. *Journal of Food Protection 72*(11), 2252-2263.

Devriese, S., Huybrechts, I., Moreau, M., Van Oyen, H. (2006). De Belgische Voedselconsumptiepeiling 1 - 2004. Brussel, Wetenschappelijk Instituut Volksgezondheid, Afdeling Epidemiologie, IPH/EPI REPORTS N° 2006 - 016. Available at: http://www.iph.fgov.be/epidemio/epinl/foodnl/table04.htm (accessed 17-11-2010).

Duffy, G., Cummins, E., Nally, P., O' Brien, S., Carney, S., Henchion, M., Mahon, D., Butler, F. (2006). *E. coli* O157:H7 in beef burgers produced in the Republic of Ireland: a quantitative microbial risk assessment. Report published by Teagasc. Ashtown, Dublin 15, Ireland, Ashtown Food Research Centre. Available at: http://www.teagasc.ie/ashtown/research/foodsafety/E.coli%20O157.QMRAbeef.pdf (accessed 17-11-2010).

EFSA (2006a). The Community Summary Report on trends and sources of zoonoses, zoonotic agents, antimicrobial agents, antimicrobial resistance and foodborne outbreaks in the European Union in 2005. *The EFSA Journal 94*, 1-288.

EFSA (2006b). Guidance of the Scientific Committee on a request from EFSA related to Uncertainties in Dietary Exposure Assessment, Request No EFSA-Q-2004-019, Adopted on 14 December 2006. *The EFSA Journal 438*, 1-54.

EFSA (2007). Opinion of the Scientific Panel on Animal Health and Welfare on a selfmandate on the Framework for EFSA AHAW Risk Assessments. *The EFSA Journal 550*, 1-46.

EFSA (2009a). The Community summary report on trends and sources of zoonoses, zoonotic agents and foodborne outbreaks in the European Union in 2008. *The EFSA Journal 1496*.

EFSA (2009b). The Community summary report on trends and sources of zoonoses and zoonotic agents in the European Union in 2007. *The EFSA Journal 223*.

EFSA (2009c). Scientific Opinion. Transparency in Risk Assessment - Scientific Aspects. Guidance of the scientific committee on the transparency in the Scientific Aspects of Risk Assessments carried out by EFSA. Part 2: General Principles (Question No EFSA-Q-2005-050B). *The EFSA Journal 1051*, 1-22.

EFSA (2010). EFSA Panel on Biological Hazards; Scientific opinion on a quantitative microbiological risk assessment of *Salmonella* in slaughter and breeder pigs. *The EFSA Journal 8(4): 1547*.

EIIP (1996). Appendix F, EIIP Recommended approach to using the data attributing rating system., EIIP, Volume VI. Available at: http://www.epa.gov/ttn/chief/eiip/techreport/volume06/viapp-f.pdf (accessed 17-11-2010).

EPA (2003). Draft guidance on the development, evaluation, and application of regulatory environmental models. Washington, D.C., US EPA Council for Regulatory Environmental Modeling. Available at: http://www.epa.gov/crem/library/CREM Guidance Draft 2012_03.pdf (accessed 17-11-2010).

FAO/WHO (2002a). Risk assessment of *Campylobacter* spp. in broiler chickens and *Vibrio* spp. in seafood. Report of a Joint FAO/WHO Expert Consultation Bangkok, Thailand 5-9 August 2002. Available at: http://www.who.int/foodsafety/publications/micro/aug2002.pdf

FAO/WHO (2002b). Risk assessments of *Salmonella* in eggs and broiler chickens. Microbiological Risk Assessment Series no. 2. Available at: http://www.who.int/foodsafety/publications/micro/en/salmonella.pdf (accessed 17-11-2010).

FAO/WHO (2003). Hazard Characterization for Pathogens in Food and Water: Guidelines. Microbiological Risk Assessment Series no. 3. Available at: http://www.fao.org/ag/agn/agns/jemra_guidelines_hazard_en.asp (accessed 17-11-2010).

FAO/WHO (2006). Food safety risk analysis. A guide for national food safety authorities. Rome, FAO Food and Nutrition Paper 87. Available at: ftp://ftp.fao.org/docrep/fao/009/a0822e/a0822e00.pdf (accessed 17-11-2010).

FAO/WHO (2008). Exposure assessment of microbiological hazards in food: Guidelines. Microbiological Risk Assessment Series no. 7. Available at: http://www.fao.org/ag/agn/agns/jemra_guidelines_exposure_en.asp (accessed 17-11-2010).

FAO/WHO (2009). Risk characterization of microbiological hazards in food: guidelines. Microbiological Risk Assessment Series no. 17. Available at: http://www.fao.org/ag/agn/agns/jemra_guidelines_risk_en.asp (accessed 17-11-2010).

FAVV (2006). Activiteitenverslag 2005. Federaal Agentschap voor de Veiligheid van de Voedselketen. Available at: http://www.favv-afsca.fgov.be/home/pub/doc/rapport/AV_05_Nl_S.pdf (accessed 17-11-2010).

FAVV (2007). Activity report 2006. Federal Agency for the Safety of the Food Chain, FASFC Report

Fleiss, J.L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin 76*, 378-382.

Frey, H.C., Patil, S.R. (2002). Identification and review of sensitivity analysis methods. *Risk Analysis 22*(3), 553-578.

Frey, H.C., Mokhtari, A., Danish, T. (2003). Evaluation of selected sensitivity analysis methods based upon applications to two food safety process risk models. Prepared by North Carolina State University for the Office of Risk Assessment and Cost-Benefit Analysis, USDA, Washington, DC. Available at: http://www.ce.ncsu.edu/risk/Phase2Final.pdf (accessed 17-11-2010).

Frey, H.C., Mokhtari, A., Zheng, J. (2004). Recommended practice regarding selection, application, and interpretation of sensitivity analysis methods applied to food safety process risk models. Prepared by North Carolina State University for the Office of Risk Assessment and Cost-Benefit Analysis, USDA, Washington, DC. Available at: http://www.ce.ncsu.edu/risk/Phase3Final.pdf (accessed 17-11-2010).

FSIS (2001). Draft risk assessment of the public health impact of *Escherichia coli* O157:H7 in ground beef. Prepared for the Food Safety and Inspection Service and United States Department of Agriculture, Washington, DC. Available at: http://www.fsis.usda.gov/OPPDE/rdad/FRPubs/00-023N/00-023NReport.pdf (accessed 17-11-2010).

Funtowicz, S.O., Ravetz, J.R. (1990). *Uncertainty and Quality in Science for Policy*. Kluwer, Dordrecht, 229 p.

Gardner, I.A. (2004). An epidemiologic critique of current microbial risk assessment practices: the importance of prevalence and test accuracy data. *Journal of Food Protection 67*(9), 2000-2007.

Ghafir, Y., China, B., Korsak, N., Dierick, K., Collard, J.M., Godard, C., De Zutter, L., Daube, G. (2005). Belgian surveillance plans to assess changes in *Salmonella* prevalence in meat at different production stages. *Journal of Food Protection 68*(11), 2269-2277.

Goossens, L.H., Cooke, R.M. (2006). Expert Judgement - calibration and combination. In Workshop on the use of Expert Judgment for decision-making (Aix-en-Provence, France, June 2005).

Goulet, V., Jacques, M., Pigeon, M. (2009). expert: Modeling without data using expert opinion. *The R Journal 1*(1), 31-36.

Grijspeerdt, K., Messens, W., Bollaerts, K., Van Dessel, P., Delhalle, L., Maes, D., Boone, I., Mintiens, K. (2007). Quantitative risk assessment of human salmonellosis through consumption of pork in Belgium: a modular risk model. In: Proceedings of the 7th international conference on the epidemiology and control of foodborne pathogens in pork, Safepork 2007, Verona, Italy, pp. 5-9.

Gwet, K. (2002). Computing inter-rater reliability with the SAS system. Statistical Methods for Inter-rater Reliability Assessment 3. Available at: http://www.ccitonline.org/jking/homepage/SAS%20kappa_docum.doc (accessed 17-11-2010).

Haas, C.N., Rose, J.B., Gerba, C.P. (1999). *Quantitative Microbial Risk Assessment*. John Wiley & Sons, Inc., New York.

Havelaar, A. (2005). Recommendations for addressing quantitative microbial risk assessment at the European level. EFSA report. Available at: http://www.efsa.europa.eu/EFSA/DocumentSet/af_qrma_16thmeet_en_3b1,0.pdf (accessed 17-11-2010).

Havelaar, A.H., Nauta, M.J., Jansen, J.T. (2004). Fine-tuning Food Safety Objectives and risk assessment. *International Journal of Food Microbiology 93*(1), 11-29.

Havelaar, A.H., Mangen, M.J., de Koeijer, A.A., Bogaardt, M.J., Evers, E.G., Jacobs-Reitsma, W.F., van Pelt, W., Wagenaar, J.A., de Wit, G.A., van der Zee, H., Nauta, M.J. (2007). Effectiveness and efficiency of controlling campylobacter on broiler chicken meat. *Risk Analysis 27*(4), 831-844.

Havelaar, A.H., Evers, E.G., Nauta, M.J. (2008). Challenges of quantitative microbial risk assessment at EU level. *Trends in Food Science & Technology 19*(Supplement 1), S26-S33.

Hill, A., England, T., Snary, E.L., Cook, A., Kelly, L.A., Evans, S.J., Wooldridge, M. (2003). A 'farm-to-consumption' risk assessment for the adverse effect to human health of *Salmonella* Typhimurium in pigs. Weybridge, Veterinary Laboratories Agency.

Hill, A.A., Snary, E.L., Arnold, M.E., Alban, L., Cook, A.J. (2008). Dynamics of *Salmonella* transmission on a British pig grower-finisher farm: a stochastic model. *Epidemiology and Infection 136*(3), 320-333.

Honingh, N. (2004). Onzekerheden in het Mest- en ammoniakmodel. Utrecht, Copernicus Institute for Sustainable Development and Innovation (Universiteit Utrecht), NW&S-I-2004-20. Available at: http://www.chem.uu.nl/nws/www/publica/I2004-20.pdf

Horst, H.S., Dijkhuizen, A.A., Huirne, R.B.M., De Leeuw, P.W. (1998). Introduction of contagious animal diseases into The Netherlands: elicitation of expert opinions. *Livestock Production Science 53*(3), 253-264.

Huysmans, K., Nollet, N., Vandebroeck, M., Desmedt, K., Henot, D. (2003). Epidemiological research of *Salmonella*, *Mycoplasma hyopneumoniae* and *Lawsonia intracellularis* on Belgian pig farms. In: Proceedings of the 10th Annual Meeting of the Flemish Society for Veterinary Epidemiology and Economics, Leuven, Belgium, pp. 47-56.

Janssen, P.H., Petersen, A.C., van der Sluijs, J.P., Risbey, J.S., Ravetz, J.R. (2005). A guidance for assessing and communicating uncertainties. *Water Science and Technology 52*(6), 125-131.

Janssen, P.H.M., Petersen, A.C., van der Sluijs, J.P., Risbey, J.S., Ravetz, J.R. (2003). RIVM/MNP Guidance for Uncertainty Assessment and Communication: Quickscan Hints & Actions List. Bilthoven, The Netherlands, RIVM/MNP. Available at: http://www.mnp.nl/bibliotheek/digitaaldepot/Guidance_QS-HA.pdf (accessed 17-11-2010).

Kennedy, J., Jackson, V., Blair, I.S., McDowell, D.A., Cowan, C., Bolton, D.J. (2005). Food safety knowledge of consumers and the microbiological and temperature status of their refrigerators. *Journal of Food Protection 68*(7), 1421-1430.

Klapwijk, P.M., Jouve, J.L., Stringer, M.F. (2000). Microbiological risk assessment in Europe: the next decade. *International Journal of Food Microbiology 58*(3), 223-230.

Kloprogge, P., van der Sluijs, J., Petersen, A.C. (2005). A method for the analysis of assumptions in assessments, exploring the value-ladenness of two indicators in the fifth Dutch environmental outlook. Bilthoven, Netherlands Environmental Assessment Agency, Report no. 550002010/2005. Available at: http://www.mnp.nl/bibliotheek/rapporten/550002010.pdf (accessed 17-11-2010).

Kloprogge, P., van der Sluijs, J., Wardekker, J. (2007). Uncertainty communication. Issues and good practice. Utrecht, the Netherlands, Copernicus Institute Sustainable Development and Innovation, Report NWS-E-2007-199. Available at: http://www.nusap.net/downloads/reports/uncertainty_communication.pdf (accessed 17-11-2010).

Kloprogge, P., van der Sluijs, J.P., Petersen, A.C. (2010). A method for the analysis of assumptions in model-based environmental assessments. *Environmental Modelling & Software 26* (3), 289-301.

Knol, A.B., Petersen, A.C., van der Sluijs, J.P., Lebret, E. (2009). Dealing with uncertainties in environmental burden of disease assessment. *Environ Health 8*, 21.

Knol, A.B., Slottje, P., van der Sluijs, J.P., Lebret, E. (2010). The use of expert elicitation in environmental health impact assessment: a seven step procedure. *Environ Health 9*(1), 19.

Korsak, N., Groven, B., Jacob, B., Daube, G., Flament, E. (2002). Prevalence of *Salmonella* along a meat pork production system, In: Smulders, F.J.M., Collins, J.D. (editors). *Food safety assurance and veterinary public health*. Wageningen Academic Publishers, Wageningen, The Netherlands, pp. 260-262.

Korsak, N., Jacob, B., Groven, B., Etienne, G., China, B., Ghafir, Y., Daube, G. (2003). *Salmonella* contamination of pigs and pork in an integrated pig production system. *Journal of Food Protection 66*(7), 1126-1133.

Krapohl, S., Zurek, K. (2006). The perils of committee governance: intergovernmental bargaining during the BSE scandal in the European Union. *European Integration online Papers 10*(2).

Krayer von Krauss, M.P., Casman, E.A., Small, M.J. (2004). Elicitation of expert judgments of uncertainty in the risk assessment of herbicide-tolerant oilseed crops. *Risk Analysis 24*(6), 1515-1527.

Krayer von Krauss, M.P. (2005). Uncertainty in policy relevant sciences. Ph.D. thesis. Technical University of Denmark, Kgs. Lyngby.

Krayer von Krauss, M.P., Kaiser, M., Almaas, V., van der Sluijs, J., Kloprogge, P. (2008). Diagnosing and prioritizing uncertainties according to their relevance for policy: the case of transgene silencing. *Science of the Total Environment 390*(1), 23-34.

Lammerding, A.M., Fazil, A. (2000). Hazard identification and exposure assessment for microbial food safety risk assessment. *International Journal of Food Microbiology 58*(3), 147-157.

Lammerding, A.M. (2007). Using microbial risk assessment (MRA) in food safety management. Summary report of a workshop held in October 2005 in Prague, Czech Republic. Brussels, ILSI Europe Report Series. Available at: http://www.ilsi.org.ar/biblioteca/ILSI_Europa_Monografias/MRAWorkshopreport[1].pdf (accessed 17-11-2010).

Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics 33*, 159-173.

Lefebvre, D., Tatry, M.-V., Schepers, F., Rodenburg, T.B., Huneau-Salaün, A., Allain, V. (2010). TOward an information system on broiler welfare: Genetic selection Aspects (TOGA). Technical report submitted to EFSA in response of Question No EFSA-Q-2009-00796 (CFP/EFSA/AHAW/2009/02)

Lehmann, E.L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, Inc, San-Fransisco.

Lindqvist, R., Sylven, S., Vagsholm, I. (2002). Quantitative microbial risk assessment exemplified by *Staphylococcus aureus* in unripened cheese made from raw milk. *International Journal of Food Microbiology 78*(1-2), 155-170.

Lo Fo Wong, D., Pires, S.M., Group, M.W.W. (2006a). Pre-harvest Microbial Risk Assessment - an inventory of European studies. Danish Institute for Food and Veterinary Research (DFVF), Copenhagen, Denmark, Med-Vet-Net Workpackage 14, Report no. 06-005. Available at: http://www.medvetnet.org/pdf/Reports/Report_06-005.pdf

Lo Fo Wong, D., van Boven, M., Clough, H.E., Evers, E.G., Greiner, M., Havelaar, A.H., Little, C.L., Nöremark, M., Snary, E. (2006b). Evaluation of modeling techniques and approaches for pre-harvest microbial risk assessment. Med-Vet-Net Workpackage 14, Report no. 06-004. Available at: http://www.medvetnet.org/pdf/Reports/Report_06-004.pdf (accessed 17-11-2010).

Macgill, S., Fewtrell, L., Chudley, J., Kay, D. (2001). Quality audit and the assessment of waterborne risk, In: Fewtrell, L., Bartram, J. (editors). *Water Quality: Guidelines, Standards and Health*. IWA Publishing, London, pp. 185-206.

Marklinder, I.M., Lindblad, M., Eriksson, L.M., Finnson, A.M., Lindqvist, R. (2004). Home storage temperatures and consumer handling of refrigerated foods in Sweden. *Journal of Food Protection 67*(11), 2570-2577.

Mead, P.S., Slutsker, L., Dietz, V., McCaig, L.F., Bresee, J.S., Shapiro, C., Griffin, P.M., Tauxe, R.V. (1999). Food-related illness and death in the United States. *Emerging Infectious Diseases 5*(5), 607-625.

METZOON (2006). Activity Report METZOON project first working year [in Dutch]. Development of a methodology for the quantitative assessment of zoonotic risks in Belgium applied to *Salmonella* in pigs. METZOON (R-04/003). University of Hasselt, University of Liège, Ghent University, Institute for Agricultural and Fisheries Research, Veterinary and Agrochemical Research Centre, Scientific Institute for Public Health, Brussels, Belgium. Available at: www.metzoon.net

MNP/RU (2008). Stakeholder Participation Guidance for the Netherlands Environmental Assessment Agency. Bilthoven and Nijmegen, Practice Guide, Netherlands Environmental Assessment Agency and RU Nijmegen, MNP publication number 550032007. Available at: www.mnp.nl/en (accessed 17-11-2010).

Morgan, M.G., Henrion, M. (1990). *Uncertainty, a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press, New York.

Nauta, M., Havelaar, A. (2008). Risk-based standards for *Campylobacter* in the broiler meat chain. *Food Control 19*, 372-381.

Nauta, M.J., Evers, E.G., Takumi, K., Havelaar, A.H. (2001). Risk assessment of Shiga-toxin producing *Escherichia coli* O157 in steak tartare in the Netherlands. Bilthoven, RIVM, RIVM report 257851003. Available at: www.rivm.nl/bibliotheek/rapporten/257851003.pdf (accessed 17-11-2010).

Nauta, M.J., Litman, S., Barker, G.C., Carlin, F. (2003). A retail and consumer phase model for exposure assessment of *Bacillus cereus*. *International Journal of Food Microbiology 83*(2), 205-218.

Nauta, M.J., Jacobs-Reitsma, W.F., Havelaar, A.H. (2007). A risk assessment model for *Campylobacter* in broiler meat. *Risk Analysis 27*(4), 845-861.

Nollet, N., Houf, K., Dewulf, J., De Kruif, A., De Zutter, L., Maes, D. (2005). *Salmonella* in sows: a longitudinal study in farrow-to-finish pig herds. *Veterinary Research 36*(4), 645-656.

Notermans, S., Dufrenne, J.B., Teunis, P., Beumer, R., te Giffel, M.C., Peeters-Weem, P. (1997). A risk assessment study of *Bacillus cereus* present in pasteurized milk. *Food Microbiology 14*, 143-151.

OMB (2002). Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by Federal Agencies. Office of Management and Budget, Executive Office of the President. Available at: http://www.whitehouse.gov/omb/fedreg/reproducible2.pdf (accessed 17-11-2010).

OMB (2004). Final information quality bulletin for peer review. Washington D.C., Office of Management and Budget. Available at: http://www.ocio.usda.gov/qi_guide/doc/peer_bulletin.pdf (accessed 17-11-2010).

Paisley, L. (2007). Quality Audit Worksheet. Available at: http://www.foot-and-mouth.org/fmd-csf-ca/community/work-package-3-risks/Quality-Audit-Worksheet_V5.xls/view?searchterm=audit (accessed 17-11-2010).

Petersen, A.C., Janssen, P.H.M., van der Sluijs, J.P., Risbey, J.S., Ravetz, J.R. (2003). RIVM/MNP Guidance for Uncertainty Assessment and Communication: Mini-Checklist & Quickscan Questionnaire. RIVM/MNP. Available at: http://www.rivm.nl/bibliotheek/digitaaldepot/Guidance_MC_QS-Q.pdf (accessed 17-11-2010).

Petersen, A.C., Cath, A., Hage, M., Kunseler, E., van der Sluijs, J.P. (in press). Post-Normal science in practice at the Netherlands Environmental Assessment Agency. *Science, Technology, & Human values*(Special issue on 'Where Now For Post-Normal Science').

Pfeifle, D., Dressel, K., Wright, D. (2006). Good practices in risk communication. Deliverable 3, A report of the STARC consortium to the European Commission, 31 August 2006. Available at: http://starc.jrc.it (accessed 17-11-2010).

Pigeon, M., Jacques, M., Goulet, V. (2009). Modeling without data using expert opinion. Available at: http://cran.r-project.org/

Ranta, J., Tuominen, P., Rauttiainen, E., Maijala, R. (2004). *Salmonella* in Finnish pork production – a quantitative risk assessment. Helsinki, Finland, National Veterinary and Food Research Institute, EELA reports 04/2004.

Ravetz, J., Funtowicz, S. (1999). Post-normal science - an insight now maturing. *Futures 31*, 641-646.

Refsgaard, J.C., Henriksen, H.J., Harrar, W.G., Scholten, H., Kassahun, A. (2005). Quality assurance in model based water management - review of existing practice and outline of new approaches. *Environmental Modelling & Software 20*(10), 1201-1215.

Refsgaard, J.C., Van der Sluijs, J., Brown, C.D., van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources 29*, 1586-1597.

Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A. (2007). Uncertainty in the Environmental Modelling Process: A Review. *Environmental Modelling & Software 22*(11), 1543-1556.

Risbey, J., Van der Sluijs, J., Ravetz, J. (2001a). A protocol for assessment of uncertainty and strength of emissions data. Dep't. of Science, Technology and Society, Utrecht University, Report no. NW&S-E-2001-10. Available at: http://www.nusap.net/downloads/reports/E2001-10.pdf (accessed 17-11-2010).

Risbey, J., van der Sluijs, J., Ravetz, J., Janssen, P. (2001b). A checklist for quality assistance in environmental modelling Utrecht, Dep't. of Science, Technology and Society, Utrecht University, Report no. NW&S-E-2001-11. Available at: http://www.nusap.net/downloads/reports/E2001-11.pdf (accessed 17-11-2010).

Risbey, J., van der Sluijs, J., Kloprogge, P., Ravetz, J., Funtowicz, S., Corral Quintana, S. (2005). Application of a checklist for quality assistance in environmental modelling to an energy model. *Environment Modelling and Assessment 10*, 63-79.

Rosenquist, H., Nielsen, N.L., Sommer, H.M., Norrung, B., Christensen, B.B. (2003). Quantitative risk assessment of human campylobacteriosis associated with thermophilic *Campylobacter* species in chickens. *International Journal of Food Microbiology 83*(1), 87-103.

Saltelli, A., Chan, K., Scott, M. (2000). *Sensitivity Analysis*. John Wiley & Sons publishers, West Sussex.

Sanchez, J., Dohoo, I.R., Christensen, J., Rajic, A. (2007). Factors influencing the prevalence of *Salmonella* spp. in swine farms: A meta-analysis approach. *Preventive Veterinary Medicine 81*(1-3), 148-177.

Sargeant, J.M., Amezcua, M., Rajic, A., Wadell, L. (2005). A guide to conducting systematic reviews in agri-food public health. Department of Clinical Epidemiology and Biostatistics, McMaster University, Ontario, Canada, and Laboratory for Foodborne Zoonoses, Public Health Agency of Canada, Guelph, Ontario, Canada. Available at: http://dsp-psd.pwgsc.gc.ca/Collection/HP5-9-2005E.pdf (accessed 17-11-2010).

Schlundt, J. (2000). Comparison of microbiological risk assessment studies published. *International Journal of Food Microbiology 58*(3), 197-202.

Scholz, R.W., Hansmann, R. (2007). Combining experts' risk judgments on technology performance of phytoremediation: self-confidence ratings, averaging procedures, and formative consensus building. *Risk Analysis 27*(1), 225-240.

Shaffer, J.P. (1995). Multiple Hypothesis Testing. *Annual Review of Psychology 45*, 561-584.

SIPH (2010). Données de surveillance du Centre National de Référence des *Salmonella* et *Shigella*, Belgique 2009. Rapport 2009. Scientific Institute for Public Health, Report n° D/2010/2505/54. Available at: http://bacterio.wiv-isp.be (accessed 17-11-2010).

Slottje, P., van der Sluijs, J.P., Knol, A.B. (2008). Expert Elicitation: Methodological suggestions for its use in environmental health impact assessments. Utrecht University, RIVM Letter report 630004001/2008. Available at: http://www.nusap.net/downloads/reports/Expert_Elicitation.pdf (accessed 17-11-2010).

Stärk, K.D., Wingstrand, A., Dahl, J., Mogelmose, V., Lo Fo Wong, D.M. (2002). Differences and similarities among experts' opinions on *Salmonella enterica* dynamics in swine pre-harvest. *Preventive Veterinary Medicine 53*(1-2), 7-20.

StatXact (1999). Release 4.0.1., Cytel Software Corporation: Cambridge, MA, USA.

Tielemans, E., Marquart, H., De Cock, J., Groenewold, M., Van Hemmen, J. (2002). A proposal for evaluation of exposure data. *Annals of Occupational Hygiene 46*(3), 287-297.

Tuominen, P. (2009). Developping risk-based food safety management. Ph.D thesis, Faculty of Veterinary Medicine, University of Helsinki, Finland.

USDA-FSIS (1998). *Salmonella* Enteritidis risk assessment. Shell eggs and egg products. Prepared for FSIS by the *Salmonella* Enteritidis Risk Assessment Team. Final report. Available at: http://www.fsis.usda.gov/ophs/risk/contents.htm (accessed 17-11-2010).

van der Fels-Klerx, H.J., Cooke, R.M., Nauta, M.N., Goossens, L.H., Havelaar, A.H. (2005). A structured expert judgment study for a model of *Campylobacter* transmission during broiler-chicken processing. *Risk Analysis 25*(1), 109-124.

van der Fels-Klerx, I.H., Goossens, L.H., Saatkamp, H.W., Horst, S.H. (2002). Elicitation of quantitative data from a heterogeneous expert panel: formal process and application in animal health. *Risk Analysis 22*(1), 67-81.

van der Gaag, M., Huirne, R.B.M. (2002). Elicitation of expert knowledge on controlling *Salmonella* in the pork chain. *Chain and network science*, 135-147.

van der Gaag, M., Vos, F., Saatkamp, H.W., van Boven, M., Beek, v., Huirne, R.B.M. (2004). A state-transition simulation model for the spread of *Salmonella* in the pork supply chain. *European Journal of Operational Research 156*, 782-798.

van der Sluijs, J.P. (2002). A way out of the credibility crisis of models used in integrated environmental assessment. *Futures 34*(2), 133-146.

van der Sluijs, J.P., Potting, J., Risbey, J., van Vuuren, D., de Vries, B., Beusen, A., Heuberger, P., Corral Quintana, S., Funtowicz, S., Kloprogge, P., Nuijten, D., Petersen, A.C., Ravetz, J. (2002). Uncertainty assessment of the IMAGE/TIMER B1 CO2 emissions scenario, using the NUSAP method. Bilthoven, the Netherlands, Dutch National Research Program on Climate Change, Report no. 410 200 104. Available at: http://www.nusap.net

van der Sluijs, J.P., Risbey, J.S., Kloprogge, P., Ravetz, J.R., Funtowicz, S.O., Corral Quintana, S., Guimarães Pereira, Â., De Marchi, B., Petersen, A.C., Janssen, P.H.M., Hoppe, R., Huijs, S.W.F. (2003). RIVM/MNP Guidance for Uncertainty Assessment and Communication: Detailed Guidance. Utrecht, Utrecht University, Copernicus Institute/Dep't. of Science, Technology and Society and RIVM. Available at: http://www.nusap.net/downloads/detailedguidance.pdf (accessed 17-11-2010).

van der Sluijs, J.P., Janssen, P.H.M., Petersen, A.C., Kloprogge, P., Risbey, J.S., Tuinstra, W., Ravetz, J.R. (2004). RIVM/MNP Guidance for Uncertainty Assessment and Communication: Tool Catalogue for Uncertainty Assessment. Utrecht, the Netherlands, Utrecht University and RIVM-MNP, Report no. NWS-E-2004-37. Available at: http://www.nusap.net/downloads/toolcatalogue.pdf (accessed 17-11-2010).

van der Sluijs, J.P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J., Risbey, J. (2005a). Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: The NUSAP System. *Risk Analysis 25*(2), 481-492.

van der Sluijs, J.P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz., J., Risbey, J. (2005b). Experiences with the NUSAP system for multidimensional uncertainty assessment. *Water Science and Technology 25*(6), 133-144.

van der Sluijs, J.P., Risbey, J.S., Ravetz, J. (2005c). Uncertainty Assessment of VOC Emissions from Paint in the Netherlands Using the NUSAP System. *Environmental Monitoring and Assessment 105*, 229-259.

van der Sluijs, J.P. (2006). Uncertainty, assumptions, and value commitments in the knowledge base of complex environmental problems, In: Ângela Guimarães Pereira, Sofia Guedes Vaz, Sylvia Tognetti (editors). *Interfaces between science and society*. Greenleaf.

van der Sluijs, J.P. (2007). Nusap Kite Diagram Maker. Available at: http://www.jvds.nl/makekite410.php (accessed 17-11-2010).

van der Sluijs, J.P., Petersen, A.C., Janssen, P.H.M., Risbey, J.S., Ravetz, J.R. (2008). Exploring the quality of evidence for complex and contested policy decisions. *Environmental Research Letters 3 024008*, (9 pp).

Van der Stede, Y., Bollaerts, K., Cortinas Abrahantes, J., Imberechts, H., Mintiens, K. (2007). Identification of *Salmonella* risk farms by serological surveillance at pre-harvest: Mission Impossible!? In: Proceedings of the 4th lustrum of the Dutch and Flemish society for Veterinary Epidemiology and Economics, December 12-13, 2007, Wageningen, the Netherlands, pp. 65-70.

van Rooij, M.W.J.W. (2005). Performance of expert judgement methods with expert modelling. Master's thesis. Department of Applied Mathematics and Risk Analysis, Delft University of Technology, the Netherlands.

Van Vlaenderen, I., Biront, P., Verleye, A., Imberechts, H., Stockmans, A. (1999). *Salmonella* uitscheiding op varkensbedrijven - Screening 1998, unpublished report. Brussels, Belgium, Veterinary and Agrochemical Research Centre.

VLA-DTU-RIVM (2010). EFSA Quantitative microbiological risk assessment on *Salmonella* in slaughter and breeder pigs: Final Report. Question No EFSA-Q-2007-00245. Grant number CFP/EFSA/BIOHAZ/2007/01. Veterinary Laboratories Agency. Available at: http://www.efsa.europa.eu/en/scdocs/scdoc/46e.htm (accessed 17-11-2010).

Vose, D. (2000). *Risk analysis - A quantitative guide*, 2nd Edition. John Wiley and Sons publishers Ltd., Chichester.

Voysey, P.A., Brown, M. (2000). Microbiological risk assessment: a new approach to food safety control. *International Journal of Food Microbiology 58*(3), 173-179.

Walker, K.D., Catalano, P., Hammitt, J.K., Evans, J.S. (2003a). Use of expert judgment in exposure assessment: part 2. Calibration of expert judgments about personal exposures to benzene. *Journal of Exposure Analysis and Environmental Epidemiology 13*(1), 1-16.

Walker, W., Harremoës, P., Rotmans, J., Van der Sluijs, J., Van Asselt, M., Jansen, P., Krayer von Krauss, M.P. (2003b). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Journal of Integrated Assessment 4*(1), 5-17.

Walls, I. (2006). Role of quantitative risk assessment and food safety objectives in managing *Listeria monocytogenes* on ready-to-eat meats. *Meat Science 74*, 66-75.

Walls, I. (2007). Framework for identification and collection of data useful for risk assessments of microbial foodborne or waterborne hazards: a report from the International Life Sciences Institute Research Foundation Advisory Committee on data collection for microbial risk assessment. *Journal of Food Protection 70*(7), 1744-1751.

Wardekker, J.A., van der Sluijs, J.P., Janssen, P.H.M., Kloprogge, P., Petersen, A.C. (2008). Uncertainty communication in environmental assessments: views from the Dutch science-policy interface. *Environmental Science & Policy 11*(7), 627-641.

WHO (2008). Guidance document on characterizing and communicating uncertainty in exposure assessment, Harmonization Project Document No. 6. IPCS Harmonization Project Document. Part 1. Available at: http://www.who.int/ipcs/methods/harmonization/areas/exposure/en/index.html (accessed 17-11-2010).

SUMMARY

Quantitative microbial risk assessment (QMRA) is being increasingly used to support decision-making for food safety issues. Decision-makers need to know whether these QMRA results can be trusted, especially when urgent and important decisions have to be made. This can be achieved by setting up a quality assurance (QA) framework for QMRA. A Belgian risk assessment project (the METZOON project) aiming to assess the risk of human salmonellosis due to the consumption of fresh minced pork meat was used as a case study to develop and implement QA methods for the evaluation of the quality of input data, expert opinion, model assumptions, and the quality of the QMRA model (the METZOON model).

The first part of this thesis consists of a literature review of available QA methods of interest in QMRA (chapter 2). In the next experimental part, different QA methods were applied to the METZOON model.

A structured expert elicitation study (chapter 4) was set up to fill in missing parameters for the METZOON model. Judgements of experts were used to derive subjective probability density functions (PDFs) to quantify the uncertainty on the model input parameters. The elicitation was based on Cooke's classical model (Cooke, 1991) which aims to achieve a rational consensus about the elicitation protocol and allowed comparing different weighting schemes for the aggregation of the experts' PDFs. Unique to this method was the fact that the performance of experts as probability assessors was measured by the experts' ability to correctly and precisely provide estimates for a set of seed variables (=variables from the experts' area of expertise for which the true values were known to the analyst). The weighting scheme using the experts' performance on a set of calibration variables was chosen to obtain the combined uncertainty distributions of lacking parameters for the METZOON model.

A novel method for the assessment of data quality, known as the NUSAP (Numeral Unit Spread Assessment Pedigree) system (chapter 5) was tested to screen the quality of the METZOON input parameters. First, an inventory with the essential characteristics of parameters including the source of information, the sampling methodology and distributional characteristics was established. Subsequently the quality of these parameters was evaluated and scored by experts using objective criteria (proxy, empirical basis, methodological rigour and validation). The NUSAP method allowed to debate on the quality of the parameters

within the members of the risk assessment team using a structured format. The quality evaluation was supported by graphical representations which facilitated decisions on the inclusion or exclusion of inputs into the model.

It is well known that assumptions and subjective choices can have a large impact on the output of a risk assessment. To assess the value-ladenness (degree of subjectivity) of assumptions in the METZOON model a structured approach based on the protocol by Kloprogge et al. (2005) was chosen (chapter 6). The key assumptions for the METZOON model were first identified and then evaluated by experts in a workshop using four criteria: the influence of situational limitations, the plausibility, the choice space and the agreement among peers. The quality of the assumptions was graphically represented (using kite diagrams, pedigree charts and diagnostic diagrams) and allowed to identify assumptions characterised by high degree of subjectivity and high expected influence on the model results, which can be considered as weak links in the model. The quality assessment of the assumptions was taken into account to modify parts of the METZOON model, and allows to increase the transparency in the QMRA process.

In a last application of a QA method, a quality audit checklist (Paisley, 2007) was used to critically review and score the quality of the METZOON model and to identify its strengths and weaknesses (chapter 7). A high total score (87%) was obtained by reviewing the METZOON model with the Paisley-checklist. A higher score would have been obtained if the model was subjected to external peer review, if a sensitivity analysis, validation of the model with recent data, updating/replacing expert judgement data with empirical data was carried out. It would also be advisable to repeat the NUSAP/Pedigree on the input data and assumptions of the final model. The checklist can be used in its current form to evaluate QMRA models and to support model improvements from the early phases of development up to the finalised model for internal as well as for external peer review of QMRAs.

The applied QA methods were found useful to improve the transparency in the QMRA process and to open the debate about the relevance (fitness for purpose) of a QMRA. A pragmatic approach by combining several QA methods is recommendable, as the application of one QA method often facilitates the application of another method. Many QA methods (NUSAP, structured expert judgement, checklists) are however not yet or insufficiently described in QMRA related guidelines (at EFSA and WHO level). Another limiting factor is the time and resources which need to be taken into account as well. To understand the degree of quality required from a QMRA a clear communication with the risk managers is required. It is therefore necessary to strengthen the training in QA methods and in the communication

of its results. Understanding the usefulness of these QA methods could improve among the risk analysis actors when they will be tested in large number of QMRAs.

Keywords: Quality assurance, NUSAP, data quality, assumptions, structured expert judgement, checklist, quantitative microbial risk assessment

RÉSUMÉ

Introduction

L’appréciation quantitative des risques microbiologiques (AQRM) est utilisée de plus en plus pour soutenir des décisions dans le domaine de la sécurité des aliments. Compte tenu de l’importance pour la santé publique, des enjeux économiques et pour faire face à d’éventuelles critiques, les décideurs souhaitent savoir si les résultats des AQRMs sont suffisamment valables et précis afin de pouvoir prendre des décisions importantes et urgentes. Ceci nécessite l’élaboration d’une structure d’assurance qualité (AQ) du processus de l’AQRM. Le projet multidisciplinaire belge METZOON (Bollaerts et al., 2009) avait pour but l’appréciation du risque de la fourche à la fourchette de la salmonellose chez l’homme suite à la consommation de viande hachée de porc. Dans le cadre de ce projet, des méthodes d’assurance qualité ont été développées et appliquées afin d’évaluer la qualité des données, des opinions des experts ainsi que des hypothèses du modèle METZOON.

Une première partie des recherches comprenait une synthèse de la littérature concernant les méthodes d’assurance qualité ayant un intérêt pour les AQRM (chapitre 2). Parmi les méthodes décrites, quatre d’entre elles ont été appliquées dans l’élaboration et l’évaluation du modèle METZOON.

Les objectifs spécifiques des recherches consistaient à :

1. Effectuer une étude pour l’élicitation structurée de jugements d’experts afin d’évaluer si et comment les jugements d’un panel hétérogène d’experts peuvent être combinés afin d’obtenir des distributions pour des paramètres manquants du modèle METZOON (chapitre 4).
2. Préévaluer la qualité des paramètres potentiels pour le modèle METZOON, en appliquant le système NUSAP (Numéraire, Unité, Dispersion, Evaluation, Pédigrée) afin de fournir une base objective pour la sélection de paramètres dans le modèle (Funtowicz et Ravetz, 1990). En outre, des outils graphiques permettant la communication de la qualité des données ont été évalués (chapitre 5).
3. Identifier les hypothèses du modèle METZOON, évaluer leur degré de subjectivité ainsi que leur impact sur les résultats du modèle (chapitre 6).

4. Evaluer une check-list pour l'analyse critique du modèle permettant l'identification de ses points forts et faibles (chapitre 7).

L'ensemble de ces méthodes ont été appliquées afin d'augmenter la transparence et la confiance dans les résultats d'une AQRM.

Résultats

1. Dans l'étude de l'élicitation de jugements d'experts, les jugements de onze experts ont été utilisés afin de dériver les fonctions de densité de probabilités subjectives (FDP) et de quantifier le degré d'incertitude sur les paramètres manquants du modèle. L'élicitation se basait sur le modèle classique de Cooke (1991) qui vise à aboutir à un consensus rationnel sur la méthode d'élicitation. Cette méthode a permis de comparer différents schémas de pondération pour l'agrégation des FDPs des experts. Ce qui caractérise cette méthode est le fait que l'aptitude des experts à exprimer leurs estimations subjectives est mesurée à partir de leur capacité à fournir des estimations correctes et précises pour une série de variables de calibration (c'est-à-dire des variables comprises dans le domaine de l'expertise des experts dont les valeurs réelles sont connues par l'évaluateur). Le schéma de pondération qui utilise la performance des experts par rapport aux variables de calibration a été utilisé pour fournir les distributions combinées des paramètres manquants du modèle METZOON.

2. L'évaluation de la qualité des paramètres a débuté par l'établissement d'un inventaire reprenant les caractéristiques essentielles de paramètres (source de l'information, méthode d'échantillonnage, caractéristiques de distribution). Ensuite, les experts du projet METZOON ont évalué la qualité de ces paramètres tout en attribuant des scores à partir de quatre critères objectifs (proxy, base empirique, rigueur de la méthode et validation). Les paramètres ont obtenu des scores plus faibles en ce qui concerne le critère « validation » comparés aux autres critères étudiés. En outre, la qualité globale des paramètres évalués par les experts (moyenne pondérée des quatre critères) était plus élevée au niveau du module de production primaire par rapport aux paramètres des modules situés plus loin dans la chaîne alimentaire (transport, abattoir, découpe, hachage, distribution et consommation). La méthode NUSAP a permis de structurer la discussion concernant la qualité des données entre les experts du projet. Les représentations graphiques de la qualité ont facilité les décisions d'exclusion ou d'inclusion des données dans le modèle.

3. Afin d'évaluer le degré de subjectivité des hypothèses du modèle METZOON, une approche structurée basée sur le protocole de Kloprogge et al. (2005) a été choisie. Après que les hypothèses-clés du modèle aient été identifiées, elles ont été évaluées par des experts

pendant une réunion de travail au moyen de quatre critères : les contraintes situationnelles, la plausibilité, les hypothèses alternatives et l'accord avec la communauté de pairs. La qualité des hypothèses a été visualisée par des graphiques (les diagrammes en cerf-volant, le graphique pédigrée et le diagramme diagnostic). Ce dernier type de diagramme permet d'identifier les hypothèses caractérisées par un degré élevé de subjectivité et ayant une influence importante sur les résultats du modèle, qui peuvent être considérées comme les chaînons faibles du modèle. Il s'agissait entre autres de l'hypothèse liée à la pertinence de l'utilisation de données de concentration de salmonelles dans le modèle METZOON. Les diagrammes en cerf-volant ont permis de représenter les scores des différents critères en même temps que d'éventuels (dés)accords entre les experts. L'évaluation de la qualité des hypothèses a été prise en compte pour ajuster le modèle METZOON et a permis d'améliorer la transparence du modèle.

4. Une check-list (basée sur celle de Paisley, 2007) a été appliquée pour évaluer le modèle METZOON. Une première partie de la check-list comprenait des questions générales (objectif de l'AQRM, identification des acteurs impliqués, pertinence des données de sortie du modèle) et des questions plus spécifiques liées à l'évaluation des incertitudes, à l'évaluation de la qualité de l'information utilisée dans l'AQRM, à la structure du modèle, aux techniques de modélisation, à la validation du modèle, à la révision par des pairs, et à la communication du risque. Les réponses à chacune des questions étaient codifiées sur une échelle de 0 à 5 et ont été additionnées de façon à produire un score final (87%) qui donne une indication de la qualité d'un modèle. Dans l'évaluation il s'est avéré qu'un score plus élevé aurait pu être obtenu si le modèle avait été soumis à une évaluation externe par des pairs, si une analyse de sensibilité avait été effectuée, si le modèle avait été validé avec des résultats plus récents ou obtenus grâce à d'échantillons plus étendus, si certaines données obtenues au moyen d'opinions d'experts avaient été remplacées par des données empiriques. Les scores obtenus dans la check-list reflètent l'état actuel du modèle. Ces scores peuvent aider à améliorer la qualité du modèle si cela est jugé nécessaire et réalisable compte tenu de diverses contraintes (exigences des décideurs, argent, temps, disponibilité d'experts etc). Une des questions-clés consiste en l'évaluation de l'aptitude à l'emploi du modèle (fitness for purpose). En ce moment, il n'est pas encore possible d'établir l'efficacité du modèle dans le processus de décision.

Conclusions et perspectives

Pour traiter les problèmes décisionnels caractérisés par des enjeux importants, un grand nombre d'incertitudes et une pluralité de valeurs, la mise en place d'une structure d'assurance qualité et l'utilisation de méthodes appropriées d'assurance qualité est nécessaire. Plus particulièrement, de nouvelles méthodes pour l'évaluation de la qualité de données (l'approche NUSAP) et de l'évaluation d'hypothèses, l'utilisation d'une méthode pour l'élicitation structurée de jugements d'experts et une check-list ont été appliquées pour évaluer le modèle METZOON. L'objectif principal de ces méthodes est d'augmenter la transparence du processus de l'AQRM et d'installer une structure pour débattre de façon critique les points forts et les lacunes du modèle. Les méthodes proposées incluent toutes un apport d'opinions d'experts. Compte tenu du fait que l'AQRM est une discipline pluridisciplinaire, il est indispensable de constituer un groupe d'experts représentatif des disciplines traitées dans l'analyse de risques et d'obtenir un consensus rationnel quant aux méthodes utilisées. La méthode NUSAP a été testée pour la première fois pour évaluer la qualité des données et des hypothèses d'un modèle AQRM. Elle a l'avantage d'utiliser des critères universels de qualité et a été acceptée comme une méthode valable pour l'évaluation de paramètres. L'évaluation des données a été facilitée grâce à un inventaire résumant leurs caractéristiques essentielles (source de l'information, méthode d'échantillonnage, caractéristiques de distribution). La majorité des paramètres ont obtenu des scores très faibles pour le critère de validation, ce qui peut être expliqué par le fait que la plupart des paramètres n'ont pas été collectés avec l'objectif d'être utilisés dans une AQRM. L'évaluation des hypothèses du modèle ainsi que la représentation graphique via le diagramme diagnostic ont facilité la réflexion des points forts et faibles du modèle. Parmi les critères utilisés dans l'évaluation des hypothèses, les experts ont éprouvé des difficultés à donner des scores quant au critère « accord avec la communauté de pairs » parce qu'il est difficile de s'imaginer les choix de la communauté de pairs pour les hypothèses discutées.

Afin d'atteindre les objectifs d'une AQRM, une approche intégrée et pragmatique de l'assurance qualité est recommandable. Une assurance qualité est optimale en combinant plusieurs méthodes puisque l'application d'une méthode d'assurance qualité facilite souvent l'application d'une autre méthode. Afin d'augmenter l'efficacité (fitness for purpose) du modèle, la communication parmi les acteurs (scientifiques, décideurs et parties prenantes) du processus d'analyse de risques devrait être renforcée. L'assurance qualité devrait débuter au

moment de la formulation du problème à modéliser et nécessite dès lors une stratégie claire de communication entre tous les acteurs de façon à cerner le problème. Le modèle conceptuel de l'AQRM peut être évalué grâce à des check-lists et une vérification interne/externe par des pairs. La phase d'acquisition des données est cruciale et nécessite une communication claire (basée sur les principes de confidentialité) avec les détenteurs de données. L'application des méthodes telles la revue systématique et le système NUSAP qui combinent l'utilisation de critères objectifs de qualité avec des jugements d'experts peuvent aider à caractériser la qualité des données ainsi qu'au processus d'inclusion et d'exclusion en tant qu'intrant dans le modèle. Le manque de données peut être comblé grâce à l'apport d'opinions d'experts via des méthodes structurées. Par la suite, l'aptitude du modèle peut être évaluée par une variété de méthodes (évaluation de l'incertitude, vérification du modèle, analyse de sensibilité, analyse des scénarios, évaluation des hypothèses et validation du modèle, check-lists). Les lacunes et les points forts découverts lors de l'application de ces méthodes permettent de savoir si le modèle est apte à l'emploi ou s'il est nécessaire de l'ajuster avant son utilisation dans le processus décisionnel.

Pour une amélioration de l'acceptation des nouvelles méthodes d'assurance qualité proposées dans cette thèse celles-ci devraient être testées dans un large éventail d'AQRMs.

Tous les acteurs actifs dans l'analyse de risques (scientifiques, décideurs, parties prenantes) devraient être sensibilisés à l'importance de l'assurance qualité, grâce à une communication renforcée, des représentations graphiques de qualité et une formation aux méthodes d'assurance-qualité.

Ceci nécessiterait également des directives sur l'utilisation des méthodes d'assurance qualité (au niveau EFSA ou FAO/OMS), des références aux logiciels et aux études-cas, ainsi que le développement de critères de qualité (label) pour les AQRMs aux enjeux importants.

En conclusion, tout en tenant compte que la mise en œuvre de méthodes d'assurance qualité exige un investissement en temps et en argent, celle-ci peut être compensée par l'augmentation de la confiance et de l'acceptabilité des résultats parmi les acteurs de l'analyse de risques.

Mots-clés : Assurance qualité, NUSAP, qualité des données, hypothèse, jugement structuré d'experts, check-list, appréciation quantitative des risques microbiologiques.

Presses de la Faculté de Médecine vétérinaire de l'Université de Liège

4000 Liège (Belgique)

D/2010/0480/35
ISBN 978-2-930404-87-5

9 782930 404875