Application of Systematic Review Methodology to Food and Feed Safety Assessments to Support Decision Making

Jonathan J. Deeks  
*University of Birmingham*

Geoff K. Frampton  
*University of Southampton*

Julie M. Glanville  
*University of York*

Mattias Greiner  
*Federal Institute for Risk Assessment*

Follow this and additional works at: [http://lib.dr.iastate.edu/vdpam_pubs](http://lib.dr.iastate.edu/vdpam_pubs)

The complete bibliographic information for this item can be found at [http://lib.dr.iastate.edu/vdpam_pubs/30](http://lib.dr.iastate.edu/vdpam_pubs/30). For information on how to cite this item, please visit [http://lib.dr.iastate.edu/howtocite.html](http://lib.dr.iastate.edu/howtocite.html).

This Article is brought to you for free and open access by the Veterinary Diagnostic and Production Animal Medicine at Iowa State University Digital Repository. It has been accepted for inclusion in Veterinary Diagnostic and Production Animal Medicine Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Application of Systematic Review Methodology to Food and Feed Safety Assessments to Support Decision Making

Abstract
Systematic reviews are commonly used in human health research to provide overviews of existing evidence pertinent to clearly formulated specific questions, using pre-specified and standardised methods to identify and critically appraise relevant research, and to collect, report and analyse data from the studies that are included in the reviews. Formal systematic reviews have rarely been used in food and feed safety risk assessments and the existing systematic review methods in other disciplines may not be directly applicable in this field. This Guidance aims to assist the application of systematic reviews to food and feed safety risk assessments in support of decision making, by describing a framework for identifying the different types of question suitable for systematic review generated by the risk assessment process and for determining the need for systematic reviews when dealing with broad food and feed safety policy problems. The Guidance provides suggestions and examples for the conduct of eight key steps in the systematic review process (preparing a review, searching for studies, selecting studies for inclusion, collecting data from included studies, assessing the methodological quality of included studies, synthesising data from the studies, presenting data and results, and interpreting the results and drawing conclusions) for questions suitable for systematic reviews, taking into account issues that may be unique to food and feed safety. Due to its methodological rigor and its objective and transparent nature, systematic review methodology and its principles could provide additional value for answering well-formulated specific questions generated by the risk assessment process or other analytical frameworks in food and feed safety. Regular updates of this Guidance are foreseen in light of experience and new evidence both in food and feed safety and systematic review methodology.

Keywords
Systematic review, food safety, feed safety, literature search, risk assessment

Disciplines
Community Health and Preventive Medicine | Food Science | Laboratory and Basic Science Research | Large or Food Animal and Equine Medicine

Comments

Authors
Jonathan J. Deeks, Geoff K. Frampton, Julie M. Glanville, Mattias Greiner, Julian P. T. Higgins, Gabor Lövei, Annette M. O'Connor, Andrew S. Pullin, and Andrijana Rajić

This article is available at Iowa State University Digital Repository: http://lib.dr.iastate.edu/vdpam_pubs/30
GUIDANCE OF EFSA

Application of systematic review methodology to food and feed safety assessments to support decision making

EFSA Guidance for those carrying out systematic reviews

European Food Safety Authority

European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

Systematic reviews are commonly used in human health research to provide overviews of existing evidence pertinent to clearly formulated specific questions, using pre-specified and standardised methods to identify and critically appraise relevant research, and to collect, report and analyse data from the studies that are included in the reviews. Formal systematic reviews have rarely been used in food and feed safety risk assessments and the existing systematic review methods in other disciplines may not be directly applicable in this field. This Guidance aims to assist the application of systematic reviews to food and feed safety risk assessments in support of decision making, by describing a framework for identifying the different types of question suitable for systematic review generated by the risk assessment process and for determining the need for systematic reviews when dealing with broad food and feed safety policy problems. The Guidance provides suggestions and examples for the conduct of eight key steps in the systematic review process (preparing a review, searching for studies, selecting studies for inclusion, collecting data from included studies, assessing the methodological quality of included studies, synthesising data from the studies, presenting data and results, and interpreting the results and drawing conclusions) for questions suitable for systematic reviews, taking into account issues that may be unique to food and feed safety. Due to its methodological rigor and its objective and transparent nature, systematic review methodology and its principles could provide additional value for answering well-formulated specific questions generated by the risk assessment process or other analytical frameworks in food and feed safety. Regular updates of this Guidance are foreseen in light of experience and new evidence both in food and feed safety and systematic review methodology.

KEY WORDS

Systematic review, food safety, feed safety, literature search, risk assessment.
SUMMARY

A systematic review (SR) is an overview of existing evidence pertinent to a clearly formulated question, which uses pre-specified and standardised methods to identify and critically appraise relevant research, and to collect, report and analyse data from the studies that are included in the review.

Despite the common use of systematic reviews in areas of human health research, formal systematic reviews have rarely been used in food and feed safety and the existing SR methods may not be directly applicable to food safety issues.

This Guidance was developed to assess the possible modification of available systematic review methods for the systematic evaluation of food and feed safety research, and to evaluate the potential use of SR methodology when doing risk assessments (RA) to support decision making in food and feed safety.

To develop this Guidance, the Assessment Methodology Unit of EFSA recruited a working group of EFSA scientific officers and external members with expertise in food and feed safety, systematic reviews (in health care, ecology, veterinary medicine, zoonotic public health, and environmental management), and in information science. This Guidance has been written for those with expertise in various areas of food and feed safety and risk assessment in support of decision making who may not be familiar with the methodology of systematic reviews.

Systematic review methodology can be implemented to answer well-formulated specific questions generated by the risk assessment process or by other analytical frameworks developed in food and feed safety in a transparent, reproducible, and rigorous evidence-based way. However, several aspects must be considered in order to decide whether specific questions obtained by simplifying broad food or feed safety policy problems are suitable for systematic review. A useful means of determining whether a question is answerable by SR is to identify the structure of the question. If the question structure can be specified in such a way that a particular primary research study design can be envisaged that would answer the question, then it is likely that a systematic review would be appropriate.

If a question is suitable for systematic review, it does not necessarily follow that a systematic review would be worthwhile or practically feasible. Considerations include: prioritisation of risk assessment model parameters for which refinement of the parameter estimates is considered most critical; the quantity and quality of available evidence; the source and potential confidentiality of the evidence; the need for transparency and/or for integrating conflicting results; and the availability of resources for carrying out the review.

This Guidance describes a general method for performing systematic reviews, taking into account issues that may be unique to the field of food and feed safety and risk assessment that should be factored into the review process.

This Guidance represents a first step towards the application of systematic review methodology in food and feed safety, and regular updates are foreseen in light of experience and developments both in food and feed safety and systematic review methodology.
TABLE OF CONTENTS

Abstract .................................................................................................................................................. 1
Summary .................................................................................................................................................. 2
Table of contents .................................................................................................................................... 3
Background as provided by EFSA ........................................................................................................ 5
Terms of reference as provided by EFSA ............................................................................................... 5
Approach to the Mandate ...................................................................................................................... 7
Objectives of the Guidance .................................................................................................................. 8
Intended users of the Guidance ............................................................................................................ 10
1. Systematic review principles ............................................................................................................ 11
2. The relevance of systematic review methods to food and feed safety assessments ....................... 13
   2.1. Identifying appropriate food and feed safety questions for systematic review ......................... 14
       2.1.1. Question types and the role of key elements of questions .................................................. 14
       2.1.2. Determining key elements of questions in food and feed safety ...................................... 15
       2.1.3. Open-framed and closed-framed questions ......................................................................... 19
   2.2. Assessing the need for and practical feasibility of a systematic review ....................................... 22
       2.2.1. Selecting questions for which a systematic review is needed .......................................... 22
       2.2.2. Assessing the volume and scope of the research evidence: scoping the literature ............ 26
3. General method for a systematic review .......................................................................................... 27
   3.1. Preparing the review ................................................................................................................... 28
       3.1.1. Developing the review protocol .......................................................................................... 28
       3.1.2. Logistics for doing the review ............................................................................................ 32
   3.2. Searching for research studies .................................................................................................... 34
       3.2.1. Searching a range of different information sources ......................................................... 34
       3.2.2. Developing and optimising the search strategy .............................................................. 35
       3.2.3. Managing references ....................................................................................................... 35
       3.2.4. Documenting and reporting the search process .............................................................. 35
   3.3. Selecting the studies .................................................................................................................. 37
       3.3.1. Piloting the study selection process .................................................................................. 38
       3.3.2. Reporting the results of the study selection process ....................................................... 38
       3.3.3. Documenting the study selection process ......................................................................... 38
   3.4. Collecting data from the included studies and creating evidence tables ....................................... 41
       3.4.1. Collecting information from studies about the characteristics that affect external applicability and internal validity ......................................................................................... 41
       3.4.2. Collecting information from studies about the results ...................................................... 41
       3.4.3. Details of any software or tool for recording the data ....................................................... 42
       3.4.4. Procedure for data collection ............................................................................................ 42
   3.5. Assessing methodological quality of the included studies ......................................................... 44
       3.5.1. Procedures for assessing methodological quality ............................................................. 45
       3.5.2. Quality of reporting .......................................................................................................... 46
   3.6. Synthesising data from included studies - Meta-analysis ......................................................... 48
   3.7. Presenting data and results ........................................................................................................ 49
       3.7.1. Key principles of data presentation in systematic reviews .............................................. 49
       3.7.2. Which results to present? .................................................................................................. 50
   3.8. Interpreting results and drawing conclusions ............................................................................. 52
4. EFSA workshop on the application of systematic review methodology to food and feed safety assessments to support decision making ................................................................. 55
   4.1. Workshop outline ....................................................................................................................... 55
   4.2. Outcomes of the open discussion on the advantages and disadvantages of SR ........................ 55
   4.3. Workshop conclusions and further developments .................................................................... 55
References ................................................................................................................................................ 57
Appendices ............................................................................................................................................. 60
Appendix A - Breaking down broad food and feed safety policy problems and identifying suitable questions for using systematic reviews - Examples ................................................................. 60
  Appendix A1 - Import risk assessment ........................................................................... 60
  Appendix A2 - Risk assessment of chemical contaminants in the food chain ............... 65
  Appendix A3 - Analytical frameworks developed for assessing nutrition-related topics ...... 69
Appendix B - Searching for research studies - Full description and Examples .................. 72
  Appendix B1 - Developing the search strategy ................................................................. 72
    Appendix B1(1) - Search terms ..................................................................................... 72
    Appendix B1(2) - Search tools and facilities ................................................................. 72
    Appendix B1(3) - Choosing the key elements to include in the search strategy ............ 74
    Appendix B1(4) - Lumping and splitting ..................................................................... 75
  Appendix B2 - Testing and translating the search strategy ............................................. 75
  Appendix B3 - Searching a range of different information sources ............................... 76
  Appendix B4 - Documenting and reporting the search process .................................... 78
    Appendix B4(1) - Documenting the search process ..................................................... 78
    Appendix B4(2) - Reporting the search process ......................................................... 78
Appendix C - Study selection process - Example ............................................................ 80
Appendix D - Data collection form - Example ................................................................. 81
Appendix E - Meta-analysis: Further information ........................................................... 82
  Appendix E1 - Introduction ............................................................................................ 82
  Appendix E2 - Further issues ......................................................................................... 83
Glossary ............................................................................................................................ 85
Abbreviations .................................................................................................................... 89
BACKGROUND AS PROVIDED BY EFSA

According to its founding Regulation (EC) No 178/2002, the European Food Safety Authority (EFSA) is strongly committed to produce scientific opinions and documents based on the best available evidence, and to ensure that those documents are systematically developed using high scientific quality, as well as transparent and efficient methodology. For these purposes, the use of systematic reviews (SR) as a standardised method to identify, select and critically appraise relevant research would allow EFSA to appraise and synthesize the current body of knowledge on targeted food safety issues. Additionally, the use of a SR methodology would lend increased credibility and transparency to findings in the field, so that decision-makers could access timely information on the most relevant scientific literature.

Systematic review and food and feed safety

Despite the common use of systematic reviews in human health related fields, formal systematic reviews have rarely been used in food safety and there are very few research groups dealing with this topic. Additionally, systematic review protocols developed for use in human health studies may not be directly applicable to evaluate food safety issues (Sargeant et al., 2005). SR methodology has been applied by EFSA in some scientific opinions, such as the assessments of diagnostic techniques for tuberculosis in deer (EFSA, 2008) and for brucellosis in bovines, sheep, and goats (EFSA, 2006) and the assessment of the default Q10 value used to describe the temperature effect on transformation rates of pesticides in soil (EFSA, 2007). Those opinions demonstrate the possibility of efficacious application of SR methodology to some issues within the food safety area. However, the issues covered in those opinions do not represent all the possible topics that can be submitted to EFSA for evaluation (intervention assessments, exposure assessments, disease incidence/prevalence estimates, programme evaluations, etc).

Systematic review and risk assessment

In order to achieve the general objective of a high level of protection of human health and life, Regulation (EC) No 178/2002 recommends that food law is based on risk analysis and that risk assessments are undertaken by EFSA in an independent, objective and transparent manner, on the basis of all the available scientific information and data. For this purpose, the use of systematic review methodology to identify, select and critically appraise relevant information, would ensure that all the steps of the risk assessment (RA) process (hazard identification, hazard characterisation, exposure assessment, risk estimate) are based on relevant and robust data, and the findings of systematic reviews could provide information as input into risk assessment models. Furthermore, systematic reviews could strengthen the potential of RA to highlight areas where there is insufficient scientific evidence or where there are common methodological imperfections in the available research and thereby provide direction and impetus for future basic and applied research in a specific food safety area (Sargeant et al., 2005) or give information for uncertainty analysis.

TERMS OF REFERENCE AS PROVIDED BY EFSA

Objectives

In view of the above, this project aims to:

- Assess the possible modification of the existing protocols of systematic reviews for human health studies, for the systematic evaluation of food and feed safety research. For this purpose, the various types of questions within the food and feed safety field must be considered (such as intervention assessments, disease incidence/prevalence estimates, diagnostic tests

---

5 See footnote 4.
comparisons, programme evaluations, exposure assessments, applications assessments, etc.) as well as the methodologies in place for searching and selecting the relevant scientific information, data collection, data quality evaluation and for analysing and synthesising the data.

- Assess the potential use of systematic review methodology within the risk assessment process. For this purpose a comprehensive evaluation is foreseen of the potential of systematic review methodology to provide information as input into risk assessments, ensuring that all the steps of the RA process (hazard identification, hazard characterisation, exposure assessment, risk estimate) are based on relevant and robust data. A thorough analysis of the strengths and limitations of utilising systematic reviews when doing risk assessments for risk management purposes is planned.

**Deliverables**

The working group will produce a Guidance document (by the third quarter of 2009) containing:

- a comprehensive and standardised model for producing systematic reviews in food and feed safety (taking into account search strategies development models, quality assessment tools, data synthesis methodologies, etc.). The strengths and limitations of the methodology within food and feed safety areas of analysis will also be described;

- an outline of the advantages and disadvantages of applying systematic review methodology to the risk assessment process when providing scientific advice to risk managers.

Additionally, the outputs of the above Guidance document will be presented during an EFSA-tailored workshop on systematic review in food and feed safety and its possible integration into the risk assessment process. The workshop, which will be organised by the Assessment Methodology Unit in the last quarter of 2009 and will be presented by the working group members, has the following primary objectives:

- to present to EFSA scientific staff and experts the SR methodology, its possible application to food and feed safety research and integration into RA;

- to discuss the challenges and the opportunities of utilising systematic reviews when doing risk assessment;

- to evaluate the possible integration of SR methodology within EFSA scientific outputs workflow.
**Approach to the Mandate**

For the development of this Guidance, the Assessment Methodology Unit of EFSA recruited a working group which comprised EFSA scientific officers and external members with expertise in food and feed safety, systematic reviews (in health care, ecology, veterinary medicine, zoonotic public health, and environmental management), and in information science. The external experts were provided with a list of the food and feed safety topics within EFSA’s remit identified via an internal survey. The Guidance was developed at three round-table meetings and several teleconferences during March to December 2009 and was presented by the working group experts at a workshop, which took place in EFSA in February 2010 and was attended by fifty participants among EFSA Panel members and scientific staff. The workshop outline, and the outcomes of the open discussion held during the workshop are summarised at the end of this Guidance.

The present Guidance outlines the basic principles of systematic review methodology and their relevance to and potential application in the areas of food and feed safety. It draws upon information from a number of specialist guides or manuals (Table 1). These sources provide more detailed information than is feasible to include in the present Guidance, but should be interpreted with care as they cover disciplines other than food and feed safety, or some specific areas of food and feed safety only.

**Table 1:** Available Guidelines on systematic reviews in various professional fields

<table>
<thead>
<tr>
<th>Guideline</th>
<th>Professional field</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence-Based Review System for the Scientific</td>
<td>Health claims in</td>
<td>FDA (Food and Drug Administration), 2009</td>
</tr>
<tr>
<td>Evaluation of Health Claims</td>
<td>food safety</td>
<td></td>
</tr>
<tr>
<td>Systematic Reviews, CRD’s Guidance for undertaking</td>
<td>Health care</td>
<td>CRD (Centre for Reviews and Dissemination), 2009</td>
</tr>
<tr>
<td>reviews in health care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cochrane Handbook for Systematic Reviews of</td>
<td>Health care</td>
<td>Higgins and Green (editors), 2009</td>
</tr>
<tr>
<td>Interventions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guidelines for Systematic Review in Conservation</td>
<td>Conservation and</td>
<td>CEBC (Centre for Evidence-Based Conservation), 2008</td>
</tr>
<tr>
<td>and Environmental Management</td>
<td>environmental</td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>management</td>
<td></td>
</tr>
<tr>
<td>A Guide to Conducting Systematic Reviews in</td>
<td>Agri-food public</td>
<td>Sargeant et al., 2005</td>
</tr>
<tr>
<td>Agri-Food Public Health</td>
<td>health</td>
<td></td>
</tr>
</tbody>
</table>
**Objectives of the Guidance**

This Guidance aims to provide a framework of basic instructions on how to apply systematic review methodology to food and feed safety assessments to support decision making.

Particularly, the Guidance seeks to give clear instructions on:

1. how to identify, within a broad food and feed safety policy problem, specific questions that are suitable for systematic review;

2. how to decide whether a systematic review is needed and whether one is practically feasible, highlighting the potential advantages and limitations of using systematic reviews in food and feed safety;

3. how to conduct a systematic review, taking into account issues that may be unique to the field of food and feed safety that should be factored into the review process.

Figure 1 gives an overview of the structure of the Guidance and how to navigate it.
Figure 1: Overview of the structure of the Guidance
Intended users of the Guidance

The Guidance has been written for those with expertise in various areas of food and feed safety and risk assessment in support of decision making, who are new to, or have little knowledge of systematic reviews. Experienced systematic reviewers might also find this Guidance useful for understanding how systematic reviews could be applied to food and feed safety when giving support to decision making.
1. **Systematic review principles**

A systematic review is an overview of existing evidence pertinent to a clearly formulated question, which uses pre-specified and standardised methods to identify and critically appraise relevant research, and to collect, report and analyse data from the studies that are included in the review\(^6\). Statistical methods to synthesise the results of the included studies (meta-analysis\(^7\)) may or may not be used in the process.

Systematic reviews are more effective when limited to addressing specific questions. A specific question is a question which is sufficiently well structured that it could be answered in a primary study without needing to be further broken down.

A systematic review differs from a narrative review in several ways (Table 2).

An explicit and documented protocol for conducting a systematic review is always developed *a priori*, defining in advance the review question and scope, the methods of the systematic review, and the eligibility criteria for the inclusion of studies into the review. This helps to reduce bias in the review, as the process is clearly specified in advance and the reviewers are committed to follow it. Additionally, the protocol might be peer reviewed before implementation of the review. The process of systematic review reduces bias in the selection of research studies by the extensiveness and reproducibility of the search strategy (conclusions are not overly influenced by the most accessible research) and the transparent reporting of how studies are selected and included in the review. The search strategy is reported in order to allow the readers to judge how much of the relevant literature is likely to have been found.

Systematic reviews assess the quality of the evidence in terms of study methodological soundness and give an indication of the strength of evidence provided by the review. In systematic reviews, emphasis may be given to the results from studies of higher quality (designs that are less prone to error and bias). This additional analytical step does not typically occur in narrative reviews (Sargeant et al., 2005).

In systematic reviews the results are explicitly synthesised to clarify the links between the original research and the reviewers' conclusions. Study results are fully reported, irrespective of the statistical significance of their results.

The methodology of the review process is adequately documented to allow others to critically appraise the judgments made in study selection and the collection, analysis, and interpretation of the results and, if necessary, to repeat or update the systematic review.

The fundamental principles of systematic review can be summarised as follows:

- Methodological rigour and coherence in the retrieval and selection of studies, assessment of their methodological quality, and the synthesis and interpretation of information.

- Transparency.

- Reproducibility.

---

\(^6\) SRs typically do not include primary collection of new data.

\(^7\) The process of synthesising research results from a number of independent studies (published or unpublished) by using statistical methods to combine results from previous separate but related studies, in order to determine overall trends and significance.
Table 2: Characteristics of systematic and narrative reviews

<table>
<thead>
<tr>
<th></th>
<th>Systematic Reviews</th>
<th>Narrative Reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study question</strong></td>
<td>Focused and explicit</td>
<td>Often broad in scope</td>
</tr>
<tr>
<td><strong>Eligibility criteria for inclusion or exclusion of studies</strong></td>
<td>Pre-defined and documented; objectively applied</td>
<td>Not always explicitly stated</td>
</tr>
<tr>
<td><strong>Description of the review method</strong></td>
<td>Reported and also predefined in a protocol</td>
<td>Seldom reported</td>
</tr>
<tr>
<td><strong>Literature search</strong></td>
<td>Structured to identify as many relevant studies as possible</td>
<td>Not always extensive</td>
</tr>
<tr>
<td><strong>Methodological quality assessment of included studies</strong></td>
<td>Included, typically using a quality assessment tool</td>
<td>Variable</td>
</tr>
<tr>
<td><strong>Reporting of study results</strong></td>
<td>Full reporting of relevant results (numerical results)</td>
<td>Selective reporting; often of study author interpretation</td>
</tr>
<tr>
<td><strong>Synthesis</strong></td>
<td>Quantitative synthesis (meta-analysis) when possible</td>
<td>Usually narrative, sometimes selective</td>
</tr>
</tbody>
</table>

**SUMMARY POINTS**

- Systematic review methods use a standardised approach to identify and critically appraise relevant research, and to collect, report and analyse data from the studies that are included in the review.
- Systematic reviews are more effective when limited to addressing specific questions.
- In systematic reviews the method is always pre-defined in a protocol, before starting the review.
- Due to their methodological rigour, transparency and reproducibility systematic reviews are different from narrative reviews.
2. The relevance of systematic review methods to food and feed safety assessments

The risk assessment (RA) process provides a method for the decision maker to develop policy under uncertainty that is logical, science based, transparent, reproducible and provided within an agreed timeframe. Pragmatically, RA is a process (a model) that breaks down a broad policy problem\(^8\) into various questions. Answers to these questions feed back into the model to provide the answer to the policy problem.

Systematic review methodology can be implemented to answer specific questions (i.e. sufficiently well structured that it could be answered in a primary study without needing to be further broken down) generated by the risk assessment process or by other analytical frameworks developed in food and feed safety in a transparent, reproducible, evidence-based way.

Several factors must be considered in order to decide whether the questions obtained by breaking down broad food or feed safety policy problems questions are actually suitable for systematic review. These factors are discussed in section 2.1.

Appendix A illustrates the approach for breaking down broad food and feed safety policy problems into various questions and identifying, amongst them, well-formulated specific questions that are amenable to systematic review. Three examples are provided in Appendix A: a specific case based on the Office International des Épizooties (OIE) (World Organisation for Animal Health) methodology for Import Risk Analysis; a generic example based on the Codex Alimentarius methodology for chemical contaminants in the food chain Risk Analysis; and a specific example of the analytical framework that can be developed for assessing a nutrition-related topic.

Although a particular question may be identified as being answerable using a systematic review, it does not necessarily follow that a SR would be worthwhile or practically feasible. Particularly, where several specific questions are identified within a broad food and feed safety policy problem it may not be feasible to perform systematic reviews for all of them. Considerations on the need and practical feasibility of a systematic review are illustrated in section 2.2.

---

\(^8\) Broad policy problem: a broad question (e.g. a risk assessment model or the analytical framework developed for assessing a broad issue) that could be refined into more specific questions.
2.1. Identifying appropriate food and feed safety questions for systematic review

2.1.1. Question types and the role of key elements of questions

Risk assessment models or other broad food and feed safety policy problems may contain one or more questions. However, only certain questions are suitable for systematic review. For example, a specific question about the efficacy of a vaccine in preventing a disease might be addressed by performing a systematic review of experimental studies (e.g. randomised trials) of the vaccine. Conversely, a question about which countries have introduced the vaccine is unlikely to be efficiently addressed by performing a systematic review since it may be difficult to identify which types of primary evidence would be relevant. Table 3 lists some of the commonly encountered types of specific questions that are suitable for systematic review. Further elaboration is provided for three examples in Appendix A.

A useful device to determine whether a question is suitable for systematic review is to identify its **key elements**. For questions on the effects of an intervention, the key elements are the population of interest (P), the intervention of interest (I), a comparator (a control or reference intervention - C) and the outcomes that are of interest (O). The acronym PICO represents these particular components. Other types of questions can be broken into different sets of key elements, as described below.

If the key elements of a question can be specified in such a way that an individual study design could be envisaged that would answer the question in a primary research setting, then it is likely that a systematic review could be performed to answer the same question. This type of question is referred to as a closed-framed question (section 2.1.3). Sometimes some of the key elements of a question can be specified but others cannot. This type of question is referred to as an open-framed question (section 2.1.3).

Key elements have other important roles in the systematic review process:

1. They help to define the eligibility criteria for including studies in the review (e.g. which study designs are appropriate to answer the review question) (section 3.1.1.1).
2. They may assist development of the search strategy, particularly the choice of the basic concepts to be included (section 3.2).
3. The presentation of the details of the studies included in the review will need to include all key elements of the question being addressed by the study. Thus the key elements provide a useful starting point for planning data collection from each study (section 3.4) and for structuring a table of study characteristics (section 3.7).

In some cases the key elements of the questions may be determined by the risk assessment process. For example, in a RA of chemical contaminants in the food chain, to assess the toxic effects of a substance on humans, the population of interest is clearly the human population; however, if no relevant human data are available the risk assessment approach could be to look at relevant animal experiments. This would refocus the population of interest to experimental animals and systematic review methodology could be used to gather evidence-based data from experimental animal studies (Appendix A2).

---

9 A study design can be defined as a specific plan or protocol for conducting a study, which allows the investigator to translate the conceptual hypothesis into an operational one (e.g. a randomised controlled trial, a cohort study, a case-control study, etc).
2.1.2. Determining key elements of questions in food and feed safety

Table 3 lists 10 types of question that are typically answerable using a systematic review. The questions in Table 3 can conveniently be categorised into one of three basic question structures:

- Effects of an intervention or exposure (including dose-dependent fate in organisms, dose-response relationships and environmental fate), in which the population (P), intervention (I) or exposure (E), comparator (C) and outcome (O) need to be specified. The acronyms PICO and PECO represent the key elements in these questions.

- Test accuracy, in which the population (P), index test(s) (I) and target condition (T) need to be specified. The acronym PIT represents the key elements in these questions.

- Descriptive questions of populations or systems, such as questions about prevalence, occurrence, consumption, and incidence in which the population (P) and the outcome of interest (O) need to be specified. The acronym PO represents the key elements in these questions.

The following sections elaborate on these three basic question structures, highlighting the considerations that facilitate decisions on whether a systematic review is feasible and, if so, how the review question can be refined. Examples are provided in Appendix A.

Table 3: Some types of question that are usually suitable for systematic review

<table>
<thead>
<tr>
<th>Type of question</th>
<th>Examples of what the question seeks to assess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of a deliberate intervention</td>
<td>• Nutritional properties of an additive in a food or feed&lt;br&gt;• Efficacy of a vaccine in preventing a disease</td>
</tr>
<tr>
<td>Effect of exposure to a potential risk factor</td>
<td>• Mutagenic effect of a chemical on cells used in mutagenicity tests</td>
</tr>
<tr>
<td>Assessment of a dose-dependent fate of a substance or dose-response relationship (toxicokinetics and toxicodynamics)</td>
<td>• Changes in toxicokinetic parameters as a function of the dose of a chemical in animals or humans&lt;br&gt;• Changes in physiological parameters or biomarkers as a function of the dose of a chemical in animals or humans (toxicodynamics)</td>
</tr>
<tr>
<td>Environmental fate</td>
<td>• Changes in the environmental distribution, degradation, leaching, or run-off of a substance into surrounding areas as a function of its concentration</td>
</tr>
<tr>
<td>Diagnostic test accuracy</td>
<td>• Ability of a test to indicate whether a condition is present or absent</td>
</tr>
<tr>
<td>Analytical accuracy of a test or measurement</td>
<td>• Extent to which a measurement technique correctly determines what the investigator intends to measure</td>
</tr>
<tr>
<td>Prevalence of a disease or condition</td>
<td>• Proportion of animals infected with a virus</td>
</tr>
<tr>
<td>Incidence of a disease or event</td>
<td>• Number of new infections per subject</td>
</tr>
<tr>
<td>Occurrence of a substance</td>
<td>• Level of e.g. a chemical in food, feed or the environment</td>
</tr>
<tr>
<td>Consumption of a substance</td>
<td>• Average intake of e.g. a foodstuff</td>
</tr>
</tbody>
</table>
2.1.2.1. Effects of an intervention or exposure (including questions about dose-dependent fate in organisms, dose-response relationships and environmental fate)

This question type seeks to assess the relationship between a factor and a population exposed to it. The key elements of the questions on effects of intervention or exposure (acronyms PICO or PECO) are:

1. **Population(s) (P).** The population of interest can be represented by:
   - groups of people or entire human communities, e.g. in the assessment of the health effect of a specified food compound;
   - groups of animals, e.g. in the evaluation of the effects of a feed additive or the efficacy of a vaccine;
   - plant species (e.g. groups of plants cultivated or not) or a plant product, e.g. in the assessment of the unintended effects of pesticides on non-target organisms;
   - a food or a feed product, e.g. in the assessment of the effects of a food processing process;
   - a system, level or sector of agriculture at a particular geographical scale. For animals, this could include a specific livestock commodity group or a level of the farm to fork continuum. For plants, this could include a specific plant commodity group (at a regional or field scale) or a level of the agricultural production chain. Within a plant commodity, it could include a specific production type (e.g. timber, propagation material, or nursery plants);
   - a particular taxon or geographic scale, for environmental risk assessments;
   - a problem, such as a vector or a plant pest.

2. **Intervention(s) or exposure(s) (I or E).** This is the factor to which the population is exposed. It could include, for example, an additive in food or feed; a vaccine; a disinfection or eradication method; a chemical or pathogen in food or feed; introduced invasive species; or a harmful impact on the environment.

3. **Comparator(s) (C).** A reference scenario against which the intervention or exposure can be compared. Some examples are:
   - a control or reference group in an experimental study;
   - lack of exposure in a study to the factor of interest;
   - different dose levels in the assessment of the dose-dependent fate of a substance (toxicokinetics and environmental fate) or a dose-response relationship (toxicodynamics).

4. **Outcome(s) (O).** Outcomes are variables for which data are collected to enable the questions of the systematic review to be answered. Usually, outcomes are measurable properties of a population that indicate the consequences of an intervention or exposure (e.g. the performance of animals exposed to a feed additive, the toxicological effect of a chemical, the performance of a target pest species on a host plant, or the presence or absence of a pest or parasite). For dose-dependent fate or toxicokinetics, outcomes may be the toxicokinetic parameters of a substance; for environmental fate it may be the time and/or spatial distribution, degradation, leaching, or run-off of substances or their metabolites.
Questions on the **effects of an intervention or exposure** seek to assess either intended or unintended effects. Examples of intended effects are the reduced shedding of *E.coli* O157 (= the outcome, O) in the faeces of weaned domestic ruminants (= the population, P) treated with pre-harvest interventions (= I) (Sargeant et al., 2007); or effects on the abundance and diversity (= the outcome, O) of reducing pesticide inputs (= the intervention, I) on populations of non-target invertebrates (= P) compared with populations in full-pesticide-input areas (= the comparator, C) (Frampton and Dorne, 2007).

Examples of unintended effects are the potential toxic, genotoxic and carcinotoxic effect (= the outcome, O) of a contaminant (= the exposure, E) in the food chain; the prevalence of *Salmonella* (= the outcome, O) in market-weight swine (= the population, P) exposed to different feeding management practices and feed characteristics (= the intervention, I), which may represent a risk factor (O’Connor et al., 2005); changes (= the outcome, O) in a food item (= the population, P) due to a food processing technique (= the intervention, I); or a vaccine (= the intervention, I) which may represent a risk factor for a particular disease (= the outcome, O) (e.g. such as early *Mannheimia haemolytica* vaccines for bovine respiratory disease - Rice et al., 2007).

Questions which describe a problem and enquire about its cause (**aetiology questions**) assess the relationship between two factors and follow the basic PECO structure. In such questions the outcome (= O) is already known (e.g. an animal disease or poor animal welfare) in a determined population (= P) and the exposure (= E) causing the outcome is what the questions seek to assess (e.g. possible determinants of an animal disease or poor welfare).

Questions about effects of an intervention or exposure include questions on **dose-dependent fate (toxicokinetics)**, **dose-response relationships (toxicodynamics)** and **environmental fate**.

Dose-dependent fate (toxicokinetics) questions assess dose-dependent changes (where dose can be thought of as the comparator, C) in the absorption, distribution, metabolism, and excretion (ADME) (= the outcome, O) of a substance and its metabolites (= the exposure, E) within an organism (= the population, P). Example outcomes include the ADME of a feed or food additive, a pesticide, or a contaminant in humans or animals. Typical toxicokinetics parameters included in the determination of ADME may include the half-life and the clearance of a chemical reflecting the overall elimination of the chemical from a test animal species used in food and feed safety (rat, mouse, dog, or rabbit) or from humans.

Dose-response questions (toxicodynamics) investigate the relationship between a factor and an effect on a population exposed to different doses of it. An example is the relationship between different doses (= the comparator, C) of a given chemical (= the exposure, E) and liver toxicity (= the outcome, O) in a determined population (= P).

Environmental fate questions assess the concentration-dependent absorption in the soil or sediments or the distribution, degradation, leaching, or run-off (= O) of substances or their metabolites (= E) in crops and in the environmental compartments surrounding these crop areas (= P). These environmental compartments include the soil supporting the crop, the soil adjacent to the crop, groundwater, or adjacent surface water bodies. As fate processes are concentration-dependent, the comparator (= C) could be an alternative scenario that influences the exposure concentration (e.g. a set of different assumptions for redistribution, adsorption, degradation, leaching, or run-off). The precise nature of the comparator would depend on the specific question.

### 2.1.2.2. Test Accuracy

Test accuracy refers to analytical accuracy or diagnostic test accuracy (sensitivity, specificity).

A question on analytical accuracy investigates the extent to which a measurement technique correctly determines what the investigator intends to measure (e.g. techniques to determine concentrations of nutrients or contaminants in specific food matrices).
A question on diagnostic test accuracy seeks to assess the ability of a test to indicate whether a condition is present or absent (e.g. a test for *Salmonella* in pigs). In some circumstances the population may be divided in more than two groups.

The key elements of test accuracy questions (acronym PIT) are:

1. **Population(s) (P)**. Since diagnostic and accuracy tests perform differently in different populations, it is important to define clearly the population of interest. The population can be an animal species or, for analytical accuracy, a food matrix.

2. **Index test(s) (I)**. The test whose performance is being evaluated.

3. **Target condition(s) (T)**. The disease or condition (e.g. a chemical in food, or a particular disease) whose presence/absence, or quantity, the index test seeks to detect or measure. If there is a reference standard, this is used to determine whether or not the target condition is present.

2.1.2.3. **Descriptive questions**

Questions on prevalence, incidence and occurrence seek to quantify a condition of interest in a given population. Examples are the prevalence and incidence of an animal or plant pathogen in food, feed, or a geographical area; or the occurrence of a chemical in food, feed, or the environment.

A question on consumption assesses how much a food component or contaminant is consumed in a specified population.

In questions about prevalence, incidence, occurrence, and consumption, the key elements are:

1. **Population (P)**. The population, organism or setting in which the condition of interest is measured. For a question about the prevalence or incidence of a pathogen or pest, the population could be an animal or plant species. For a question about the occurrence of a chemical, the population could be a food or feed product, a spatially-defined environmental compartment or a species of organism. For a consumption question the population could be humans or animals.

2. **Condition of interest (= outcome, O)**. What is assessed or measured in the population. For prevalence or incidence questions the condition of interest is often a disease; for occurrence questions it may be a chemical substance or pathogen; for consumption questions it is often the substance consumed (e.g. a food material or food contaminant).
2.1.3. **Open-framed and closed-framed questions**

A systematic review is an overview of existing primary research studies pertinent to a specific question. To determine whether a systematic review can answer a question, it is necessary to consider the structure of the question. A **closed-framed** question has a well-formulated structure, presenting all relevant key elements. An **open-framed** question is a question that lacks specification of some of the key elements. A useful way to distinguish between closed-framed and open-framed questions is that it is usually possible to envisage a primary research study design (which may or may not be feasible or ethical) to answer a closed-framed question, but rarely possible to foresee a primary research study to directly answer an open-framed question.

Thus, closed-framed specific questions allow the reviewer to set *a priori* clear eligibility criteria for studies, including a statement of appropriate (and thus eligible) study designs (section 3.1.1.1), and to create focused search strategies (section 3.2).

On the contrary, open-framed questions require a more flexible approach. In some cases they may be easily refined and turned into closed-framed questions, by specifying the missing key elements. For instance, a question such as “Is vaccination against *E coli* O157H effective?” asks about the effect of an intervention, without saying anything about the type of outcome to be evaluated, the population of interest or the comparator. This question could be reformulated as “Is vaccination with currently commercially available vaccines against *E coli* O157 (intervention) associated with a decrease in faecal shedding of *E. coli* O157 (outcome) compared to non-vaccination (comparator) in weaned domestic ruminants (population)?”. This is a closed-framed, well-formulated question, for which appropriate (eligible) study designs can be defined, and would therefore be amenable to systematic review.

In other cases, open-framed questions may not always readily translate into closed-framed questions and therefore they would not be suitable for systematic review. An example is the question “What diagnostic tools are available to determine virus x?”. As it is not possible to envisage a specific study design to answer this question, the initial literature search will be very broad, and the criteria for selecting the studies (e.g. diagnostic tools’ features) and the search strategy are likely to be redefined once the reviewer becomes more familiar with the related literature, leading, potentially, to further searches. However, once the available diagnostic tools have been identified, their accuracy may be assessed by one or more systematic reviews.

A food and feed safety policy problem may be broad, comprising various closed-framed and open-framed questions. The reviewer should explore the questions carefully to choose the right approach for answering them. Appendix A provides some examples of the approach for breaking down broad food and feed safety policy problems into specific questions and identifying, amongst them, well-formulated questions that are amenable to systematic review.

Examples of question types potentially occurring in food and feed safety are given in Table 4. For some of them, the related key elements and some open-framed variants are also illustrated.
Table 4: Examples of some open-framed and closed-framed question types potentially occurring in food and feed safety, illustrating their key elements

<table>
<thead>
<tr>
<th>Key elements</th>
<th>Effect of intervention or exposure</th>
<th>Test accuracy</th>
<th>Descriptive questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(P)opulation(s)</td>
<td>(P)opulation(s)</td>
<td>(P)opulation(s)</td>
</tr>
<tr>
<td>(I)ntervention(s) or (E)xposure(s)</td>
<td>(I)ndex test(s)</td>
<td>(O)utcome = Condition(s) of interest</td>
<td></td>
</tr>
<tr>
<td>(C)omparator(s) (not always specified)</td>
<td>(T)arget condition(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(O)utcome(s)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Examples of closed-framed questions amenable to SR**

1. In population P, does intervention I affect outcome O when assessed using comparator C?
   - Do different concentrations C of chemical E have adverse effect O in ecosystem P?
   - Is E a risk factor for O in population P?
   - Is E the cause of O in population P?
   - Is there a dose-response relationship between intervention I (at different does C) and outcome O in population P?
   - What is the sensitivity (specificity) of test I for target condition T in population P?
   - What is the prevalence or incidence of disease O in population P?
   - How much of O occurs in population (foodstuff, environment, etc) P?
   - How much of O does population P consume?

**Examples of open-framed questions where some key elements are missing and must be identified in order to perform a SR**

- In which populations does exposure I result in outcome O? (determine P, C)
- In which populations does test I accurately measure condition T? (determine P)
- In which locations (or species) (or foodstuffs) has O been observed? (determine P)
- What are the vector species for pest O? (determine P)
- What are the best intervention to reduce effect O in population P? (determine I, C)
- What are interventions I are available for problem P? (determine O)
- What causes outcome O? (determine P, E)
- What are the effects of intervention I? (determine P, O)
- What happens after exposure E? (determine P, O)
- Does disease O occur in species P? (yes/no question not amenable to SR)
SUMMARY POINTS

- Food and feed safety assessments in support of decision making often involve broad policy problems, made up of several questions.

- Identifying the key elements (main components) of each question helps in deciding whether the question can be answered by conducting a systematic review; these key elements include, among others, the population, intervention or exposure, comparator, and outcome.

- A specific question that includes all of the possible key elements that would be needed to design a primary research study to answer it is likely to be answerable by conducting a systematic review; this is referred to as closed-framed question.

- A question with missing key elements, referred to as an open-framed question, may not be directly answerable by conducting a systematic review; in some (but not all) cases, open-framed questions may be translated into closed-framed questions by specifying the missing key elements.

- Questions that are answerable by systematic review are likely to fall into one of three basic types: effect of an intervention or exposure, test accuracy, or descriptive questions.
2.2. Assessing the need for and practical feasibility of a systematic review

Although a question may be suitable for systematic review, it does not necessarily follow that a systematic review will be worthwhile or practically feasible. Particularly when dealing with broad policy problems in food and feed safety, it may be impractical to perform systematic reviews for all specific questions that are identified. Scoping the literature may be used to assess the state of the evidence for a review question and thus support decisions about the need and practical feasibility of a systematic review (section 2.2.2).

2.2.1. Selecting questions for which a systematic review is needed

Several considerations may help to decide which of the identified questions should be addressed using systematic review. These include: assessing the likely impact of the evidence (e.g. prioritising the structural parameters considered most critical for a model); assessing the quantity and quality of available evidence; considering the source and potential confidentiality of the evidence; considering the need for transparency and/or for integrating conflicting results; and evaluating the resources needed for carrying out the review. These considerations are illustrated in the following sections.

2.2.1.1. The impact of the evidence: example for the uncertainty and impact of parameters in risk modelling

Risk assessments are based on conceptual models agreed by international guidelines (EFSA, 2009a). According to these conceptual models, components such as the occurrence, release, exposure and consequences are to be assessed. Multiple, specific questions may arise in each of these components. Mathematical models may be used to reflect the structure (e.g., risk pathways and interactions) and information on model input quantities (parameters) for the conceptual models. Usually, a distinction is made between deterministic and probabilistic (stochastic) models, based on whether known variability and uncertainty is ignored or included, respectively. In principle, the criteria for considering systematic review in the context of risk modelling are no different from other application areas. In particular, it may be an appropriate conclusion that no SR is required for any model input parameter.

However, a high degree of transparency and the full use of all available scientific information is a prerequisite in risk modelling. Therefore, it is recommended to consider SR in the phase when the conceptual model is built and the required input information is identified. The following aspects may serve as check-list for each model input quantity. Similar criteria have been used to assess the uncertainty of model assumptions by Van der Sluijs et al. (2005).

Table 5: Criteria to identify candidate questions for SR in the context of risk modelling (adapted from Van der Sluijs et al., 2005)

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Explanation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a: Anticipated local effect in the model</td>
<td>Magnitude and direction of direct impact of the parameter in the model</td>
<td>The toxic effects or the fate of a chemical or test substance in the human body may show considerable interindividual variability. Factors such as genetic polymorphisms or other factors that may affect the toxicokinetics and/or toxicodynamics for a given chemical substance could be investigated using SR</td>
</tr>
<tr>
<td>1b: Anticipated structural effect in the model</td>
<td>Position of the parameter in the model may be central (many pathways are involved) or marginal (few pathways are involved)</td>
<td>Various pathways apply for assessing the risk associated with import of commodity $x$. Testing of $x$ for hazard $h$ at import is common to all pathways. The diagnostic sensitivity of the test is structurally the most important model parameter, for which a SR may be appropriate</td>
</tr>
<tr>
<td>2. Intersubjectivity</td>
<td>Variability among</td>
<td>High degree of variability in the interpretation of</td>
</tr>
</tbody>
</table>
Good practice in risk modelling will require the assessment and reporting of uncertainty. The assessment of both the uncertainty of each parameter and of its impact in the model (e.g. using sensitivity analyses) can produce a priority list of model parameters for which refinement of the parameter estimates is considered most critical. Systematic review could then be considered for refining the most important parameter(s) identified, if no other immediate solutions (such as generation of new data) are available.

2.2.1.2. The quantity and quality of evidence

Systematic reviews can be helpful either when there is a large amount of evidence available or when the evidence is scarce.

When the evidence is extensive, systematic reviews can be particularly useful in formally and systematically summarising the evidence and providing more precise estimates of effects or parameters with enhanced statistical power (meta-analysis) compared with any individual study. If the evidence is scarce, systematic reviews can be particularly helpful to formally identify knowledge gaps. If scoping (section 2.2.2) suggests that evidence is scarce, a systematic review may be able to identify evidence not previously known to exist.

The amount of evidence has implications in terms of the time and resources required (e.g. the budget and available expertise - section 3.1.2); if resources are limited and the amount of evidence is large, methodological restrictions may be required to enable completion of a SR (e.g. limitations may be imposed in the search strategy - sections 2.2.1.5 and 3.2). If it is anticipated that few studies will be found, the reviewer will have to consider how important it is to formally document the scarcity of available information, bearing in mind that unforeseen relevant studies might be identified.

A systematic review may have value regardless of whether the evidence the review identifies is of high or low quality. If the evidence located is of high quality the review may be able to produce an estimate of effect that is unbiased and more precise than those available from the individual studies.

If the research located is of poor quality then the review will document the limitations and flaws with the existing evidence, formally identify knowledge gaps, and make informed proposals for designs for future research. Again, a decision as to whether to proceed where evidence is likely to be weak will depend on the value of characterising an apparent knowledge gap.
2.2.1.3. The source and confidentiality of the evidence

Systematic reviews are designed primarily to retrieve evidence that is publicly available. The availability of the evidence therefore influences the feasibility of a systematic review.

When a review question is answerable using publicly available literature\(^{10}\), an extensive and systematic literature search may be performed and the SR process may be planned since the beginning.

If a review is based on primary research studies provided by an external party\(^{11}\), the full systematic review process is less likely to be applicable. However, in such cases the reliability of the evidence provided, assessed and synthesised by the external party must be evaluated in the light of the best review practice.

2.2.1.4. The need for transparency and for integrating conflicting results

For controversial topics, it may be critical that the process by which the evidence was defined, located, assessed and synthesised is described fully, irrespective of the nature of that evidence. Healthy debate is often more appropriately based on discussion of methods and assumptions than on discussion of empirical findings, and systematic review facilitates this by making these aspects transparent.

Systematic reviews can be useful in understanding apparently conflicting results. For example, differences in the results might be explained by limitations of certain studies brought to light by the critical appraisal; the results might be discovered to lie on a continuum (identified by locating other studies through an extensive search); or differences in the results might be explained by certain characteristics of the studies (identified for example using meta-analytic methods such as meta-regression).

2.2.1.5. The resource requirements in terms of deadlines and budget

Major implications of adopting systematic review methodology are the length of time it takes to complete a systematic review and the need to support a review team with relevant expertise. Thus, the resources associated with completing a review may not be trivial. In some cases it may not be feasible to adhere to all the components of systematic review methodology.

In EFSA, for instance, the feasibility of SR methodology frequently depends on the output deadline as well as the resources and amount of evidence available. Often EFSA mandates contain multiple terms of reference addressing broad policy problems that, once focussed, may require numerous systematic reviews, which would be difficult to achieve with short deadlines. When an EFSA Statement is developed as a fast-track response in order to address an urgent matter, the time constraint might hamper the use of an exhaustive SR process. However, the SR core steps may still be followed and the questions may be answered systematically. For example, if the searches reveal large numbers of records whose assessment will require more resources than are available, the search may be refocused according to clearly pre-defined criteria (e.g. publication date, research designs, or language limits).

In some circumstances, performing a systematic review can imply high costs. When a preliminary literature search indicates that a large amount of evidence is available for a review question and will require ordering and processing of a large number of documents, the associated cost must be assessed and taken into account in the available budget.

---

\(^{10}\) Such as it may occur, in the EFSA context, for EFSA Generic Opinions, Guidance documents, Statements, Scientific and Technical Reports, which may be based upon published or unpublished primary research studies.

\(^{11}\) In the case of EFSA Application Opinions or Conclusions on Pesticides Peer Review and Reasoned Opinions, the evidence is submitted to EFSA respectively by an independent applicant or a rapporteur Member State.
If methodological restrictions are applied due to limitations in the resources, they must be clearly documented when writing the methods section of the review, in order to ensure transparency and reproducibility.
2.2.2. Assessing the volume and scope of the research evidence: scoping the literature

Scoping is an approach to searching the literature that can be used to identify and assess the state of the evidence (quantity of research and types of studies) for a review question. Scoping can take place at one or more stages of systematic review process, including:

a. to support decisions about whether it is possible or worthwhile proceeding with a systematic review;

b. to assess the volume of research which may need to be processed and guide the estimate of resources required;

c. to inform the development of the review protocol (section 3.1.1).

There is no universally agreed definition of scoping (Davis et al., 2009). Scoping methods can be undertaken to various degrees of rigour, from non-systematic to systematic and the approach adopted will depend on the resources available. All may assist with estimation of the size, type and quality of the available literature but with varying levels of confidence:

- A brief and ad hoc approach which uses selected highly focused search terms in one or more bibliographic databases. This approach will provide some key relevant papers and may give some idea of available reviews and the size of the literature.

- A systematic search followed by a rapid assessment of the titles and abstracts of a sample of the records retrieved to assess the number of potentially relevant records in the sample. This will involve a sensitive search (using a range of synonyms and related terms; section 3.2.2) in a small number of selected bibliographic databases and other information sources, with a brief assessment of the quantity and focus of the retrieved studies. This type of scoping can be used to obtain a crude estimate of the volume of research. This can be achieved by selecting a sample of records from a result set and judging how many might be relevant. The number deemed relevant can then be extrapolated to the whole result set if this seems reliable. For example, if there are 1000 records retrieved and five of the first 100 records seem relevant, this means that in the 1000 records there may be 50 relevant records. Additional information about the scale of research which might need to be assessed may be provided by looking at the number of studies identified in other published reviews.

SUMMARY POINTS

- Although a question may be suitable for systematic review, it does not necessarily follow that a systematic review will be worthwhile or practically feasible.

- A number of factors influence whether a systematic review is likely to be feasible and worthwhile.

- These factors include (among others) the likely impact of the evidence in risk modelling, the quantity and quality of the evidence, the sources of evidence (whether publicly available or submitted for assessment by an applicant), the need for transparency and for integrating conflicting results, and the availability of resources for conducting a systematic review (time, budget and expertise).

- Scoping is an approach to searching the literature that can be used to support decisions about whether it is possible or worthwhile proceeding with a systematic review; assess the volume of research which may need to be processed; and guide the estimate of resources required.
3. General method for a systematic review

The core steps of a systematic review are illustrated in Figure 2 and described in section 3.1-3.8. Each step must be carefully documented in the SR to ensure transparency and reproducibility.

**Figure 2:** Core steps for performing a systematic review (adapted from the Cochrane Handbook for Systematic Reviews of Interventions, Higgins and Green (editors), 2009)
3.1. Preparing the review

3.1.1. Developing the review protocol

In a systematic review, the methods to be used in all steps of the review process are made explicit *a priori* in a protocol, which includes a background section, specification of the review question, the objective and the inclusion criteria (section 3.1.1.1). The key areas to cover in the review protocol are listed in Table 6.

Specifying the methods in advance reduces the risk of introducing bias into the review. For example, the reviewers are less likely to be influenced by their knowledge of study authors or by study findings.

The protocol is a component of an open, consultative and iterative approach to undertaking reviews that involves the review team and, if necessary, relevant stakeholders.

The protocol can be made available (by submission and peer review) to a reputable repository (e.g. in human health research, The Cochrane Library\(^\text{12}\); in environmental management, the Collaboration for Environmental Evidence Library\(^\text{13}\)), as a contribution of evidence to the wider community, and to encourage constructive criticism from other reviewers and research users at the stage at which it is most likely to help improve the final review.

Some commissioning and funding bodies may require that they formally approve the protocol. For commissioned reviews it can be useful to communicate with the commissioner at the protocol development stage, to ensure that it meets the commissioning requirements, before the review starts.

Scoping the literature (section 2.2.2) may be used to inform the development of the protocol.

**Table 6:** Key areas to cover in the review protocol

<table>
<thead>
<tr>
<th>Areas covered in the protocol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>Conceptual framework relevant to the review question; reasons for doing the review</td>
</tr>
<tr>
<td>Review question, objective and inclusion criteria</td>
<td>Clear definition of the review question and objective; pre-definition of criteria for study inclusion or exclusion (section 3.1.1.1)</td>
</tr>
<tr>
<td>Methods for searching for research studies (section 3.2)</td>
<td>Development of a preliminary search strategy (i.e. combination of search terms) and identification of information sources that will be searched; decisions about language restrictions, publication status and software for managing the references); in reviews of one year or more duration (or in rapidly evolving fields), indication for repeating the searches towards the end of the process</td>
</tr>
<tr>
<td>Methods for selecting the studies (section 3.3)</td>
<td>Explanation of the process by which decisions on study selection will be made (i.e. how many experts will screen titles, abstracts and full texts; expertise of the reviewers; whether the examination of the studies will be done independently by the reviewers; how potential disagreements on study eligibility will be solved; and whether the assessment will be blinded or unmasked)</td>
</tr>
<tr>
<td>Methods for collecting the data from the included studies</td>
<td>Description of the information that will be collected from included studies; details of any tools for recording the data (e.g. data forms and software); the procedure for data collection, including the number of...</td>
</tr>
</tbody>
</table>

\(^{12}\) [http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME?CRETRY=1&SRETRY=0]

\(^{13}\) [http://www.environmentalevidence.org/Library.htm]
Areas covered in the protocol | Description
--- | ---
reviewers involved and how any discrepancies will be resolved; specification if authors of primary studies will be contacted to provide missing or additional data; if foreign language papers are to be included, specification of translation arrangements

• assessing the methodological quality of the included studies (section 3.5) | Description of the method of study appraisal, including examples of the specific quality assessment criteria; details of how the study appraisal is to be used, e.g. if the results will inform sensitivity analyses; the process for conducting the appraisal of study quality (the number of reviewers involved and how any disagreements will be resolved); if resources are limited, decisions about potential simplifications of the quality assessment tools

• synthesising the data from the included studies (section 3.6) | Explanation of the strategy for data synthesis (although it is difficult to anticipate all the statistical issues that may arise, as analyses will depend on what data are available); how heterogeneity will be explored and quantified; under what circumstances a meta-analysis would be considered appropriate and whether a fixed or random-effects model or both would be used; if appropriate, the approach to narrative synthesis; the outcomes of interest and what effect measures will be used; any planned subgroup or sensitivity analyses or investigation of publication bias

Sometimes it may be necessary to apply amendments to the protocol. For example, consideration of the primary research may raise questions which were not anticipated at the protocol stage and which clarify the review question. Where modifications to the review question are applied, the likely impact on the literature search should be assessed. If changes are needed to the protocol as the review progresses, it must be noted in the review's final report and the rationale for making changes must be made clear.

3.1.1.1. Defining the review question and developing the eligibility criteria for including studies

The protocol includes a clear definition of the review question (i.e. formulation of the question) and a clarification of its scope.

Formulation of the question (indicating issues such as the question type and the key elements of the question, i.e. the PICO elements) is described in detail above (section 2.1). This process is crucial since all other aspects of the review follow directly from it. Specifically, formulating the question helps in: defining the scope of the review; setting the inclusion criteria for studies in the review; guiding the search strategy for identifying the relevant studies; critically appraising the studies; and analysing any variation among the results (Higgins and Green (editors), 2009).

An important attribute of systematic reviews is that criteria for study inclusion are clearly pre-specified. This is perhaps the most notable difference between a narrative review and a systematic review (Sagoo et al., 2009).

The validity of a systematic review depends on the validity of the primary studies included. A systematic review of biased studies will itself be biased. The criteria for identifying studies for inclusion in the review needs to include a statement of appropriate (and thus eligible) study designs used to address the review question. Deciding on which study designs to use includes three considerations:

1. Which study designs can address the specified question? The question type (with PICO, PECO, PIT or PO key elements) will determine a set of possible designs that could have been
used in primary studies.

2. Which study designs will produce the best evidence? Study designs vary in the degree to which they minimise bias, and it will be desirable to restrict inclusion to the study designs least likely to be biased and most likely to give valid answers.

3. Which study designs are available? Sometimes no or very few studies of the best design exist, but evidence from weaker designs is more readily available.

There is often a trade-off made when defining the study design eligibility criteria. On one hand there is a desire to produce a review which yields an unbiased estimate of an effect or parameter (achieved by restricting only to the few study designs that protect against bias) whilst on the other there is a desire that the review yields a precise estimate of the effect or parameter (achieved by including more studies, often of weaker designs). The decision of the minimal threshold for inclusion of studies is thus difficult, and must weigh up the likely magnitude of biases present in the included studies against the need to produce a timely answer to the review question.

This is the first point in the systematic review process at which study design and associated study validity are considered. Decisions made at this eligibility stage will determine which research is included in the systematic review (detailed assessments of study validity are also made later in systematic review process when critically appraising the methodological quality of the included research - section 3.5).

In some fields (e.g. areas of health research), consensus exists for the ranking of study designs according to the likelihood of bias (often called hierarchies of evidence). If no relevant evidence hierarchy exists, it will be necessary during development of the protocol for the review to delineate the strengths and limitations of the alternative study designs and rank or group them according to the degree to which they are susceptible to bias. This process needs to involve both methodologists (who understand the limitations and biases associated with the possible study designs) and domain experts (who will know the way in which biases occur in the field of interest). Below we illustrate two existing hierarchies which can be used for PECO, PICO and PIT types of study.

---

Eligible study designs for questions on the effects of an intervention or exposure

Hierarchies of evidence exist for studies evaluating interventions (PICO studies), where outcomes are compared between a group of study subjects which receive an intervention with a group which receive an alternative control intervention. Studies evaluating exposure or ascertaining aetiology (PECO studies) are also part of this hierarchy as they have similarities in comparing outcomes in two groups according to whether or not they received a particular exposure. Study designs in the hierarchy include:

- Randomised studies, where the study subjects are randomly allocated to either intervention or control, or, in a challenge study, to be exposed. Randomisation is used to allocate study subjects to groups as on average it creates groups which are comparable with respect to all factors other than the intervention or exposure they receive. Thus differences in outcomes between the groups are likely to be due to the difference in interventions or exposure rather than anything else, allowing conclusions of cause and effect to be made.

- Non-randomised experimental studies, in which the allocation of study subjects to intervention or exposure is decided by the investigator (thus such studies often are classified as experimental) but is not undertaken using randomisation. It is possible that a non-random allocation process produces groups which are not comparable, and hence the final difference in the measured outcomes between the groups may reflect differences other than those caused by the intervention or exposure.

- Prospective cohort studies, which recruit a single group of study subjects and follow them,
noting which intervention or exposure they receive, and what their outcomes are. Because the allocation to interventions or exposure is not determined randomly and not influenced by the investigator (such studies are often described as being observational), allocation may relate to the characteristics of the study subjects, or factors such as location or time period. Thus a comparison of outcomes between the groups will be confounded by the other differences that exist. Statistical methods (such as regression modelling) may be used to attempt to statistically correct for differences between groups.

- Retrospective case-control studies, where study subjects who have experienced the outcome and those who have not are recruited separately. The proportion of each group who previously received the intervention or exposure is discerned through inspection of records or other historical enquiry, and the proportions compared. Because the two groups are recruited separately it is possible that they may not be comparable in many ways. Also, the ability to retrieve or recall data on the intervention or exposure may be influenced by whether or not study subjects experience the outcome, as recall biases may exist.

- Cross-sectional studies, where the outcomes and interventions or exposures are ascertained in the study subjects at the same point in time. If a relationship is observed between the outcome and the intervention or exposure, it is not possible to assess whether the outcome is caused by the intervention or exposure, the intervention or exposure is a consequence of the outcome, or both are related to a third unknown factor.

- Case series, which usually report only on study subjects who all experienced the intervention or exposure, and show that their subsequent outcomes are favourable or otherwise. Case series usually do not formally compare outcomes with a comparator group, and thus cannot provide estimates of the effect of the intervention or exposure. They often are early stage studies which indicate that a potential effect could exist and should be evaluated further. They are usually excluded from systematic reviews.

*Eligible study designs for evaluating diagnostic test accuracy*

Studies of diagnostic test accuracy need to compare results of one or more index tests with a classification of whether or not study subjects do or do not have the condition of interest. Studies are usually cross-sectional designs, but a distinction can be made between two study designs according to the way in which groups of subjects (human or animal) are recruited and classified:

- Studies which recruit a single series of subjects at the point where the test will be used in practice (and hence meet some stated criteria), all of whom receive the index test(s) and for whom a reference standard diagnosis is made (which may be a single test or a combination of tests). In some study design taxonomies such studies are referred to as cohort studies, which is not really appropriate as they are cross-sectional. Such studies may be referred to as single gate design as all subjects enter the study by the same route, or consecutive series.

- Studies which recruit two separate series of subjects, in one of which subjects are known to have the condition of interest and one in which subjects are known (or presumed) not to have the condition. Often a reference standard is not undertaken, with subjects instead classified on the basis of existing information. Such studies are likely to overestimate both sensitivity and specificity as the subjects in both groups will not be representative of those in whom the test will be used in practice. Subjects known to have the condition are more likely to have advanced or severe disease (which is usually easier to diagnose); those who are in the disease free group are likely to be healthy (which is easy to identify) and not have other diseases which may relate or mimic the condition under investigation (which are harder to identify). Such studies are sometimes referred to as two-gate designs reflecting the separate recruitment of cases and controls, or as diagnostic case-control designs.
3.1.2. Logistics for doing the review

3.1.2.1. Establishing a multidisciplinary team

The initial step when starting a systematic review is to form a multidisciplinary team, which will develop the protocol and conduct the review.

The review team should include expertise in the relevant topic area, information retrieval, statistics, and systematic review methods. The role of the information specialist is fundamental to develop appropriate search strategies, identify appropriate and relevant information sources and guarantee the extensiveness of the information retrieved (section 3.2).

For broad policy problems in food and feed safety, the multidisciplinary team may also develop the risk assessment model or the analytical framework, identify the questions that are amenable to systematic review and decide when a systematic review is needed or practically feasible. Therefore, the input of a modelling expert should also be foreseen (section 2.2.1.1).

Sometimes it may desirable to solicit input from external experts; in such cases, the process of obtaining external inputs should be defined before starting the review.

3.1.2.2. Setting the project timetable and the resources

The review team must determine the appropriate time frame for undertaking the review by considering the tasks involved and the time required for each of them. The tasks may vary widely from one review to another, depending upon the topic of the review, the amount of existing evidence, the methods used (e.g. the extent of efforts to obtain unpublished information), the number of meetings foreseen or the need for involving stakeholders.

Other resources that might be required for undertaking the review, in addition to the reviewers’ time, include bibliographic software programs for managing the references and documenting the study selection process (e.g. EndNote, ProCite, Reference Manager, or RefWorks) (section 3.2.3) and programs specifically designed for carrying out systematic reviews (e.g. DistillerSR, EPPI-Reviewer, RevMan, or TrialStat SRS). The latter may be particularly useful as they allow the various steps of the SR process to be shared among members of a review team, who may be in different geographical locations. Such systematic review programs enable the importing of bibliographic records from different sources, checking for duplicates and storing abstracts and full texts; allocating references to the reviewers for evaluation and recording their comments (e.g. inclusion or exclusion decisions - section 3.3); and keeping track of which papers still need to be retrieved, which are on order and from which libraries they are available.

The size and expertise of the review team, the time required and any software tools needed for undertaking the review will obviously affect the financial cost of the review. Thus the budget must be carefully planned before starting a systematic review.
SUMMARY POINTS

- The preparation of a systematic review involves several planning steps; these include the development of a protocol, establishment of a multidisciplinary review team, and setting of the review timetable and budget.

- Specifying the methods in advance in the protocol reduces the risk of introducing bias into the review and assists critical assessment and reproduction of the review methodology.

- The protocol should explain and define the review question and objective; the criteria for study inclusion or exclusion; and describe the methods for searching research studies, selecting the studies, collecting the data from the included studies, assessing the methodological quality of the included studies, and synthesising the data from the included studies.

- The types of study eligible for inclusion in the review should be determined and clearly specified in the protocol, indicating the quantity and quality of evidence desirable; a hierarchy of evidence illustrating the relative strengths and limitations of different study designs may assist these decisions.

- The systematic review team should include expertise in the review topic, information science, systematic review methods, and (if appropriate) statistical methods and risk assessment.

- Financial resources allocated to the review should include the costs of supporting an appropriately qualified review team as well as all costs associated with the location and retrieval of studies (bibliographic database and journal subscription charges, bibliographic reference management and systematic review software costs, and other library charges).
3.2. Searching for research studies

Conducting a thorough and extensive search to identify as many studies as possible relevant to the review question is a key step in the systematic review process, which helps in minimising bias. Technically this step should be performed by an information specialist (in collaboration with the review team) because the information specialist has expert knowledge of (a) structuring searches to capture research questions, (b) the different characteristics of various bibliographic databases and database interfaces and (c) how to adapt searches to work efficiently for different sources of information.

The SR process as far as identifying research evidence is concerned involves four aspects:

1. Identifying the information sources which are likely to yield relevant studies. These sources are likely to have been identified during a scoping stage, from discussions with the review team and through exploring other systematic reviews.

2. Developing the search strategy (search terms and their combination) to capture the review question and identify relevant studies. The development of the search strategy is likely to begin when planning the review during a scoping exercise and in discussion within the review team. Testing the strategy helps to identify the best possible yield (i.e. the largest number of relevant studies) and quantify the potential number of bibliographic records which may be missed as a result of the chosen strategy.

3. Managing the references and the documents retrieved.

4. Documenting and reporting the searches (with a flow chart and narrative description), in order to make the search process as transparent as possible and to enable it to be evaluated and reproduced.

An extensive search of studies is required for two main reasons: to minimise the effects of publication bias and to compensate for limitations of research reporting and indexing. In the biomedical literature, there is evidence for a wide range of publication biases (Hopewell et al., 2007a; Hopewell et al., 2007b; Hopewell et al., 2009). There is some evidence that publication bias is also present in food safety or feed science (Moles et al., 2003; Nielen et al., 2006; Berteaux et al., 2007; Duffield et al., 2008; Haxton and Findlay, 2008; Ceballos et al., 2009). Publication bias can be seen, for example, when positive results are more likely to be published than negative results. To compensate for publication bias, sensitive searches, which involve trying to identify unpublished studies (e.g. collections of reports and working papers, dissertations, theses, and conference abstracts - chapter 3.2.1 and Appendix B) and studies in languages other than English, are undertaken to inform systematic reviews. Sensitive searches seek to identify as many studies as possible that meet the eligibility criteria by searching a range of resources (extensive searching). For each information source a sensitive search strategy specific to that information source is required because bibliographic databases are selective (not comprehensive) and use different syntax and indexing terms to characterise studies. Developing a sensitive search involves identifying and compiling a range of synonyms and related terms to be searched in the title, abstract and indexing fields of database records. The aim of a sensitive search is to maximise the opportunities to capture relevant research reports.

3.2.1. Searching a range of different information sources

The choice of information sources to be searched is decided by discussion between the review team and the information specialist. The objective in choosing the sources is to identify a wide range of possible forms of research publication and also unpublished works so as to minimise publication bias. Publications in journals and books are relatively easy to identify in electronic bibliographic databases. Research findings in reports, working papers, dissertations and conference proceedings are often more difficult to identify and their identification may be resource intensive and require substantial topic expertise. These types of unpublished research reporting, often referred to as “grey literature”, may be less consistently archived and indexed in bibliographic databases. Ongoing and recently completed
Information on approaches for locating and selecting bibliographic databases and other information resources which index a range of publication types is provided in Appendix B.

### 3.2.2. Developing and optimising the search strategy

Search strategies are *ad hoc* combinations of keywords (search terms) designed to retrieve as many studies as possible which are relevant to the review question. Strategies can be used within bibliographic databases which have sophisticated search interfaces, or text-based indexes, or used within simple search interfaces such as Google. Further guidance with examples of how to develop and optimise a search strategy is presented in Appendix B.

Constructing an effective combination of search terms involves breaking down the review question into “concepts”. The key elements (e.g. the PICO elements of an intervention or exposure question - section 2.1.1) can form the structure of the search strategy. However, not all key elements are always necessarily required in the strategy. The search strategy may also try to capture other concepts (as well as the key elements) such as the study design and other limits that may be required, e.g. dates, languages or geographic locations. Examples of these issues are presented in Appendix B.

### 3.2.3. Managing references

Extensive literature searches can generate large volumes of bibliographic records. It is recommended that bibliographic software (e.g. Endnote, ProCite, Reference Manager or RefWorks) is used to manage (i.e. store and classify) the references downloaded from bibliographic databases. This can assist documentation of the search process and streamline document management. For example, bibliographic reference management software usually enables direct links to word processing software which can make the production, formatting and updating of reference lists for reports and journal papers more efficient. In addition, an electronic library of references allows reference information to be shared by the whole systematic review team and makes duplicate references easier to identify and delete. Data fields within the reference records can be used to record decisions (e.g. about study inclusion or exclusion) or queries about the publication, and to categorise references by agreed topics if desired.

Some bibliographic database providers, such as Web of Science or Ovid, provide facilities to download records in a format which allows easy importing into reference management software. Other databases and providers offer fewer facilities for downloading. In some cases, references obtained from internet sites may need to be copied and pasted into the bibliographic software.

When an electronic library of references is established, it is important to set in advance clear rules about which team members can add or amend records in the library, and to develop consistent terminology to record decisions. It is usually preferable to have one person from the team responsible for the library of references.

### 3.2.4. Documenting and reporting the search process

Guidance in chapter 6 of the Cochrane Handbook (Lefebvre et al., 2008) indicates that the following issues are important when documenting and reporting the search process for systematic reviews. The search process should be documented in enough detail to ensure that it can be reproduced and that search strategies can be rerun, i.e. the search terms and search term combinations for all databases should be clearly reported. The search strategies ideally should be copied and pasted exactly as run and included in full, together with the number of records retrieved. Retyping search strategies should be avoided if possible because it can introduce errors. The number of records retrieved should be recorded in the Results section of the review. It is important to save electronic or printed copies of any information found on the internet, such as information about ongoing studies, as this information may no longer be accessible at the time the review is written. Notes should be kept of key decisions which might impact on the review findings, for example any effects of choosing specific subject headings or...
introducing search limits.

Examples of how to report the search process are presented in Appendix B.

<table>
<thead>
<tr>
<th>SUMMARY POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Conducting an extensive search to identify as many studies as possible relevant to the review question is a key step in systematic review process, as it seeks to minimise the effects of publication bias and compensates for limitations of research reporting and indexing.</td>
</tr>
<tr>
<td>● The search should be performed by the information specialist.</td>
</tr>
<tr>
<td>● Identifying research evidence involves four aspects: identifying the information sources; developing the search strategy; managing the references and the documents retrieved; documenting and reporting the searches.</td>
</tr>
<tr>
<td>● The objective in choosing the sources is to identify a wide range of possible forms of research publication and also unpublished works so as to minimise publication bias.</td>
</tr>
<tr>
<td>● An effective search strategy captures studies relevant to the review question by breaking down the review question into “concepts”. These “concepts” often reflect some of the key elements (e.g. the PICO elements), but may also try to capture other issues such as the study design, date or specific limits.</td>
</tr>
<tr>
<td>● Documenting and reporting the searches is crucial to ensure transparency and reproducibility.</td>
</tr>
</tbody>
</table>
3.3. Selecting the studies

Once searching is completed, relevant studies must be efficiently assessed for inclusion against criteria that have been defined *a priori* in the protocol.

Studies (not papers or reports) represent the unit of interest. Duplicate publications may occur in which the same study is reported in several articles, abstracts or other reports. In some cases more than one study may be reported, or partially reported, in a single publication or report. These situations should be carefully identified to avoid double-counting or missing relevant results. Records that refer to the same study may be marked (e.g. using keywords) in an electronic bibliographic library to clarify how the various publications and reports are linked and assist the identification of any duplicate records.

Some searching methods provide access to full papers directly, such as hand searching journals and contacting research groups, in which case assessment for inclusion is a one stage process. For studies retrieved from electronic databases, normally the selection process is conducted in two stages:

1. **Screening of titles and abstracts for relevance to the study question.** First, screening is made of titles and, where available, abstracts identified from the searches. Normally a screening checklist is developed according to the key elements of the question (e.g. the PICO elements) identified in the protocol. The screening tool could, for example, contain questions about the population or problem, intervention, comparator and outcome. The screening tool enables a decision to be made for each record (title and/or abstract) whether the record is relevant or irrelevant. If no abstract is available (i.e. there is only a title) or the abstract is too vague, it may be unclear whether the record is relevant to the review question. Such situations may be resolved by discussion among two or more independent reviewers, or it may be necessary to retrieve the full text version to allow a decision to be made.

   At this initial screening stage it may or may not be clear whether individual records are duplicated or linked. For records that are clearly linked (i.e. if they report the same study) it may be appropriate to apply the screening process simultaneously to a group of linked records. As it is not always clear from titles and abstracts alone whether records refer to the same study, the grouping together of linked records could instead be left until the next screening step (full text).

2. **Examining full-text reports for the eligibility of studies.** For records that pass relevance screening based on titles and abstracts, or in cases when a definite decision cannot be made based on the title and/or abstract alone, the full paper or report must be obtained for detailed assessment against the inclusion criteria. Obtaining the full text of all articles can be very time consuming and a realistic deadline may have to be imposed and a record kept of those articles not obtained. At this stage it may be appropriate to communicate with the original investigators, to clarify a study’s eligibility. Where several records refer to the same study these should be grouped together and screened together as one study unit.

   At all stages of the screening process all studies should be independently assessed by more than one reviewer, in order to prevent the introduction of errors and personal biases.

   Reviewers’ expertise in the topic area (domain) is essential; nevertheless, to reduce the risk of pre-formed opinions that can bias the assessments, it may be an advantage if one of the reviewers is not a domain expert.

   Where disagreements between reviewers occur, the process for resolving them should be documented and specified in the protocol. A third independent reviewer may help to resolve differences of opinion. In some cases, additional information may be necessary for a decision to be reached on a full-text article and the study may be classified in the review as “awaiting assessment”, until the additional information is obtained from the study authors. If contacting the authors is not practical, the study in
question could be excluded and listed as “potentially relevant”. The effects of including such “potentially relevant” studies on the results of the review may be explored using sensitivity analysis (i.e. comparing the results of the systematic review both with and without the potentially relevant studies included).

Blind assessment at each screening step may be possible by removing identifying information such as authorship, institutions, journal titles and year of publication, but may not be warranted given the time and effort required to disguise the source of each article. An unmasked assessment by two independent reviewers is usually acceptable (CRD, 2009).

If resources and time allow, the lists of included and excluded studies may be posted on a dedicated website with a request, within a set deadline, for feedback on any missing studies.

3.3.1. Piloting the study selection process

Initially, the selection process has to be validated for reliability and reproducibility. This validation normally is made by two reviewers, who apply the selection criteria to the same randomly-selected sub-sample of studies. Independent checking of sub-samples of studies can be done for both stages of the selection process; at title or abstract screening and at full text screening. Piloting may also give an indication of the likely time needed for the full screening process.

To check for consistency in the interpretation of the screening criteria, reviewer relevance decisions can be compared by performing a kappa analysis (a measure of chance-corrected agreement)\(^\text{14}\). However, comparison of a value of kappa with arbitrary cut-points is unlikely to convey the real impact of any disagreements on the review. For example, disagreement about the eligibility of a large, well conducted, study will have more substantial implications for the review than disagreement about a small study with risks of bias (Higgins and Green (editors), 2009).

Overall, if agreement is not achieved, then a revision of the screening criteria or an improvement of their coding will be necessary.

3.3.2. Reporting the results of the study selection process

The number of studies selected for inclusion at each stage of the screening process can be reported in a flow chart, which represents a simple and useful way of documenting the study selection process.

A list of studies excluded from the review based on screening full text should also be reported where possible, giving the reasons for exclusion. In general, such information is not reported for studies that are excluded based on the screening of abstracts and titles.

3.3.3. Documenting the study selection process

Each stage of the study selection process must be well documented, in order to make it assessable and reproducible. Particularly, the following information should be clearly documented in the methods section of both the protocol and the review:

1. expertise of the reviewers (e.g., whether a domain expert, information specialist, or statistician);

2. whether the examination of the studies was done independently by the reviewers;

3. how disagreements were handled;

\(^\text{14}\) Formal measures of agreement are available to describe the extent to which assessments by multiple reviewers (inter-assessor reliability) agree. For a description of how a kappa statistic may be calculated for measuring agreement between two reviewers making simple inclusion/exclusion decisions, see section 7.2.6 in Higgins and Green (editors), 2009).
4. whether the assessment was blinded or unmasked.

The process of study selection is summarised in 1See Appendix B3

Figure 3 (adapted from Sagoo et al., 2009).

\[\text{See Appendix B3}\]

\textbf{Figure 3:} Study selection process flow chart (adapted from Sagoo et al., 2009)
SUMMARY POINTS

- In systematic reviews studies are selected for inclusion according to clearly pre-specified criteria.
- Normally the selection process is conducted in two stages: screening of titles and abstracts for relevance to the study question and examining full-text records for the eligibility of studies.
- Independent assessment by more than one reviewer (at all stages of the selection process) reduces the introduction of errors and personal biases.
- Initially, the selection process has to be validated for reliability and reproducibility.
- The study selection process must be reported using e.g. a flowchart. A list of studies excluded from the review based on screening full text should also be reported where possible, giving the reasons for exclusion.
- Each stage of the study selection process must be well documented, in order to make it assessable and reproducible.
3.4. Collecting data from the included studies and creating evidence tables

For each study included in a systematic review, the guiding principle for data collection, regardless of the topic area, should be to determine study findings and to report study characteristics that influence the external applicability and internal validity of the findings. Methods of presenting these study characteristics in tabular form when writing the systematic review are described in section 3.7 (Presenting data and results).

The systematic collection of data from each primary research study may be extensive but this step has a key role in ensuring the reproducibility of the systematic review. During data collection and reporting, some elements of subjective judgement may apply (e.g. if a reviewer has to make an assumption about a reported parameter). To ensure consistency of interpretation and reporting, criteria for such judgements should be defined a priori by the review team. The review team may decide to modify or update any aspect of the data collection process if this is deemed necessary. However, any such change to the data collection process would have to be applied to all the included studies.

The data collection step in the systematic review forms the basis for research synthesis methods, such as meta-analysis (section 3.6 and Appendix E), and should be done with research synthesis in mind. Data collection requirements vary from review to review and should be tailored to the review question and the planned analyses specified a priori in the review protocol.

3.4.1. Collecting information from studies about the characteristics that affect external applicability and internal validity

For many topics the characteristics of the population studied are relevant to interpreting the outcome and may be a source of heterogeneity in the effect measure. The members of the review team have primary responsibility for identifying these characteristics and ensuring that this information is collected for each study. For example, in livestock studies the production system, age of the animals and their stocking density might affect the external population to which the results could be applicable. If this is considered to be the case, then the production system, age of the animals and their stocking density should be collected from each study and reported in the systematic review. It will also be the responsibility of the review team to specify the form and detail of data to be reported. The collected information may be numerical, fixed text such as yes/no, or free text.

In some cases the method of study execution could be associated with the potential for bias. Study characteristics related to study conduct might therefore be a source of heterogeneity in the outcome. For example, for a systematic review evaluating the occurrence of a contaminant in feed, different analytical methods might influence the estimate of occurrence. If the systematic review team feels that the detection methods may be relevant to interpreting the prevalence estimates, this information about the detection methods should be collected from all studies. Such information could include the level of validation and performance characterisation (analytical sensitivity and specificity, limit of detection, limit of quantification, repeatability, reproducibility, diagnostic sensitivity and specificity) of the measurement methods used.

3.4.2. Collecting information from studies about the results

Information about results will depend on the food or feed safety issue and the main data to be collected for a systematic review will depend upon the outcome(s) specified a priori in the review protocol. For quantitative data on effect or parameter estimates, it is good practice, if possible, to provide a point estimate (e.g. mean or median) together with an estimate of variability (e.g. standard deviation, standard error or confidence interval). Sample sizes of all study groups should be indicated, including any changes through time (e.g. if study subjects drop out from the study at different time points). Reviewers should be alert to the possibility that study sample sizes differ from those that would be ideal for avoiding bias (e.g. if a study is randomised but not all of the randomised study subjects are accounted for). Data that may be expressed as proportions (e.g. prevalence) should be presented as both the numerator and denominator, to allow for differences in the size of the populations upon which the proportions are based.
3.4.3. Details of any software or tool for recording the data

Data collection is usually done using a specifically designed form, which may be in paper and/or electronic format. Ideally, such a form should allow efficient collection of data in a standardised way, which in turn allows fast compilation with a reduced risk of typographical errors. An example of a data collection form used for a systematic review of a diagnostic test is given in Appendix D. A well-structured data collection form could be useful also as a template for tables that report the results of the systematic review (section 3.7).

The format of the data collection form will depend upon the systematic review question and its design may be influenced by the key elements of the question (e.g. population, intervention, comparator and/or outcome). The form should be structured so that it captures the characteristics identified as important for assessing external applicability and internal validity. Instructions for completion should be provided and each data field should have decision rules about coding data in order to avoid ambiguity and to aid consistent completion.

If the review team has decided to contact study authors directly for missing data or to translate studies, the process of collecting information may differ from that used for journals (e.g. perhaps an interview format will be used when contacting authors directly), but the information collected from all studies should be the same.

After creating the data collection form it should be piloted on a sample of included studies to ensure that all relevant information is captured and the data collection procedure is well-understood and consistent among the reviewers.

If possible, the data collection form should be described, or an example provided when reporting the systematic review, to provide an unambiguous account of the data collection process. Depending upon the nature of the systematic review, data collection forms which have been completed and checked may be presented in an appendix alongside the main systematic review (section 3.7). This has the advantage of ensuring transparent and thorough reporting of all collected data, but may not be feasible if many studies are included in the review.

3.4.4. Procedure for data collection

The protocol will describe how many reviewers will be responsible for data collection and how any discrepancies of opinion will be resolved. It should also be described how any conflicts of interest will be handled. A conflict of interest occurs, for example, if a reviewer has contributed to any of the studies included in the review or has any stakes in the outcome of the study. Systematic reviews usually use parallel review with at least two reviewers completing the data collection form. Parallel review refers to the process of reviewers independently collecting data and conferring later to determine the similarity of the results extracted. An alternative approach is sequential review, which refers to collection initially by one reviewer and verification by another reviewer.

Reviewers may be randomly assigned the studies from which to collect data. Blinding reviewers to the journal and author details may be recommended. Sometimes a non-random allocation process may be employed. For example if the review team includes speakers of different languages it may be appropriate to allocate foreign-language studies to specific reviewers.

After the information has been collected the next steps in the systematic review process are to assess the methodological quality of the included studies (section 3.5); synthesise data from the included studies (section 3.6); clearly present the data from the studies together with the results of the systematic review (section 3.7); and interpret the results and draw conclusions (section 3.8). Careful attention to the accuracy of data collection will avoid a need to revisit papers for clarification of the study characteristics.
SUMMARY POINTS

- The systematic collection of data from each primary research study has a key role in ensuring the reproducibility of the systematic review.

- The guiding principle for data collection, regardless of the topic area, should be to determine study findings and to report study characteristics that influence the external applicability and internal validity and relevance of the findings.

- Data collection requirements should be tailored to the review question and the planned analyses specified a priori in the review protocol.

- To ensure consistency of interpretation and reporting, criteria for any subjective judgements should be defined a priori by the review team.

- Any change to the data collection process has to be applied to all the included studies.

- The data collection step in the systematic review forms the basis for research synthesis methods such as meta-analysis and should be done with research synthesis in mind.
3.5. **Assessing methodological quality of the included studies**

An important aspect of a systematic review is the consideration of the validity of the individual studies. Figure 4 illustrates that there are many stages of the review at which such considerations occur. First, the eligibility criteria will specify which study designs should be included, and this - either implicitly or explicitly - determines a threshold related to the validity of the studies’ findings. Although formal methodological quality assessment might occasionally be used to exclude studies that do not meet certain criteria, this is not standard practice and differential quality is more usually assessed at the synthesis stage through sensitivity analysis (CRD, 2009). If studies are to be excluded because of their methodology, this should be part of the study design eligibility criteria discussed in Section 3.1.1.1.

![Diagram](image)

**Figure 4:** Where methodological quality might be addressed in a systematic review

In this section we consider the step within the systematic review process in which the methodological quality of each included study is critically appraised. Methodological quality is defined here as aspects of the design, execution, analysis and reporting of a study that may lead it to give a biased result, so that there is a risk that its findings differ systematically from the truth. In a systematic review each study should undergo a standardised assessment, checking whether or not it meets a predefined list of methodological characteristics, to assess the degree to which it is susceptible to bias.
There are some common types of bias that can occur in many different study designs. In health care research these are often classified as selection, performance, detection, attrition and reporting biases. Elaborating on these dimensions of bias may be helpful:

- Where study subjects recruited to a study systematically differ from those to whom the results are likely to be applied, a study is described as having a selection bias. This can occur for several reasons - either methodological, because the sample was selected in an unrepresentative way - or related to the context in which the study was undertaken not matching with the area to which its results are going to be applied. Some texts describe these issues as relating to external validity, generalisability, or applicability.

- A second form of selection bias arises in studies which compare two or more groups, such as an intervention versus a control, or exposure versus no exposure. If the way in which study subjects selected to go into the different groups creates groups which differ in other characteristics, then the estimate of the effect of the risk factor or intervention made will be potentially confounded. Randomisation is the best way of avoiding selection bias, and is recommended whenever possible. Statistical methods for case-mix adjustment can sometimes be used to partially correct for confounders, but their success will depend on the availability of good measures of the confounders, which is often limited.

- If study subjects are misclassified as being exposed when in fact they were not (or vice versa), or treated when in fact they did not receive treatment, then the comparison between groups will be biased. This could occur through use of poor data sources for exposure records or poor compliance of the study subjects. Also, if those who are exposed or treated receive additional undocumented exposures or treatments then the attribution of the difference in groups to a single treatment will be erroneous. These are types of performance bias. To prevent performance bias, steps should be taken to ensure treatment or exposure fidelity of the study subjects and if possible the blinding of both study subjects and treatment providers.

- Detection bias refers to a potential artefact in the assessment of outcomes caused, for example, by the use of a particular diagnostic technique or type of equipment (disease rates may be over- or underestimated in different populations, regions or periods because of different diagnostic technologies used). Correct assessment of outcomes, including the blinding of outcome assessors, is necessary for preventing detection bias.

- All studies are commonly troubled by missing data and study subjects who are lost during follow-up. Missing data that relate to the exposure, intervention or outcome will lead to bias in the analysis. This is known as attrition bias. Statistical methods of multiple imputation are becoming more widespread allowing for the uncertainty related to missing data to be better understood.

- Reporting bias refers to an unrepresentative or incomplete selection of the facts which are reported. This could include selective reporting of outcomes, selective choice of statistical methods (such as the use of a particular categorisation technique), or reporting only on certain subsets of the study subjects. Such selective reporting may give the most favourable perspective on a research hypothesis. However, reporting bias may be very difficult to detect, particularly if a study does not have a clear protocol or detailed methods section explaining a priori which methods and outcome assessments were planned.

### 3.5.1. Procedures for assessing methodological quality

Assessment of methodological quality involves using tools to identify those aspects of study design, conduct, or analysis which induce a possible risk of bias. Part of the systematic evaluation involves ensuring that all included studies are assessed in a standardised manner.

Tools to assess methodological quality usually contain a series of items that focus on particular aspects
of study design and execution (such as the method used to allocate study subjects to groups, the use of blinded assessments, and the completeness of the data).

Tools for assessing quality are often categorised into checklists (which are a list of questions and produce a list of areas of concern for each study) or scores (which produce a numerical rating for each study). The use of scores is often controversial as there rarely is a scientific rationale for the differing weights given to each aspect of the bias assessment.

The Cochrane Handbook for Systematic Reviews of Interventions (Higgins and Green (editors), 2009) and the Centre for Reviews and Dissemination (CRD) Guidance for Undertaking Reviews in Health Care (CRD, 2009) give examples of tools for assessing methodological quality that have been extensively applied in health research. There are many different checklists available, which can be modified to meet the requirements of the review. The Cochrane “risk of bias” tool can be used to assess risk of bias (Higgins and Green (editors), 2009). The most recent version of the Cochrane Handbook also contains guidance on dealing with non-randomised studies in systematic reviews of interventions. A useful checklist for observational studies was published as part of the US Agency for Healthcare Research and Quality (AHRQ) Systems to Rate the Strength of Scientific Evidence (AHRQ, 2002). Sargeant et al. (2005) have developed a series of quality assessment checklists adapted to the different study types that can be used for agri-food public health topics, but each individual systematic review question will probably require the development of a more specific checklist tailored to its own requirements. For evaluating studies of diagnostic test accuracy the QUADAS tool is recommended by the Cochrane Collaboration.

### 3.5.2. Quality of reporting

Poor reporting of intervention studies has been documented in human medicine and also in veterinary medicine (Sargeant et al., 2009). Quality of reporting does not necessarily reflect the quality of the underlying methods or data. It is important to be accurate and distinguish between failure to report a methodological quality criterion and failure to meet a criterion.

In health research there are a number of initiatives aimed at improving the quality of reporting of primary research. These include the CONSORT statement for experimental studies, the STROBE statement for observational studies and the STARD statement for studies of diagnostic accuracy. These and other guidance statements, together with their routine updates, have been collected together under the EQUATOR Network. Although in the area of health research, the EQUATOR Network and the guidance that it encompasses may be transferable to other disciplines including food and feed safety. For experimental studies in food safety, more specific guidance is REFLECT-LFS (Reporting Guidelines for Randomized Controlled Trials in Livestock and Food Safety) (Sargeant et al., 2010).

---

15 <http://www.equator-network.org>
SUMMARY POINTS

- In a systematic review each study should undergo a standardised assessment, checking whether or not it meets a predefined list of methodological characteristics, to assess the degree to which it is susceptible to bias.

- There are many stages of the review at which the validity of the individual studies is considered. This section focuses on the step within systematic review process in which the methodological quality of each included study is critically appraised.

- Common types of bias that can occur in many different study designs are often classified as selection, performance, detection, attrition and reporting biases.

- Assessment of methodological quality involves using tools (e.g. checklists) to identify those aspects of study design, execution, or analysis which induce a possible risk of bias.

- It is important to distinguish between the quality of a study and the quality of reporting the study, although both may be correlated.
3.6. Synthesising data from included studies - Meta-analysis

Meta-analysis refers to the quantitative synthesis of results from multiple independent studies using statistical methods to obtain an overall estimate of a parameter or effect. Meta-analysis has several advantages compared to obtaining parameter or effect estimates from individual studies:

- Being based on a larger total sample size, it may provide a more precise estimate of a parameter or effect;
- The statistical methods used in meta-analysis enable the uncertainty or confidence of parameter or effect estimates to be calculated and displayed (e.g. as confidence intervals);
- Sensitivity analyses may be conducted by selectively including or excluding particular studies from the analysis (e.g. according to methodological quality) to explore their contribution to the overall outcome;
- Individual studies may be weighted in the analysis according to their sample size;
- Graphical methods for displaying the results of meta analyses are available (forest plots) which can display all the individual parameter or effect estimates together with the overall effect; this may help to identify the reasons for any differences in results between the primary studies.

Some potential limitations of meta-analysis are:

- The statistical methods underpinning the analysis approach need to be reasonably well understood to avoid misapplication of methods (e.g. to data that are too dissimilar (heterogeneous) to warrant quantitative synthesis);
- Meta-analysis may propagate bias if applied to studies of poor quality;
- Primary studies may not provide sufficient quantitative information to permit meta-analysis.

In cases where meta-analysis is not feasible, quantitative results from a range of studies may be presented in tables and/or charts, and interpreted and discussed narratively (section 3.7).

Further information about the potential use of meta-analysis, including issues relevant to systematic review of food and feed safety, is provided in Appendix E.

**SUMMARY POINTS**

- Data generated through a systematic review may be suitable to conduct a meta-analysis, which is a statistical synthesis of estimates from multiple independent studies. Meta-analyses provide more statistical power in the estimation of parameters compared to the primary studies.
- Meta-analysis may not be applicable in case of severe heterogeneity in the primary results. (Meta) regression techniques can sometimes be useful to explain heterogeneity, e.g. through investigation of the impact of study design and other factors.
- In cases where meta-analysis is not feasible, quantitative results from a range of studies may be presented in tables and/or charts, and interpreted and discussed narratively.
3.7. Presenting data and results

3.7.1. Key principles of data presentation in systematic reviews

To assist interpretation of systematic reviews and ensure transparency of the systematic review process it is important that the characteristics of the included studies, the data collected from them, and results of the analyses conducted by the systematic review team are clearly presented.

It is often convenient and appropriate to present the results in tabular form. Summary tables have a number of advantages compared to narrative reporting:

1. Large amounts of information may be concisely presented in a structured way.
2. Information may be easier to find within a systematic review if presented in tables rather than text.
3. The information may be structured into groups or subgroups, facilitating comparisons of relevance to the systematic review (e.g. studies may be grouped according to their design, authors, location, year of publication, sample size, population characteristics or other variables).
4. Data tables may facilitate a continuity of structure throughout the systematic review. For example, it may be helpful to structure the presentation of data according to the same key elements (e.g. populations, outcomes, interventions, exposures) that are used to design the data collection form (section 3.4).
5. Tabulation of information can assist checking for accuracy and may help to avoid selective reporting (as missing information may be more easily identified in a table than in a section of text).

A potential disadvantage of data tabulation is that if a large amount of information is presented in a very large table or split across multiple tables, this may appear overwhelming to readers of the systematic review and nullify some of the advantages noted above (this problem would also apply to narrative reporting of very large amounts of information).

A useful approach could be to present the most detailed information in tables in an appendix to the systematic review, enabling the results section of the review to focus on the most important of the variables under consideration. Carefully designed data collection forms (section 3.4) could be included directly in an appendix to the systematic review as an efficient way to achieve this.

In some situations, graphical presentation of results may be appropriate. If charts (e.g. line charts, histograms, or more complex charts) are to be included it will be necessary to consider how they will be used by readers of the systematic review. While charts may usefully summarise complex information, they may be unhelpful if the information presented is important but inaccessible to readers. For example, parameter estimates such as means, standard deviations, and confidence intervals cannot be accurately extracted visually from charts.

The results section of a systematic review should always provide a narrative statement of the results. Where information is presented in tables and charts, the narrative description may not need to be very extensive. It should clearly cross reference all the relevant tables and charts, but should avoid duplication of information given in these, unless necessary to clarify key parameters. This applies for each section of the results, i.e. the description of study characteristics (section 3.7.2.1), the description of the data collected (section 3.7.2.2) and the description of the results of analyses (section 3.7.2.3).

16 The Health Technology Assessment monograph series provides examples of systematic reviews in health research that usually employ this approach and may serve as a useful reference (<http://www.hta.ac.uk/project/htapubs.asp>).
3.7.2. Which results to present?

There are three main types of information that should be presented in the results section of a systematic review:

1. the characteristics of the primary studies that are included in the systematic review;
2. the data that the review team have collected from the primary studies to analyse; and
3. the results of analyses carried out on those data by the review team (which may include the results of meta analyses).

The principles of good reporting practice mentioned above (section 3.7.1) apply to each of the types of information being reported.

3.7.2.1. Characteristics of the primary studies

The information that should be reported about the primary studies will depend upon the objectives of the systematic review. It is usual practice in systematic reviews to provide the following basic information:

1. Authors and year of publication.
2. Type of study.
3. Location.
4. Sample size.
5. A summary of the key elements (e.g. population, intervention(s) or exposure scenario(s), comparator(s), and outcome(s) reported) (section 2.1).

It is particularly important to ensure that aspects of the primary studies which could influence the interpretation and analysis of results are reported here (section 3.4.1). If studies have complex populations, interventions or exposure scenarios it may be appropriate to describe these more fully in an appendix (e.g. within the data collection forms referred to above - section 3.4.3).

3.7.2.2. Data collected from the primary studies

For consistency of reporting, data that are collected from the primary studies should be presented in a manner that is consistent with the presentation of study characteristics. For example, if study characteristics are presented in chronological order by publication year, a similar approach would be appropriate for the presentation of the results. If any assumptions have been made in extracting (or imputing) information from the primary studies these should be clearly stated. The information presented should agree with that stated in the protocol and methods section of the systematic review (i.e. the rationale for how and why data were extracted should have been previously explained).

Data should be presented in such a way as to enable others to conduct secondary analysis or synthesis if appropriate (even if the data do not seem amenable to quantitative syntheses, e.g. if the data quality is highly questionable or the number of studies is small compared to the expected heterogeneity, the data may still be of value to other secondary research studies, whose objectives may not be foreseen).

3.7.2.3. Results of analyses conducted by the systematic review team

Depending upon the nature of the evidence available, there may be little or no quantitative analysis of the data extracted from the primary studies, or a detailed meta-analysis may be conducted (section 3.6 and Appendix E).
The information presented should agree with that stated in the protocol and methods section of the systematic review (i.e. if results are presented for meta-analyses, the statistical techniques used should have been previously explained).

Whether a meta-analysis is conducted or not, a narrative statement should always be provided to explain the overall synthesis of the results.

It is common practice in systematic reviews to conduct an assessment of methodological quality to characterise the included primary studies in terms of their likely susceptibility to bias (section 3.5). Results of such an assessment may be presented here and may inform the final summing up of the results (e.g. a meta-analysis may be conducted omitting studies of lower quality).

**SUMMARY POINTS**

- In systematic reviews the characteristics of the included studies, the data collected from them, and results of the analyses conducted are clearly presented to assist interpretation and ensure transparency of the process.

- Presenting the results in tabular form has a number of advantages compared to narrative reporting.

- Three main types of information should be presented in the results section of a systematic review: the characteristics of the primary studies that are included in the systematic review; the data that the review team collected from the primary studies to analyse; and the results of analyses carried out on those data by the review team (which may include the results of meta-analyses).
3.8. **Interpreting results and drawing conclusions**

A well-structured Discussion and a clear presentation of the reviewers’ Conclusions are important parts of the review that can assist decision making.

The following issues could be addressed in the Discussion or Conclusions sections of the systematic review (cross-referring to the relevant information presented in the results section):

1. The quantity of evidence. Reference could be made to the total number of papers screened and of those included and the total number of subjects (reported in the results section) in order to describe the weight of the evidence gathered.

2. The quality of the evidence. The Discussion should include an assessment of the quality of the body of evidence for each individual outcome, involving considerations of study methodological quality, heterogeneity, precision of parameter or effect estimates, and risks of bias. The quality of a body of evidence may be decreased by limitations in the design and implementation of the studies; based on the methodological quality assessment (section 3.5), the reviewers must make judgements about study limitations for each main outcome. When studies yield widely differing estimates of a parameter or effect (heterogeneity or inconsistency of results), reviewers should seek explanations for that heterogeneity.

3. Interpretation of the results. This should include interpretation of both the statistical significance and the biological significance of the findings with a clear explanation of all assumptions made. Sensitivity analyses may help to assess the robustness of the conclusions and give an additional measure of the confidence or uncertainty attached to the main conclusions. In cases where very few relevant data are found, the characterisation and reporting of the knowledge gaps may be useful in supporting research recommendations.

4. Any potential limitations of the review process.

5. Agreements or disagreements with other studies or reviews.

Whether incorporated in the Discussion section or presented separately, it is essential that the Conclusions are clearly worded, based solely on the evidence reviewed, and provide a focused answer to the question(s) asked for the systematic review.

It is possible that unforeseen secondary or complementary questions or hypotheses may arise from the results of a systematic review. Care should be taken not to place emphasis on any questions, hypotheses or results that are not among those that were planned *a priori* to be addressed in the SR. Such “additional” findings should not form the basis of any conclusions or recommendations arising, but may be mentioned in the Discussion section.

Specific gaps in the evidence should be highlighted and recommendations for further research may be included. Where methodological issues have been identified in existing studies, suggestions for future approaches may be made. Where possible, research recommendations should be listed in order of priority, with an explanation. This can also assist in planning an update of the review.
SUMMARY POINTS

- A well-structured Discussion and a clear presentation of the reviewers’ Conclusions are important parts of the review that should assist decision making.

- Discussion or Conclusions could include description of the quantity and quality of evidence underpinning the review question; interpretation of the results; any potential limitations of the review process; and agreements or disagreements with other research.

- Recommendations for further research may be included. Where methodological issues have been identified in existing studies, suggestions for future approaches may be made.
SUMMARY POINTS AND FURTHER DEVELOPMENTS

SUMMARY POINTS

Due to its methodological rigor, objective and transparent nature, systematic review methodology and its principles are appropriate for answering well-formulated specific questions generated by the risk assessment process or other analytical frameworks in food and feed safety.

Risk models or other broad food and feed safety policy problems may contain one or more specific questions. As only certain questions are answerable by performing a systematic review, the use of systematic review must be evaluated for each specific question within a mandate addressing a broad policy problem in food and feed safety.

If a specific question is suitable for systematic review, this does not necessarily mean that a SR is worthwhile or practically feasible. Several considerations may be made to help decide which questions could be answered by conducting a systematic review. These considerations include prioritisation of model parameters for which refinement of the parameter estimates is considered most critical (e.g. in risk assessment); assessment of the quantity and quality of available evidence; the source and potential confidentiality of the evidence; the need for transparency and/or for integrating conflicting results; and the resources needed for carrying out the review.

This Guidance presents a method for performing systematic reviews for suitable food and feed safety questions, taking into account issues that may be unique to the field of food and feed safety.

FURTHER DEVELOPMENTS

As this Guidance represents a first step towards the application of systematic review methodology in food and feed safety, regular updates are recommended in light of experience and new evidence both in food and feed safety and systematic review methodology.

Feedback from users of the Guidance is foreseen and for this purpose the creation of a repository for frequently asked questions on the use of systematic reviews in food and feed safety would be useful. This would benefit users of the Guidance and inform future updates.

When systematic reviews are conducted in food and feed safety, the protocols of the systematic reviews could be made available on a dedicated website, to encourage feedback from reviewers and stakeholders. The creation of an archive of systematic reviews in food and feed safety is recommended to avoid duplication of efforts and to promote exchange of information on current challenges in this field.
4. EFSA workshop on the application of systematic review methodology to food and feed safety assessments to support decision making

4.1. Workshop outline

As part of this EFSA Mandate (section “Terms of reference as provided by EFSA”), on 23, 24 and 25 February 2010 the Assessment Methodology Unit organised an EFSA-tailored workshop on the application of systematic review methodology to food and feed safety assessments to support decision making. Fifty participants attended the workshop, comprising Panel members and scientific staff.

The workshop was structured to contain theoretical and practical sessions and an open discussion, led by members of the working group who developed this EFSA Guidance. During the practical sessions, the attendees developed and worked on examples of broad policy problems from different areas of food and feed safety (environmental risk assessment, chemicals risk assessment and zoonotic public health) and considered their applicability for systematic review.

During the first two days the programme focused on presenting (and receiving feedback on) the Guidance document (whose draft version had been previously circulated to the workshop attendees), namely: the core principles of systematic reviews; how to identify food and feed safety questions suitable for SR; how to perform a SR, considering the specific requirements of food and feed safety questions; and how to assess when a SR is needed and practically feasible, when performing food and feed safety assessments to support decision making.

On the last day the attendees discussed the advantages and disadvantages of systematic reviews to food and feed safety risk assessments in support of decision making and discussed when and how systematic review could be implemented in EFSA, i.e. how far the systematic review process could be integrated in the scientific outputs workflow.

4.2. Outcomes of the open discussion on the advantages and disadvantages of SR

The outcomes of the discussion on the advantages and disadvantages of using the SR review method when doing risk assessment in food and feed safety to support decision making can be summarised as follows:

1. Advantages: the attendees emphasised that most of the principles of systematic reviews (methodological rigour, transparency, and reproducibility) are already in use in EFSA; however, the standardised structure of the systematic review process is not always used. Being methodologically rigorous, objective and well documented, systematic reviews could enhance transparency and help reduce criticism of food and feed safety assessments. Extensive and systematic information searches could reduce the risk of excluding potentially relevant data from the assessments. Thus, systematic reviews could be particularly valuable in case of sensitive issues. For applications (health claims) and dossiers, systematic reviews could ensure that all potentially relevant information is considered, properly assessed and synthesised by the external party.

2. Disadvantages: systematic reviews are resource-intensive in terms of expertise required and time. The method may be limited to questions for which primary research (i.e. studies that generate primary data) is available and therefore may address only parts of an EFSA mandate or RA model. In the case of applications for health claims, the use of systematic reviews may considerably increase the cost of the product assessment. Therefore, the use of SR should be carefully assessed on a case-by-case base.

4.3. Workshop conclusions and further developments

Systematic reviews could be performed in the EFSA framework at different levels, depending on the questions to be answered, the timeframe and the available resources. When a full SR is not applicable, the systematic review process could be adapted (i.e. each step of the process could be implemented
with some compromises, e.g. less extensive searches, one reviewer only, etc), in order to increase methodological rigour, transparency, and reproducibility when producing scientific outputs.

The outcomes and the conclusions of the workshop were presented at the EFSA Scientific Committee plenary meeting in April 2010.
REFERENCES


EFSA (European Food Safety Authority), 2009c. Scientific Opinion of the Panel on Animal Health and Welfare (AHAW) on a request from the Commission on porcine brucellosis (Brucella suis). The EFSA Journal (2009), 1144, pp. 2-112.

EFSA (European Food Safety Authority), 2009d. Guidance of the Scientific Committee on a request from EFSA on the use of the benchmark dose approach in risk assessment. The EFSA Journal (2009)


O’Connor AM, Denagamage T, Sargeant JM, Rajic´ A, McKean J, 2005. Feeding management practices and feed characteristics associated with Salmonella prevalence in live and slaughtered
market-weight finisher swine: A systematic review and summation of evidence from 1950 to 2005.
Preventive Veterinary Medicine, 87, pp. 213-228.


APPENDICES

Appendix A - Breaking down broad food and feed safety policy problems and identifying suitable questions for using systematic reviews - Examples

In this Appendix three examples of broad food and feed safety policy problems are presented:


2. a generic example based on the Codex Alimentarius (FAO/WHO, 2005) methodology for chemical contaminants in the food chain Risk Analysis;

3. a specific example of the analytical framework that can be developed for assessing a nutrition-related topic.

The same approach for breaking down broad food and feed safety policy problems into specific questions and identifying, amongst them, well-formulated questions that are amenable to systematic review is applied to the three examples.

Appendix A1 - Import risk assessment

The epizootic hemorrhagic disease virus (EHDV) (EFSA, 2009b) import risk assessment represents a broad assessment that includes several conceptual, qualitative, or quantitative models where various specific questions can be identified.

The risk assessment is performed using the methodology of the import risk analysis described by the Office International des Épizooties (OIE, 2004), which includes three phases:

1. hazard identification. Epizootic hemorrhagic disease virus (EHDV) is the causal agent of Epizootic hemorrhagic disease (EHD);

2. release assessment (describing the possible introduction pathways of the virus);

3. exposure assessment (estimating the probability that a susceptible host becomes infected).

For the purposes of this example only the release assessment is considered and, within it, the pathway of introduction with legal import of infectious livestock (Figure 5).
Figure 5: Pathways of introduction of EHDV

The pathway of introduction through legal importation of infectious livestock from third countries is developed in Figure 6 (from EFSA, 2009b).

Legend:
P_f: Probability of Freedom
p: Annual prevalence in the country of origin
P_{IWQ}: Probability of being infectious without quarantine
P_{IAQ}: Probability of being infectious with quarantine

1 - Test Sensitivity (1 - Se)

Figure 6: Release assessment (from EFSA, 2009b)

In relation to the above route of introduction an assessment of the significance of the presence, origin and occurrence of EHDV in livestock animals in the EU neighbour countries should be performed. This assessment includes: the description of the hazard; the identification of species susceptible to EHDV and their geographical distribution; the description of geographical occurrence and distribution
of EHDV; and, specifically in EU neighbouring countries, the pathogenesis of EHDV. The diagnostic tools available and their accuracy should also be assessed.

This risk assessment comprises a number of questions. To establish which specific questions are potentially answerable by conducting a systematic review, it is helpful to identify all the specific questions, determine their key elements and establish whether the questions are open-framed or closed-framed (section 2.1.3) (Table 7).
### Table 7: List of questions included in an import risk assessment model (release assessment). For each question, the characteristics of the question and suitability for systematic review are illustrated

<table>
<thead>
<tr>
<th>List of specific questions included in the RA model</th>
<th>Characteristics of the question (question type, key elements, open-framed/closed-framed question)</th>
<th>Is the question answerable using SR?</th>
</tr>
</thead>
<tbody>
<tr>
<td>In which species has EHDV been observed? (list of susceptible species)</td>
<td>Specific, OPEN-FRAMED question (variant of the Prevalence type). Key elements: Population = susceptible species (it must be determined) Outcome (conduction of interest) = EHDV (it is given)</td>
<td>(YES) But the eligibility criteria (including the appropriate study designs) will be very broad and the literature search may retrieve many reports since the only usable term may be the virus name (and synonyms or terms indicative of the virus). It may be preferable to address this question using just a systematic literature search to help refine the question.</td>
</tr>
<tr>
<td>In which countries have the susceptible species been observed? (geographic distribution of the susceptible species identified, particularly in the EU neighbouring countries)</td>
<td>Specific, OPEN-FRAMED question (as above). Key elements: Population = the location (it must be determined) Outcome (conduction of interest) = susceptible species (it is given)</td>
<td>(YES) Same as above.</td>
</tr>
<tr>
<td>In which countries has EHDV been observed?</td>
<td>Specific, OPEN-FRAMED question (as above). Key elements: Population = the location (it must be determined) Outcome (conduction of interest) = EHDV (it is given)</td>
<td>(YES) Same as above.</td>
</tr>
<tr>
<td>With what prevalence does EHDV occur in (e.g.) cattle in country Y? (prevalence)</td>
<td>Specific, CLOSED-FRAMED question (Prevalence type). Key elements (both determined): Population = cattle and geographic area Outcome (conduction of interest) = EHDV</td>
<td>YES This is a well-formulated question, answerable by systematic review.</td>
</tr>
<tr>
<td>What diagnostic tools are available to determine EHDV?</td>
<td>Specific, OPEN-FRAMED question (variant of Test Accuracy type). Key elements: Population = susceptible species (it is given) Index test = diagnostic tools (to be determined) Target condition = EHDV (it is given)</td>
<td>NO The question is answerable by doing a systematic literature search (chapter 3.2). Once the existing diagnostic tests are identified, their accuracy can be assessed (i.e. the question is converted into the question below).</td>
</tr>
<tr>
<td>What is the sensitivity (specificity) of diagnostic</td>
<td>Specific, CLOSED-FRAMED question (Test Accuracy type).</td>
<td>YES This is a well-formulated question, answerable by systematic review.</td>
</tr>
<tr>
<td>List of specific questions included in the RA model</td>
<td>Characteristics of the question (question type, key elements, open-framed/closed-framed question)</td>
<td>Is the question answerable using SR?</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
</tr>
</tbody>
</table>
| test I?                                          | Accuracy type).  
Key elements (all determined):  
Population = susceptible species  
Index test = diagnostic test I  
Target condition = EHDV | systematic review. |
| In which vector species has EHDV been observed (per virus strain)? (list of vectors) | Specific, OPEN-FRAMED question (variant of the Prevalence type).  
Key elements:  
Population = vectors (to be determined)  
Outcome (conduction of interest) = EHDV (it is given) | (YES) But the eligibility criteria (including the appropriate study designs) will be very broad and the literature search may retrieve many reports. It may be preferable to address this question using just a systematic literature search to help refine the question. |
| In which geographical areas/seasons have the vectors been observed? | Specific, OPEN-FRAMED question (as above).  
Key elements:  
Population = location (to be determined)  
Outcome (conduction of interest) = vectors (it is given) | (YES) Same as above. |
| What control measures are available against the vectors? | Specific, OPEN-FRAMED question (variant of Effect of Intervention type).  
Key elements:  
Population = vectors (it is given)  
Intervention = control measures (to be determined)  
Comparator = (to be determined)  
Outcome (conduction of interest) = elimination of vectors (it is given) | NO The question is answerable by doing a systematic literature search (chapter 3.2).  
Once the existing control measures are identified, their efficacy can be assessed. |
Appendix A2 - Risk assessment of chemical contaminants in the food chain

The risk assessment of a chemical contaminant in food and the implication of its presence for human health illustrate how the four pillars of risk assessment (hazard identification, hazard characterisation, exposure assessment and risk characterisation) can be broken down into specific questions to answer a broad policy problem. This section addresses how such specific questions can be tackled using systematic literature searches and when systematic review approach is possible. A generic example is given here to illustrate the potential for application of systematic review methodology. This example could be adapted to specific RA questions on a case by case basis.

In this RA, the specific questions and their key elements are determined by the RA process. For example, to assess the toxic effects of a substance on humans the population of interest is clearly the human population. However, if no relevant human data are available the RA approach could be to look at animal experiments. This refocuses the population of interest to experimental animals.

A list of all specific questions identified from a risk assessment of a chemical contaminant in food is provided in Table 8. For each question, the characteristics in terms of question type, whether open-framed or closed-framed, the key elements present, and whether the question is answerable using SR methodology are illustrated. It is important to note that if a question is answerable using a SR, this does not necessarily mean that a SR is warranted or practically feasible. The reasons for doing or not a systematic review are described in section 2.2.
Table 8: List of specific questions included in the four steps of risk assessment to assess the potential human health effects of the presence of a chemical contaminant in food. For each specific question, the characteristics of the question and suitability for using a systematic review are illustrated.

<table>
<thead>
<tr>
<th>List of specific questions included in the RA model</th>
<th>Characteristics of the question (question type, key elements, open-framed/closed-framed question)</th>
<th>Is the question suitable for using SR?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hazard identification</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does chemical $x$ have a mutagenic effect on cells used in mutagenicity tests?</td>
<td>Specific, CLOSED-FRAMED question (Effect of exposure type). Key elements: Population = cells used in mutagenicity tests Exposure = chemical $x$ Comparator = non-exposure Outcome = mutagenicity</td>
<td>YES This is a well-formulated question, answerable by conducting a SR.</td>
</tr>
<tr>
<td>Does chemical $x$ have a genotoxic effect or cancer induction in rat liver?</td>
<td>Specific, CLOSED-FRAMED question (as above). Key elements Population = rat liver Exposure = chemical $x$ Comparator = non-exposure Outcome = genotoxic effect or cancer induction</td>
<td>YES This is a well-formulated question, answerable by conducting a SR.</td>
</tr>
<tr>
<td>Does chemical $x$ have a genotoxic effect or cancer induction in human liver?</td>
<td>Specific, CLOSED-FRAMED question (as above). Key elements Population = human liver Exposure = chemical $x$ Comparator = non-exposure Outcome = genotoxic effect or cancer induction</td>
<td>YES This is a well-formulated question, answerable by conducting a SR.</td>
</tr>
<tr>
<td><strong>Hazard characterisation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is the fate of chemical $x$ (absorption, distribution, metabolism, excretion) in the rat?</td>
<td>Specific, CLOSED-FRAMED question (effect of an exposure (fate) type). Key elements Population = rat Exposure = chemical $x$;</td>
<td>YES This is a well-formulated question, answerable by conducting a SR.</td>
</tr>
</tbody>
</table>
### List of specific questions included in the RA model

<table>
<thead>
<tr>
<th>Question</th>
<th>Characteristics of the question (question type, key elements, open-framed/closed-framed question)</th>
<th>Is the question suitable for using SR?</th>
</tr>
</thead>
</table>
| What is the fate of chemical $x$ (absorption, distribution, metabolism, excretion) in humans? | Comparator = different dose levels  
Outcome = toxicokinetic parameters  
Specific, CLOSED-FRAMED question (as above).  
Key elements  
Population = humans  
Exposure = chemical $x$  
Comparator = different dose levels  
Outcome = toxicokinetic parameters | YES  
This is a well-formulated question, answerable by conducting a SR. |
| Is there a dose-response relationship between chemical $x$ and liver toxicity in the rat? (Dose-response assessment) | Comparator = different dose levels  
Outcome = liver toxicity  
Specific, CLOSED-FRAMED question (effect of an exposure (dose-response) type).  
Key elements  
Population = rat  
Exposure = chemical $x$  
Comparator = different dose levels  
Outcome = liver toxicity | YES  
This is a well-formulated question, answerable by conducting a SR. |
| Is there a dose-response relationship between chemical $x$ and liver toxicity in humans? | Comparator = different dose levels  
Outcome = liver toxicity  
Specific, CLOSED-FRAMED question (as above).  
Key elements  
Population = humans  
Exposure = chemical $x$  
Comparator = different dose levels  
Outcome = liver toxicity | YES  
This is a well-formulated question, answerable by conducting a SR. |
| What is the acceptable or tolerable daily intake (ADI/TDI) for chemical $x$ in humans (or animal test species)? | Comparator = different dose levels  
Outcome = benchmark dose (BMD), no observed adverse effect level (NOAEL), lowest observed adverse effect level (LOAEL).  
Key elements:  
Population = humans (or animal test species);  
Exposure = chemical $x$;  
Comparator = different dose levels; Outcome = benchmark dose (BMD), no observed adverse effect level (NOAEL), lowest observed adverse effect level (LOAEL). | YES  
However, normally this question would be answered by reviewing previous ADI/TDI or margin of exposure data for a chemical in humans derived by public agencies. |

### Exposure Assessment

<table>
<thead>
<tr>
<th>Question</th>
<th>Characteristics of the question (question type, key elements, open-framed/closed-framed question)</th>
<th>Is the question suitable for using SR?</th>
</tr>
</thead>
</table>
| How much of chemical $x$ occurs in the different food | Specific, CLOSED-FRAMED question | YES  
This is a well-formulated question, answerable by |
<table>
<thead>
<tr>
<th>List of specific questions included in the RA model</th>
<th>Characteristics of the question (question type, key elements, open-framed/closed-framed question)</th>
<th>Is the question suitable for using SR?</th>
</tr>
</thead>
<tbody>
<tr>
<td>commodities?</td>
<td>Occurrence type. Key elements Population = food commodity Quantity of interest (O) = quantity of chemical x</td>
<td>conducting a SR.</td>
</tr>
<tr>
<td>How much of the food commodity is consumed by humans?</td>
<td>Specific, CLOSED-FRAMED question, Consumption type. Key elements: Quantity of interest (O) = quantity of food commodity consumed Population = humans</td>
<td>YES This is a well-formulated question, answerable by conducting a SR.</td>
</tr>
<tr>
<td>What is the risk associated with human exposure to chemical x?</td>
<td>A question about risk characterisation reflects a broad question. This is because risk characterisation integrates answers from questions about hazard identification, hazard characterisation, and exposure assessment (each of which may be separately answerable by SR, as indicated above).</td>
<td>YES In principle a SR could be applied to synthesise outcomes of the risk characterisations from previous risk assessment studies. An alternative approach is to conduct a SR to answer questions about hazard identification, hazard characterisation and exposure assessment (as indicated above) and then use the answers as input for a de novo risk characterisation.</td>
</tr>
</tbody>
</table>
Appendix A3 - Analytical frameworks developed for assessing nutrition-related topics

This section provides an example of how specific questions addressable by systematic review can be identified within a broad food safety policy problem, such as the analytical framework developed for assessing a nutrition-related topic.

The example illustrated considers the health effects of (n-3) fatty acids on arrhythmogenic mechanisms in animal and isolated organ or cell culture studies (adapted from Lichtenstein et al., 2008). This is a broad problem about the effects of an exposure, as there are multiple possible target populations, and multiple experimental models and outcomes of potential interest.

For addressing this topic an analytical framework is developed to reflect the known biological mechanisms involving the nutrients which guides in integrating the various types of information available into a coherent picture (i.e. a conceptual model). In some cases, analytical frameworks serve as a guide for the integration of information from multiple types of data and offer visual maps of linkages among the populations of interest, exposures, modifying factors, biological roles of a nutrient, and outcomes of interest, outlining the chain of logic that evidence must support to link the exposure to the outcomes.

Defining these relationships helps to identify all possible specific questions included in the broad problem and, among them, those that can be answered using systematic review. The analytical framework developed for addressing the above mentioned example on (n-3) fatty acids and cardiovascular disease is illustrated in Figure 7.
Figure 7: Analytical framework (conceptual model) developed for (n-3) fatty acid exposure and cardiovascular disease (adapted from Lichtenstein et al., 2008)

Figure 7 shows how the analytical framework developed for (n-3) fatty acid exposure and cardiovascular disease depicts several questions that could be answerable using a SR. These could include questions about the effects of (n-3) fatty acids on three different target populations considering various types of potential outcomes (i.e. PECO questions - section 2.1.2.1), and a question on the exposure (of one or more of the populations) to the fatty acids, which implies the assessment of occurrence and consumption (i.e. PO questions - section 2.1.2.3). Different type of evidence is necessary to answer these questions. For example, for questions about the effects of effects of (n-3) fatty acids, the reviewer must include evidence from whole animal studies that (n-3) fatty acids affect arrhythmogenic outcomes as well as evidence from cell cultures and tissue studies that (n-3) fatty acids directly affect cell organelles involved in electrogensis. The results of all specific questions are then combined to give an answer to the broad problem.

The use of key elements (e.g. PICO) can help to define the structure of the various questions included in the analytical framework. An example is provided in Table 9, which shows possible choices of PICO elements in questions about the effects of an exposure. Different questions could be formulated by combining the items in each of the PICO categories shown in Table 9.
**Table 9:** Examples of the PICO elements in a question about the effects of exposure in a nutrition-related topic (from Lichtenstein et al., 2008)\(^\text{17}\)

<table>
<thead>
<tr>
<th>Population (P)</th>
<th>Intervention (I)</th>
<th>Comparator (C)</th>
<th>Outcome (O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population (primary prevention)</td>
<td>Fish oil</td>
<td>Isocaloric fat placebo</td>
<td>All cause mortality</td>
</tr>
<tr>
<td>Population with history of myocardial infarction (secondary prevention)</td>
<td>Fish oil (eicosapentaenoic acid - EPA, 20:5 n-3 - and docosahexaenoic acid - DHA, 22:6 n-3) supplement</td>
<td>No placebo</td>
<td>Cardiac death</td>
</tr>
</tbody>
</table>

\(^\text{17}\) The entries in the table are not meant to be exhaustive.
Appendix B - Searching for research studies - Full description and Examples

Appendix B1 - Developing the search strategy

Developing a search strategy, which may begin during a scoping stage or during the protocol development, may take several iterations during which the strategy is drafted, tested in key databases, the results assessed for relevance and the strategy then revised.

Search strategies are structured using one or more of the key elements of the research question (e.g. PICO or PECO) and take account of several factors:

- the choice of key elements to be used in the search;
- the search terms we identify to capture the key elements and the terms used by authors and indexers to describe their documents in database records;
- the search tools and facilities offered by the individual information resources such as Boolean operators (AND, OR, NOT) or truncation.

Appendix B1(1) - Search terms

Search terms are the words used to capture a search topic. Database records tend to provide limited information to search: words in the title and abstract (provided by authors) and sometimes subject indexing (provided by database publishers). With such limited information available, database search strategies tend to be designed to be highly sensitive. Sensitive searches involve the use of many synonyms and related terms to retrieve as many potentially relevant studies as possible. As a consequence, many studies may be retrieved that do not subsequently meet the inclusion criteria.

Search terms should take account of synonyms, abbreviations, geographical variations, changes in terminology over time and spelling variants (including US and British English variants and common mis-spellings) that may be used in the studies.

For example, the search question “Is vaccine x effective in cows against disease y?” has “cows” as the population key element. To search for “cows” requires the identification of a range of relevant terms to ensure a sensitive search:

Cow, Cows, Cattle, Bovine, Calf, Calves, Bull, Bulls, Heifer, Heifers, Livestock …

These terms can be used to search the titles and abstracts of database records and in the full text of internet publications and are known as “free text” terms. Searches should also include available subject indexing terms. These are keywords from a standardised or controlled vocabulary which have been added to the records by the database producer, to enhance retrieval of records that are described by authors using very different vocabulary. Indexing terms can be identified using the database thesaurus (if available) or by searching for obvious records and noting the indexing terms within those records. For example, CABI indexes records on cattle using subject headings such as “Cattle breeds”.

Appendix B1(2) - Search tools and facilities

Search terms and key elements can be combined using Boolean operators (AND, OR, NOT).

- The OR operator is inclusive and is used to capture relevant synonyms, making the search results larger. For example, retrieving all the records with any of the “cow” key element terms into a single result set is achieved using OR:

Cow OR Cows OR Cattle OR Bovine OR Calf OR Calves OR Bull OR Bulls OR Heifer OR Heifers OR Livestock
AND is usually used to combine key elements to narrow down (focus) the search, so that all the key components are present in each of the records retrieved. Combining the “cow” key element (set one) with the “vaccine” key element terms (set 2) would be achieved as follows:

1. Cow OR Cows\(^\text{18}\) OR Cattle OR Bovine OR Calf OR Calves OR Bull OR bulls OR bullocks OR Heifer OR Heifers OR Livestock
2. “bluetongue vaccine” OR BTV OR bovilis
3. 1 AND 2

NOT is used to exclude records from the search. NOT should be used with care because it may have a larger effect than anticipated; a record may well discuss both the concept of interest and the one to be excluded.

Other database-specific search facilities may enhance the search strategy:

- Truncation allows search term variants to be retrieved. For example using a truncation symbol ($) reduces the number of terms which need to be typed:

  Cow OR Cows\(^\text{18}\) OR Cattle OR Bovine OR Calf OR Calves OR Bull OR Bulls OR Bullock$ OR Heifer$ OR Livestock

  “bluetongue vaccin$”

- Wildcard symbols within terms (if available) can retrieve variant spellings:

  e.g. for hemorrhagic disease: H*emorrhagic disease would retrieve both American and English spelling of h(a)emorrhagic

- Some databases offer hierarchical indexing with the facility to “explode” general (higher level) thesaurus terms to retrieve more specific (lower level) thesaurus terms efficiently.

- Some databases offer proximity operators which can improve search efficiency. ISI Web of Science offers the SAME operator to search for words in same sentence:

  vaccine$vaccin* SAME cows

  This can improve the precision (focus) of the search because it identifies words within the same sentence whereas AND only requires words to be in the same record.

- Set combination allows structured combinations of search terms. The sets are numbered and can be combined using the Boolean operators:

  1. (Cow OR Cows OR Cattle OR Bovine OR Calf OR Calves OR Bull OR Bulls OR bullock$ OR Heifer$ OR Livestock)/TI,AB
  2. Cattle breeds/DE
  3. 1 OR 2
  4. (“bluetongue vaccin$” OR BTV OR Bovilis)/TI,AB

\(^{18}\) The truncation for “cow” is not used due to the opportunity for false drops from such a short word stem (i.e. it could produce false drops like e.g. “cowslips”, “cowsheds”, “cowherds”, or “cowbells”).
5. Bluetongue vaccine/DE
6. 4 OR 5

Key: /TI,AB refers to words in the title and abstract for the database in this example
/DE refers to indexing terms for the database in this example

Appendix B1(3) - Choosing the key elements to include in the search strategy

One or more key elements can be used to structure the search. The most sensitive search uses only one key element such as the population. This type of search may be helpful for projects where several questions need to be answered on the same subject or where there is only a small literature. Where the search yields many records the addition of a second or third key element is likely to be considered. Often the most logical key elements to test are the population and exposure or intervention.

Sometimes it can be difficult to capture a key element using search terms, because:

- it may use similar vocabulary to one of the other key elements;
- it may be expressed as numbers or units of measurement which are often difficult to capture efficiently as search terms;
- the potential search terms may be unknown, for example all possible outcomes of an intervention may be unknown at the time of the search.

Study design may be used as a search term (combined with other key elements using the AND operator) but care should be taken because of the likelihood of inconsistent reporting of study design by authors, which could lead to relevant studies being missed. Study design-focused search filters, such as those used in the health care field, might be tested for relevance in food and feed safety databases.

Depending on the research question and the resources available (e.g. time, budget, database subscriptions), limits may also be applied to the search strategy, for example:

- dates (particularly relevant if updating a systematic review);
- languages;
- exclusion of some publication types (e.g. letters, editorials, notes and comments). However, these publications can provide clues to other published or unpublished research, so their removal should be considered carefully.
- other options for limiting the searches may be available in specific databases, such as the “Concept Codes” or “Methods & Equipment” options in BIOSIS.

These decisions risk reintroducing biases, so their use should be noted in the documentation and their effect on the search results should be discussed.

An example strategy is shown below for a fictional database. It uses two key elements: Population = “cows” and Exposure = “bluetongue vaccine”. The results of combining the two key elements are then

---

19 They usually consist of a series of database index terms relating to study type combined with free text terms describing the methods used in conducting that type of research. See e.g. <http://www.york.ac.uk/inst/crd/intertasc/> for examples of search filters for various study designs.
limited by date, language and type of publication:

1. (Cow OR Cows OR Cattle OR Bovine OR Calf OR Calves OR Bull OR Bulls OR Bullock$ OR Heifer$ OR Livestock)/TI,AB

2. Cattle breeds/DE

3. 1 OR 2

4. (“bluetongue vaccin$” OR BTV OR Bovilis)/TI,AB

5. Bluetongue vaccine/DE

6. 4 OR 5

7. AND 6

8. 1995-2009/PY

9. 7 AND 8

10. 9 AND (English or French or Italian)/LA

11. 10 NOT (editorial or comment or letter)/PT

Key: /TI,AB word in the title and abstract

/DE indexing term

/PY publication year

/LA language

/PT publication types

**Appendix B1(4) - Lumping and splitting**

Where there are many questions it may be more practical to prepare one search (per database) to capture all the questions rather than a series of individual searches (per database) to identify evidence for each question.

Grouping questions into one search is known as “lumping” and treating questions separately is known as “splitting”. The benefits of lumping are that the search is likely to be more sensitive than a series of more focused searches, the searches are less time consuming and fewer duplicate records are produced. The disadvantages of lumping are that a single set of search results is returned which then needs to be assessed for relevance from the perspective of a series of questions. Splitting may be desirable when a very large literature is being searched so that the separate search questions are required to produce manageable numbers of records to assess.

**Appendix B2 - Testing and translating the search strategy**

The draft search strategy can be tested in one database by assessing whether it retrieves papers that are already known to the team and are present in the database. A sample of the results can be examined to identify additional search terms or to highlight potential limitations. Testing may be repeated several times until a strategy can be agreed. The final strategy should be peer reviewed by a colleague to check for spelling mistakes, incorrect set combination, etc (CRD, 2009).
The agreed generic search strategy will need to be “translated” to run efficiently on all the bibliographic databases that have been selected. Translating a strategy focuses on any necessary changes to the search terms and the search syntax:

- Free text terms can usually be re-used in other databases, whereas subject indexing differs and needs to be identified for each database. Databases may offer more than one indexing scheme.

- A different database search interface may require changes to the search syntax such as different truncation symbols, wildcard symbols, proximity operators, field names etc.

If a database has no abstracts, the translation may involve making the search even more sensitive to compensate for the fact that there is only the title to search. To compensate for reduced information to search, more synonyms and additional broader terms may be required. Also, fewer key elements may be required because it may be too stringent to require that all are mentioned in a title.

Information sources with simple web interfaces offer a restricted range of search options. For example, there may be only a single search entry line (as with Google) or restricted options to combine terms and key elements. In this situation strategies may need to be translated in a pragmatic way and made simpler. If Boolean operators are not available, an alternative approach may be to run a number of separate searches.

**Appendix B3 - Searching a range of different information sources**

Identifying research evidence requires searching a range of sources, whose choice will depend on the review topic. In addition to searching electronic bibliographic databases, other approaches to identifying research evidence may be used. These can involve searching a range of sources and adopting a range of search approaches. Evidence from other fields (Hopewell et al., 2007a) indicates that identifying published and unpublished research reports outside of large bibliographic databases may yield additional studies which can mitigate the effects of publication bias.

Examples of information sources are represented by:

1. journals and books recorded in electronic bibliographic databases;
2. full text journals;
3. journal tables of contents;
4. grey literature;
5. reference lists;
6. citations;
7. websites;
8. ongoing (and recently completed) research and research results registers;
9. relevant research centres and experts.

1. **Journals and books recorded in electronic bibliographic databases**

A range of large general bibliographic databases are available to retrieve records of journal articles and books such as Web of Science, Current Contents and CAB. The selection of general and specialist databases (e.g. IUCLID\(^{20}\)) to search will depend upon the review topic and will build on the knowledge of information specialists. There is no agreed standard for what constitutes an acceptable

search in terms of the number of databases searched.

2. **Full text journals**

Searching or handsearching electronic or paper full text journals may identify very recent publications that have not yet been indexed in electronic bibliographic databases or which may never be indexed by electronic databases, as well as journal articles that have been missed by database searches. Searching involves using the search engine provided by the publication. Handsearching involves scanning the publication cover to cover.

The choice of journals to (hand)search may be made by analysing the results of database searches, to identify the journals that contain the largest number of relevant studies.

3. **Journal tables of contents**

Looking at tables of contents can retrieve highly current information and may be used for current awareness during the life of the research project. Tables of contents may be searched using a search interface or handsearched (as described above). Tables of contents can often be delivered direct to e-mail accounts or via other electronic alerting methods (e.g. RSS feeds).

4. **Searching the grey literature**

- Collections of reports and working papers such as OAISTER (via WorldCat21) offer access to millions of reports, dissertations and theses in all subjects.

- Dissertations and theses can be identified from databases such as the Index to Theses and Proquest Dissertations and Theses. Less systematically, dissertations may also be identified from internet searches.

- Conference abstracts or proceedings can be identified from databases such as Web of Science Conference Proceedings Citation Index. Conference abstracts can also be identified by searching the websites of specific conferences.

5. **Checking reference lists**

Scanning the reference lists at the end of relevant publications including reviews and guidelines can identify studies which have not been retrieved through other methods.

6. **Searching citations**

Citation searching is searching for publications which have cited key papers included in the systematic review. This approach can identify related, and sometimes highly relevant, papers. Resources offering citation searches include the Science Citation Index (via Web of Knowledge), Google Scholar and MEDLINE.

7. **Web searching**

This may involve several approaches including the following:

- using specialist internet search engines, e.g. Intute22;

- Google Scholar23;

---

21 <http://www.worldcat.org/>
22 <http://www.intute.ac.uk/>
• searching Google using site limits (e.g. search UK universities by using “site:ac.uk” or Australian universities by using “site:edu.au” in the same search line as the subject search). Advanced Google24 search concepts may be helpful;

• searching websites of relevant organisations (e.g. funders, research organisations, commercial, stakeholders, or other agencies) and professional networks.

8. **Searching for ongoing (and recently completed) research and research results registers**

Identifying project registers specific to the search question can assist with identifying grey literature. Useful databases might include: CORDIS (Europe)25, the US Agricultural Research Service website26, and the (English) DEFRA project database27.

9. **Contacting experts and inviting contributions**

Contacting relevant research centres, experts and librarians by email may yield additional reports of research. Requests for information should include a list of resources and studies already identified to help experts to respond helpfully. Posting the search process and inclusion and exclusion criteria on a project website, along with bibliographic details of studies may help to encourage feedback.

**Appendix B4 - Documenting and reporting the search process**

**Appendix B4(1) - Documenting the search process**

As each search is conducted, the following details should be recorded by the information specialist:

• The name of the database;

• The dates of the search for each database and the period searched;

• Any language or publication status restrictions;

• The full search strategy (all terms and set combinations) and the numbers of records retrieved.

**Appendix B4(2) - Reporting the search process**

1. **Reporting the search process in the Methods section:**

• List all resources (databases, grey literature, hand searched journals etc) searched.

• The dates of the last search for each database and the period searched.

• Note any language or publication status restrictions.

• List individuals or organizations contacted.

• List any other sources searched (e.g. reference lists, the internet).

2. **Reporting the search process in the Results section:**

---

23 <http://scholar.google.com/>
24 <http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=136861>
26 <http://www.ars.usda.gov/research/projects.htm>
27 <http://randd.defra.gov.uk/>
The number of records retrieved by each search strategy should be included in the Results section and described using text and flow charts, for example:

The searches identified 5000 records. After deduplication and the removal of obviously irrelevant records 400 records remained for assessment. After assessment for relevance 100 reports were deemed likely to be relevant and were assessed in detail. A total of 23 records proved to be relevant studies and were extracted.

3. **Reporting the search process in an Appendix:**

The full search strategies including all the search terms (as noted in Documenting the search process above) are typically reported in an Appendix, so that the Methods section can be kept as concise as possible.
Appendix C - Study selection process - Example

As an example of the study selection process, Figure 8 illustrates the sequential workflow for selecting the full texts (phase 2 of the study selection process) in the EFSA opinion on diagnostic tests for *Brucella suis* (EFSA, 2009c). The workflow was organised using Microsoft®-Word form templates, read out as text files and processed and analysed using R software (R Development Core Team, 2009) and code generated for this purpose (available on request from the authors of the report).

The first reviewer collected data. Any non-plausible entries were reported back (FR1) while plausible data sheets were forwarded by the study centre (SC) to a second reviewer. After passing a further plausibility test (FR2), the identity of the reviewers was disclosed and the second reviewer was responsible for resolution of any disagreements. Final approval resulted in acceptance into the study data base (DB) or could have led to rejection of studies (actually, no studies were rejected in this particular example).

**Figure 8**: Sequential workflow for conducting the literature review based on full studies (stage 2) with random allocation of a first and second reviewer to each paper (from EFSA, 2009c)

The review process was organised in two stages involving six reviewers who were also members of the working group. Each paper was allocated randomly to two reviewers. In the first stage, only the title and abstract were used to select each article for further review. Papers were only excluded where both reviewers independently voted for exclusion. In the second stage (illustrated in Figure 8), the papers were reviewed sequentially. The reviewer randomly allocated as “first reviewer” for a given paper completed the review and completed a data collection form. The data collected included the bibliographic information, information on the diagnostic tests evaluated, reference populations used and study results as well as inclusion/exclusion codes for the studies and workflow checkboxes. The completed data collection form was sent to the allocated “second reviewer”, who checked all data entries and discussed and resolved any disagreements with the first reviewer. In cases of unresolved disagreements (which did not occur in this example) the working group would have been responsible for making the final decision.
### Appendix D - Data collection form - Example

<table>
<thead>
<tr>
<th>(5) RefID</th>
<th>Enter Reference identification number (integer between 1 and 1002)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2) ExCode</td>
<td>Code applicable to the whole paper (exact 1 box per line must be ticked for each valid test).</td>
</tr>
<tr>
<td>(3) Of interest for</td>
<td>Tick-the-box if paper is relevant for another chapter of the report. Enter chapter for which paper is of interest (separately by serological, diagnostic, epidemiological, econometric, etc). Note: this box is not to use; note that publication date is no exclusion criterion</td>
</tr>
<tr>
<td>(4) Multiple tests</td>
<td>Tick-the-box if paper reports cross-calculated results of 2 or more relevant tests in 2 or more populations or in latent class analysis was used to estimate Se and Sp of tests. Box may be ticked if Se and Sp estimates are excluded.</td>
</tr>
<tr>
<td>(5) Name of test</td>
<td>Enter clear and brief name of test (max. 12 letters). Repeat for up to 5 tests. Use new sheet for more than 5 tests.</td>
</tr>
<tr>
<td>(6) Test principle</td>
<td>Tick-the-box for the appropriate test principle (exactly 1 box per line must be ticked for each valid test).</td>
</tr>
<tr>
<td>(7) standardised</td>
<td>Tick-the-box if antigen and whole method is standardised according to OIE, EU or other (applies to CFT, SAT, ELISA, RBT).</td>
</tr>
<tr>
<td>(8) cytoto</td>
<td>Tick-the-box if cytotoxic fraction used as test antigen</td>
</tr>
<tr>
<td>(9) Other</td>
<td>Tick-the-box if other monochrome proteins were used as test antigen</td>
</tr>
<tr>
<td>(10) Other antigen</td>
<td>Tick-the-box if any other antigen including whole cell, any LPS preparation, OPS</td>
</tr>
<tr>
<td>(11) Cost</td>
<td>Tick-the-box if reagents or lab services commercially available</td>
</tr>
<tr>
<td>(12) ExCode</td>
<td>Exclusion code applicable to one particular test. If test is not relevant (add brief justification).</td>
</tr>
<tr>
<td>(13) Short name</td>
<td>Enter clear and brief name of reference population (RefPop) as used in the paper (max. 15 letters). Use new sheet for more than 5.</td>
</tr>
<tr>
<td>(14) ExCode</td>
<td>Exclusion code applicable to one particular RefPop.</td>
</tr>
<tr>
<td>(15) Parameter</td>
<td>Se: Tick-the-box if Brucella infection is considered present in the population and sample used is estimated sensitivity.</td>
</tr>
<tr>
<td>(16) Study type</td>
<td>Ep: (epidemiological study); Tick-the-box if the sample was selected from a natural population of domestic pigs.</td>
</tr>
<tr>
<td>(17) Infection status</td>
<td>Tick-the-box if the sample was taken from animals experimentally infected with B. suis</td>
</tr>
<tr>
<td>(18) Area of origin</td>
<td>EU: Tick-the-box if sample is from any of the EU Member States.</td>
</tr>
<tr>
<td>(19) Gold standard</td>
<td>gold standard is inappropriate (e.g. new test is part of the “gold standard”).</td>
</tr>
<tr>
<td>(20) Confirmation</td>
<td>Tick-the-box if at least some animals in the sample were diagnosed with Brucella abortus</td>
</tr>
<tr>
<td>(21) Yersinia status</td>
<td>Yersinia enterocolitica is not estimated or status is unknown</td>
</tr>
<tr>
<td>(22) Days post inf</td>
<td>Enter mean number of days post infection (dpi) if experimental study was used.</td>
</tr>
<tr>
<td>(23) ExCode</td>
<td>Exclusion code applicable to one particular estimation of Se or Sp.</td>
</tr>
<tr>
<td>(24) Page/Table</td>
<td>Enter page(s) or Table(s) number from where the result has been read.</td>
</tr>
<tr>
<td>(25) Correct results</td>
<td>Enter number of correct positive (negative) test results for estimation of Se (Sp). This field may not be empty.</td>
</tr>
<tr>
<td>(26) Inconclusive</td>
<td>Enter number of non-positive non-negative test results (intermediate zone).</td>
</tr>
<tr>
<td>(27) Sample size</td>
<td>Enter number of infected (non-infected) animals used for estimation of Se (Sp). This field may not be empty.</td>
</tr>
<tr>
<td>(28) Parameter</td>
<td>Enter parameter value Se or Sp with one decimal; eg. 95.5</td>
</tr>
<tr>
<td>(29) Comment</td>
<td>Enter comments about potential biases, etc. This data field will not be used in analysis.</td>
</tr>
</tbody>
</table>

---

**Figure 9:** Data collection form for systematic review of diagnostic tests for *Brucella suis* (EFSA, 2009c)
Appendix E - Meta-analysis: Further information

Appendix E1 - Introduction

Many systematic reviews contain statistical syntheses (meta-analyses) of the results of independent studies. By combining information from multiple studies, such syntheses of evidence can provide more precise estimates of the quantities of interest than those derived from the individual studies included within a review. They also facilitate investigations of the consistency of evidence across studies, allow the exploration of differences across studies and can sometimes answer questions not addressed by the individual studies.

The term “meta-analysis” can be interpreted in different ways. Some people use it to refer to any kind of synthesis of evidence from multiple studies. In this context it is not necessary for all studies to be of a similar design or to have addressed the same substantive question. Many methods can be used for the analysis, depending on the nature of the studies, ranging from a simple weighted average of estimates from each study to complex syntheses of diverse evidence using multi-level models, covariates and interactions. The methods used for the synthesis would depend on the types of question being addressed (e.g. about the effects of an intervention, or about prevalence), the study designs (e.g. randomised experiments or cross-sectional studies) and the availability of data (e.g. summary data from published reports, or complete original data from the studies).

In contrast, other people use the term “meta-analysis” to refer to the special case of the combination of multiple estimates of the same underlying quantity or quantities from studies with similar designs (Glass, 1976). For this narrower interpretation, the basic principles of the analysis are as follows.

1. Respect within-study comparisons where possible. In essence, each study is analysed separately and summary measures are combined (e.g. averaged) across studies.

2. Combine results in a way that gives more precise studies more weight in the analysis. The estimates are usually weighted by the inverse of their variance. Indeed the simplest meta-analysis method, known as a fixed-effects meta-analysis is simply a weighted average of estimates (of the same thing) from each study, using the inverses of the variances of the estimates as weights.

3. Allow for the possibility that studies will not be estimating the same underlying quantities, for example through incorporation of study-level covariates (a method known as meta-regression) or allowing for inter-study variation using a random-effects model. A simple random-effects meta-analysis can be performed using an inverse-variance weighted average, in which with the variances include a component for inter-study variation in addition to the sampling (or error) variance of the estimate.

The third point refers to heterogeneity, which can arise from many sources such as variation in study populations, environmental conditions or analytical techniques in different studies. Statistical tests for heterogeneity have commonly been used, but they are problematic and should be interpreted very cautiously as they are too weak to detect heterogeneity unless there are many studies. It is preferable to try and quantify heterogeneity than to test for its presence.

Meta-regression methods can be useful to explore whether study-level characteristics explain variation in quantities across studies, although they can also be prone to low power as well as high false-positive rates. Furthermore, aggregation of individual-level covariates at the study level (e.g. regional averages or age group averages) can be problematic, since inter-study relationships may not reflect within-study relationships, resulting in a severe fallacy problem in the interpretation (Thompson and Higgins, 2002).

In addition to these three principles, attention should be paid to the potential limitations of the studies included (study quality) and to the potential for the data to be subject to reporting biases such as...
publication bias. All these considerations are relevant also in more complex syntheses. When all data for individual subjects are available, meta-analyses may be performed by extending methods used for analysing individual studies. The principles described above should be followed.

Appendix E2 - Further issues

The following are some special issues in meta-analysis that may arise in the food and feed safety area.

1. The structure of variability and clustering of data is an important aspect to consider in meta-analysis. Indeed, the list of potential covariates affecting the study results can be vast (including age, gender, body weight). Such covariates may explain a part of the inter-individual variability or the inter-study variability. The variability structure present in human or environmental data is often more complex at the large scales at which food and feed safety questions usually apply (including variations between population ethnic subgroups or temporal and spatial variations in variables). The data are often hierarchically structured, with various levels of clusters (e.g. country, slaughterhouse, etc). Some of this complexity may or may not be captured and handled by appropriate study designs. When the data are available, hierarchical models can often help to account for such complexity, which is an important aspect to be evaluated in risk assessment.

2. In particular, time and spatial variations of the collected data are often of particular relevance and importance in food and feed safety questions. Cross-sectional studies can be problematic regarding time variations such as seasonal or periodic effects or time trends. For example, in the case of hazard characterisation with long-term effects, exposure at the time of the study may not reflect the long-term exposure. Similarly, cohort studies may not represent well the spatial variation. As a result, careful data selection is necessary prior to any analysis. The level of precision and accuracy expected from food and feed questions should be commensurate with the relevant temporal and spatial scales investigated (National Research Council of the National Academies, 2009).

3. The exposure groups are often unbalanced with respect to influential factors, especially in the case of observational studies. Therefore, the data collected are often subject to confounding factors that may interfere with the outcome variable of interest. For example, in the case of dose-response assessments, socioeconomic factors may be confounded with exposure when high exposure is correlated with poorer living conditions, which may also increase the risk for ill health. This difficulty may be addressed by ensuring that the confounding factors are included as covariates in a meta-regression model.

4. Sometimes there may be an absence of desirable comparators. For example, in the case of exposure assessment, it may be that zero exposure data are not available. The response at zero exposure might have to be estimated by low-dose extrapolation based on a dose-response model, resulting in uncertain (model-dependent) estimates of the response in an unexposed population, and hence the outcome of the analysis is likely to be strongly model-dependent. Such weaknesses of the analysis need to be acknowledged and possibly quantified (e.g. using sensitivity analyses).

5. Missing or censored data are a typical problem for statistical analyses. They can be important in meta-analysis, especially for non-randomised studies and for national surveys where the proportion of missing or censored data can be above 50 % (e.g. more than 80 % in chemical exposure studies) and where the missingness is usually far from being at random. Such issues arise especially in meta-analyses of data for individual subjects. Statistical approaches have been developed to handle such missing data (e.g. multiple imputation, mixed effect models) and censored data (e.g. using adequate maximum likelihood approaches, or Kaplan Meier estimators). Another critical issue is the heterogeneity between studies or between laboratories associated with the treatment of those missing values (e.g. different limits of detection - LoDs). Such heterogeneity also occurs in meta-analysis of aggregated data and can be handled.
similarly to other common heterogeneity issues (e.g. using appropriate random effects). However, it is often useful for risk managers to compare results using more naive imputations based on worst and best cases scenarios.
### GLOSSARY

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acceptable daily intake (ADI)</strong></td>
<td>Acceptable daily intakes and tolerable daily intakes (ADI and TDI) are health-based guidance values and correspond to a dose or level of exposure below which no deleterious effect is measurable (threshold) and is “without appreciable health risk” when consumed daily over a lifetime (WHO, 1987). ADIs are derived for chemicals intentionally added to food or raw commodities and TDIs for chemical contaminants which are undesirable in food and feed. Food additives, flavourings and food-contact materials constitute the largest classes of chemicals that are intentionally added to food, whereas chemicals resulting from intentional treatment of raw commodities include mostly pesticides or biocides (e.g., herbicides, fungicides and insecticides) and veterinary residues (e.g. antibiotics).</td>
</tr>
<tr>
<td><strong>Attrition bias</strong></td>
<td>Systematic differences between groups in withdrawals from a study.</td>
</tr>
<tr>
<td><strong>Benchmark dose (BMD)</strong></td>
<td>The statistical lower bound on a dose corresponding to a specified level of risk (Allen et al., 1994).</td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>A systematic error or deviation from the truth, in results or inferences. A common classification scheme for bias includes selection bias, performance bias, attrition bias, detection bias and reporting bias.</td>
</tr>
<tr>
<td><strong>Bibliographic databases</strong></td>
<td>Databases that provide descriptive records of items such as books and articles.</td>
</tr>
<tr>
<td><strong>Bibliographic software</strong></td>
<td>Computer software that assists with the organisation of bibliographic references. There are many different software programs (e.g. EndNote, Reference Manager, Procite); most allow for the import of references from bibliographic databases and the automated production of reference lists.</td>
</tr>
<tr>
<td><strong>Boolean operator</strong></td>
<td>Boolean operators are used to combine terms when conducting electronic searches. Examples include “AND” (used to narrow a search), “OR” (used to broaden a search) and “NOT” (used to exclude terms from a search).</td>
</tr>
<tr>
<td><strong>Broad policy problem</strong></td>
<td>Broad policy problem: a broad question (e.g. a risk assessment model or the analytical framework developed for assessing a broad issue) that could be refined into more specific questions (as opposite to “Specific question”).</td>
</tr>
<tr>
<td><strong>Case series</strong></td>
<td>A study reporting observations on a series of study subjects, usually all receiving the same intervention, with no control group.</td>
</tr>
<tr>
<td><strong>Case-control study</strong></td>
<td>An observational study that compares subjects with a specific disease or outcome of interest (cases) with a suitable control group of subjects without that disease or outcome, and which seeks to find associations between the outcome and prior exposure to particular risk factors.</td>
</tr>
</tbody>
</table>
| **Closed-framed question**   | A specific question with a well-formulated structure, presenting all relevant key elements for the application of systematic review methods. To answer a closed-framed
question it is usually possible to envisage a primary research study design (which may or may not be feasible or ethical) (see also “Key elements of a question” and “Open-framed” question).

**Cohort study**

An observational study in which a defined group of subjects is observed over time and a comparison made between those who did and those who did not receive an intervention.

**Comparator(s) (C).**

A reference scenario against which the intervention or exposure can be compared (see “Intervention” and “Exposure”)

**Confounding**

A situation in which a measure of the effect is distorted because of an association between the intervention (or exposure) with other factor(s) that influence the outcome under investigation.

**Cross-sectional study**

A study that examines the relationship between diseases (or other health related characteristics) and other variables of interest as they exist in a defined population at a particular time.

**Detection bias**

A potential artefact in the assessment of outcomes caused, for example, by the use of a particular diagnostic technique or type of equipment.

**Exposure(s)**

In the context of SR, it is the factor(s) to which the population is exposed (e.g. a chemical or pathogen in food or feed, introduced invasive species, or a harmful impact on the environment).

**Estimate of effect**

The observed relationship between an intervention and an outcome expressed as, for example, odds ratio, risk difference, risk ratio, hazard ratio, standardised mean difference, weighted mean difference, or number needed to treat.

**Evidence**

Everything that can be used to independently determine or demonstrate the truth of an assertion. Scientific evidence is evidence which serves to either support or counter a scientific theory or hypothesis.

**External validity**

The degree to which the results of a study hold true in other settings (generalisability).

**Free text terms**

In literature searching, the use of everyday words and phrases, as opposed to index terms, to search bibliographic databases.

**Generalisability**

See “External validity”.

**Grey literature**

Types of publication which are less systematically recorded in bibliographic tools such as catalogues and databases than journals and books.

**Handsearching**

The process of searching a journal page by page to identify relevant articles.

**Heterogeneity**

In systematic reviews heterogeneity refers to variability or differences between studies.
<table>
<thead>
<tr>
<th><strong>Index test(s)</strong></th>
<th>The test(s) whose performance is being evaluated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal validity</strong></td>
<td>The degree to which a result of a study is likely to be true and free of bias (systematic errors), and hence the degree to which inferences drawn from the study are likely to be justified.</td>
</tr>
<tr>
<td><strong>Intervention(s)</strong></td>
<td>The factor(s) to which the population is exposed (e.g. an additive in food or feed, a vaccine, or a disinfection or eradication method).</td>
</tr>
<tr>
<td><strong>Kappa statistic</strong></td>
<td>A measure of inter-reviewer agreement.</td>
</tr>
<tr>
<td><strong>Key elements of a question</strong></td>
<td>Elements of a review question that, if well defined, help to answer it (e.g. selecting the eligibility criteria for studies, developing the search strategy, selecting the studies, or collecting the data). The key elements vary depending on the question type. For questions about effects of an intervention or exposure, the key elements are the population (P), the intervention (I) or exposure (E), the comparator (C) and the outcome (O) (together represented as PICO or PECO). For test accuracy questions, the key elements are the population (P), the index test (I) and the target condition (T) (together PIT). For descriptive questions (prevalence, incidence, occurrence and consumption), the key elements are the population (P) and the condition of interest (O) (together PO).</td>
</tr>
<tr>
<td><strong>Limit of detection (LoD)</strong></td>
<td>The LOD represents the minimum concentration or mass of an analyte that can be detected with a given confidence for a given analytical procedure. More formally, the LOD can be defined as the lowest concentration level that can be determined to be statistically different from a blank, customarily using confidence levels equal to 95 % or 99 % (EFSA, 2010).</td>
</tr>
<tr>
<td><strong>Lowest observed adverse effect level (LOAEL)</strong></td>
<td>See “No observed adverse effect level” (NOAEL).</td>
</tr>
<tr>
<td><strong>Margin of exposure (MOE)</strong></td>
<td>The Margin of exposure approach is applied for the risk characterisation of compounds which are both genotoxic and carcinogenic and uses a reference point, often taken from an animal study and corresponding to a dose that causes a low but measurable response in animals. This reference point is then compared with various dietary intake estimates in humans, taking into account differences in consumption patterns (EFSA, 2005).</td>
</tr>
<tr>
<td><strong>Meta-analysis</strong></td>
<td>The process of synthesising research results from a number of independent studies (published or unpublished) by using statistical methods to combine results from previous separate but related studies, in order to determine overall trends and significance.</td>
</tr>
<tr>
<td><strong>No observed adverse effect level (NOAEL)</strong></td>
<td>The highest dose tested without observation of an adverse effect in the particular experiment. The numerical value of the NOAEL is thus dependent upon the selection of dose levels when the study was designed and on the ability of the study to detect adverse effects. Since studies with low power...</td>
</tr>
</tbody>
</table>
Systematic review methodology and food and feed safety risk assessment

(e.g. small group sizes) and/or insensitive methods are able to detect only relatively large effects, these tend to result in higher NOAELs. If there is a significant effect at all dose levels, the lowest dose used in the study may be set as the lowest-observed-adverse-effect-level (LOAEL) (EFSA, 2009d).

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observational study</td>
<td>A study in which the investigators observe and measure but do not seek to intervene.</td>
</tr>
<tr>
<td>Open-framed question</td>
<td>A specific question in which not all key elements are specified. It is rarely possible to envisage a primary research study design to answer an open-framed question. Open-framed questions might not be suitable for systematic review.</td>
</tr>
<tr>
<td>Outcome(s)</td>
<td>Variable(s) for which data are collected to enable the questions of the systematic review to be answered</td>
</tr>
<tr>
<td>PECO</td>
<td>Acronym summarising the population (P), exposure (E), comparator (C) and outcome (O) in a question about an exposure effect (see also “Key elements of a question”).</td>
</tr>
<tr>
<td>Performance bias</td>
<td>It occurs when study subjects are misclassified as being exposed when in fact they were not (or vice versa), or treated when in fact they did not receive treatment, or if those who are exposed or treated receive additional undocumented exposures or treatments.</td>
</tr>
<tr>
<td>PICO</td>
<td>Acronym summarising the population (P), intervention (I), comparator (C) and outcome (O) in a question about an intervention effect (see also “Key elements of a question”).</td>
</tr>
<tr>
<td>PIT</td>
<td>Acronym summarising the population (P), index test (I), and target population (T) in a question about test accuracy (see also “Key elements of a question”).</td>
</tr>
<tr>
<td>PO</td>
<td>Acronym summarising the population (P) and outcome (O) in a descriptive question (see also “Key elements of a question”).</td>
</tr>
<tr>
<td>Primary study</td>
<td>The original study in which data were collected. The term is sometimes used to distinguish such studies from secondary studies that re-examine previously collected data (e.g. a review).</td>
</tr>
<tr>
<td>Publication bias</td>
<td>Bias arising when the publication or non-publication of research findings is dependent on the nature and direction of the results.</td>
</tr>
<tr>
<td>Quality of a body of evidence</td>
<td>The extent to which one can be confident that an estimate of effect or association is close to the quantity of specific interest.</td>
</tr>
<tr>
<td>Question formulation</td>
<td>The process of identifying the question type, the key elements of a specific question and the best study design necessary for answering it (see also “Key elements of a question”).</td>
</tr>
<tr>
<td>Question type</td>
<td>The category of question that the review seeks to answer (e.g. effects of intervention or exposure; test accuracy; or descriptive questions such as prevalence, incidence,</td>
</tr>
</tbody>
</table>
Randomised controlled trial(s) (RCT(s)) Experiment(s) in which investigators use randomisation to allocate study subjects into the groups that are being compared.

Reporting bias Bias arising when the dissemination of research findings is influenced by the nature and direction of results.

Scoping the literature An approach to searching the literature that can be used to support decisions about whether it is possible or worthwhile proceeding with a systematic review; to assess the volume of research which may need to be processed and guide the estimate of resources required; and to inform the development of the review protocol. The rigour with which this process is undertaken may vary.

Search strategy The exact terms and their combinations used to search a bibliographic database.

Selection bias Bias arising from systematic differences in baseline characteristics between the groups that are compared.

Sensitivity analysis An analysis used to test the robustness of findings and determine how sensitive results are to the data that were included and/or the way that analyses were done.

Simplifying a complex problem The process of breaking down a complex problem into all possible specific questions.

Specific question A question which is sufficiently well structured that it could be answered in a primary study without needing to be further broken down (as opposite to a “Broad policy problem”).

Structure of the question Composition of the question, given by the key elements (see also “Key elements of a question”).

Study design A protocol for conducting a study, which allows the investigator to translate the conceptual hypothesis into an operational one (e.g. Cohort study, Case-control study, Cross-sectional study, Randomised controlled trial, Non-randomised controlled trial, Case series, Case study).

Tolerable daily intake (TDI) See Acceptable daily intake (ADI).

**ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADI</td>
<td>Acceptable daily intake</td>
</tr>
<tr>
<td>BMD</td>
<td>Benchmark dose</td>
</tr>
<tr>
<td>EFSA</td>
<td>European Food Safety Authority</td>
</tr>
<tr>
<td>EHD</td>
<td>Epizootic hemorrhagic disease</td>
</tr>
<tr>
<td>EHDV</td>
<td>Epizootic hemorrhagic disease virus</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>LOAEL</td>
<td>Lowest observed adverse effect level</td>
</tr>
<tr>
<td>LoDs</td>
<td>Limits of detection</td>
</tr>
<tr>
<td>MOA</td>
<td>Mode of action</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>MOE</td>
<td>Margin of exposure</td>
</tr>
<tr>
<td>NOAEL</td>
<td>No observed adverse effect level</td>
</tr>
<tr>
<td>OIE</td>
<td>Office international des épizooties (World Organisation for Animal Health)</td>
</tr>
<tr>
<td>PECO</td>
<td>Population, Exposure, Outcome, Comparator</td>
</tr>
<tr>
<td>PICO</td>
<td>Population, Intervention, Outcome, Comparator</td>
</tr>
<tr>
<td>PIT</td>
<td>Population, Index test, Target condition</td>
</tr>
<tr>
<td>PO</td>
<td>Population, Outcome</td>
</tr>
<tr>
<td>RA</td>
<td>Risk Assessment</td>
</tr>
<tr>
<td>SR</td>
<td>Systematic review</td>
</tr>
<tr>
<td>TDI</td>
<td>Tolerable daily intake</td>
</tr>
</tbody>
</table>