TECHNICAL REPORT

APPROVED: 26 June 2015 PUBLISHED: 01 July 2015

Tools for critically appraising different study designs, systematic review and literature searches

European Food Safety Authority

Abstract

A Critical Appraisal Tool (CAT) allows the methodological quality of a study/process to be assessed, which, in turn, influences the reliability of the evidence produced by such a study/process. The European Food Safety Authority (EFSA) Science Strategy 2012–2016 emphasises the importance of strengthening the scientific basis for risk assessment and monitoring. Under this framework, EFSA started the PROMETHEUS project (PROmoting METHods for Evidence Use in Scientific assessments), aiming to further improve the methods for "dealing with data and evidence" in EFSA scientific assessments and to increase their consistency. CATs providing structured and consistent guidance on assessing the methodological quality of a study/process play a key role in this context. This report provides a series of CATs containing a comprehensive list of items to appraise the following: (i) systematic reviews of intervention studies; (ii) randomised controlled trials in humans; (iii) genetically modified plant equivalence studies; and (iv) extensive literature searches. They focus on the risk of bias/appropriateness of the design /conduct of the item under consideration and not on how the item has been reported. The report does not provide detailed guidance on how the items should be appraised. This may be considered by EFSA in a second step, together with the definition of approaches for the prioritisation of the items under assessment.

© European Food Safety Authority, 2015

Key words: critical appraisal tool, systematic review, randomised controlled trial, genetically modified plant, equivalence, extensive literature search

Requestor: EFSA

Question number: EFSA-Q-2014-00907

Correspondence: amu@efsa.europa.eu

Acknowledgements: EFSA wishes to thank the members of the AMU Critical Appraisal Tools task force, Elisa Aiassa, Fulvio Barizzone, Hermann Broll, Laura Ciccolallo, Andrea Gervelmeyer, Laura Martino, Camilla Smeraldi, Jordi Tarres-Call and Silvia Valtueña Martínez, for the support provided to this scientific output. In addition, EFSA wishes to thank Lara Congiu and Julie Glanville for the support provided in developing the Critical Appraisal Tool on extensive literature searches.

Suggested citation: European Food Safety Authority, 2015. Tools for critically appraising different study designs, systematic review and literature searches. EFSA supporting publication 2015:EN-836. 65 pp.

© European Food Safety Authority, 2015

Reproduction is authorised provided the source is acknowledged.

Summary

A Critical Appraisal Tool (CAT) allows the methodological quality of a study/process to be assessed, which, in turn, influences the reliability of the evidence produced by such a study/process. CATs help to minimise subjectivity in the appraisal and maximise transparency. Nevertheless, they always require expert judgement in their use.

Regulation (EC) No 178/2002 recommends that assessments undertaken by the European Food Safety Authority (EFSA) be conducted in an independent, objective and transparent manner, on the basis of all available scientific information and data. Moreover, the EFSA Science Strategy 2012–2016 emphasises the importance of strengthening the scientific basis for risk assessment and monitoring. Under this framework, EFSA started the PROMETHEUS project (PROmoting METHods for Evidence Use in Scientific assessments) (2014–2016), aiming to further improve the methods for "dealing with data and evidence" (i.e. collecting/extracting, validating/appraising, analysing and integrating data and evidence) in EFSA scientific assessments and to increase their consistency. In this context, CATs providing structured and consistent guidance on assessing the methodological quality of a study/process play a key role.

In the scientific community, there are various on-going initiatives on CATs. However, many of the available tools do not necessarily have a straightforward application in relation to the needs of EFSA (e.g. they have been developed for clinical trials or are specific for only a subset of studies relevant to EFSA). In order to address EFSA's needs on food safety assessment, it was decided that a series of CATs, containing a comprehensive list of items, should be provided to appraise the following: (i) systematic reviews of interventions studies (SR CAT); (ii) randomised controlled trials in humans (RCT CAT); (iii) genetically modified plant equivalence studies (GMO CAT); and (iv) extensive literature searches (ELS CAT).

Detailed reporting of a study/process is important for assessing its methodological quality. In fact, a lack of transparent and relevant information can lead to (i) delays in the appraisal process if it is decided that clarification is required from the authors and/or (ii) increased uncertainty in the assessment when clarification is not asked for/received. However, the quality of the reporting has previously been addressed by the EFSA Guidance on Statistical Reporting (2014) and it is outside the scope of the current document.

The CATs are provided as appendices to the main document and should be applied by outcome or endpoint to each individual study/process included in the assessment. This allows a consistent classification of the studies/processes according to their methodological quality. Ideally, they should be used by a multidisciplinary team of experts that includes methodologists (e.g. experts in information science, epidemiology, statistics) and domain experts in the field under assessment. The CATs are intended to be relevant for various users in the EFSA context (e.g. EFSA staff, experts, applicants, contractors, Member States). Where appropriate, they could be used as a reference for developing tailor-made CATs for specific fields, being aware that such modifications could hamper comparability across assessments. They have to be considered as working documents and advances in empirical methodological research will be reflected in further improvements to these instruments. Additional practice and studies are needed with a focus on the reproducibility and construct validity of these tools.

The report does not provide detailed guidance for applying the CATs. This may be considered by EFSA in a second step, together with the definition of approaches for the prioritisation of the items under assessment.

Table of contents

1. Introduction

1.1. Background as provided by EFSA

Regulation (EC) No 178/2002[1] recommends that assessments undertaken by the European Food Safety Authority (EFSA) shall be conducted in an independent, objective and transparent manner, on the basis of all available scientific information and data. Moreover, the EFSA Science Strategy 2012–2016[2] emphasises the importance of strengthening the scientific basis for risk assessment and risk monitoring. Under this framework, EFSA started the PROMETHEUS project (PROmoting METHods for Evidence Use in Scientific assessments) (2014–2016), which aims to further improve the methods for "dealing with data and evidence" (i.e. collecting/extracting, validating/appraising, analysing and integrating data and evidence) in EFSA scientific assessments and to increase their consistency. In this context, Critical Appraisal Tools (CATs) providing structured and consistent guidance on how to assess the methodological quality of a study play a key role.

In the scientific community there are various on-going initiatives on CATs.[3] However, many of the available tools do not necessarily have a straightforward application for the needs of EFSA (e.g. they are developed for clinical trials or are specific for only a subset of studies relevant for EFSA). Therefore, the Assessment and Methodological support Unit (AMU) developed specific CATs for appraising: i) systematic reviews of interventions (SR), ii) randomised controlled trials (RCT) relevant for food and feed safety assessments; iii) Genetically Modified (GM) plant equivalence studies and iv) the methodological quality of extensive literature searches (ELS). These tools have been used to enhance standardisation, consistency and transparency of AMU appraisals and shall now be streamlined and published in an EFSA Technical Report in order to make them available to all EFSA units and panels.

1.2. Terms of Reference as provided by EFSA

Discuss and streamline the existing AMU CATs on systematic reviews of interventions (SR), randomised controlled trials (RCT), Genetically Modified (GM) plant equivalence studies and extensive literature searches (ELS).

Publish the revised CATs in a Technical Report.

The Technical Report should be practical and applicable to the different relevant food and feed safety fields. In particular, the report should include:

- the list of elements that are considered by AMU for critically appraising SR of interventions, RCT, GM plant equivalence studies and ELS and indications on how the Critical Appraisal Tools are currently used;
- a glossary of relevant terms.

2. Approach followed for developing the document

An internal task force was set up to discuss and streamline the CATs already developed by the AMU.

In addition to AMU staff, the task force was composed of EFSA staff from areas for which these CATs were considered particularly relevant.

With the aim of having an effective discussion, the task force members were assigned some specific CATs to review according to their background. The reviewers considered the comprehensiveness of the items included in the CATs, their clarity and the proposed ways to judge/characterise them. Each CAT was tested in parallel by two members of the task force using two concrete examples. The specific CATs were amended according to the feedback received.

[1] OJ L 31, 1.2.2002, p. 1.
[2] http://www.efsa.europa.eu/en/corporate/pub/sciencestrategy12.htm
[3] For example, SYRCLE's risk of bias tool for animal studies (Hooijmans et al., 2014); the Cochrane Collaboration's tool for assessing risk of bias in randomised trials (Higgins et al., 2011); or the OHAT risk of bias tool (Rooney et al., 2014; OHAT/NTP, 2015).

3. What is a Critical Appraisal Tool (CAT)?

A CAT allows the *methodological quality* of a study (or a process) to be assessed, which influences the *reliability* of the evidence produced by such a study. Reliability of a piece of evidence refers to: (i) *precision*, i.e. the extent to which random error is minimised and the outcome of the process is reproducible over time; and (ii) *accuracy* (also referred to as internal validity), i.e. the extent to which systematic error (bias) is minimised[4] (EFSA, 2015).

Detailed reporting of a study/process is important for assessing its methodological quality. In fact, a lack of transparent and relevant information can lead to (i) delays in the appraisal process if it is decided that clarification is required from the authors and/or (ii) increased uncertainty in the assessment when clarification is not asked for/received. However, the quality of the reporting has been addressed elsewhere by EFSA (EFSA, 2014) and it is not an aspect inherent to methodological quality; thus, it is outside the scope of the current document.

Tools for appraising the methodological quality of studies need to be design specific. For instance, the items to be considered when appraising an RCT are not the same as those to be considered for an observational study (e.g. randomisation). For the same study design, CATs should be applied by outcome or endpoint, because the methodological quality of a study in which multiple outcomes are assessed may differ depending on the outcome considered (Higgins et al., 2011). For instance, some outcomes may be analysed using appropriate methods and some others in the same study may not. CATs should be applied to each individual study included in the assessment to allow for consistent classification of the studies according to their methodological quality.

While helping to minimise subjectivity in the appraisal and maximise transparency, CATs always require domain and methodological expert judgement in their use (see also section 5 below).

4. Objective of this document

This report aims to provide a series of CATs containing a comprehensive list of items (called "appraisal questions") for appraising the following:

- systematic reviews of intervention studies (SR);
- randomised controlled trials (RCTs) in humans;
- genetically modified (GM) plant equivalence studies;
- the process for conducting extensive literature searches (ELS).

For each appraisal question, the focus is on the risk of bias/appropriateness of the design/conduct of the item under consideration and not on how the item was reported (see section 3).

However, the report does not provide detailed guidance on performing the appraisal, which may be considered by EFSA in a second step.

The CATs provided have to be considered as working documents and advances in empirical methodological research will be reflected in further improvements to these instruments. Additional practice and studies are needed with a focus on the reproducibility and construct validity[5] of these tools.

The CATs are provided as appendices to this document.

5. Intended users

Ideally the CATs should be used by a multidisciplinary team of experts that includes methodologists (e.g. experts in information science, epidemiology, statistics) and domain experts in the field of the study/process under assessment.

[4] Risk of bias also addresses aspects such as the sensitivity and specificity of the detection method used in an assessment.
[5] The extent to which scores on a particular instrument relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured (Terwee et al., 2007).

The elements contained in the CATs could be relevant for various users in the EFSA context (e.g. EFSA staff, experts, applicants, contractors, Member States). Where appropriate, they could be used as a reference for developing tailor-made CATs for specific fields, being aware that such modifications could hamper comparability across assessments.

6. General structure of the CATs and legend

The CATs are provided as appendices to this document in a tabular format with five columns and the items are grouped by topic (e.g. methods, sample selection) identified by capital letters.

The Systematic Review, Randomised Controlled Trials and Extensive Literature Search CATs contain the following columns:

1. **#: the number of the item under assessment;**
2. **appraisal question:** the question to be answered to evaluate each individual item included an the CAT. It contains the description of the item to be evaluated, including some examples or a description of how the item should have been implemented in the study;
3. **information as reported:** in this column, information from the study/ELS under assessment is quoted or summarised;
4. **appraisal**: in this column, a concise answer to the appraisal question (see column 2) is provided. For each item considered in the appraisal, the appraisal scales shown in Table 1: or Table 2: will be used. The scales provide different ways to appraise the items under consideration according to their domain of pertinence (risk of bias or appropriateness). As the appraisal of the methodological quality of a study is topic specific, the guidance for judgement for each individual item should be tailored according to the topic of the study under assessment before starting the appraisal process. For instance, unblinded outcome assessors in an RCT would normally be appraised as having a high risk of bias. However, when an outcome is measured using a method that prevents subjectivity in the measurement, a lack of blinding could be judged as probably low risk of bias;

Table 1: Proposed appraisal scale for risk of bias (modified from OHAT/NTP, 2015)

Appraisal	Definition
Definitively low risk of bias	There is direct evidence in the study of low risk of bias practices.
Probably low risk of bias	There is indirect evidence in the study of low risk of bias practices OR it is deemed that deviations from low risk of bias practices for the item to be appraised would not appreciably bias the final results.
Probably high risk of bias	There is indirect evidence of high risk of bias practices OR there is insufficient information (e.g. not reported or "NR") provided about relevant risk of bias practices.
Definitively high risk of bias	There is direct evidence of high risk of bias practices.

Table 2: Proposed appraisal scale for appropriateness

Appraisal	Definition
Definitively appropriate	There is direct evidence in the study of appropriate practices.
Probably appropriate	There is indirect evidence of appropriate practices OR it is deemed that deviations from appropriate practices for the item to be appraised would not appreciably modify the final results.
Probably not appropriate	There is indirect evidence of inappropriate practices OR there is insufficient information (e.g. not reported or "NR") provided about appropriate practices.
Definitively not appropriate	There is direct evidence of inappropriate practices.
Not applicable	An item is not appraisable when a previous practice on which it is dependent was not performed. For instance, one of the questions of the SR CAT concerns the appraisal of the methodological quality of the studies. The subsequent question concerns the process that has been used when appraising the methodological quality (see items E1 and E2 of SR CAT in Appendix A –). If the methodological quality of the studies has not been appraised, the subsequent question (item) becomes not applicable.

5. **rationale for the appraisal**: in this column, the rationale supporting the appraisal is reported.

The GMO CAT contains the following columns:

1. **#**: the number of the item under assessment;
2. **appraisal question**: the question to be answered to evaluate each individual item included in the CAT. It contains the description of the item to be evaluated, including some examples or a description of how the item shall be implemented in the study;
3. **rationale of the appraisal and possible consequences of flaws identified:**
   - information from the review under assessment is quoted or summarised under the sub-heading "Information as provided";
   - the rationale of the assessment and the possible consequences of flaws identified are illustrated under the sub-heading "Explanation of the assessment and possible consequences". Weaknesses and unclear/missing items along with strengths are also described.
4. **answer to the appraisal question**: concise answer to the appraisal question (see column 2):
   - 'yes, fully'. The study is well performed as far as this item is concerned;
   - 'no'. The study is NOT at all well performed as far as this item is concerned;
   - 'partially'. The study is well performed as far as some aspects related to this item are concerned. However, other aspects are not handled appropriately or adequately;
   - 'unclear'. This option is applicable in two cases:
     - the item is poorly reported or not reported at all, making it difficult to assess it (and further information is needed from authors—see "action required" below);

     OR

     - the item seems appropriate from a methodological point of view, but domain expertise is not present in the team and confirmation is needed (see "action required" below).
5. **action required**: the action required, i.e. whether or not it is necessary to consult domain experts and/or the authors of the review under assessment:
   - *none*. Sufficient information is available for performing the assessment of this item;

- *ask confirmation from domain experts*. Sufficient information is available for performing the assessment of this item from a methodological point of view. However, confirmation from domain experts is needed;
- *ask authors for further information*. Insufficient information is available for performing the assessment of this item; therefore, there is a need for clarifications or additional information from the authors of the study.

7. Recommendations

It is recommended that:

- detailed guidelines be developed for applying the CATs;
- approaches be explored and implemented on how to prioritise the items to assess (i.e. according to the specific design and context, some items can be more relevant in terms of risk of bias and/or precision than others and consequently have a higher impact on the overall assessment).

References

EFSA (European Food Safety Authority), 2014. Guidance on Statistical Reporting. EFSA Journal 2014;12(12):3908, 18 pp. doi:10.2903/j.efsa.2014.3908

EFSA (European Food Safety Authority), 2015. Principles and process for dealing with data and evidence in scientific assessments. EFSA Journal 2015;13(5):4121, 36 pp. doi:10.2903/j.efsa.2015.4121

Higgins JPT, Altman DG and Sterne JAC, 2011. Assessing risk of bias in included studies. In: Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011]. Eds Higgins JPT and Green S. The Cochrane Collaboration. Available online: www.cochrane-handbook.org

Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M and Langendam MW, 2014. SYRCLE's risk of bias tool for animal studies. BMC Medical Research Methodology, 14, 43.

OHAT/NTP (Office of Health Assessment and Translation/National Toxicology Program), 2015. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Available online: http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf

Rooney AA, Boyles AL, Wolfe MS, Bucher JR and Thayer KA, 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environmental Health Perspectives, 122, 711–718.

Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM and de Vet HC, 2007. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology, 60, 34–42.

Glossary and Abbreviations

TERM	DEFINITION
Bias	A systematic error or deviation from the truth, in results or inferences
Bibliographic record	An entry representing a specific item in a library catalogue or bibliographic database, containing all the data elements necessary for a full description, presented in a specific bibliographic format. In modern cataloguing, the standard format is machine-readable (*example*: the MARC record), but prior to the use of computers, the traditional format was the catalogue card
Boolean operator	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)
Controlled terms	See Controlled vocabulary
Controlled vocabulary	An established list of preferred terms from which a cataloguer or indexer must select when assigning subject headings or descriptors in a bibliographic record, to indicate, for example, the content of the work in a library catalogue, index, or bibliographic database. Synonyms are included as lead-in vocabulary, with instructions to see or use the authorised heading. For example, if the authorised subject heading for works about dogs is "Dogs," then all items about dogs will be assigned the heading "Dogs," including a work titled All about Canines. A cross-reference to the heading "Dogs" will be made from the term "Canines" to ensure that anyone looking for information about dogs under "Canines" will be directed to the correct heading. Controlled vocabulary is usually listed alphabetically in a subject headings list or thesaurus of indexing terms. The process of creating and maintaining a list of preferred indexing terms is called vocabulary control
Endpoint	See Outcome(s)
Extensive literature search (ELS)	A literature search process structured in a way to identify as many studies relevant to a review question as needed. It is tailored in order to address the trade-off between sensitivity and specificity depending on the context of the review question. The fundamental characteristics of an ELS are: (1) use of tailored search strings, and (2) tailored use of literature sources (i.e. bibliographic databases and other sources accessed via electronic or hand-searching – for example, websites, journal tables of content, theses repositories, etc.)
Free text search	A search of a bibliographic database in which natural language words and phrases appearing in the text of the documents indexed, or in their bibliographic descriptions, are used as search terms, rather than terms selected from a list of controlled vocabulary (authorised subject headings or descriptors)
GMO	An organism or microorganism whose genetic material has been altered by means of genetic engineering. Techniques of genetic engineering to produce a genetically modified organism (GMO) are described in Annex 1 A of Directive 2001/18/EC
Grey literature	Types of publication which are less systematically recorded in bibliographic tools such as catalogues and databases than journals and books
Intervention questions	A question that seeks to assess the effect of an intervention, which is 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)

Key elements of a question	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 question, 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)
Metadata	It means, "data about data." Structured information describing information resources/objects for a variety of purposes
Outcome(s)	Variable(s) for which data are collected to enable the questions of the study to be answered
PECO(S)	Acronym summarising the population (P), exposure (E), comparator (C) and outcome (O) in a question about an exposure effect. S stands for study design. See also "Key elements of a question"
PICO(S)	Acronym summarising the population (P), intervention (I), comparator (C) and outcome (O) in a question about an intervention effect. S stands for study design. See also 'Key elements of a question'
PIT	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".
PO	Acronym summarising the population (P) and outcome (O) in a descriptive question. See also "Key elements of a question"
Primary research study	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).
Record	Summary information about a full-text document or conference presentation, typically included in a bibliographic database, which may include a bibliographic reference and one or more of the following: an abstract or summary of the scientific content, additional categorisations or indexing terms
Reference	A conventional word or phrase used in a work to refer the reader to another part of the text (see above or see below) or a similar word or phrase used in an index, catalogue, or reference work to direct the user from one heading or entry to another (see or see also). Also refers to any Latin phrase used in footnotes, endnotes, and bibliographies to refer the reader to works previously quoted or cited, for example, ibid. and op. cit. Sometimes used synonymously with citation. Also refers to a letter written in support of a person's application for employment or housing, usually by someone familiar with the applicant's qualifications or reputation, or to a person who agrees to be contacted for such a recommendation, usually by telephone.

Appendix A – Critical appraisal tool for assessing quality of Systematic Reviews of intervention studies (SR CAT)

1. Systematic Reviews: main definitions and the EFSA context

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. Statistical methods to synthesise the results of the included studies (meta-analysis) may or may not be used in the process.

The systematic review process consists of at least the following fundamental steps: preparing the review protocol and, in particular, formulating the review question and setting the eligibility criteria for studies; performing extensive literature searches (ELS) to identify relevant studies; selecting studies for inclusion; extracting data from included studies, including exploring sources of (and coping with) heterogeneity; assessing the methodological quality of included studies; analysing and synthesising data from the studies; presenting data and results; and interpreting the results and drawing conclusions. In a systematic review the method applied is clearly reported and documented.

Systematic reviews are performed to answer specified questions aiming to estimate risk assessment parameters such as might be evaluated in well-defined primary research studies. These questions are (i) questions on the effect of an exposure or intervention; (ii) questions on the sensitivity or specificity of a test; or (iii) questions on the prevalence of a condition or the incidence of an outcome).

The key elements of the review question are the components of the question that specify what information must be provided in a primary study to estimate the parameter under assessment and hence answer the question. 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).

This document focuses on how to appraise the methodological quality of systematic reviews of interventions (type (i) above).

As mentioned above, in the context of a systematic review an extensive literature search has to be performed. In order to appraise it the ELS CAT provided in Appendix D –should be applied and an overall judgement on the appropriateness of the extensive literature search performed should be given (see item 4 of the SR CAT).

In the EFSA context, there are three main situations when it is necessary to appraise the methodological quality of a SR to inform the risk assessment process (both generic risk assessments and evaluation of applications for authorisation of products):

- when the results of an already completed SR are used *per se* in the risk assessment;
- when EFSA needs to perform a systematic review to answer a review question and, being SR(s) on the same subject available, EFSA considers updating the existing one(s) after having critically appraised its/their quality;
- when EFSA outsources SRs and needs to appraise the quality of the SRs performed by the external contractors.

2. SR CAT

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
A.	**Review question and eligibility criteria for study selection**			
A1	**Was the review question clearly formulated?** Formulating the review question means properly translating the review question into "PICO" elements: • Population(s); • Intervention(s); • Comparator(s); • Outcome(s) (including adverse effects where relevant). Have the relevant PICO elements been identified and these elements are adequately characterised? For instance, the objective of a systematic review might be related to the impact of Enzootic Bovine Leukosis in dairy herds considering the loss of production due to the infection and the presence of lymphomas. In this case the review question could be defined in this way: "In dairy cattle raised for the production of milk in modern production systems or in university farms, including any size system, with or without confinement housing systems (P), what is the impact of infection with Enzootic Bovine Leukosis virus as measured by AGID or ELISA (P24 or GP51) or viral detection by PCR (E), compared with seronegative animals (C), for the measures of milk production, mortality, morbidity and presence of lymphomas (O), in studies where the unit of concern and analysis is the individual animal?" If the PICO elements and study design(s) are not appropriate to answer the review question, there is a risk of excluding relevant studies and/or including irrelevant studies.	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
A2	**Were the eligibility criteria related to *study* characteristics appropriate to answer the review question and clearly defined *a priori*?** Appropriately defining a priori the study(s) that shall be selected for inclusion in the review implies that they have the appropriate study design and characteristics allowing to answer the review question. For instance for efficacy studies it should be stated if the reviewer(s) chose to include only randomised, double-blind, placebo controlled studies, or e.g. allocation concealment was also considered as inclusion criteria (for other study designs alternative items will be relevant). In addition, study-specific elements should be considered such as e.g. reference population, dosage of the treatment, duration of the study. *If the eligibility criteria for study selection are not defined a priori, there is a risk of introducing a bias due to selective inclusion/exclusion of studies influenced by the results of such studies*	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively low risk of bias ☐ Probably low risk of bias ☐ Probably high risk of bias ☐ Definitively high risk of bias	*Please provide the main supporting information for your appraisal*
A3	**If *report* characteristics were used as eligibility criteria, are they appropriate to meet the review question? If applied, were these criteria defined *a priori?*** In principle all studies should be considered including peer-reviewed studies and grey literature. If limits are applied such as publication type, language, years, geographical/political area, etc… these should be documented, justified and defined *a priori*.	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
B.	**Search process**			
B1	**Was the extensive literature search performed in an appropriate way?** Please refer to ELS CAT and give an overall appraisal of the appropriateness of the ELS done.	*Not Applicable*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the link to the evaluation performed using the EFSA ELS Critical Appraisal Tool*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
C.	**Study selection process**			
C1	**Were preventative steps taken to minimise bias and errors in the study selection process?** Study selection is normally a 2-step process: (1) rapid assessment of titles and abstracts to exclude obviously irrelevant records and (2) examination of full-text documents. The assessment should be done evaluating: • whether the selection was carried out by at least 2 mutually independent reviewers, in parallel. It is also acceptable to have only 1 reviewer in the first step and 2 reviewers in the 2nd step only (i.e. examination of full-text documents), provided that in the first step an over-inclusive approach was applied; • the presence of a clearly defined consensus procedure for disagreements (e.g. discussion or involvement of another reviewer). Ideally the eligibility criteria should be pilot-tested by the reviewers on a subset of records and re-formulated if prone to misinterpretation.	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
C2	**Were the results of the study selection process consistent with the eligibility criteria previously defined?** Assessment should be done considering the following items: • total number of records identified; • number of records and studies excluded at each stage; • list of excluded studies, and primary reasons for exclusion after examination of full-text documents (e.g. not relevant, duplicate study, etc.). Such information is usually provided in a flow diagram.	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
D.	**Data extraction from the included studies**			

D1	**Was data extraction carried out appropriately and adequately? Was the approach defined *a priori*?**
	
	Assessment should be done evaluating:
	• if data extracted are relevant for the PICO elements:
	  ○ if relevant data on study subjects, intervention and control groups were identified and extracted;
	  ○ if relevant study outcomes and related measures of precision were extracted (i.e. either summary data on each intervention group or effect estimates with confidence intervals) and, if not available, proxy information was used.
	• if sources of heterogeneity related to the PICO elements (defined as "clinical" heterogeneity in the health care context) across the included studies were explored and the relevant data extracted;
	• if sources of methodological heterogeneity (i.e. related to the study design characteristics and risk of bias) across the included studies were explored and the relevant data extracted (e.g. parallel vs. cross-over, follow-up period);
	• if any assumptions related to data extraction were justified;
	• if data extraction was harmonised across the different studies from which the data were extracted (e.g. measurement unit, data coding).
	Ideally, the data extraction form should include what is reported in the study and, if translation was necessary, how the data was "translated" for the purposes of the review for a given parameter.
	Ideally, information on sources of funding in the individual studies should be extracted (to evaluate the risk of publication bias).
		Please insert a quotation or a brief summary of what has been reported in the systematic review	☐ Definitively appropriate
			☐ Probably appropriate
			☐ Probably not appropriate
			☐ Definitively not appropriate
			☐ Not Applicable
				Please provide the main supporting information for your appraisal

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
D2	**Were preventative steps taken to minimise bias and errors in the data extraction process?** Assessment should be done evaluating: • if data extraction was carried out by at least 2 mutually independent reviewers, in parallel; • the presence of a clearly defined consensus procedure for disagreements (e.g. discussion or involvement of another reviewer); • the presence of a procedure for obtaining and/or confirming data from researchers; • if duplicate studies were identified and data extracted only once, with appropriate explanations. Ideally the data extraction forms should be pilot- tested by the reviewers on a subset of records and improved if prone to misinterpretation.	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
E.	**Assessment of the methodological quality of the studies included in the review**			
E1	**Was the methodological quality of the individual studies, which were included in the review, appropriately and adequately appraised? Was the approach defined *a priori?*** Assessment should be done evaluating: • if methodological flaws and biases relevant to the study design(s) were identified and adequately assessed; • if quality assessment was harmonised across the different studies by the specific outcomes included in the review. The elements to consider for appraising the methodological quality of the studies included in the review vary depending upon the study designs. Different critical appraisal tools are available and all should be considered in view of the subject of the review and, if needed, adapted.[6]	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
E2	**Were steps taken to minimise bias and errors when appraising the methodological quality of the studies included in the review?** Assessment should be done evaluating: • if quality assessment was carried out by at least 2 mutually independent reviewers, in parallel; • the presence of a clearly defined consensus procedure for disagreements (e.g. discussion or involvement of another reviewer). Ideally the appraisal tool should be pilot-tested, also to maximise consistency of application between different reviewers.	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

[6] A tool for appraising RCTs in the EFSA context is the AMU RCT CAT (Appendix B –). Other CATs currently available in the scientific community are e.g. SYRCLE's risk of bias tool for animal studies (Hooijmans et al. 2014); ACROBAT-NRSI: A Cochrane Risk Of Bias Assessment Tool for Non-Randomized Studies of Interventions (Sterne et al. 2014); Cochrane Collaboration's tool for assessing risk of bias in randomised trials (Higgins et al. 2011); or OHAT risk of bias tool (Rooney et al., 2014; OHAT/NTP, 2015).

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
F.	**Data analysis and synthesis of results**			
F1	**Was the analysis and synthesis of the individual effect estimates properly undertaken? Was the approach defined *a priori?*** • if a meta-analysis was carried out, it should be assessed if it was appropriate to calculate the pooled estimate. The meta-analysis should be appraised evaluating: ○ if the effect measure chosen was appropriate for the data type of the outcomes measured in the individual studies (e.g. for continuous variables, differences of means; for dichotomous variables, risk ratios); ○ if the appropriate pair-wise comparisons of intervention groups were made;[7] ○ if the units of analysis were correctly identified (i.e. the analysis must take into account the level at which randomisation occurred as in most circumstances the number of observations in the analysis should match the number of "units" that were randomised). Thus carefulness is required for study designs like cluster-randomised trials, cross-over trial or multiple sites trials); ○ if statistical heterogeneity was appropriately identified and quantified (e.g. Q test, I2); ○ if a consistent model was chosen and justified (i.e. Fixed Effect Model/Random Effect Model); ○ if missing data, when existing (i.e. entire studies, outcomes, summary data for an outcome, individual participants from summary data) were properly handled and the related assumptions clarified. • if the results could not be pooled using meta-analysis, it should be appraised if: ○ a systematic approach was applied, i.e. the same elements were described for each study; ○ the studies were organised in groups or clusters, to enable identification of patterns in results (for this purpose tables and graphics - i.e. forest plots, can be provided). • if the approach to data analysis and synthesis was defined a priori.	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

[7] From the Cochrane Handbook: "*The comparisons addressed in the review should relate clearly and directly to the questions or hypotheses that are posed when the review is formulated*" and can be translated into the following questions: - What are the experimental and control (comparator) interventions of interest? - Does the intervention have variations (e.g. dosage/intensity, mode of delivery, personnel who deliver it, frequency of delivery, duration of delivery, timing of delivery)?

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
F2	**If there was the need or opportunity to perform additional analyses (e.g. sensitivity or subgroup analyses, meta-regression), were they performed? Was the approach defined *a priori?*** Assessment should be done evaluating: • if, when enough data were available for subgroup analysis or meta-regression, such analyses were performed. The reasons for performing or not performing the analyses should be explained and the results clearly discussed. For instance, normally additional analyses should be performed in order to account for the appraisal of the methodological quality in the synthesis of the results; • if sensitivity analysis was conducted to test the robustness of the results with respect to any assumptions and decisions that were made in the review (e.g. related to the search, the eligibility criteria, the approach taken to handle missing data, sources of funding of the individual studies, etc); • if the approach (i.e. when to perform a sub-group and/or sensitivity analysis and how) was defined a priori.	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
G.	**Reporting bias**			
G1	**Was risk of publication bias addressed? Was the approach defined *a priori?*** Risk of publication bias can be addressed in the review by using a combination of graphical aids (e.g. funnel plot, other available tests) and/or statistical tests (e.g. Egger regression test). The approach should be defined a priori. For assessing publication bias, the authors of the review could also consider the source of funding of the individual studies included in the review. Under this appraisal item other types of reporting biases could be assessed, e.g.: • publication bias: the publication or non-publication of research findings, depending on the nature and direction of the results; • time lag bias: the rapid or delayed publication of research findings, depending on the nature and direction of the results; • multiple (duplicate) publication bias: the multiple or singular publication of research findings, depending on the nature and direction of the results; • location bias: the publication of research findings in journals with different ease of access or levels of indexing in standard databases, depending on the nature and direction of results; • citation bias: the citation or non-citation of research findings, depending on the nature and direction of the results; • language bias: the publication of research findings in a particular language, depending on the nature and direction of the results; • outcome reporting bias: the selective reporting of some outcomes but not others, depending on the nature and direction of the results.	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
H.	**Interpretation of Results and Conclusions**			
H1	**Did the conclusions reflect the results of the review and any limitation in the process?** Assessment should be done evaluating if the conclusions took into account the following aspects: • the pooled and/or individual studies estimates (e.g. via appropriate forest plot); • if performed, the results of the additional analyses; • the results of the assessment of methodological quality of the included studies. • any limitations in the process, e.g. limitations of the search, limited number of reviewers involved in the selection and data extraction process, etc. If the quality of the body of evidence was evaluated in the review, this should be reflected by the conclusions.	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
I.	**Additional considerations**			
I1	**If a protocol was provided, are appropriate justifications given for any described deviations from the protocol?**	*Please insert a quotation or a brief summary of what has been reported in the systematic review*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
I2	**Have any competing interests been identified?**	*Please insert a quotation or a brief summary of what has been reported in the systematic review*		
I3	**Add here any aspects that should be outlined and are not covered above.**	• Add • Add • etc		

3. References

The SR CAT was developed using and integrating the following documents:

Berlin JA, 1997. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. Lancet, 350, 185–186.

CASP 2010. 10 questions to help you make sense of reviews.. pp. http://www.casp-uk.net/wp-content/uploads/2011/11/CASP_Systematic_Review_Appraisal_Checklist_14oct10.pdf

CEBM (Centre of Evidence Based Medicine), Systematic Review Critical Appraisal Sheet. pp. http://www.cebm.net/critical-appraisal/

CRD, 2009. Systematic Reviews, CRD's Guidance for undertaking reviews in health care. Centre for Reviews and Dissemination,

Detsky AS, Naylor CD, O'Rourke K, McGeer AJ and L'Abbe KA, 1992. Incorporating variations in the quality of individual randomised trials into meta-analysis. J Clin Epidemiol, 45, 255–265.

EFSA (European Food Safety Authorithy), 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. 2010; 8(6):1637, 90 pp. doi:10.2903/j.efsa.2010.163 http://www.efsa.europa.eu/en/efsajournal/pub/1637.htm

EFSA (European Food Safety Authority), 2011. Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under regulation (EC) No 1107/2009. 2011;9(2):2092, 49 pp. doi:10.2903/j.efsa.2011.2092 http://www.efsa.europa.eu/en/efsajournal/pub/2092.htm#

Higgins JPT and Green S, 2011. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. www.cochrane-handbook.org

Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M and Langendam MW, 2014. SYRCLE's risk of bias tool for animal studies. BMC Med Res Methodol, 14, 43.

Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M and Moher D, 1996. The quality of randomised controlled trials included in meta-analyses and systematic reviews: how often and how is it assessed? Proceedings of the 4th Cochrane Colloquium, Adelaide, Australa.

Khan KS, Daya S and Jadad A, 1996. The importance of quality of primary studies in producing unbiased systematic reviews. Arch Intern Med, 156, 661–666.

Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, Liberati A, Linde K and Penna A, 1996. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. Lancet, 347, 363–366.

Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P and Klassen TP, 1998. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet, 352, 609–613.

Moher D, Liberati A, Tetzlaff J, Altman DG and Group P, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med, 6, e1000097.

NNR5 working group 2011. A guide for conducting systematic literature reviews for the 5th edition of the Nordic Nutrition recommendations. Copenhagen: Nordic Council of Ministers pp. http://www.diva-portal.org/smash/get/diva2:716939/FULLTEXT01.pdf

OHAT/NTP (Office of Health Assessment and Translation/National Toxicology Program), 2015. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. pp. http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf

Rooney AA, Boyles AL, Wolfe MS, Bucher JR and Thayer KA, 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect, 122, 711–718.

Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D and Bouter LM, 2007. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol, 7, 10.

Sterne JAC Higgins JPT Reeves BC on behalf of the development group for ACROBAT-NRSI 2014. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomised Studies of Interventions (ACROBAT-NRSI). 56 pp. http://www.riskofbias.info

Task Force on Systematic Review and Guidelines 2013. Assessing the quality and applicability of systematic reviews (AQASR). Austin, TX - Centre on Knowledge Translation for Disability and Rehabilitation Research. pp. http://ktdrr.org/ktlibrary/articles_pubs/ncddrwork/aqasr/

Appendix B – Critical appraisal tool for assessing quality of Randomised Controlled Trials in humans (RCT CAT)

1. Randomised Controlled Trials: main definitions and the EFSA context

A Randomised Controlled Trial (RCT from now onward) is a study in which experimental units are allocated at random to interventions. One of these interventions is the standard of comparison or control. The control may be a standard practice, a placebo, or no intervention at all. RCTs seek to measure and compare the outcomes after the experimental units receive the interventions. Because the outcomes are measured, RCTs are quantitative studies. A RCT is considered the gold standard methodology to determining whether a cause–and–effect relationship exists between an intervention and an outcome (modified from Webster's New World™ Medical Dictionary, 3rd Edition, 2008).

In the EFSA context, RCTs are mainly encountered in the context of human nutrition. The current CAT does not take into account the relevance of the study to the assessment question instead it is specifically aimed at appraising the intrinsic methodological quality of the RCT.

2. RCT CAT

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
A.	**METHODS- Overview**			
A1	**Are general and specific objectives of the trial consistent with the research question?** General objective: testing for difference, non-inferiority or equivalence. Specific objectives: hypotheses to be tested. In general a study should report first the research question, then the general and specific objectives. In case the research question is not directly mentioned in the paper the reply to this question should be "Not Applicable".	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
A2	**Is the trial design appropriate to meet the objective?** E.g. cross-over/parallel, completely randomised/randomised block design, study duration etc.	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
B.	**METHODS - Participants**			
B1	**Are the eligibility (inclusion/exclusion) criteria for participants consistent with the study objectives?** For instance the objective of a study may be to examine the effect on a specific outcome of the daily intake of a particular substance in diabetic patients. In this case the participants should be enrolled among people affected by diabetes mellitus as confirmed by specific criteria.	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
B2	**Are the experimental settings appropriate to meet the objectives of the trial?** Examples of the items to be considered are: • location; • start and end dates; • interruptions to the trial; • delayed start at some locations/centres.	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
C.	**METHODS – Interventions**			
C1	**Were the interventions (including administration route and dosage) defined consistently with the study objectives and appropriate?** Examples of the items to be considered are: • treatment and control arms; • dose; • administration route; • administration timing; • duration.	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
D.	**METHODS – Outcomes**			
D1	**Are the outcomes appropriate to meet the objectives of the RCT?** Examples of the items to be considered are: • list of all outcomes measured in the trial; • primary outcomes; • secondary outcomes; • how were they measured? • when were they measured? • any changes with reasons after trial commenced.	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
E.	**METHODS - Sample size calculation**			

E1	**Is sample size calculation appropriate for primary objective(s) and estimates/hypotheses to be tested?**
	
	In case of hypotheses testing:
	• power analysis carried out a priori considering the following elements:
	  ○ desired confidence level;
	  ○ desired minimum power of the test;
	  ○ minimum effect size considered biologically relevant?
	  ○ expected variability of the effect.
	• is the issue of multiplicity (if any) addressed while planning the sample size?
	• are drop out considered for sample size calculation?
		Please insert a quotation or a brief summary of what has been reported in the RCT	☐ Definitively appropriate
			☐ Probably appropriate
			☐ Probably not appropriate
			☐ Definitively not appropriate
			☐ Not Applicable
				Please provide the main supporting information for your appraisal

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
F.	**METHODS – Randomisation (sequence generation and allocation concealment)**			

F1	**Is the random allocation appropriate?**
	Examples of the items to be considered are:
	• methods for generating random allocation sequence e.g.
	  ○ algorithm used should generate an unpredictable allocation sequence;
	  ○ type of randomisation should prevent unbalances in the allocation (e.g. completely randomised design, randomised block design).
	• implementation of the randomisation: the eligibility check of the subjects to be enrolled in the study should be done before randomisation.
		Please insert a quotation or a brief summary of what has been reported in the RCT	☐ Definitively low risk of bias
			☐ Probably low risk of bias
			☐ Probably high risk of bias
			☐ Definitively high risk of bias
				Please provide the main supporting information for your appraisal

F2	**Is allocation concealment appropriate?** Assignment mechanism should prevent foreknowledge of treatment (i.e. prevent participants selection bias) until allocation. Note that allocation concealment can always be successfully implemented.	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively low risk of bias ☐ Probably low risk of bias ☐ Probably high risk of bias ☐ Definitively high risk of bias	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
G.	**METHODS – Blinding**			
G1	**Were the participants and trial personnel appropriately blinded?** • assignment mechanism should prevent knowledge of treatment after allocation (i.e. prevent performance bias) by participants and trial personnel; • trial can be open label (no blinding), single blinded (only participants) or double blinded (participants and trial personnel). Note that it is not always possible to carry out a blinded study. For example, when the intervention consists of a substance with peculiar organoleptic characteristics (e.g. colour, taste or smell) which cannot be masked or replicate in the control or when the substance has remarkably visible effects (e.g. changes in the urine, faeces, skin) for which participants and trial personnel can assume to which group the subjects were randomised to. If an open label trial is performed an adequate justification should be provided by the authors.	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively low risk of bias ☐ Probably low risk of bias ☐ Probably high risk of bias ☐ Definitively high risk of bias	*Please provide the main supporting information for your appraisal*
G2	**Were the outcome assessors appropriately blinded?** Assignment mechanism should prevent knowledge of treatment after allocation (i.e. prevent performance bias) by outcome assessors. Examples of outcome assessors are: • primary data collectors (e.g. interview staff responsible for measurement or collection of outcome data); • secondary assessors (e.g. external outcome adjudication committees).	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively low risk of bias ☐ Probably low risk of bias ☐ Probably high risk of bias ☐ Definitively high risk of bias	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
H.	**METHODS – Analysis populations**			
H1	**Was the population approach taken to analyse data appropriate? Was the approach planned a priori?** • intention to treat (ITT8); • per protocol (PP9) • complete cases.10 A justification should be provided for the choice.	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

[8] Intention-to-treat: analysis carried out including all randomised participants and retaining all of them in the group to which they were allocated.
[9] Per protocol: analysis carried out by excluding participants who did not adequately adhere to the protocol (e.g. those who did not meet the inclusion, did not take all the intended treatment, or received a different treatment or no intervention).
[10] Complete cases: analysis carried out only on those whose outcome is known (in case on missing outcome data and no imputation envisaged).

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
I.	**METHODS - Data analysis**			
I1	**Was the need for any data transformation appropriately evaluated? If transformation was applied, is it appropriate? Was it planned a priori?** • Is there any evidence that transformation should have been used/not used? • If transformation applied, was the correct formula used? • If transformation applied, were data back-transformed appropriately?	Please insert a quotation or a brief summary of what has been reported in the RCT	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
I2	**Were appropriate methods used to detect and handle outliers? Were criteria for detection defined a priori?** • Was the presence of outliers appropriately checked (at least with graphical methods)? • Were outliers appropriately handled?	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
I3	**If any model was used to analyse data, is it appropriate? Was it planned a priori?** • assumptions underlying the analysis; • specification of the model consistent with variable types, objectives, assumptions etc.; • estimates of the effect; • indications of the goodness of fit of the model (if applicable).	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
I4	**If any method was used to handle missing data, is it appropriate? Was it planned a priori?** • Are the assumptions on the missing mechanism realistic in the specific context (e.g. missing at random – MAR, MNAR)? • Is the method used to handle missing data consistent with the assumptions?	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
I5	**Were appropriate tests used to assess the effects of the intervention? Was it planned a priori? Was the issue of multiple testing (multiplicity) appropriately handled?** Test suitable for: • type of variables in the study; • more than 2 intervention groups (if it is the case); • multiple primary endpoints (if it is the case); multiple measurements of primary endpoints (if it is the case).	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
I6	**Were potential discrepancies among groups at baseline adequately investigated and appropriately taken into consideration in the analyses?**	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively low risk of bias ☐ Probably low risk of bias ☐ Probably high risk of bias ☐ Definitively high risk of bias	*Please provide the main supporting information for your appraisal*
I7	**If any, were the additional analyses (e.g subgroups analysis, interim analyses, sensitivity analysis) performed appropriately? Were they planned a priori?** • foreseen prior to the start of the experiment; • consistently with objectives of the study.	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
I8	*Only for crossover trials* **Are the following issues adequately addressed?** • carry over effect; • period effect; • sequence effect; • treatment by period interactions.	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
J.	**RESULTS**			
J1	**Is the number of sampling units and measurements taken on them used in the analyses consistent with the approach to the analysis declared in the study?** The number of observations actually used in each analysis should be consistent with what occurred to the subjects, the measurement of endpoints along the study (subjects randomised, who received intended treatment, drop-outs etc…) and the approach to the analysis. E.g. if an intention to treat approach was declared, the number of observations used in the analysis must be consistent with the subjects allocated to the treatment groups.	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively low risk of bias ☐ Probably low risk of bias ☐ Probably high risk of bias ☐ Definitively high risk of bias	*Please provide the main supporting information for your appraisal*
J2	**Were the results appropriately presented?** • Was the uncertainty around the estimates appropriately investigated and appropriately taken into account in the conclusions? • Was the appropriate effect type taken into consideration in drawing conclusions (e.g. absolute and/or relative effect)? • Was the full list of endpoints taken into consideration for the conclusions? • Are the results of the analyses conducted post-hoc identifiable?	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
J3	**Were the results of the analysis interpreted in line with the stated hypotheses and objectives of the study? Are the conclusions consistent with the actual evidence that was produced?**	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
J4	**Was biological relevance of the results clearly discussed?**	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
K.	**Additional considerations**			
K1	**Was the Protocol registered before the starting of the trial and was the trial performed and reported accordingly?**	*Please insert a quotation or a brief summary of what has been reported in the RCT*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
K2	**Have any competing interests been identified?**	*Please insert a quotation or a brief summary of what has been reported in the RCT*		
K3	**Add here any aspects that should be outlined and are not covered above.**	• Add • Add • etc		

3. References

The RCT CAT was developed using and integrating the following documents:

Altman DG, 1996. Better reporting of randomised controlled trials: the CONSORT statement. BMJ, 313, 570–571.

Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D and Stroup DF, 1996. Improving the quality of reporting of randomised controlled trials. The CONSORT statement. JAMA, 276, 637–639.

Bornhoft G, Maxion-Bergemann S, Wolf U, Kienle GS, Michalsen A, Vollmar HC, Gilbertson S and Matthiessen PF, 2006. Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity. BMC Med Res Methodol, 6, 56.

Chalmers TC, Smith H, Jr., Blackburn B, Silverman B, Schroeder B, Reitman D and Ambroz A, 1981. A method for assessing the quality of a randomised control trial. Control Clin Trials, 2, 31–49.

Chalmers TC, Celano P, Sacks HS and Smith H, Jr., 1983. Bias in treatment assignment in controlled clinical trials. N Engl J Med, 309, 1358–1361.

Cho MK and Bero LA, 1996. The quality of drug studies published in symposium proceedings. Ann Intern Med, 124, 485–489.

Colditz GA, Miller JN and Mosteller F, 1989. How study design affects outcomes in comparisons of therapy. I: Medical. Statistics in medicine, 8, 441–454.

Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C and Antes G, 1997. Language bias in randomised controlled trials published in English and German. Lancet, 350, 326–329.

Gluud LL, 2006. Bias in clinical intervention research. Am J Epidemiol, 163, 493–501.

Greenhalgh T, 1997. How to read a paper. Papers that report diagnostic or screening tests. BMJ: British Medical Journal, 315, 540.

Higgins JPT and Green S, 2011. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. www.cochrane-handbook.org

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ and McQuay HJ, 1996. Assessing the quality of reports of randomised clinical trials: is blinding necessary? Control Clin Trials, 17, 1–12.

Moher D, Jadad AR, Nichol G, Penman M, Tugwell P and Walsh S, 1995. Assessing the quality of randomised controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials, 16, 62–73.

Moher D, Jadad AR and Tugwell P, 1996. Assessing the quality of randomised controlled trials. Current issues and future directions. Int J Technol Assess Health Care, 12, 195–208.

Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M and Altman DG, 2010. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ, 340, c869.

Ohlsson A and Lacy JB, 1995. Quality assessments of randomised controlled trials: evaluation by the Chalmers versus the Jadad method. Proceedings of the 3rd Annual Cochrane Colloquium.

Rothwell PM, 2005. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet, 365, 82–93.

Schulz KF, Chalmers I, Hayes RJ and Altman DG, 1995. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA, 273, 408–412.

Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Gluud C, Martin RM, Wood AJ and Sterne JA, 2008. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ, 336, 601–605.

Appendix C – Critical appraisal tool for assessing quality of GM plant equivalence studies (GMO CAT)

1. GM plant equivalence studies: main definitions and the EFSA context

The GMO CAT was developed in order to facilitate a harmonised and integrated approach in the assessment of the methodological quality of the dossiers received by EFSA from applicants requested to demonstrate the agronomical and compositional equivalence of GMO plants and comparators using the methodology set up by the GMO Panel (EFSA 2010, EFSA 2011, Van der Voet et al. 2011).

It should be noted that the different reports usually delivered by the applicant for compositional and agronomic-phenotypic characteristics should be assessed filling in only one check-list eventually distinguishing the comments when the evaluation is different for the various aspects (agronomical and compositional).

The following definitions apply in the context of the GMO CAT:

TERM	DEFINITION
Test material	Plant material (usually grain, seeds or kernels) analysed in field trials. In the context of this CAT the test material comprises the genetically modified plant, comparator and non-GM reference varieties.
Reference varieties	Non-GM reference varieties with a 'history of safe use', used to establish in the field trials the natural variation for the analysed parameters. The choice of non-GM reference varieties should be appropriate for the chosen sites and should be justified explicitly.
Comparator	It is a non-GM line with a genetic background as close as possible to the GM plant under assessment in case of sexually propagated crops, or isogenic varieties in case of vegetatively propagated crops. These traditionally cultivated crops can thus serve as comparators when assessing the safety of GM plants and derived food and feed.
Field trials	Field trials are experimental activities carried out outside the usual laboratory situation at different locations in order to study the agronomic, phenotypic and composition of genetically modified plants.
Site	The locations included in the field trials.
Block	At a particular site of field trials, blocks are defined areas (plots) of planted material of each individual test material to be analysed. Typically the plots of each test sample are present only once per block.
Plot	Plots are defined areas cultivated with a particular test material of the same source usually in repeated blocks at a specific location of field trials.
Replicate	Replicates are the multiple cultivation of the same test material at the same location of field trials.

The following mathematical notations apply in the context of the GMO CAT:

NOTATION	DEFINITION
$\mathbf{m_G, m_C\ m_R}$	mean of GMO, comparator and reference varieties (treated as a single test material)
lsd(GC;1;95)	least significant difference between GMO and comparator in model 1 at 95 % confidence level
lsd(GR;1;95)	least significant difference between GMO and reference varieties in model 1 at

95 % confidence level

lsd(GC;2;97.5) least significant difference between GMO and comparator in model 2 at 97.5 % confidence level

lsd(GR;2;97.5) least significant difference between GMO and reference varieties in model 2 at 97.5 % confidence level

EL_L EL_U Lower and Upper Equivalence Limits

2. GMO CAT

#	Checklist item	Assessment explanation	Assessment	Action required
A. METHODS- Overview				
A1	**Is the trial design appropriate to meet the objective? (e.g. completely randomised design, completely randomised block design)** Guidance: Section 3.1.3.2, paragraph "Experimental design", pg 13–15. Opinion: section 2.1, pg 9–14.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
A2	**Are the experimental setting and location appropriately reported? Does the information provided allow replication of the study?** • location; • start and end dates; • interruptions to the trial (if any); • delayed start at some locations (if any); • definition of the test materials (GMO/comparator/reference varieties); • basic field operations (planting, fertilisers, pesticides, climatic conditions). Guidance: Section 3.1.3.2, paragraph "Experimental design", pg 14.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required

#	Checklist item	Assessment explanation	Assessment	Action required
B. METHODS - Trial design				
B1	**How many sites was the trial performed at? Was this a sufficient number? Are they representative of the relevant meteorological, soil and agronomic conditions in the geographical area?** • the trial should be conducted at least at 8 sites; • if sites cover a very restricted geographical range, the trial has to be replicated over more than 1 year. Guidance: Section 3.1.3.2, paragraph "Experimental design", pg 14. Opinion: section 2.3.3, pg 12–13.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
B2	**How many non-GM reference varieties were used? Was this a sufficient number?** There should be at least 3 non-GM reference varieties per site and at least 6 in the complete trial. If not, justification should be provided. Guidance: Section 3.1.3.2, paragraph "Experimental design", pg 14. Opinion: section 2.3.2, pg 12–13.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
B3	**How many replications were carried out at each site? Was this a sufficient number?** The number of replications should ensure adequate power. It is not advisable to have less than 15 residual degrees of freedom per trial. • normal rule: 4 replications per site; • if only 2 non-GM reference varieties: 6 replications; • if only one commercial variety: 8 replications. Guidance: Section 3.1.3.2, paragraph "Experimental design", pg 13–15. Opinion: section 2.3.1, pg 11–12.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required

#	Checklist item	Assessment explanation	Assessment	Action required

B4	**What is the design? Does it meet the requirements of the Guidance/Opinion?**
	
	All GMO, comparator and non-GM reference varieties should be fully randomised to the plots at each site. Any additional test material (e.g. when herbicide tolerant system is assessed) and/or comparator (e.g. negative segregant) should be fully randomised and replicated.
	
	Only in case of randomised block design
	• each block should contain all the GMO varieties (if more than one), their comparators and non-GM reference varieties;
	• when testing more than 1 GMO, if the number of plots per block is more than 16, a partially balanced incomplete block design may be used by excluding some of the GMO plants and their comparators from each block provided that:
	  ○ each comparator should always occur together with its respective GM plant in the same block;
	  ○ all the reference varieties should appear in each of the incomplete blocks and be fully randomised with the GM plants and their comparators.
	Guidance: Section 3.1.3.2, paragraph "Experimental design", pg 14–15.
	Opinion: section 2.1, pg 9–10, section 2.3.1, pg 11–12, section 2.4, pg 13–14, section 2.5 pg 14.
		Information as provided
		
		Assessment explanation and consequences
			☐ Yes fully
			☐ No
			☐ Partially
			☐ Unclear
				☐ None
				☐ Confirmation required
				☐ Further information required

B5	**Is the method for generating the random sequence appropriate? Is allocation sequence concealment appropriate?** • blocking and block size; • stratification (in any).	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
B6	**Were the personnel in charge of the field maintenance and the outcome assessors appropriately blinded?** Was blinding feasible? If not, is this clearly stated?	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required

#	Checklist item	Assessment explanation	Assessment	Action required
C. METHODS – Endpoints				
C1	**Are the response variables (endpoints) clearly defined?** List of all endpoints measured in the trial. Guidance: Section 3.1.3.3, pg 19, section 3.1.3.4, pg 20 and section 3.1.3.5, pg 20.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
D. METHODS - Data analysis				
D1	**Was the statistical analysis plan provided and if so was it complete?** • What version? • When was it signed? • Any changes from the statistical plan? • Any changes after sign off?	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
D2	**Was there any data transformation?** • Is there any evidence that transformation should have been used (e.g. to change scale, to normalise a distribution)? • If used, was the problem resolved? • Was it done correctly? Was the correct formula applied? • Were the data back-transformed correctly? Guidance: Section 3.1.3.2, paragraph "Statistical analysis", pg 16. Opinion: section 3.1, pg 15–16.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required

#	Checklist item	Assessment explanation	Assessment	Action required
D3	**Were methods used to detect outliers?** • Was the presence of outliers checked at least with graphical methods? • How were outliers handled? • If outliers were found and removed, were sensitivity analyses performed? Opinion: section 3.1, pg 15–16.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
D4	**Was the following information appropriately reported for each model used:** • assumptions underlying the analysis; • full specification of the models in terms of effects; • results of any test of interaction between the test materials and sites; • outcomes of the models in terms of the Sum of Squares, Mean Square and estimated degrees of freedom for each fixed effect; • contribution to the variance of each random effect; • whenever possible, an indicator of the goodness of fit of the model should be provided. Guidance: Section 3.1.3.2, paragraph "Statistical analysis", pg 17.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
D5	**Does the linear mixed model used to calculate the confidence intervals for the difference test and for the equivalence test meet the requirements of the Guidance/Opinion? (Model 1)** • fixed factors: It should have as many levels as test materials and be used to compare them. The set of non-GM reference varieties is considered as a single level of the fixed factor. For the difference and equivalence tests, the component of the fixed factor of interest is the single degree-of-freedom contrast between, respectively, the GM plant and its comparator, and the GM plant and the set of non-GM reference varieties. • random factors (the model should include the following but not necessarily be restricted to): ○ interaction between test materials and I (indicator variable for non-GM reference varieties); ○ variation between sites; ○ variation between blocks within sites. Guidance: Section 3.1.3.2, paragraph "Statistical analysis", pg 16. Opinion: section 3.2.3 pg 19–20, section 3.2.4, pg 20–22, section 3.3.2, pg 23–24.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required

#	Checklist item	Assessment explanation	Assessment	Action required
D6	**Were the equivalence limits established in accordance to the requirements of the Guidance/Opinion?** Limits established using non-GM reference varieties in the experiment: a model (model 2) should be used. Model 2 has to be identical to model 1 but replacing the random component due to the interaction between test materials and I (indicator variable) with a random component due to test materials (GMO, comparator and each of the reference varieties). Equivalence limits should be computed as 95 % confidence interval around the mean of the reference varieties: **$m_R \pm lsd(GR;2;97.5)$** • limits established using literature; • they should be set up and analysis conducted in accordance with the requirements described in section 3.3.3 of the Opinion; • limits established subjectively; • they should be set up and analysis conducted in accordance with the requirements described in section 3.3.4 of the Opinion. Guidance: Section 3.1.3.2, paragraph "Statistical analysis", pg 16. Opinion: section 3.2.3 pg 19–20, section 3.3 pg 22–26.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
D7	**Was the possible site-specific effects (genotype by site interaction) appropriately investigated?** In case of significant difference and/or lack of equivalence for any particular endpoint, is found for an endpoint it is recommended to assess whether there are interactions between any of the test material and sites (possibly using a simple standard ANOVA approach) and conduct site-specific analyses. Details should be given, for each endpoint analysed, listing: (a) the assumptions underlying the analysis, and, when appropriate: (b) degrees of freedom, (c) the estimated residual variation for each source of variation, and variance components, (d) any other relevant statistics. Guidance: Section 3.1.3.2, paragraph "Statistical analysis", pg 19. Opinion: section 4.2 pg 28–29.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required

#	Checklist item	Assessment explanation	Assessment	Action required
D8	**Were Confidence Intervals (CIs) for testing difference and equivalence computed in accordance to the requirements of the Guidance/Opinion?** • in case of limits established using non-GM reference varieties in the experiment: ○ CI for testing difference: a two-sided 90 % confidence interval should be calculated around the difference between GMO and comparator using model 1 as follows: **$(m_G-m_C)\pm lsd(GC;1;95)$** ○ CI for testing equivalence: a two-sided 90 % confidence interval should be calculated around the difference between GMO and reference varieties using model 1 as follows: **(mG-mR)±lsd(GR;1;95)** • in case of limits established using literature: ○ analysis should be conducted in accordance with the requirements described in section 3.2.2 of the Opinion. • in case of limits established subjectively: ○ analysis should be conducted in accordance with the requirements described in section 3.3.4 of the Opinion. Guidance: section 3.1.3.2, paragraph "Statistical analysis", pg 15–19. Opinion: section 3.2.2 pg17–19 and section 3.3.2 pg 23–24.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
D9	*Only in case of limits established using non-GM reference varieties* **Were difference and equivalence tests performed in accordance to the current requirements of the Guidance (EFSA 2011)?** • **Test for difference**: the null hypothesis of 'no difference between GMO and comparator' should be tested assessing whether the CI established according to item 17a **$(m_G-m_C)\pm lsd(GC;1;95)$** includes '0' • **Test for equivalence**: the equivalence between GMO and reference varieties should be tested assessing whether the CI established according to item 17b **$(m_G-m_R)\pm lsd(GR;1;95)$** lies within the equivalence limits established according to methodology item 15. Guidance: section 3.1.3.2, paragraph "Statistical analysis", pg 15–19.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required

#	Checklist item	Assessment explanation	Assessment	Action required
D10	**If any, were the site-specific or any other subgroup analysis performed appropriately?** Foreseen prior to the start of the experiment in the event of observing significant interactions. Guidance: Section 3.1.3.2, paragraph "Statistical analysis", pg 19.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
D11	**What methods (if any) were used to handle missing data? Were they appropriate?** • Are the reasons for missing data stated? • Are the assumptions on the missing mechanism clearly stated (e.g. missing at random – MAR, MNAR)? • Are the stated assumptions on missing data justified?	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
D12	**What methods (if any) were used to handle the issue of multiplicity? Were they appropriate?** • for more than 2 testing materials; • multiple primary endpoints; • multiple measurements of primary endpoints. Opinion: section 3.2.4 pg 20–22 and section 4.4 pg 29–30. Note: The issue should be discussed. The debate on the best way to handle multiplicity is still ongoing in the scientific community. Therefore it should be handled flexibly.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required

#	Checklist item	Assessment explanation	Assessment	Action required
E. RESULTS				
E1	**Were the estimated results appropriately provided?** • For each endpoint, do the results (in terms of main effects or absolute value of the endpoint) present a point and interval estimate (e.g. 95 % confidence interval)? • For binary outcomes, are both the absolute and relative values and/or effect sizes presented? • Were results presented for the full list of endpoints? • For each endpoint (measured as continuous variables): were the mean and standard error of GMO, comparators and their difference provided for the whole trial? Were they provided by site when Genotype by environment interaction was found? • Are the results of the analyses conducted post-hoc identifiable? Opinion: section 1.3 pg 8–9 and section 3.2.2 pg 17–19.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
E2	**Were appropriate graphs provided to facilitate visual interpretation of the difference and equivalence testing (recommended)?** It is highly recommended to show difference and equivalence tests simultaneously on a single graph. To this purpose the equivalence limits have to be adjusted as follows: $(m_G-m_C)+[(m_R-m_G)\pm lsd(GR;2;97.5)]* lsd(GC;1;95)/lsd(GR;1;95)$ • scaling the basic equivalence limits so that the confidence limits required for the difference and equivalence tests have the same width; • shifting means to facilitate display of the adjusted limits on the scale that has comparator mean as its baseline zero value. When another test material is used as additional comparator, its confidence limits and its adjusted equivalence limits should be displayed on the same graph by referring this to the same zero baseline defined by the main comparator. Guidance: Section 3.1.3.2, paragraph "Statistical analysis", pg 17–19. Opinion: section 4.1 pg 26–28.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required

#	Checklist item	Assessment explanation	Assessment	Action required
E3	**Were the results of the analyses interpreted in line with the stated hypotheses and objectives of the study?** • Are the conclusions based on the actual evidence that was produced? • Were conclusions based on the position of the CI in reference to the line of no difference, and the equivalence limits? • Were the endpoints classified according to the four 'equivalence categories' mentioned in the Guidance/Opinion: o the GM plant is equivalent to the set of non-GM reference varieties; o the equivalence between the GM plant and the set of non-GM reference varieties is more likely than not. Further evaluation may be required o the equivalence between the GM plant and the set of non-GM reference varieties is less likely than not. Further evaluation is required; o the GM plant is considered 'not equivalent' to the set of non-GM reference varieties. In case the equivalence limits could not be established because of no variability in the reference varieties then no conclusions can be drawn on equivalence and only difference can be assessed. Guidance: Section 3.1.3.2, paragraph "Statistical analysis", pg 17–19. Opinion: section 4.1 pg 26–28.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required
E4	*Only in case the equivalence limits could not be established using non-GM reference varieties* **Was biological relevance clearly discussed?** The applicant should report and discuss all significant differences observed between the GMO plant, its comparator and, where applicable, any other test material, focusing on their biological relevance Guidance: Section 3.1.3.2, paragraph "Statistical analysis", pg 17.	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required

#	Checklist item	Assessment explanation	Assessment	Action required
F. OTHER				
F1	**Highlight any other issues of interest here**	Information as provided Assessment explanation and consequences	☐ Yes fully ☐ No ☐ Partially ☐ Unclear	☐ None ☐ Confirmation required ☐ Further information required

3. References

The GMO CAT was developed using and integrating the following documents:

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), 2010. Statistical considerations for the safety evaluation of GMOs. EFSA Journal 2010; 8(1):1250, 59 pp. doi:10.2903/j.efsa.2010.1250

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), 2011. Guidance for risk assessment of food and feed from genetically modified plants. EFSA Journal 2011; 9(5): 2150, 37 pp. doi:10.2903/j.efsa.2011.2150

European Community 2001. Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. Official Journal of the European Communities L106, 1–39 pp. http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32001L0018

European Union 2013. Commission Implementing Regulation (EU) No 503/2013 of 3 April 2013 on applications for authorisation of genetically modified food and feed in accordance with Regulation (EC) No 1829/2003of the European Parliament and of the Council and amending Commission Regulations (EC) No 641/2004 and (EC) No 1981/2006. Official Journal of the European Union, L157:1–47 pp.http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:157: FULL: EN: PDF

van der Voet H, Perry JN, Amzal B and Paoletti C, 2011. A statistical assessment of differences and equivalences between genetically modified and reference plant varieties. BMC Biotechnol, 11, 15.

Appendix D – Critical appraisal tool for assessing quality of Extensive Literature Searches (ELS CAT)

1. Extensive Literature Searches: main definitions and the EFSA context

An extensive literature search (ELS) is a literature search process structured in a way to identify as many studies relevant to a review question as needed. It is tailored in order to address the trade-off between sensitivity and specificity depending on the context of the review question.

The fundamental aspects of an extensive literature search are: (1) the tailored search strategy (i.e. a collection of search terms combined together and used to interrogate an information source to identify relevant records); (2) and the extensive list of information sources used (i.e. electronic bibliographic databases and other sources accessed via electronic or hand-searching - for example, websites, journal tables of content, theses repositories, etc.). The ELS process is clearly documented and reported to enhance transparency.

The output of an ELS is a (potentially extensive) collection of evidence (to be screened for relevance).

In the EFSA context, there are two main situations when extensive literature searches are conducted:

1. As part of systematic reviews (SR) (see glossary) that are performed to answer specified questions aiming to estimate risk assessment parameters such as might be evaluated in well-defined primary research studies (i.e. (i) questions on the effect of an exposure or intervention; (ii) questions on the sensitivity or specificity of a test; or (iii) questions on the prevalence of a condition or the incidence of an outcome). It is important to note that in EFSA it often occurs that ELSs followed by a study selection process are outsourced to provide the units or panels with sets of evidence relevant to specific subjects, while the subsequent steps of the SR (study appraisal, analysis, synthesis, etc.) are performed in-house by the EFSA panel or unit.
2. When it is necessary to answer questions that aim to develop scenarios of a risk assessment, rather than to estimate a parameter. Typical examples are developing an inventory of items (e.g. control measures for a specific disease, analytical techniques to determine the concentration of chemical, countries that have implemented a specific method).

In view of the above, the following terminology must be clarified:

- *Key elements*: the components of the review question that specify what information must be provided in a primary study to estimate the parameter under assessment and hence answer the question. The key elements vary depending on the question type: (i) for questions on the effect of an exposure or intervention: PECO/PICO [11] key elements; (ii) for questions on the sensitivity or specificity of a test: PIT[12] key elements; and (iii) for questions on the prevalence of a condition or the incidence of an outcome: PO[13] key elements.
- *Search concept*: a key element which might be used in the search. If the question aims to estimate risk assessment parameters using primary research studies, the search concepts might be one or more of the key elements of the studies that shall be searched (i.e. PICO/PECO, PIT, and PO) *and* (where possible) the study design. If the question aims to create an inventory of items (e.g. available diagnostic tests for a specific disease) the search concepts are likely to be the disease and the concept of diagnosis (variously expressed).

his CAT focuses in particular on two aspects: (i) the search strategy and (ii) the information sources used. The critical appraisal tool could help also the reporting of an ELS process.

[11] PECO: Population, Exposure, Comparison, Outcome. PICO: Population, Intervention, Comparison, Outcome.
[12] PIT: Population, Index test and Target condition.
[13] PO: Population and Outcome of interest.

2. ELS CAT

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
A.	**Assessing the search strategy**			
A1	**Was the review question appropriately translated into search concepts?** Assessment should be done considering whether the review question was clearly defined and translated into correct search concepts.[14] In many cases where the search aims to retrieve primary research studies, the review question should be translated into clear and appropriate key elements[15] (i.e. PICO/PECO, PIT, and PO) combined (where necessary or possible) with study design.	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
A2	**Was the search string an optimal combination of the search concepts for sensitivity and precision?[16]** Assessment should consider whether: • "too many" search concepts were used (for example, in a PICO question, if all the four key elements have been used and the number of results yielded is very limited this indicates potentially low sensitivity); • any of the search concepts were too narrow or too broad (Ideally separate searches should be conducted for each considered search concept and then combined.) The process for defining the definitive search should be documented and discussed and the search agreed should be documented; • the search appears to retrieve too many or too few records.	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

[14] Search concept: a key element which might be used in the search. See definition in "1. Extensive Literature Searches: main definitions and the EFSA context".
[15] Key elements: the components of the review question that specify what information must be provided in a primary study to estimate the parameter under assessment and hence answer the question. See definition in "1. Extensive Literature Searches: main definitions and the EFSA context".
[16] Precision refers to the proportion of relevant records among all the records retrieved by a search strategy (relevant records retrieved/all records retrieved). Sensitivity refers to the proportion of relevant records retrieved by a search strategy (relevant records retrieved/total relevant records) (Glanville et al., Technical Manual for Performing Electronic Literature Searches in Food and Feed Safety).

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
A3	**Were the appropriate free-text terms (i.e. (terms in the title and abstract) identified for each search concept?** Assessment should consider whether, for each search element, the search included: • all possible synonyms (e.g. welfare, wellbeing, etc); • related terms, e.g. pesticide, pest control, etc.; • acronyms, e.g. Bluetongue, BTV, etc. If an acronym or abbreviation is used, a full text term or substantial part of a full text term should also be present; • spelling variants, e.g. behaviour, behaviour; anaemia/anaemia; • old and new terminology, e.g. aspartame/Phenylalanine; • brand and generic names, e.g. imidacloprid, Gaucho; • lay and scientific terminology e.g. aspartame/1-Methyl N-L-alpha-aspartyl-L-phenylalanate; • common typos (e.g. mitomycin/mitomicin). In addition, the following aspects should be considered: • translation issues,[17] which may lead to new translated search terms or to limitations. However language in principle should not be a limitation of the ELS. If such a limitation is applied a justification should be provided; • if apparently irrelevant or excessively broad free text terms were used.	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

[17] it refers to the inclusion of non-English databases/sub databases or e-journals or to the use of specific non-English terms in the search strategy when it is known that a particular topic/subject may include such terms. This could increase the search sensitivity leading to obtain more references.

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
A4	**Were appropriate controlled terms (subject headings) identified for each search concept and information source used (when applicable)?[18]** Assessment should consider whether: • the subject headings/indexing terms used are relevant; • any subject headings/indexing terms are missing; • any subject headings/indexing terms are too broad or too narrow; • any subject headings/indexing terms were exploded where necessary and vice versa;[19] • if no subject headings were used, the reason for omission was explained; • the use of any subheadings[20] is helpful (i.e. not too focused); • the use of any floating subheadings[21] would have been helpful and, if used, was appropriate; • if no text words were used, the reason for this omission was explained.	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
A5	**Was a pilot study carried out (when applicable)?** The search should be pilot tested whenever the literature available is not too small. Pilot testing could be done, for example, by assessing the relevance of a subset of records retrieved and checking if records already known to be relevant were captured by the search. Pilot testing results can be used to revise the search terms or identify new ones.	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

[18] For an explanation see "Controlled Vocabulary" in Glossary and Abbreviations. Please note that controlled vocabulary is not available in all bibliographic databases.
[19] The use of headings and sub-headings or major terms and narrow terms is applied differently from database to database (e.g. pubmed/Mesh or CABI/cab Thesaurus). Before performing searches it is considered best practice to verify this feature in the concerned / used database.
[20] Subheadings are used by some subject indexing schemes such as MeSH and EMTREE and can be added to a subject indexing term in order to focus it. For example, in Ovid MEDLINE the search construction 'Probiotics/ae' will restrict the probiotics subject indexing term to those records where the adverse effects of probiotics are addressed. Ae is the subheading which denotes adverse effects. Glanville et al., Technical Manual for Performing Electronic Literature Searches in Food and Feed Safety.
[21] Floating subheadings means searching subheadings on their own, unattached to specific subject headings. For instance, in Ovid MEDLINE, the search construction 'Probiotics/and ae.fs.' will find records which are indexed with the term Probiotics, and also have adverse effects applied as a subheading to any of the indexing terms in the record. This is less precise than 'Probiotics/ae' (see prevous footnote) but may increase the sensitivity of the search. Glanville et al., Technical Manual for Performing Electronic Literature Searches in Food and Feed Safety .

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
A6	**Was the appropriate spelling used?** Assessment should consider whether there were any spelling errors (NOTE: some spelling errors may be deliberately included – see item A3).	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
A7	**Was the appropriate syntax used?** Assessment should consider whether there were any errors in system syntax (e.g. truncation symbols). For instance, if the search missed truncation or truncated at the wrong point or if free text terms where appropriately joined (Boolean operator OR) with the relevant controlled terminology (subject headings). Syntax varies depending on the service provider. For instance, the most common truncation symbol is "*", but also "$" or "?" may be used depending on the database. So the help files of the service provider need to be checked.	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
A8	**Were the appropriate line numbers used?** Ideally separate searches should be conducted for each search concept. This will produce different line (search) numbers in the bibliographic database used for the search. These line numbers have then to be combined in the correct order to obtain an appropriate result. Assessment should consider whether line numbers were combined in a correct way.	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
A9	**Was the use of Boolean and proximity operators appropriate?** Assessment should consider whether: • there were any mistakes in the use of Boolean or proximity operators. For instance if AND has been inadvertently replaced by OR (or vice versa);[22] • there were any mistakes in the use of nesting with brackets; • if NOT is used, was there any unintended exclusion; or would another mechanism have been a more suitable alternative;[23] • precision could be improved by using proximity operators (e.g. adjacent, near, within, same) instead of AND;[24] • the width of any proximity operators is too wide or not wide enough; • the potential importance of word order, when using such operators, has been accounted for.	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

[22] AND is used to combine two different concepts, e.g. (xanthomonas citri) AND (citrus fruit). AND will narrow the search: the results must include ALL stated concepts. OR is used to search for similar concepts, e.g. (xanthomonas citri) OR (citrus canker). OR will widen the search: the results will include a MINIMUM OF ONE of the named concepts.

[23] NOT is used to restrict the search, e.g. pig* NOT pigeon. The results will exclude ALL records containing the excluded term even those containing the term searched. From this, NOT should be used with caution because it may have a larger exclusion effect than anticipated, as it may exclude records of interest that coincidentally discuss both terms.

[24] SAME or NEAR are used to combine two different concepts adding a notion of proximity, e.g. bisphenol A NEAR bottle. SAME will narrow the search: the results must include ALL your stated concepts in the same sentence (for instance).

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
A10	**Were limits appropriately used?** Assessment should consider whether: • any of the limits used seem unwarranted; • any potentially helpful limits are missing.	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*
A11	**Were search filters[25] (if used to identify study designs) appropriately used?** Assessment should consider whether: • any filters used are appropriate for the topic; • any helpful and relevant available filters are missing.	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

[25] Search filters are pre-tested, and sometimes validated, search strategies which are designed to retrieve specific types of study or topic (such as a specific population) from a named database. They usually consist of a set of indexing and free text terms which describe the study design or topic of interest. The filter is added to the search strategy designed to retrieve the review's key elements, in order to restrict the results to the required study design or topic of interest. Glanville et al., Technical Manual for Performing Electronic Literature Searches in Food and Feed Safety.

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
A12	**Was the search strategy correctly adapted for each database used?** The searcher may adapt the search strategy for additional databases and/or interfaces. Adaptations should be provided for review. The adaptations should be assessed to ensure that they are correct (e.g. truncation symbols, controlled vocabulary, lemmatization[26] option, etc).	*Please insert a quotation or a brief summary of what has been reported in the ELS*	☐ Definitively appropriate ☐ Probably appropriate ☐ Probably not appropriate ☐ Definitively not appropriate ☐ Not Applicable	*Please provide the main supporting information for your appraisal*

[26] According to the Oxford English Dictionary lemmatize means "to sort (words as they occur in a text) so as to group together those that are inflected or variant forms of the same word". The lemmatization function in a bibliographic database allows capturing together variants of the same word; this is true overall for the British and American English. For instance when lemmatization is activated searching for the free text word "colour" (British English) will capture references where the word "color" (American English) is used.

#	Appraisal question	Information as reported	Appraisal	Rationale for the appraisal
B.	**Assessing the information sources searched**			

B1	**Assess if the search was extensive enough, i.e. assess if the right (relevant and reliable) combinations of information sources were searched**
	More than a single database should be searched. Searches of information sources for different types of publication would help to demonstrate that the search had been extensive:
	• Major bibliographic databases (journals and books)
	• Information sources recording:
	  ○ dissertations;
	  ○ conference reports;
	  ○ reports;
	  ○ ongoing research/research registers.
	In addition, one or more of the following search techniques should be reported:
	• Handsearching[27]
	• Citation searches28
	• Checking websites of relevant organisations
		Please insert a quotation or a brief summary of what has been reported in the ELS	☐ Definitively appropriate
			☐ Probably appropriate
			☐ Probably not appropriate
			☐ Definitively not appropriate
			☐ Not Applicable
				Please provide the main supporting information for your appraisal

[27] Handsearching: manually scanning a publication cover-to-cover to identify all eligible reports of trials. Hand searching can also be carried out electronically by browsing through documents online. Glanville et al., Technical Manual for Performing Electronic Literature Searches in Food and Feed Safety

[28] Citation searches: identify papers in which the original paper has been cited (this allow to look forward in time from the publication of a relevant article and identify additional studies) or searching the reference list of a relevant paper (this technique is called "snowballing" and allows identifying studies published before the source paper). For further details see section 3.3.1. of Glanville et al., Technical Manual for Performing Electronic Literature Searches in Food and Feed Safety.

3. References

The ELS CAT was developed using and integrating the following documents:

EFSA (European Food Safety Authorithy), 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. 2010; 8(6):1637, 90 pp. doi:10.2903/j.efsa.2010.163

Glanville J, Wood H, Arber M, Varley D, Frampton G and Brazier H Technical Manual for Performing Electronic Literature Searches. EFSA External Report CFT/EFSA/SAS/2011/03 pp. http://www.yhec.co.uk/yhec-content/uploads/2015/03/Glanville-EFSA-search-manual-Sept-2013.pdf

Hausner E, Waffenschmidt S, Kaiser T and Simon M, 2012. Routine development of objectively derived search strategies. Syst Rev, 1, 19.

McGowan J, Sampson M and Lefebvre C, 2010. An evidence based checklist for the peer review of electronic search strategies (PRESS EBC). Evidence Based Library and Information Practice, 5, 149–154.

Sampson M, McGowan J, Lefebvre C, Moher D and Grimshaw J 2008. PRESS: peer review of electronic search strategies. Technology report number 477. Ottawa: Canadian Agency for Drugs and Technologies in Health. 41 pp. https://www.cadth.ca/media/pdf/477_PRESS-Peer-Review-Electronic-Search-Strategies_tr_e.pdf

Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D and Lefebvre C, 2009. An evidence-based practice guideline for the peer review of electronic search strategies. J Clin Epidemiol, 62, 944–952.

South Central Healthcare Librarians (review date January 2013), 2011. The literature search process: guidance for NHS researchers. pp. http://www.workforce.southcentral.nhs.uk/ pdf/ Lit_search_ protocols_ 2011.pdf