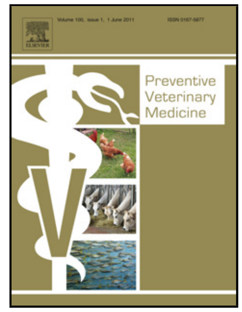
Journal Pre-proof



The RISKSUR EVA Tool (Survtool): a tool for the integrated evaluation of animal health surveillance systems

Marisa Peyre, Linda Hoinville, James Njoroge, Angus Cameron, Daniel Traon, Flavie Goutard, Clémentine Calba, Vladimir Grosbois, Alexis Delabouglise, Viktor Varant, Julian Drewe, Dirk Pfeiffer, Barbara Häsler

PII: S0167-5877(18)30102-8

DOI: https://doi.org/10.1016/j.prevetmed.2019.104777

Reference: PREVET 104777

To appear in: *Preventive Veterinary Medicine*

Received Date: 7 February 2018

Revised Date: 12 March 2019

Accepted Date: 17 September 2019

The RISKSUR EVA Tool (Survtool): a tool for the integrated evaluation of animal health surveillance systems

Marisa Peyre[a*], Linda Hoinville[b], James Njoroge[c], Angus Cameron[d], Daniel Traon[e], Flavie Goutard[a], Clémentine Calba[a], Vladimir Grosbois[a], Alexis Delabouglise[a,b], Viktor Varant[c], Julian Drewe[b], Dirk Pfeiffer[b], Barbara Häsler[f]

[a] Animal et Gestion Intégrée des Risques (AGIRs), Centre de Coopération Internationale en Recherche Agronomique pour le Dévelopement (CIRAD), Montpellier, France–, TA C 22/E Campus International Baillarguet, 343988 Montpellier Cedex 5, France

[b] Veterinary Epidemiology, Economics and Public Health Group, Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, AL9 7TA, UK

[c] Tracetracker, Prinsens gate 5 0152, Oslo, Norway

[d] AusVet Animal Health Services, Wentworth Falls, New South Wales, Australia

[e] Arcadia International E.E.I.G. Chaussée de Louvain, 1220 1200 Brussels-Belgium

[f] Veterinary Epidemiology, Economics and Public Health Group and Leverhulme Centre for Integrative Research on Agriculture and Health, Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, AL9 7TA, UK

[*]**Corresponding author**:

Email: marisa.peyre@cirad.fr

CIRAD, Campus de Baillarguet

TA C-22/E, 34938, Montpellier Cedex-5

Phone: +33467593916

Abstract

Information about infectious diseases at the global level relies on effective, efficient and sustainable national and international surveillance systems. Surveillance systems need to be regularly evaluated to ensure their effectiveness, the quality of the data and information provided, as well as to be able to allocate resources efficiently. Currently available frameworks for evaluation of surveillance systems in animal or human health often treat technical, process and socio-economic aspects separately instead of integrating them. The surveillance evaluation (EVA) tool, a support tool for the evaluation of animal health surveillance systems, was developed to provide guidance for integrated evaluation of animal health surveillance including economic evaluation. The tool was developed by international experts in surveillance and evaluation in an iterative process of development, testing and revision; taking into account existing frameworks and guidance, scientific literature and expert opinion elicitation. The EVA tool encompasses a web interface for users to develop an evaluation plan, a Wiki classroom to provide theoretical information on all required concepts and a generic evaluation work plan to facilitate implementation and reporting of outputs to decision makers. The tool was used to plan and conduct epidemiological and economic evaluations of surveillance for classical and African swine fever, bovine virus diarrhoea, avian influenza, and Salmonella Dublin in five European countries. These practical applications highlighted the importance of a comprehensive evaluation approach to improve the quality of the evaluation outputs (economic evaluation; multiple attributes assessment) and demonstrated the usefulness of the guidance provided by the EVA tool. At the same time they showed that comprehensive evaluations might be constrained by practical issues (e.g. confidentiality concerns, data availability) and resource scarcity. In the long term, the EVA tool is expected to increase professional evaluation capacity and help optimising health surveillance system efficiency and resource allocation for both public and private actors of the surveillance systems.

Keywords: animal health; assessment; disease; surveillance; evaluation; health economics; decision tool

1. Introduction

The development of efficient, effective and sustainable surveillance systems, in particular to detect emerging and exotic diseases in a timely manner, has gained importance in recent years (Anthony et al., 2012). Surveillance systems provide useful information for effective disease prevention and control thereby improving food system productivity and food security, animal welfare, economic development and access to international trade. Information about infectious diseases at a global scale relies on national and international surveillance systems. The resources, capacity and reliability of national public and/or private surveillance systems can vary considerably, especially in countries characterized by weak economies, political instability (Jebara, 2004) and/or a limited surveillance tradition. To make best use of available resources, it is critical to perform timely and relevant evaluations of surveillance systems (Drewe et al., 2012). Evaluation is an essential step in the policy cycle and allows transparent interpretation of outputs, more objective decision-making and resource allocation as well as improvements in system design and enhanced acceptance of system outputs by stakeholders (Jann and Wegrich, 2007). Given the almost continuous changes occurring in disease epidemiology and therefore in the surveillance system activities, it is essential to regularly (re-)evaluate surveillance effectiveness and efficiency taking into account surveillance system organisation, effectiveness evaluation attributes, and economic assessment criteria and methods. This requires the design of comprehensive, practical, and affordable evaluation plans for timely assessment of not only the effectiveness and efficiency of a surveillance programme but also underlying determinants (e.g. acceptability) which are linked to the system.

As described by Calba et al. (2015a) and Drewe et al. (2012), available frameworks and/or guidance for evaluation of surveillance systems in the animal and human health fields provide robust foundations, but could be expanded towards a more comprehensive, integrated approach. Indeed, none of the available guides provides a framework for a comprehensive evaluation that includes functional, process, technical and economic aspects simultaneously (Calba et al., 2015a; Calba et al., 2013a). Consequently, there is a need to integrate existing evaluation frameworks, practical methods and tools for the assessment of surveillance attributes and to provide a standardised evaluation terminology. Specific evaluation of surveillance (as opposed to the evaluation of disease interventions) has been performed only on limited occasions and a variety of approaches and methods are used without a generally agreed protocol (Drewe et al., 2012). Indeed more than 25 attributes have been described for the evaluation of animal health surveillance systems, making a complete evaluation – if all attributes are used – time-consuming and expensive. In some cases no methods have been described for the measurement of these attributes and only a fraction of these evaluation attributes have been included in evaluation process templates and in practical case studies (Drewe et al., 2011; Hoinville et al., 2011; Calba et al., 2013b).

There are always three main parts in the evaluation process of surveillance: planning, implementation and reporting (Calba et al., 2015a). Guidance and support is needed for those three parts, and especially for the definition of the evaluation plan, involving: i) the description of the surveillance system/component under evaluation; ii) the socio-economic context and the rationale for evaluation of the surveillance; iii) the definition of a precise evaluation question, and iv) the choice of evaluation attributes to be measured. The choice of evaluation attributes and methods to measure them will depend on the type of evaluation considered (e.g. evaluation of the process vs. evaluation of the outputs of surveillance) and on the surveillance system and its socio-economic context (e.g. availability of resources, level of structuration of the animal production industry, political will to develop or sustain animal breeding and production etc…) (Peyre et al., 2011 ; Peyre et al., 2017).

A recent mapping survey highlighted the fact that decision-makers in seven European countries interviewed considered economic criteria to be important in decision-making for surveillance (Haesler et al., 2014). Yet, economic evaluations of surveillance (EES) are sparse (Drewe et al., 2012; Undurraga et al., 2017; Wall et al., 2017) and available guidelines for the evaluation of surveillance fail to provide guidance on systematic economic appraisal (Calba et al., 2015a). The use of economic evaluation to inform surveillance system design has been limited so far, mainly due to a lack of appropriate guidance to allow for practical use of these methods and understanding and trust in their outputs from decision makers (Calba et al., 2015a).

The RISKSUR consortium (www.fp7-risksur.eu) investigated novel approaches for cost-effective surveillance and developed a web-based surveillance design and evaluation tool directed at users with advanced surveillance knowledge and skills. The main objective of the surveillance evaluation (EVA) tool was to develop a practical framework to guide users in planning and implementing integrated epidemiological and economic evaluations of surveillance systems. The EVA tool was developed building on existing evaluation frameworks, methods and tools taking into account input from expert meetings and discussions. The RISKSUR surveillance design framework complements the EVA tool to support the design, review and documentation of surveillance systems (Dorea et al., 2017). The EVA tool development process, characteristics and application using practical case studies are described and discussed in this paper.

2. Methods

A five stage process was used to develop the RISKSUR EVA tool from the initial development of the evaluation approach to its validation and refinement: i) technical workshops of international surveillance experts (researcher and users) to develop and agree on a first conceptual model including key elements of an evaluation plan; ii) expert opinion elicitation to review and score the relevance of evaluation attributes and the methods to assess them; iii) application of the model to the evaluation of practical case studies (referred to as development case studies); iv) development of the web tool; v) application of the web tool to case studies (referred to as validation case studies of the feasibility and operationally of the evaluation model and tool). This process is described in the following sections.

2.1 Technical workshop to develop the conceptual model

A one day technical workshop was held in June 2013 (Montpellier, France), attended by 15 experts from six European countries. The experts were selected based on their field of expertise in surveillance, surveillance design, evaluation, economic evaluation and surveillance evaluation (Figure 1). The objectives of the workshop were to review and select the key elements to be included in an integrated evaluation framework looking at the system performance, process and value/impact (Figure 2). At the end of the workshop, a conceptual model of the EVA tool was proposed and a first version of the EVA tool was subsequently developed by the enlarged RISKSUR consortium (including the initial group of experts) based on this model (RISKSUR Consortium, 2013).

2.2 Expert opinion elicitation process

Two independent expert opinion elicitation rounds were implemented among the experts involved in the first technical workshop (Figure 1) to i) define the list of evaluation attributes and their relevance level according to defined surveillance contexts; ii) identify and validate attribute assessment methods. This process is described below:

Evaluation attribute list and relevance

A list of evaluation attributes was developed by the RISKSUR project team based on data retrieved from the literature (List 1) (Calba et al., 2013b; Drewe et al., 2012). A three stage expert opinion elicitation process was then conducted (Figure 3). First, a technical workshop with 11 experts was held to review the evaluation attribute list including definitions and to identify the attributes only relevant for the evaluation process (Bilal, 2001) (Figure 4) resulting in List 2. Second, based on the outputs of the first technical workshop, a second workshop with 15 experts was held to define each attribute relevance level according to a specific evaluation context (i.e. a combination of evaluation objective and evaluation question). The relevance levels were defined in a qualitative manner using 3 categories: low, medium and high relevance. The experts were asked to justification for their choices. The inputs were compiled and reviewed by the project team to identify discrepancies between experts and to produce a generic statement on the relevance level of each attribute. Four different degrees of agreement/disagreement were identified: 1) full agreement; 2) moderate disagreements; 3) disagreement; 4) strong disagreement. Finally, a third technical workshop with the same 15 experts was held subsequently to reach a consensus on the attributes with strong disagreement and disagreement to generate a final list.

Attribute assessment methods (including economic analysis)

A literature review was conducted to identify available evaluation attribute assessment methods including economic analysis techniques. A specific search algorithm was used to retrieve methods for each evaluation attributes, combining keywords for the attribute and for methods, e.g. (sensitivity + (methods OR tools)). Information on the type of method (qualitative, semi-quantitative, quantitative), the purpose of use, the target users, a descriptive summary, data and expertise requirements and strengths and limits were retrieved from the published literature. This information was compiled by the RISKSUR project team and sent out for review to the experts who have developed the method and/or applied it in the field, to ensure that the compiled data were valid and relevant.

2.3. Development case studies

The first version of the EVA tool was applied to six development case studies (Table 1). The case studies were selected to ensure representativeness of the different surveillance objectives; target species and hazards under surveillance. The development case studies allowed to test the logic of the tool and to develop it further. They also provided data on the evaluation attribute list, their degree of relevance according to a specific context, attribute assessment methods and challenges linked to specific economic assessment techniques. The information was part of an iterative process of framework development and relevant feedback was included in the expert opinion elicitation processes described above (Figure 3).

2.4. EVA Tool web application

An online web version of the EVA tool was developed by Tracetracker Ltd. The online tool was linked to a Wikispace Classroom application that provides theoretical evaluation concepts and EVA tool user manual. Wikispace is a social writing platform for education, which is free and allows providing training type information to be shared and enriched by professionals (http://www.wikispaces.com/content/classroom/about). The operationality of the web tool was validated using the validation case studies described below.

2.5 Validation case studies

Three case studies were selected to perform an integrated epidemiological and economic evaluation and thereby validate the feasibility, operationally and usefulness of the EVA web tool: early detection of avian influenza in the UK; freedom from Classical Swine

Fever in wild boars in Germany; and case detection of salmonella Dublin in cattle in Sweden. Those validation case studies allowed assessing the functionality and interest of the EVA tool web application.

3. Results

3.1 Outputs from technical expert workshops

The EVA tool framework was developed based on other existing frameworks, guidelines and methods available in the literature (Calba et al., 2015). During the first technical workshop, the experts identified all common critical elements and essential evaluation steps from those guidelines to be included in the EVA tool with the aim to provide an harmonised approach to surveillance evaluation based on the validated and available reference guides (including guidelines from Center for Diseases Control (CDC), WHO and OIE (Calba et al.2015)).

To ensure framing of the evaluation, critical elements of the evaluation context need to be captured to be able to define the specific evaluation question. The expert workshops highlighted the critical importance of the evaluation context (especially the surveillance objective and the evaluation needs) and the evaluation question to define the relevance of evaluation attributes to be included in the evaluation process. The conceptual model of the EVA tool addressed these needs by defining the four fundamental steps: what is my situation; why am I doing an evaluation; what to evaluate and how.

The experts identified thirteen critical elements of the context (surveillance system and evaluation needs) that were deemed essential to frame the evaluation, define the evaluation question, analyse and discuss the outputs of the evaluation (Table 2).

A list of 11 evaluation questions were defined by the experts to account for diverse evaluation needs (Table 3). A decision tree pathway was also developed to assist the user with the choice of the evaluation question. In this pathway, the users are guided through a series of questions (eleven in the longest pathway) to define their evaluation priorities (e.g. system or component evaluation; previous knowledge of effectiveness; need for economic analysis) and to identify the most relevant evaluation question (. At the end of the pathway, the user is directed to the evaluation question list and the tool will pre-select the appropriate question in order to assist the evaluators in their final choice. Users who feel comfortable with the selection of the evaluation question, can go directly to the list with evaluation questions.

3.2 General overview of the EVA tool framework

The EVA tool is freely available online (http://webtools.fp7-risksur.eu) and is shared under the principles of the a non commercial Creative Commons licence 2017 (i.e. the tool can be freely used and shared for any non-commercial purposes but appropriate credit should be given, providing link to the licence and changes made should be indicated). The tool is linked to the surveillance evaluation Wikispace (http://surveillance-evaluation.wikispaces.com) which is also freely available upon registration as a member.

The tool has been organized into three main sections to capture all the elements critical to an evaluation process and highlighted by the experts during the iterative development process of the tool (Figure 5): section 1) a general introduction to the tool and essential information on evaluation concepts, including evaluation attributes and economic methods to promote the understanding of the evaluation process and economic evaluation; section 2) guidance on how to define an evaluation plan based on Steps 1 and 2 with data entry on the evaluation context and the evaluation question and Step 3 and 4 where the tool facilitates the selection of relevant evaluation attributes and assessment methods (including economic analysis); and section 3) guidance on how to perform the evaluation and how to report the outputs of the evaluation to decision makers.

3.3 Relevance of evaluation attributes

A total of 19 evaluation attributes were included in the final list consolidated within the RISKSUR project team (Table 4). The differences in relevance of evaluation attributes mainly depended on the surveillance objective (e.g. early detection; freedom from disease; case finding), the evaluation question (e.g. value attributes, organisational attributes) and in some situations on the surveillance design (e.g. risk-based surveillance) (the full table of attribute relevance could be accessed here: https://survtools.org/wiki/surveillance-evaluation/doku.php?id=evaluation-attributes-selection-process).

From the second stage of expert consultation process full agreement on relevance level of three attributes was reached (acceptability; precision; simplicity) and moderate disagreement for four attributes (negative predictive value (NPV); positive predictive value (PPV); sensitivity; risk-based criteria). Disagreement was observed for seven attributes: availability & sustainability; cost; compatibility; false alarm rate; multiple hazard; representativeness; timeliness. Strong disagreement was only observed for two attributes: bias and coverage. Disagreements between experts were observed for two other attributes (robustness and surveillance system organization) but these were caused by misunderstanding of the two attribute definitions. The definitions were subsequently revised. A consensus between experts was reached during the last stage of the expert consultation process to produce the final list presented in Table 4.

3.4 Guidance on the evaluation attribute assessment methods

A list of 70 different methods and/or specific applications of a method were retrieved from the scientific literature. Their characteristics including advantage, limits and competences required to apply the methods were validated by the relevant experts and included in the EVA tool and the Wikispace. The number of methods validated for each evaluation attribute is indicated in Table 4.

Novel methods which were developed as part of the RISKSUR project to assess the risk-based definition criteria (EVA Risk); acceptability and engagement and benefits (AccePT method) and effectiveness were also included in the EVA tool (Calba et al., 2015b; Grosbois et al., 2014).

3.5 Guidance on economic evaluation concepts and methods

The EVA tool further promotes the understanding and use of economic evaluation by explaining relevant economic theory and challenges underpinning economic evaluation of surveillance. The relationship between surveillance, intervention and loss avoidance along with the value of information, and non-monetary benefits are described and linked to economic analysis methods commonly used in animal health. In order to promote best practices in economic evaluation of surveillance, guidance and practical information on economic evaluation is provided both in the tool itself and the Wikispace. A series of relevant questions that allow defining an economic evaluation question has been developed to help frame the evaluation context and the evaluation questions according to this context. Out of the 11 evaluation questions defined in the tool, 5 are economic evaluation questions covering three common types of economic evaluation methods: least-cost assessment, cost-effectiveness and cost-benefit analysis (Table 3). These economic analysis techniques are listed and described in the tool and linked to the economic evaluation methods described in detail in the evaluation Wikispace.

3.6 Guidance on how to report the evaluation outputs back to the decision makers

Detailed guidance and a roadmap on how to report the evaluation outputs to decision makers has been integrated in the EVA tool and evaluation Wikispace.

3.7 The EVA Wikispace: a dynamic platform on evaluation concepts and guidance

The EVA Wikispace was developed to gather and share extensive information and references/links to support the successful use and further development of the EVA tool (http://surveillance-evaluation.wikispaces.com). This information sharing space allows engaging the scientific community by allowing users to edit and add information and therefore ensure relevance of the tool by updating it with the latest developments in the field of animal health surveillance evaluation. The EVA wiki is organised in a similar way as the EVA tool but provides additional sections on important elements of evaluation and economic evaluation concepts along with background and practical information on the EVA tool and application examples.

3.8 Application of the EVA tool to case studies

The application of the tool for economic evaluation of surveillance for classical and African swine fever, bovine virus diarrhoea, avian influenza, and Salmonella Dublin in five European countries provided important feedback on the relevance, functionality, advantages, feasibility and limits of the EVA tool for surveillance evaluation.

All evaluation questions selected were deemed feasible and could be addressed using available methods and data.

For each case study, 4-9 evaluation attributes were identified by the EVA tool as highly relevant for the evaluation and the users included 2-9 in their evaluations (Table 1). This choice was reported to be mainly due to practical and timing issues (e.g. time to collect additional data to assess acceptability and engagement). All case studies conducted an assessment of the costs in comparison to one or more effectiveness criteria; one case study translated the effectiveness measures into a monetary benefit for inclusion in a cost-benefit analysis. Because all case studies looked a new designs to either complement or replace old designs, the analyses were prospective / *ex ante*. Users reported difficulties in the estimation of fixed and variable costs, non-monetary benefits, co-benefits resulting from using synergies, and the selection of meaningful effectiveness measures for inclusion in economic analysis.

Interestingly the limits identified in the case studies were linked to the application of the evaluation method rather than the use of the tool itself to define the evaluation process.

4. Discussion

The EVA tool was developed to provide practical guidance on how to design integrated evaluation protocols for surveillance, conduct an evaluation and how to communicate the findings to facilitate decision-making.

The EVA tool provides a practical evaluation framework, which guides users on the implementation of the evaluation and provides essential elements for the interpretation of the results. Within the RISKSUR project a complementary tool (surveillance design framework) was also developed to support design or re-design of a surveillance system (Dorea et al., 2018 under publication). As for the EVA tool, the design framework does not take decisions for the users but provides specific guidance to facilitate the design or re-design of surveillance system according to the user's specific needs. The design framework is also complemented by a web interface and a Wikispace classroom (http://surveillance-design-framework.wikispaces.com). The combined set ofRISKSUR tools covers all the essential steps in the decision making cycle for strategic planning of animal health surveillance (design – evaluation – re-design) (Dorea et al., 2017). Itpromotes understanding of critical concepts, suitable methods, data and time requirements and is expected to nurture the use of economic evaluation of surveillance, which is still in its infancy (Haesler et al., 2015).

The evaluation question is the most important aspect of the evaluation process. Evaluation is intrinsically linked to action; it makes little sense, and is of limited interest, to perform an evaluation without a specific objective for action or at least the willingness to consider action (the outcome may be decide that no action is currently needed). In order to guide the evaluator in the selection of an appropriate evaluation question, a list of evaluation question was developed along with a selection guidance pathway and integrated within the

EVA tool. However, this list might not be exhaustive and could be reviewed based on feedback from users of the tool and/or comments made on the EVA wiki.

Until recently, recommendations on the choice of attributes to evaluate animal health surveillance systems have been based on case study application and methodological experience from public health evaluation (Calba et al., 2015a). In 2011, Hoinville et al. (2013) provided a comprehensive list of evaluation attributes relevant to the evaluation of animal health surveillance as an output of the first International Conference on Animal Health Surveillance (ICAHS 2014). Drewe and al. (2015) provided an attribute selection matrix to aid with ranking of evaluation attributes according to the surveillance objective of the system under evaluation. However, these studies only provided limited information on the relevance of the evaluation attributes according to a specific context. Indeed ranking of evaluation attributes was shown to be a challenging process as it depends on many factors and degree of interactions (Peyre et al., 2014). Within the RIKSUR project and with the development of the EVA tool, we further contributed to this work by i) identifying which attributes of the system are important for the evaluation process rather than for the design process; ii) identifying the contextual factors impacting on the priority of evaluation attributes; iii) assessing the links between the attributes, and iv) by promoting the use of a comprehensive list of evaluation attributes with expert defined relevance level rather than a selection of attributes. In order to ensure maximum flexibility of the decision support tool without withdrawing information from the user and account for the difficulties in reaching expert consensus during the process, it was decided that the choice of the evaluation attributes to be included in the evaluation process will ultimately be determined by the user, but the tool provides some basic suggestions that can be considered by the user and overridden manually if necessary.

A key innovative feature of the EVA tool is the provision of user-friendly and practical guidance to support the design and conduct of economic evaluation of surveillance. Economic theory underpinning economic evaluation of surveillance is explained and challenges highlighted that accrue from application of differing paradigms. In particular, the three-variable relationship between surveillance, intervention and loss avoidance; value of information, and non-monetary benefits are elaborated and linked to economic analysis methods commonly used in animal health. We identified and explained the use of the most common economic evaluation criteria according to the different surveillance objectives and evaluation questions. This represents an added value in the guidance to decision maker (technical advisers) to facilitate/promote the use of economic evaluation.

The tool has also been developed as a collaborative tool to enable regular update by users and to ensure its sustainability and relevance over time.

Evaluation itself is only a means to an end: it is a tool that helps to see what is happening so that the surveillance system can be improved. Evaluation aims to generate a reflexion to promote changes. The purpose of evaluations is to provide feedback to decision makers about program operations and their (cost-)effectiveness so that their decisions can be as fully informed as possible. The ad hoc evaluation exercise is completed by a deep analysis of the results which are placed in the global context of policy and/or operational interventions. Potentially this analysis would lead to the identification of improvement measures at different levels. Experienced administrators and evaluators know that this does not often happen. Evaluations may be undertaken because they are required, and the ad hoc evaluation reports are subsequently not analysed in full details. This may occur for several reasons, including: failing to address directly the policy makers' or program administrators' principal questions (wrong selection of the evaluation question); lack of communicating the evaluation results in a way that can be readily understood by non-evaluation experts (what to communicate?); lack of clear understanding of which the primary and secondary audiences of the results are (who to target?); not matching the results of the evaluation with decision makers' planning during which policy or programmatic operational decisions are made (when to target)?; evaluation findings may be perceived as too challenging to implement by stakeholders if no preparatory work is associated to the evaluation results. This could lead to resistance in implementing changes (are proposal subject to high acceptability and appropriation?).

The EVA tool also promotes the application of an integrated evaluation process. The tool generates a balanced suggestion of valuation attributes and measurement methods to assess not only effectiveness (e.g. sensitivity) but also functional aspects influencing the overall performance of a surveillance system (e.g. acceptability, flexibility) and economic efficiency. The functional attributes are of critical importance to generate meaningful recommendations for all stakeholders (Figure 2). The purpose of an evaluation and the research that goes into it is not just to tell whether or not the surveillance has been a success or not. The real value of evaluation lies in its ability to help identifying and correct problems – as well as to celebrate progress. Further reflection on how to make the surveillance even better and more effective is still required. The results for process and impact should be analysed, and changes made where they will gain greater effectiveness or efficiency. This integrated approach should ensure uptake of the evaluation outcomes by helping the technical adviser to position them back into the complex process of decision-making.

The application of the EVA tool to practical case studies highlighted the importance of considering comprehensive evaluation to improve the quality of the evaluation outputs (economic evaluation; multiple attributes assessment); and at the same time identified practical issues and resource constraints to do so. The cost-effectiveness analysis (CEA) of both Salmonella Dublin and Bovine Virus Diarrhoea in cattle demonstrated the challenges associated with interpreting these kinds of outcomes. For example, the results for Salmonella Dublin demonstrated that one surveillance design was cheaper than the other one in detecting cases, but it was not clear what the value was of one detected case and consequently how much money could be invested to detect these cases.

Similarly, in the BVD case study, decision makers could not provide clear information on what level of effectiveness would be desirable. While the analysis provided information on the prevalence, distribution, and risk factors, it was difficult to judge whether the estimated accuracy generated enough economic value to recover the additional costs related to coordination and centralisation of data..

The CSF case study highlighted the importance of considering more than one evaluation attributes to provide meaningful results and to discriminate between the different surveillance designs under evaluation. Indeed, most of surveillance designs (including the current one) reached the target effectiveness value defined in terms of surveillance system sensitivity. However, the timeliness, simplicity and acceptability differed between the different designs under evaluation. The combined analysis of all these different attributes allowed identifying the most effective and least-cost design (Schulz et al., 2017). The findings from the case studies illustrated limitations in terms of CEA of surveillance and identified common pitfalls. The feedback was used to underline the importance of reflecting carefully on the attributes included in the CEA and to ask oneself what the outputs will mean in terms of value and whether they will help to make a recommendation to decision-makers from an economic point of view. Consequently, further information and references were added to the Wikispace to explain relevant concepts in more detail. This feedback was important to be able to refine the tools and provide further guidance for users. Indeed, by making the evaluator aware of essential information on the limitations of the process (i.e. what could or should be done), the robustness of the evaluation can be increased by generating higher confidence of decision makers in the evaluation outputs and recommendations.

Conclusion

The EVA tool was developed to integrate different evaluation dimensions in a structured way and to guide the users in the development and implementation of their evaluation plans for surveillance. The objective of the tool was to promote the use of comprehensive evaluation including economic evaluation by providing detailed information on the available methods and relevance according to a specific evaluation question and context. As such, the EVA tool contributes to the implementation of robust and standardised evaluations of surveillance activities and thereby helps to produce evidence-based information relevant for surveillance decision-makers. This in turn promotes data quality and stakeholder trust in animal health status of one country. In the long term, this will increase professional capacity and help to address the problem of resource allocation for surveillance to the benefit of all.

The final step will be to update periodically the EVA tool based on feedback from future users which has been made possible by using a collaborative wiki web platform.

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Frame-work Programme for research, technological development and demonstration under grant agreement N° 310806 (RISKSUR). The authors would like to acknowledge all the experts of the RISKSUR consortium who provided valuable support and information, especially Katharina Staerk, Betty Bisdorff, Birgit Schauer, Katja Schulz, Christophe Staubach, Ann Lindberg, Fernanda Dorea, Arianna Comin and Timothée Vergne.

References

Anthony, S., Fauci, M.D., Morens, M.D., 2012. The Perpetual Challenge of Infectious Diseases. N Engl J Med; 366:454-°©- 461February 2, 2012DOI: 10.1056/NEJMra1108296.

Bilal, M. A. (2001) Elicitation of expert opinions for uncertainty and risks, 1sted. CRC Press, Boca Raton, Florida.

Calba, C., Comin, A., Conraths, F., Drewe, J., Goutard, F., Grosbois, V., Häsler, B., Horeth-Bontgen, D., Hoinville, L., Lindberg, A., Peyre, M., Pfeiffer, D., Rodriguez Prieto, V., Rushton, J., Staerk, K., Schauer, B., Traon, D., Vergne,, T.,2013a. Evaluation methods of surveillance systems and current practices. RISKSUR Deliverable 1.2. Available at: http://www.fp7-risksur.eu/progress/public-deliverables. Accessed on 29th September 2017.

Calba, C., Drewe, J., Goutard, F., Grosbois, V., Häsler, B., Hoinville, L., Peyre, M., Pfeiffer, D., Vergne, T.,2013b. The evaluation attributes used for evaluating animal health surveillance systems. RISKSUR Deliverable 1.3. Available at: http://www.fp7-risksur.eu/progress/public-deliverables. Accessed on 29th September 2017.

Calba, C., Cameron, A., Goutard, F., Grosbois, V., Häsler, B., Hoinville, L., Lindberg, A., Pinto, J., Peyre, M., Pfeiffer, D., Traon, D., Vergne, T.,2013c. The EVA tool : an integrated approach for evaluation of animal health surveillance systems. RISKSUR

Deliverable 1.4. Available at: http://www.fp7-risksur.eu/progress/public-deliverables. Accessed on 29th September 2017.

Calba, C., Goutard, F., Hoinville, L., Hendrikx, P., Lindberg, A., Saegerman, C., Peyre, M., 2015a. Surveillance systems evaluation: a review of the existing guides. BMC Public Health; 15:448. DOI: 10.1186/s12889-015-1791-5.

Calba, C., Antoine-Moussiaux, N., Charrier, F., Hendrikx, P., Saegerman, C., Peyre, M., Goutard, F. L., 2015b. Applying participatory approaches in the evaluation of surveillance systems: A pilot study on African swine fever surveillance in Corsica. Preventive and Veterinary Medicine 122, 389–398. doi:10.1016/j.prevetmed.2015.10.001.

Drewe, J., Hoinville,L., Cook, A., Floyd,T., Stärk, K., 2012. Evaluation of animal and public health surveillance systems: a systematic review. Epidemiology and infection, 140:575-590.

Drewe, J.A., Hoinville, L.J., Cook, A.J.C., Floyd, T., Gunn, G. and Stärk, K.D.C., 2015. SERVAL: A new framework for the evaluation of animal health surveillance. *Transboundary and Emerging Diseases* 62: 33-45

Häsler, B. Howe K., Peyre M., Vergne T., Calba C., Bisdorff B., Comin A., Lindberg A., Brouwer A., Snow L., Schulz K., Staubach C., Martínez-Avilés M., Traon D., Hoinville L., Stärk K., Pfeiffer D., Rushton J. Economic evaluation of animal health surveillance – moving from infancy to adolescence? ISVEE 2015, November 3-7, 2015 Merida, Yucatan, Mexico.

Häsler, B., et al. 2014. Mapping of surveillance and livestock systems, infrastructure, trade flows and decision-making processes to explore the potential of surveillance at a systems level. 2nd ICAHS conference, Cuba, May 2014.

Hoinville, L.J., Alban, L., Drewe, J.A., Gibbens, J.C., Gustafson, L., Häsler, B., Saegerman, C., Salman, M., Stärk, K.D.C., 2013. Proposed terms and concepts for describing and evaluating animal-health surveillance systems, Preventive Veterinary Medicine, Volume 112, Issues 1–2, Pages 1-12, http://dx.doi.org/10.1016/j.prevetmed.2013.06.006.

Jann, W., Wegrich, K., 2007. Theories of the policy cycle. In: Handbook of public policy analysis: theory, politics, and methods, Public administration and public policy.

Jebara, K.B., 2004. Surveillance, detection and response: managing emerging diseases at national and international levels. Rev Sci Tech. 23(2):709-15.

Peyre, M., Hendrikx, P., Do Huu, D., Goutard, F., Desvaux, S., Roger, F., others, 2011. Evaluation of surveillance systems in animal health: the need to adapt the tools to the contexts of developing countries, results from a regional workshop in South East Asia. Epidémiologie Santé Anim. 59, 415–417.

Peyre, M., Hoinville, L., Haesler, B., Lindberg, A., Bisdorff, B., Dórea, F., Wahlström, H., Frössling, J., Calba, C., Grosbois, V., Goutard, F., 2014. Network analysis of surveillance system evaluation attributes: a way towards improvement of the evaluation process. In Proceedings of the 2nd International Conference on Animal Health Surveillance (ICAHS2), The Havana, Cuba, 7-9 May.

Peyre, M., Haesler, B., Goutard, F., 2015. Case study selection for economic evaluation framework development and validation. RISKSUR Deliverable 5.20. Available at: http://www.fp7-risksur.eu/progress/public-deliverables. Accessed on 4th January 2016.

http://www.animalhealthsurveillance.org/index.php?n=Main.ConferenceProgram

Peyre, M.-I., Pham, H.T.T., Calba, C., Schulz, K., Delabouglise, A., Goutard, F.L., Roger, F., Antoine-Moussiaux, 2017. Animal health surveillance constraints in North and South:

same-same but different? [WWW Document]. Proc. ICAHS - 3rd Int. Conf. Anim. Health Surveill. Beyound Anim. Health Surveill. Rotorua New-Zealand 1-5 May 2017.

RISKSUR Consortium, 2013. The Eva tool: an integrated approach for the evaluation of animal health surveillance systems. Research brief, No. 1.4. Available at: http://www.fp7-risksur.eu/progress/public-deliverables. Accessed on 29th September 2017.

RISKSUR Consortium, 2014. Mapping of surveillance systems, animal populations, trade flows, critical infrastructure and decision making processes in several European countries. Research brief, No. 1.2. Available at: http://www.fp7-risksur.eu/progress/public-deliverables. Accessed on 29th September 2017.

Schulz, K., Peyre, M., Staubach, C., Schauer, B., Schulz, J., Calba, C., Häsler, B., Conraths, F.J., 2017. Surveillance strategies for Classical Swine Fever in wild boar – a comprehensive evaluation study to ensure powerful surveillance. Sci. Rep. 7. https://doi.org/10.1038/srep43871

Undurraga, E.A., Meltzer, M.I., Tran, C.H., Atkins, C.Y., Etheart, M.D., Millien, M.F., Adrien, P., Wallace, R.M., 2017. Cost-Effectiveness Evaluation of a Novel Integrated Bite Case Management Program for the Control of Human Rabies, Haiti 2014-2015. Am. J. Trop. Med. Hyg. 96, 1307–1317. https://doi.org/10.4269/ajtmh.16-0785

Wall, B.A., Arnold, M.E., Radia, D., Gilbert, W., Ortiz-Pelaez, A., Stärk, K.D., Klink, E.V., Guitian, J., 2017. Evidence for more cost-effective surveillance options for bovine spongiform encephalopathy (BSE) and scrapie in Great Britain. Eurosurveillance 22, 30594. https://doi.org/10.2807/1560-7917.ES.2017.22.32.30594

Figure captions

Figure 1. Field of expertise and number of experts per field involved in the EVA tool development process.

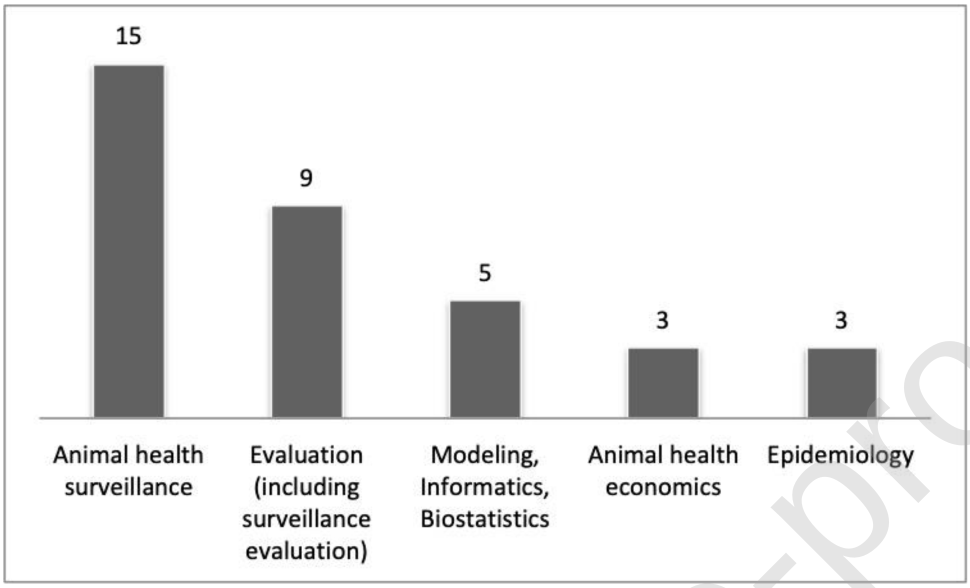


Figure 2. Scale and complexity of different levels of health surveillance sytsem evaluation: technical (looking at the performances of the system); process (looking at the factors affecting system performances); comprehensive (looking at the value of the system).

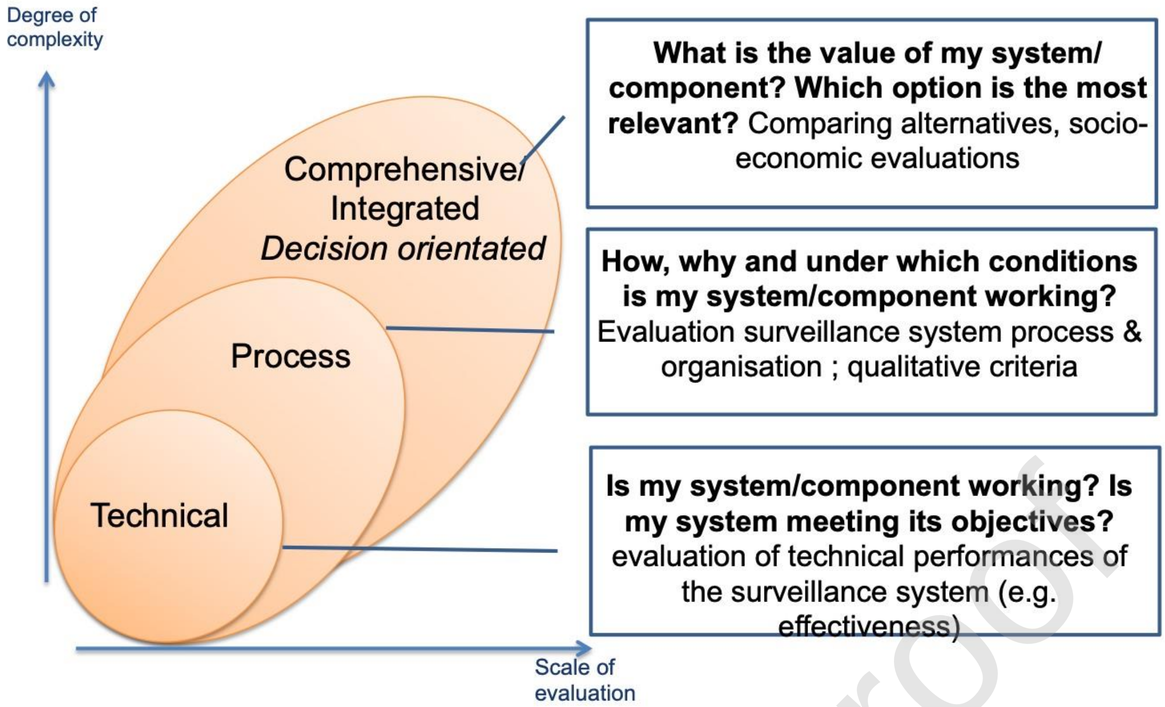


Figure 3. Expert opinion process that was used to define and agree on the list and relevance of evaluation attributes.

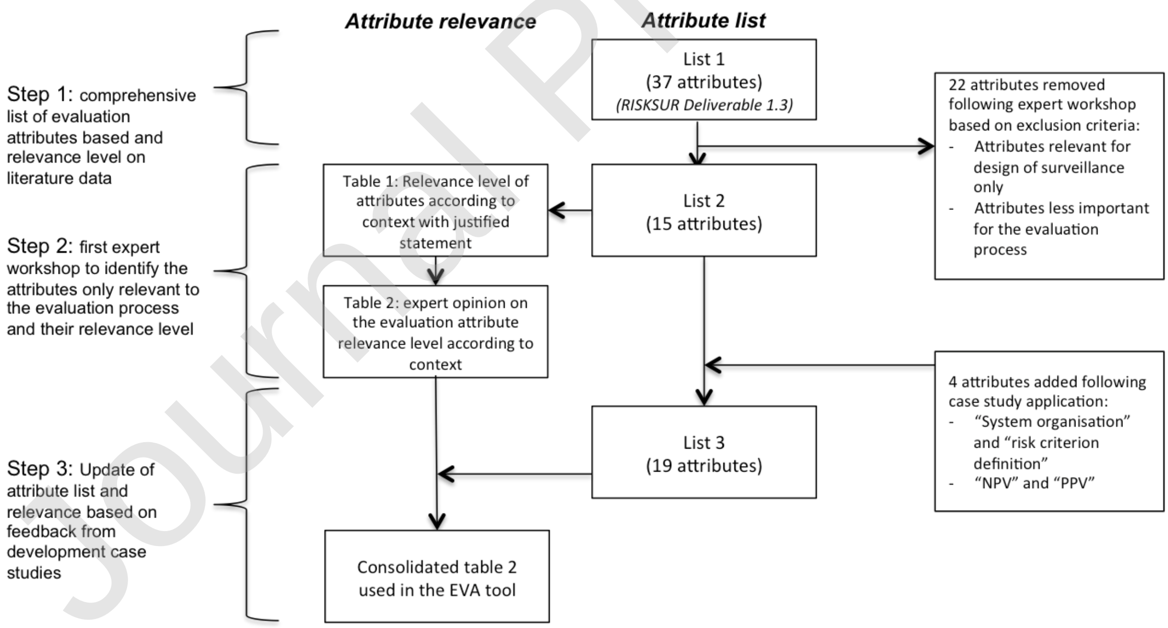


Figure 4. Evaluation process cycle, adapted from the better evaluation initiative rainbow framework (http://betterevaluation.org)

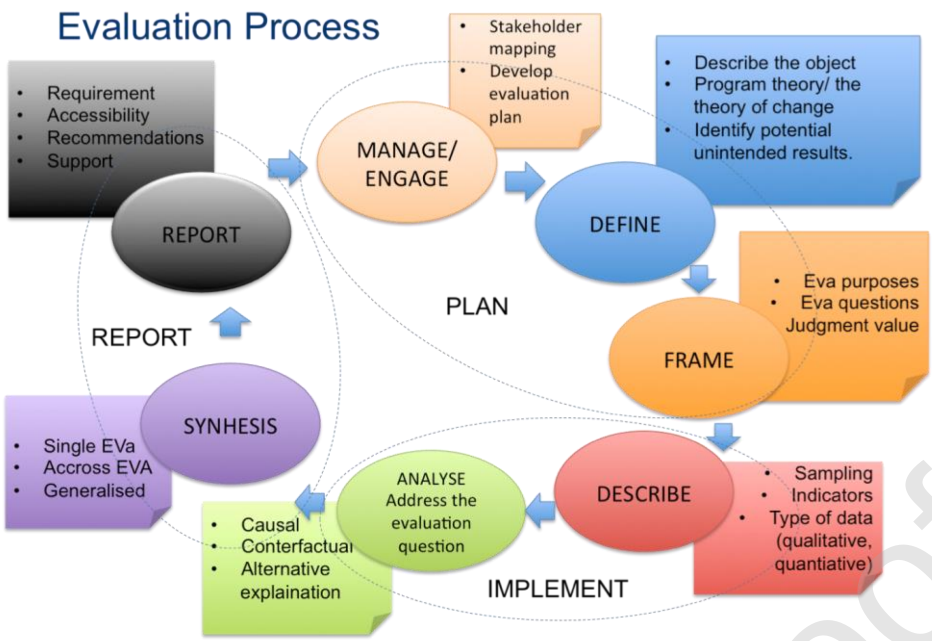


Figure 5. General organisation of the EVA tool: section 1) general introduction to evaluation concepts and economic methods; section 2) guidance on how to define an evaluation plan; and section 3) guidance on how to perform the evaluation and how to report the outputs of the evaluation to decision makers.

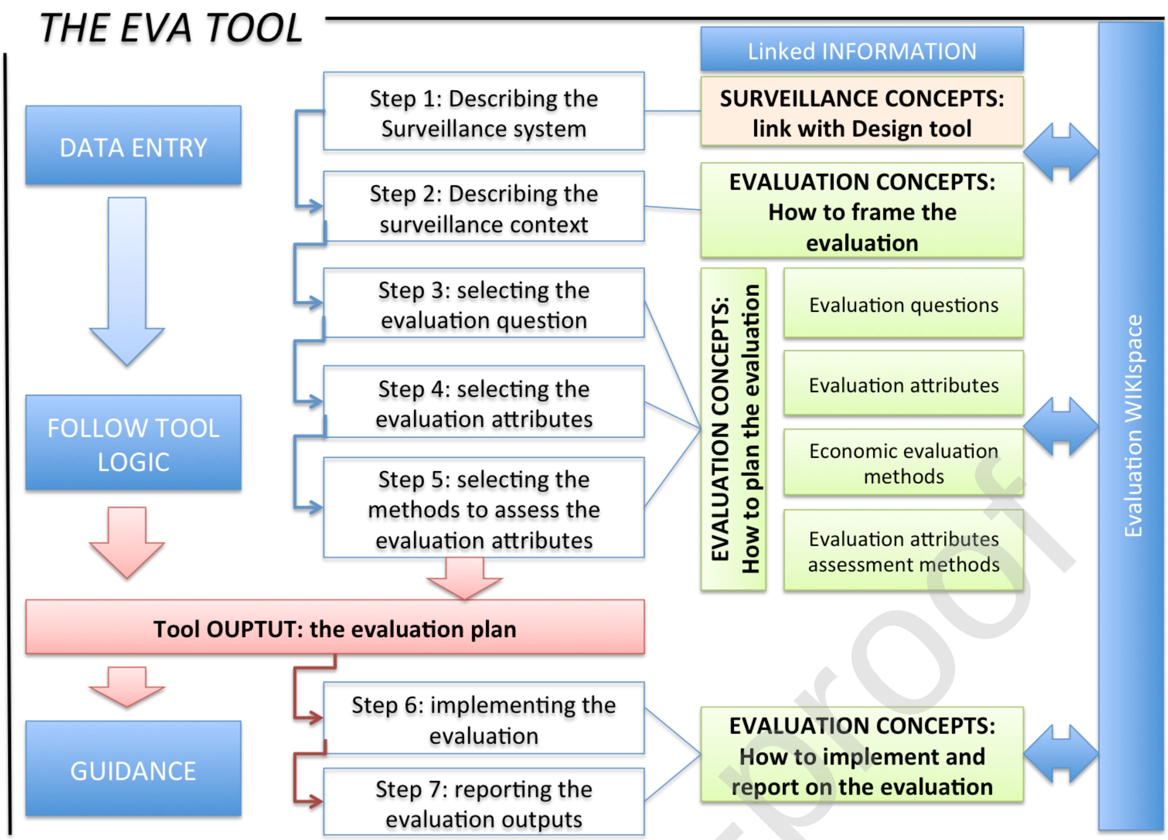

THE EVA TOOL
Linked INFORMATION
DATA ENTRY
Step 1: Describing the Surveillance system
SURVEILLANCE CONCEPTS: link with Design tool
Step 2: Describing the surveillance context
EVALUATION CONCEPTS: How to frame the evaluation
Step 3: selecting the evaluation question
Step 4: selecting the evaluation attributes
Step 5: selecting the methods to assess the evaluation attributes
FOLLOW TOOL LOGIC
EVALUATION CONCEPTS: How to plan the evaluation
Evaluation questions
Evaluation attributes
Economic evaluation methods
Evaluation attributes assessment methods
Evaluation WIKIspace
Tool OUPTUT: the evaluation plan
GUIDANCE
Step 6: implementing the evaluation
Step 7: reporting the evaluation outputs
EVALUATION CONCEPTS: How to implement and report on the evaluation

Table 1. Overview of the case studies applied to develop and validate the EVA tool

Case studies	Case detection of salmonella Dublin in cattle in Sweden	Early detection of avian influenza in the United Kingdom (UK)	Case detection of bovine virus diarrhoea virus in the UK	Demonstrate freedom from Classical Swine Fever in wild boars in Germany	Demonstrate freedom from bluetongue in ruminants in Germany	Measuring prevalence of highly pathogenic avian influenza in Egypt
Hazard under surveillance*	*Salmonella*	AI	BVD	CSF	BT	HPAI
Target species	cattle	laying hens	cattle	wild boar	ruminants	poultry
Surveillance goal:						
Case finding	✓	–	✓	–	–	-
Demonstrate freedom	–	–	–	✓	✓	-
Early detection	–	✓	–	–	–	-
Prevalence estimate	✓	–	✓	–	–	✓
Level	country	country	country	region	country	country
Surveillance structure:						
multi-component	✓	✓	✓	–	✓	✓
single component	–	–	–	✓	–	-
Use of the case study:						
EVA tool development	✓	✓	✓	✓	✓	✓
EVA tool validation	✓	✓	-	✓	-	-
Organisational attribute						
Surveillance system Organisation	✓	✓	✓	✓		✓
Functional attributes evaluated:						
Acceptability	–	–	✓	✓	–	–
Availability	–	–	✓	–	–	–
Engagement	–	–	✓	–	–	–
Simplicity	–	–	✓	–	–	–
Sustainability	–	–	✓	–	–	–
Performance attributes evaluated:						
Coverage	–	–	✓	–	–	–
Detection fraction	✓	–	–	–	–	–
Precision	–	–	✓	–	–	–
Sensitivity (other than	–	✓	–	✓	–	✓

detection fraction)						
Timeliness	−	✓	−	✓	−	−
Economic attributes evaluated:						
Cost	✓	−	✓	✓	−	−
Economic efficiency	✓	✓	−	✓	−	−

*PRRS = Porcine Respiratory and Reproductive Syndrome, AD = Aujeszky's Disease, CSF = Classical Swine Fever, AI = Avian Influenza, ASF = African Swine Fever, BVD = Bovine Viral Diarrhoea, BHV1 = Bovine Herpes Virus 1, BT = Bluetongue

Table 2. List of evaluation context elements included in the EVA tool and their relevance in the framing of the evaluation process

Context elements	Relevance
Surveillance objective	Impacts on the selection of evaluation attributes
Hazard name	Provides information about the disease under evaluation which will impact the complexity of the evaluation (e.g. between animal disease and zoonotic diseases)
Geographical area	Provides information about the scale of the evaluation
Legal requirements	Provides information about the need to meet an effectiveness target or not
Strengths and weaknesses of the current approach	Provide summary information about the rationale behind the decision to evaluate - help the evaluator to frame the evaluation question
Stakeholder concerns about current approach	Provide information about the involvement and interest of decision makers in the evaluation process - help the evaluator to frame the evaluation question
Alternative strategies to consider	Provides information about the type of evaluation required (based on a counterfactual or not)
Do you want to evaluate or change the system or some components in the	Provide information about the level of evaluation

system ?	
How many components will you include in this evaluation?	Provides information about the number of counterfactual considered
Are you considering risk-based options?	Relevant for the inclusion of the attribute risk-based criteria definition in the evaluation plan
Will you consider the costs of surveillance in your evaluation?	Provides information about the interest of economic evaluation
Do you know the current cost of your system and/or components?	Provides information about the data required
Do you have a budget constraint?	Provides information for the economic evaluation (meeting a budget target or not)

***Table 3**. List of evaluation questions developed within the EVA tool and evaluation criteria and methods linked to each question*

Evaluation question	Evaluation criteria	Evaluation method
Evaluation at the component level		
Q1. Assess whether one or more surveillance component(s) is/are capable of meeting a specified technical effectiveness target	Effectiveness	Effectiveness attribute assessment
Q2. Assess the technical effectiveness of one or more surveillance components		
Q3. Assess the costs of surveillance components (out of two or more) that achieve a defined effectiveness target, where effectiveness is already known	Effectiveness Cost	Least cost assessment
Q4. Assess the costs and effectiveness of surveillance components (out of two or more) to determine which achieves a defined effectiveness target at least cost, the effectiveness needs to be determined		
Q5 –Q7. Assess whether a surveillance component generates a net benefit, the biggest net benefit or the the biggest under a budget constraint for society, industry, or animal holder(s):		
Benefit to be measured in monetary terms	Effectiveness, Monetary benefit Cost	Cost benefit assessment
Benefit to be measured in non-monetary terms or to be expressed as an effectiveness measure	Effectiveness Non-monetary benefit Cost	Cost effectiveness assessment

Benefit to be measured in both monetary and non-monetary terms (or to be expressed as an effectiveness measure)	Monetary benefit Non-monetary benefit/effectiveness Cost	Cost benefit and cost effectiveness assessment
Evaluation at the system level		
Q8. Assess the functional aspects of surveillance which may influence effectiveness	Effectiveness	Functional attribute assessment
Q9. Assess the technical effectiveness of one or more surveillance components and the functional aspects of surveillance that may influence effectiveness		Effectiveness and functional attribute assessment
Q10. Assess the technical effectiveness of the surveillance system		Effectiveness attribute assessment
Q11. Assess the surveillance structure, function and processes		Process assessment

Table 4. Final list of evaluation attributes consolidated within RISKSUR project and number of related assessment methods

Category*	**Attribute name**	**Attribute definition**	**Assessment methods validated by experts ***
Functional	Availability and sustainability	The ability to be operational when needed (availability) and the robustness and ability of system to be ongoing in the long term (sustainability).	2
Functional	Acceptability and engagement	Willingness of persons and organisations to participate in the surveillance system, the degree to which each of these users is involved in the surveillance. (Could also assess their beliefs about the benefits or adverse consequences of their participation in the system including the provision of compensation for the consequence of disease detection.	4
Functional	Simplicity	Refers to the surveillance system structure, ease of operation and flow of data through the system.	4
Functional	Flexibility, adaptability	The ability to adapt to changing information needs or operating conditions with little additional time, personnel or allocated funds. The extent to which the system can accommodate collection of information about new health-hazards or additional/alternative types of data; changes in	4

		case definitions or technology; and variations in funding sources or reporting methods should be assessed.	
Functional	Compatibility	Compatibility with and ability to integrate data from other sources and surveillance components e.g. One health surveillance (part of data collection and data management)	0
Functional	Multiple hazard	Whether the system captures information about more than one hazard	1
Organisational	Risk-based criteria definition	Validity and relevance of the risk criteria selected and the approach/method used for their identification	0
Organisational	Surveillance system Organisation	An assessment of the organisational structures and management of the surveillance system including the existence of clear, relevant objectives, the existence of steering and technical committees whose members have relevant expertise and clearly defined roles and responsibilities, stakeholder involvement and the existence of effective processes for data management and dissemination of information.	6
Effectiveness	Coverage	The proportion of the population of interest (target population) that is included in the surveillance activity.	2
Effectiveness	Representativeness	The extent to which the features of the population of interest are reflected by the population included in the surveillance activity, these features may include herd size, production type, age, sex or geographical location or time	7

		of sampling (important for some systems e.g. for vector borne disease)	
Effectiveness	False alarm rate (inverse of specificity)	Proportion of negative events (e.g. non-outbreak periods) incorrectly classified as events (outbreaks). This is the inverse of the specificity but is more easily understood than specificity.	5
Effectiveness	Bias= Accuracy	The extent to which a prevalence estimate produced by the surveillance system deviates from the true prevalence value. Bias is reduced as representativeness is increased	7
Effectiveness	Precision	How closely defined a numerical estimate is. A precise estimate has a narrow confidence interval. Precision is influenced by prevalence, sample size and surveillance approach used.	2
Effectiveness	Timeliness	Timeliness can be defined in various ways • This is usually defined as the time between any two defined steps in a surveillance system, the time points chosen are likely to vary depending on the purpose of the surveillance activity. • For planning purposes timeliness can also be defined as whether surveillance detects changes in time for risk mitigation measures to reduce the likelihood of further spread The precise definition of timeliness chosen should be stated as part of the evaluation process. Some suggested	7

		definitions for the RISKSUR project are; **For early detection and demonstrating freedom** • Measured using time - Time between introduction of infection and detection of outbreak or presence by surveillance system • Measured using case numbers - Number of animals/farms infected when outbreak or infection detected **For case detection to facilitate control** • Measured using time - Time between infection of animal (or farm) and their detection • Measured using case numbers – Number of other animals / farms infected before case detected **For detecting a change in prevalence** • Measured using time - Time between increase in prevalence and detection of increase • Measured using case numbers - Number of additional animals/farms infected when prevalence increase is identified.	
Effectiveness	Sensitivity (detection probability and detection fraction)	Sensitivity of a surveillance system can be considered on three levels. • **Surveillance sensitivity (case detection probability)** refers to the proportion of individual animals or herds in the population of interest that have	13

		the health-related condition of interest that the surveillance system is able to detect. Sensitivity could be measured in terms of **detection fraction** (number of case detected divided by the coverage level) in a context of non-exhaustive coverage. • **Surveillance sensitivity (outbreak detection)** refers to the probability that the surveillance system will detect a significant increase (outbreak) of disease. This may be an increase in the level of a disease that is not currently present in the population or the occurrence of any cases of disease that is not currently present. • **Surveillance sensitivity (presence)** –refers to the probability that disease will be detected if present at a certain level (prevalence) in the population.	
Effectiveness	PPV	Probability that health event is present given that health event is detected	2
Effectiveness	NPV	The probability that no health event is present given that no health event is detected	1
Effectiveness	Robustness	The ability of the surveillance system to produce acceptable outcomes over a range of assumptions about uncertainty by maximising the reliability of an adequate outcome. Robustness can be assessed using info-gap models.	0

Value	Cost	The concept of economic cost includes 1) the losses due to disease (e.g. reduced milk yield, mortality), and 2) the resources required to detect the disease by a system (e.g. time, services, consumables for surveillance). In economic evaluation, the resources used to detect disease are compared with the disease losses with the aim to identify an optimal balance where a higher economic efficiency is achieved. Estimation of the total economic cost stemming from losses and expenditures is called a disease impact assessment. Estimation of the resource expenditures only is called a cost analysis.	6 (including 2 non published from RISKSUR members)
Value	Benefit	The benefit of surveillance quantifies the monetary and non-monetary positive direct and indirect consequences produced by the surveillance system and assesses whether users are satisfied that their requirements have been met. This includes financial savings, better use of resources and any losses avoided due to the existence of the system and the information it provides. These avoided losses may include the avoidance of • Animal production losses • Human mortality and morbidity • Decrease in consumer confidence • Threatened livelihoods • Harmed ecosystems • Utility loss Often, the benefit of surveillance estimated as losses avoided can only be realised by implementing an intervention. Hence, it is necessary to also assess the effect of the intervention and look at surveillance,	6

		intervention and loss avoidance as a three-variable relationship. Further benefits of surveillance include maintained or increased trade, improved ability to react in case of an outbreak of disease, maintaining a structured network of professionals able to react appropriately against a (future) threat, maintaining a critical level of infrastructure for disease control, increased understanding about a disease, and improved ability to react in case of an outbreak of disease.	

* Functional= attributes aimed to evaluate the system function ; effectiveness=attributes aimed to evaluate the system performances; organisational= attributes aimed to evaluate the system management and process