Project no. 310806

RISKSUR

Providing a new generation of methodologies and tools for cost-effective risk-based animal health surveillance systems for the benefit of livestock producers, decision makers and consumers

Task 1.3: Prioritisation and weighting of surveillance characteristics and evaluation attributes as well as factors relevant to the selection of evaluation criteria

Report for the deliverable No. 1.3: Current surveillance characteristics and evaluation criteria

The evaluation attributes used for evaluating animal health surveillance systems

KBBE
Collaborative Project
FP7-KBBE-2012-6

www.fp7-risksur.eu

Report prepared by (alphabetic order):
Clémentine Calba (CIRAD), Julian Drewe (RVC), Flavie Goutard (CIRAD), Vladimir Grosbois (CIRAD), Barbara Häsler (RVC), Linda Hoinville (AHVLA), Marisa Peyre (CIRAD), Dirk Pfeiffer (RVC), Timothée Vergne (RVC).

Report reviewed internally by:
Angus Cameron (RVC)

Date: 16/08/2013
Type: Deliverable

TABLE OF CONTENTS

Abstract

One of the major outputs of the RISKSUR project is to develop a tool allowing a comprehensive evaluation of any health surveillance system (including its economic efficiency). As an initial step before formalising the RISKSUR evaluation tool we needed to focus our efforts on the characterisation of the attributes used for evaluating animal health surveillance systems. In this report, we present the list of attributes that were thought to be relevant to be included in the RISKSUR evaluation tool along with their definitions. Then, we report the results of a network analysis used to analyse the relationships between evaluation attributes and to better understand the network of influence they can have on each other. We also propose 3 prioritised lists of effectiveness attributes and economic criteria as a function of evaluation questions and objectives of surveillance systems. The network analysis of the relationships of evaluation attributes and the prioritisation workshops clearly highlighted the difficulty to agree on reduced lists of useful attributes since they all impact on most of the others. From this observation, we developed an integrated approach for evaluating the effectiveness of surveillance systems, and introduced two global effectiveness measures to be included in the RISKSUR evaluation tool. Finally, this document reports the results from a survey among decision makers and technical advisors investigating which attributes they use for deciding whether to implement surveillance and how to implement it. The evaluation attributes appeared to be used only for deciding how to implement a surveillance system. The most frequently cited attributes were effectiveness attributes and economic criteria.

1 General Introduction

The RISKSUR project aims at developing new methodologies for cost-effective risk-based animal health surveillance systems. An important part of RISKSUR is therefore devoted to the development of a tool for systematic evaluations of surveillance systems (for details see Deliverable 1.4). Being related either to public health or to animal health surveillance systems, existing evaluation frameworks and tools generally focus their assessment on a given set of attributes, each of which needing to be assessed for conducting the general evaluation (German, 2000; German et al., 2001; Hendrikx et al., 2011; Drewe et al., 2013). As a first step towards the development of the RISKSUR evaluation tool, this report presents the work that has been carried out to better understand the surveillance attributes used for the evaluation and the complex inter-relations between each other, to identify potential gaps in evaluation attributes and to suggest new approaches for optimising the use of attributes in the evaluation of the effectiveness of the surveillance.

This report is divided into 5 different sections. The first section presents an intensive work that was initiated in 2011 and intensified within the framework of the RISKSUR project. This work aimed at listing, defining and grouping all the attributes that could be used for evaluating animal health surveillance systems. The second section is devoted to the understanding of the relationships between evaluation attributes using a network analysis approach, for identifying which attributes impact the most on others and which are the most impacted by others. The third section proposes a prioritisation of evaluation attributes for the assessment of surveillance systems given the evaluation question and the surveillance objective, and highlights the limitations of such prioritisation and the need to develop an integrated measure of the effectiveness of surveillance. Motivated by this finding, the fourth section introduces a novel global approach for evaluating the effectiveness of surveillance by integrating the decision-making process into the evaluation question. Finally, the fifth section presents the results of a survey among decision-makers to investigate how they use evaluation attributes for making decisions.

2 Definition, classification and assessment of evaluation attributes

Highlights

In total, 35 evaluation attributes have been identified in the literature. This section provides definitions for each of these attributes. In order to better understand the role of each of these attributes it was proposed to group them into four categories: organisational, functional, effectiveness and value attributes. A classification into *first line* attributes and *second line* attributes has also been suggested. *First line* attributes being the attributes needed to be assessed as part of evaluating the performance of surveillance systems and *second line* attributes being all the attributes that directly or indirectly influence *first line* attributes.

2.1 Introduction to evaluation attributes

When evaluating public and animal health surveillance systems it has been recommended that their performance is assessed using a number of evaluation attributes (German et al., 2001; Drewe et al., 2012). Some of these measurable characteristics have also been referred to as evaluation criteria or critical control points (Hendrikx et al., 2011). When defining the inputs for the RISKSUR evaluation tool[1] (EVA) the term "evaluation attributes" has been used to refer to the measurable characteristics that can be used to assess a surveillance system. The term "economic efficiency criteria" has been used for economic measures that can be used to compare the performance of different systems in relation to their cost and make a judgement about which provides best value for the investment made, *i.e.* which surveillance systems are beneficial from an economic point of view.

All evaluation attributes are closely linked, each attribute being influenced by the value of other attributes and in turn influencing the value of others. In addition the choice of attributes to be assessed in any particular evaluation exercise depends on the context of the evaluation and in particular the specific evaluation question that is defined at the outset of the evaluation. In order to better understand the linkages between attributes and the selection of attributes for different situations we have separated the evaluation attributes into groups based on which aspect of the surveillance system performance they assess. Further investigation of the linkage between attributes is described in Section 3 of this report.

2.2 Identifying and defining evaluation attributes

The evaluation attributes included in the list to be used for development of the EVA tool was initially based on those listed in the summary of evaluation terminology developed at a workshop prior

[1] The EVA tool is an integrated approach for evaluation of animal health surveillance systems. This tool will be developed as part of the RISKSUR project (WP5). A first description of that tool is available in Deliverable 1.4.

to the ICAHS conference in 2011 (Hoinville, 2012). Discussions were organised within the RISKSUR evaluation theme working group to identify additional attributes that were not already included in the 2011 list and could be relevant for the evaluation of animal health surveillance systems. The development of the list of relevant evaluation attributes is therefore the result of an evolutionary process that lasted for around two years. Table 1 summarises the full list of identified evaluation attributes along with the evaluation frameworks where they were used. The definition of all these attributes was also discussed within the RISKSUR evaluation theme working group with inputs obtained on the final document from the other members of the RISKSUR consortium. The definitions of all evaluation attributes listed in Table 1 are provided in Appendices 1 to 5. This list of evaluation attributes presented in Table 1 and the definitions presented in the Appendices represent the most complete version at the time of report writing, and will often be cited in the other sections of this document.

Table 1: Most up-to-date list of evaluation attributes that will be used in the RISKSUR project

ICAHS Grouping		Attribute	Reference
Organisational attributes	Management process	Organisation and management	Hoinville 2012, Hendrikx et al 2011
		Training provision	Hoinville 2012, Hendrikx et al 2011
		Performance indicators and evaluation	Hoinville 2012, Hendrikx et al 2011
		Resource availability	Hoinville 2012
	Technical process	Data collection	Hoinville 2012, Drewe et al 2013, Hendrikx et al 2011
		Sampling strategy	Hoinville 2012, Drewe et al 2013 , Hendrikx et al 2011
		Data storage and management	Hoinville 2012, Drewe et al 2013 , Hendrikx et al 2011
		Internal communication	Hoinville 2012, Drewe et al 2013 , Hendrikx et al 2011
		External communication and dissemination	Hoinville 2012, Drewe et al 2013 , Hendrikx et al 2011
		Laboratory testing and analyses	Hoinville 2012, Drewe et al 2013 , Hendrikx et al 2011
		Data analysis	Hoinville 2012, Drewe et al 2013 , Hendrikx et al 2011
		Quality assurance	Hoinville 2012, Drewe et al 2013 , Hendrikx et al 2011
Functional attributes	Function	Stability and sustainability (includes reliability, availability, sustainability)	Hoinville 2012, Drewe et al 2013 , Hendrikx et al 2011
		Acceptability and engagement	Hoinville 2012, Drewe et al 2013, Hendrikx et al 2011
		Simplicity	Hoinville 2012, Hendrikx et al 2011
		Flexibility	Hoinville 2012, Drewe et al 2013 , Hendrikx et al 2011
		Portability	Buehler et al 2004, Meynard et al 2008, Drewe et al 2011
		Interoperability	Meynard et al 2008, Drewe et al 2011
	Data quality	Data completeness and correctness	Hoinville 2012, Drewe et al 2013
		Historical data	Hoinville 2012, Drewe et al 2013
Effectiveness attributes	Inclusion	Coverage	Hoinville 2012, Drewe et al 2013
		Representativeness	Hoinville 2012, Drewe et al 2013 , Hendrikx et al 2011
		Multiple utility	Hoinville 2012, Drewe et al 2013
	Evidence quality	False alarm rate (complement to 1 of the specificity)	Hoinville 2012
		Sensitivity	Hoinville 2012, Drewe et al 2013 , Hendrikx et al 2011
		Timeliness	Hoinville 2012, Drewe et al 2013, Hendrikx et al 2011
		Bias	Hoinville 2012, Drewe et al 2013
		Precision	Hoinville 2012, Drewe et al 2013
		PPV	German et al 2001, Buehler et al 2004, HSCC, Drewe et al 2011
		NPV	Buehler et al 2004, Drewe et al 2011
		Repeatability	Hoinville 2012, Drewe et al 2013
		Robustness	Described in Souza-Monteiro et al. (2012)
Value attributes		Cost	Hoinville 2012, Drewe et al 2013, Howe et al., 2013
		Technical impact	Hoinville 2012, Drewe et al 2013
		Benefit	Hoinville 2012, Drewe et al 2013, Howe et al., 2013

2.3 Methods used for the assessment of evaluation attributes

The evaluation of animal health surveillance systems implies the assessment of selected evaluation attributes, which inform about the quality, effectiveness and efficiency of the surveillance system. The outputs generated by the assessment are used as a basis for making recommendations on how to improve the surveillance system. The initial review of the existing guidelines and frameworks used for the evaluation of surveillance systems performed within the RISKSUR project (Deliverable 1.2) has emphasized several gaps in the methods and tools used for the assessment of some of these attributes. These attributes are assessed using quantitative, qualitative, or semi-quantitative methods, according to the type of attribute. Depending on the context and the resources and data available, these methods will not always be straightforward to implement. Throughout the project, the methods and tools used to assess existing evaluation attributes will be specified, as well as their field of application, the data required for their implementation, their advantages and their limits. This information will be used for further development of the RISKSUR evaluation support tool. This work will be presented in Deliverable 5.1.

The identified methods and tools will be included in the EVA tool which will provide the users with all information required on which method to use to assess the attributes and how to implement them: the data required, the field of application, the time and resources required, a practical description of the implementation (step by step), the outputs provided, and the advantages and limitations. Information on the resources (human and financial) required to implement the assessment will also be provided.

2.4 Grouping of evaluation attributes

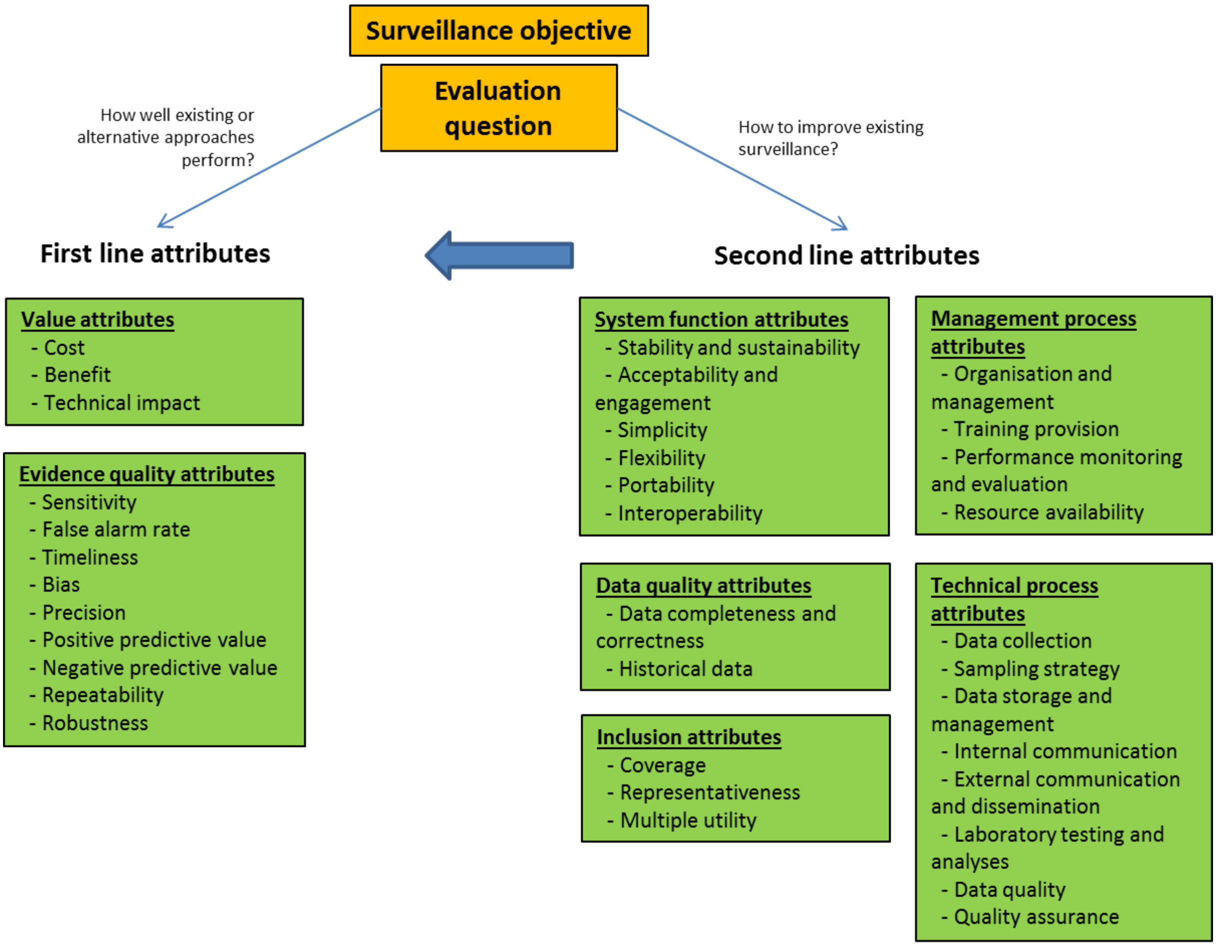
At the ICAHS workshop, a grouping of attributes was proposed based on which aspect of the surveillance activity they related to. The attribute groups were then reviewed during the preparation of the final list of evaluation attributes by the RISKSUR consortium. The final groups identified were:

- Organisational attributes
    - Attributes aimed at evaluating the management processes
    - Attributes aimed at evaluating the technical processes
- Functional attributes
    - Attributes aimed at evaluating the system function
    - Attributes aimed at evaluating the quality of the data collected
- Attributes related to surveillance effectiveness
    - Attributes aimed at evaluating inclusion
    - Attributes aimed at evaluating the quality of the evidence provided
- Attributes assessing surveillance value

Early in the RISKSUR project, it was suggested that these groups of attributes could be presented as a hierarchy in which attributes in one group are likely to influence the value of attributes in other groups. As a consequence, it was suggested to define "*first line evaluation attributes*" and "*second line evaluation attributes*". First line evaluation attributes would be represented by the evidence quality attributes (sensitivity, false alarm rate, timeliness, bias, precision, negative and positive predictive values and repeatability) and the value attributes (cost, benefits, technical impact and robustness). Second line attributes would be all the other attributes that impact directly or indirectly on the first line attributes (including organisational and functional attributes). Evaluating the performance of a surveillance system would therefore require assessing the first line attributes, and practical improvement of a surveillance system would imply modifying some second line attributes which ultimately would result in improved first line attributes (Figure 1).

Discussions are still on-going to determine whether technical impact and robustness should be considered as value attributes (and therefore first line attributes). This ambiguity is expected to be eliminated once the practical applications have been developed with these attributes being better defined in practical terms and their role being clarified.

Figure 1: Grouping of evaluation attributes as first line and second line evaluation attributes



3 Relationship between performance attributes and structural and functional attributes

Highlights

Network analysis has been used based on expert opinion to understand the links between evaluation attributes and the relative importance of each of them. Eleven experts were asked to assess the links between attributes for different surveillance systems related either to an objective of early detection, of freedom from disease or of case finding. Comparative analyses of centrality indices and outputs from principal component analysis were performed. The analysis showed a very strong inter-relationship between most attributes, highlighting the potential difficulty to discriminate attributes during the prioritisation exercise that is presented in the next section. This study also provided critical information for the improvement of evaluation attribute definitions and categories. The results helped formulating recommendations for decision makers on the surveillance system evaluation process.

3.1 Introduction

There is no standardised way to implement evaluation of surveillance and to assess the evaluation attributes mentioned above, which makes the comparison of evaluation outputs difficult. The comparability is also restricted by the wide range of effectiveness criteria used, as highlighted in the RISKSUR Deliverable 1.2. With the recent development of innovative methods for animal health surveillance (risk-based approaches, syndromic surveillance, etc.), expert discussions are ongoing to agree upon the definition of surveillance terms to enhance transparency and facilitate the exchange of data (Hoinville et al., 2013). This work has been extended within the framework of RISKSUR with the aim to agree on relevant evaluation attributes and their definitions (see Section 2 of this report). Because a comprehensive evaluation of surveillance systems including a large number of attributes could be very complex and expensive, it is critical to assess the relative importance of the attributes in order to provide recommendations on evaluation priorities for better allocation of resources and time.

Ranking methods using expert opinion have been used for the prioritisation of attributes (Drewe et al., 2013) but because many attributes are associated with each other, network analysis appears to be the most suitable method in order to represent the complex relations between attributes. Network analysis approaches were developed to study complex socio-economic interactions like influence or knowledge spreading (Jackson, 2008). The approach allows studying the interactions of nodes of the network (e.g. individual actors, features) through their links (e.g. social contact, influence). In the field of animal health, it has mainly been used to study the spread of disease but recent studies have looked into its application to assess the circulation of animal health information between the actors involved in a surveillance system (Jackson, 2008). The use of network analysis allows highlighting critical nodes in terms of a high degree of linking with other attributes and their intermediary position in the network by using measure-of-centrality indices (in-degree, out-degree, in-closeness, out-closeness and random-walk betweenness).

The objectives of this pilot study were two-fold: i) to assess the list of attributes described in Section 2 (Table 1) and their definitions, and ii) to appraise the feasibility and the adequacy of applying network analysis in the selection of key attributes in the evaluation process in different surveillance contexts.

3.2 Materials and Methods

3.2.1 Attribute comparison

The link between each pair of attributes listed in Table 1 was assessed with respect to the direct relationship between the respective two attributes. The linkage structure was expressed using a pairwise comparison matrix in a Microsoft Excel spreadsheet based on all evaluation attributes (12 about the surveillance organization, 8 about the system function, 12 about the effectiveness, 3 about the surveillance value). The direction of the effect and the strength of the correlation (negative or positive) were expressed semi-quantitatively. The assessment was based on expert opinion, with each expert being asked to indicate for each pairwise attribute comparison their opinion about the presence and the direction of the correlation (+1 = positive correlation, -1 = negative correlation, 1 = unable to attribute correlation and 0 = no correlation).

3.2.2 Case studies and use of expert opinion

The pairwise comparison matrix was completed for 3 case studies so that multiple surveillance purposes and objectives were represented, as summarized in Table 2. A minimum of 3 surveillance experts was assigned to each case study and asked to assess the pairwise relationships expressed in the matrix. Experts were sent a detailed protocol on how to complete the matrix along with a description of the case studies and the matrix template in Excel. The experts were asked to comment on their understanding of the attribute definition, their uncertainty about the link attribution and finally on the difficulties encountered during the exercise. The variability of responses between experts was assessed by comparing the different pairwise comparison matrices produced for each case study and by a qualitative assessment of the comments made by the experts.

Table 2: Description of the case studies used in the pairwise attribute comparison exercise

Case study		Surveillance purpose	Surveillance objective
1	HPAI surveillance in poultry in the United Kingdom	Early detection of absent disease	Early detection for rapid response
2	Salmonella surveillance in pigs in Sweden	Surveillance of an endemic disease	Monitoring prevalence and case detection for disease control
3	HPAI surveillance in poultry in Vietnam	Surveillance of an endemic disease	Case detection for disease control

3.2.3 Data analysis

The pairwise comparison matrix expressing the correlation between different attributes was used as input to a social network analysis. The objective with this analysis was to express the importance of each surveillance attribute and their inter-connectedness. Centrality indices were estimated and principal component analysis was conducted for each case study to assess the importance of each attribute in the network. The position of each attribute was then reviewed according to comments provided by the experts during the exercise. All analyses were performed using R version 2.14.1 (R Development Core Team, 2011) and the R package for social network analysis ("sna"), version 2.3-1.

3.3 Results

Because of the complexity of some case studies, the output produced by the experts varied. For case study 1, two experts out of three selected a risk-based approach and one selected a passive surveillance approach, completing three different matrices. For case study 2, four experts reviewed the surveillance system as a group so only one matrix was completed. Finally, for case study 3, two groups of two experts reviewed the surveillance system producing two different matrices.

3.3.1 A complex network of sub-components

The 6 matrices produced were translated into a graphical representation for visual comparison and description. One example is provided in Figure 2. The study confirmed that most attributes have links with several other attributes. The network analysis of these links emphasized the complexity of the interactions between functional, organizational and effectiveness attributes. It highlighted the influence of functional and organizational attributes on the effectiveness attributes (evaluation of performances) of the system. All networks were made of two principal components: one highly correlated to effectiveness attributes and the other to both functional and organizational attributes.

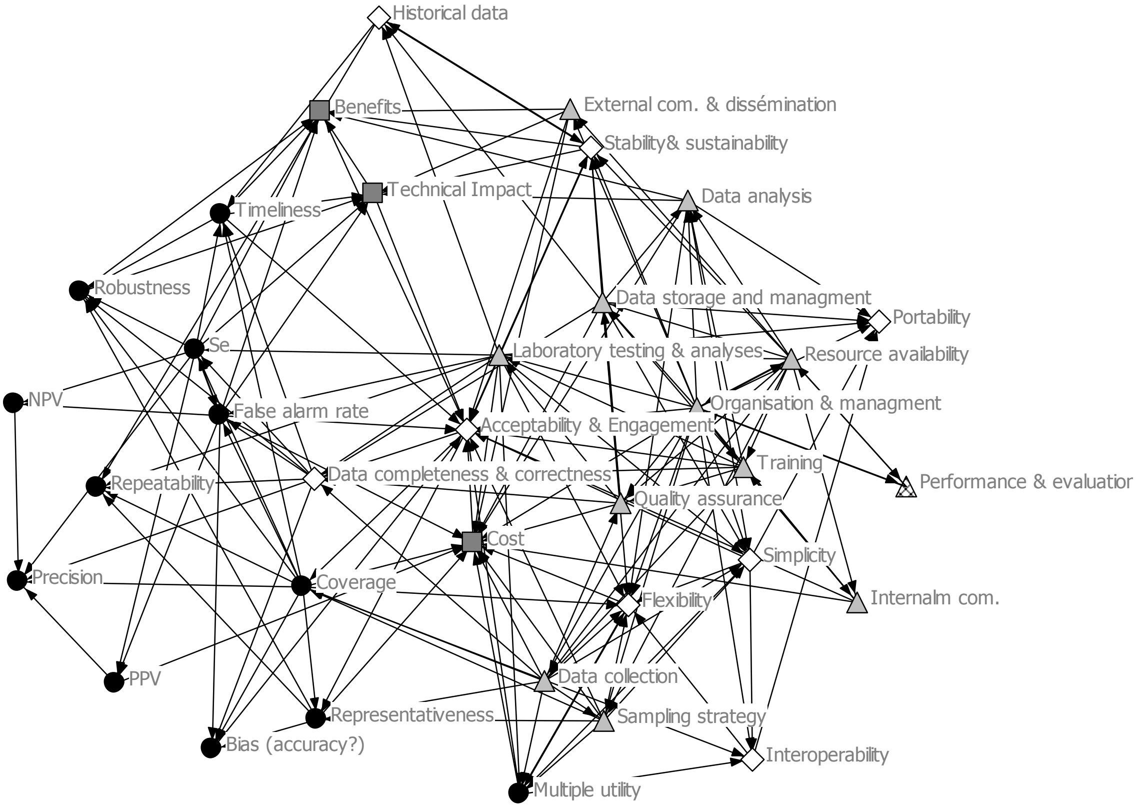


Figure 2: Graphical representation of case-study 1, highly pathogenic avian influenza (HPAI) surveillance in poultry in the United Kingdom, passive component of the surveillance system. The colours of the nodes represent the categories of attributes: organizational (grey triangle); functional (white diamond); effectiveness (black circle); value (grey square).

3.3.2 Centrality measures

"Acceptability and engagement", "stability", "training", "coverage" and "flexibility" were the 5 attributes with the highest betweenness in the networks, representing the attributes with the most direct and indirect links with the others. These results highlight that these attributes are highly connected with the other attributes suggesting that they might play an important role in the evaluation process.

"Organization and management", "coverage", "resources", and "data completeness", "data analysis" and "data collection" were the ones with the highest out-degree, meaning that they are the attributes mainly influencing the level of the other attributes. "Acceptability", "costs", "benefits", "stability" and "technical impact" were the one with the highest in-degree, meaning that they are the attributes the mostly impacted by variation in other attributes.

3.3.3 Variability between experts

For some attributes the variability of response between experts (measured by the differences in centrality indices values) was very high especially for "technical impact"; "organization and management"; "performance and evaluation"). This variability was due to two main reasons:

- The analysis of the data collected during this exercise suggested differences in the understanding and interpretation of some attribute definitions, which did not arise during the attribute characterisation process. The issues were linked to the interpretation of attributes such as “benefit”, “technical impact” (assessing the system’s performance); “flexibility”, “portability” (assessing the system’s function); “internal communication”, “laboratory management” (assessing the technical processes)
- The specific context of risk-based surveillance had great influence on the organisation of the network. Changes in the surveillance component affected all links relating to effectiveness attributes (beside “repeatability” and “robustness”), but also some links between functional attributes (“flexibility”, “portability” and “interoperability”) and the position of the cost attribute.

3.4 Discussion

Network analysis was tested as a method to classify attributes used for the evaluation of surveillance systems and to characterize the links between them. Preliminary analyses of the data have shown high variability between expert responses. This heterogeneity of the results between experts highlighted important issues related to the definition and group classification of the attributes. For example, “technical impact” and “benefit” attributes were considered by the experts as different from the other attributes as they refer to the use of results from the surveillance system rather than to the evaluation of the system itself. Another example is the “multiple utility” attribute which formed an independent component from the other effectiveness attributes and could therefore be considered as misclassified under this category. “Multiple utility” was found to be more closely linked to organizational and functional attributes.

The first output of this expert opinion study was to provide critical information on the improvement of evaluation attribute definitions and categories, highlighting the fact that the current list is useful but should be revised for the RISKSUR evaluation tool.

Some differences in the outputs produced by the experts were due to differences in the interpretation of the case study definitions which were found to be not detailed enough and allowed for variation in interpretation. This issue could be corrected by providing clearer guidance, examples and full case study details.

Another important output of this study was the difference in evaluation attribute network organizations between risk-based surveillance and the other types of surveillance.

Following this study, the next steps will be:

- To revise the list of attributes and their definition to avoid bias as a result of misinterpretation,
- To finalise this preliminary set of analyses by looking at each subcomponent of the networks separately (i.e. organisational and functional attribute networks versus effectiveness attribute network),
- To consolidate the results by involving additional experts or asking the same experts to assess the links for different contexts or surveillance components,
- To reach a consensus between the experts on the different networks,

- To develop recommendations on which attributes to include/exclude from the evaluation process and how to consider their interactions. This will be done by identifying “outlier” attributes (aside of the main network(s) component). Those attributes might only play a minor role in the evaluation process and could therefore be excluded. Moreover this work will provide robust data on the links between attributes and therefore the impact of functional and organisational attributes on the performance of the system.

4 Prioritising evaluation attributes

Highlights

The objective of this prioritisation was to provide a guide for informing the selection of attributes for evaluation given the evaluation question and the surveillance objective. Two workshops were organised with several RISKSUR members for conducting these prioritisation exercises. It was agreed that the evaluation question would inform whether the effectiveness, the cost or the benefits of the surveillance system would need to be evaluated, and that the surveillance objective would determine which effectiveness measure would be useful is effectiveness needs to be evaluated. However, because of the strong inter-relationship between attributes, identifying which effectiveness attributes are more useful than others was very challenging and complete consensus was not reached. Therefore, as in other evaluation frameworks, we emphasized that the prioritisation outputs need to be considered as a guide rather than as a rigid framework. To overcome such limitations, it was suggested to develop a standardised effectiveness measure that could be used for any surveillance objective.

Consistent with the approach adopted for other frameworks (Drewe et al., 2013), it was assumed that the relevance of individual attributes for evaluating surveillance systems in animal health depended on the surveillance objective (early detection, freedom of disease, prevalence monitoring and case finding). To develop existing evaluation frameworks further, the RISKSUR project assumed that the evaluation question also played an important role when defining which attribute to assess. For example, an evaluation interested in identifying the cheapest surveillance system amongst 3 alternatives given a predefined sensitivity will assess different attributes than an evaluation aimed at determining whether or not a surveillance system is economically justified (i.e. the benefits are greater than the costs). To inform which attribute(s) to assess when evaluating a surveillance system, it was therefore essential to prioritise the relevance of the attributes given the surveillance objective and the evaluation question. To work on this issue, two prioritisation workshops were organised with RISKSUR members.

4.1 Prioritisation workshop 1: Madrid (SVEPM)

In March 2013, a pilot prioritisation was organised with 13 consortium members to assess how the surveillance objectives impact on the relevance of attributes. This exercise was an exploratory task aimed at identifying gaps and inconsistencies in the prioritisation methodology to inform further prioritisation activities.

For three surveillance objectives (early detection, freedom from disease and case finding), experts were asked to rank the five attributes they believed to be the most useful for evaluating a surveillance system. A list of fifteen attributes, perceived to be first line attributes at the time of the workshop, were available for ranking, including sensitivity, false alarm rate (1/sensitivity), timeliness, bias, precision, positive and negative predictive values, repeatability, robustness, cost, technical impact, economic optimisation, economic acceptability, economic minimisation and benefit. This list of first line attributes has then been updated and Figure 1 presents the current list of first line attributes. The attribute definitions available at that moment in the project were provided to the experts (these definitions have then been updated).

This first exercise highlighted inconsistencies in the definition for some attributes. The comments made were used to update the definition of the attributes (definitions provided in the Appendices 1-5 are the most up-to-date). Also, it was found difficult to prioritise some of the economic criteria. This was because some of them were not economic criteria but rather methods linked to economic criteria. Consequently, the definitions used in economic evaluations were also revised and a clear distinction was made between value attributes, economic criteria and economic methods. The value attributes perceived to be the most relevant and therefore retained for further prioritisation were costs and benefits. Finally, discussions highlighted the need to identify a method that would allow taking into account both the surveillance objective and the evaluation question when prioritising attributes.

4.2 Prioritisation workshop 2: London (RVC)

In July 2013, a team of four consortium members (from RVC and CIRAD) met at the Royal Veterinary College to test three prioritisation techniques and select the most suitable for a final prioritisation workshop to be organised later in the project. The three techniques were simple categorisation (consisting in associating directly each attribute to a level of usefulness), scoring of attributes' usefulness (consisting in distributing a total of 100 points between all the attributes according to their usefulness) and the pair-wise comparison of attributes (consisting in comparing the usefulness of each attribute in regard to each of the others what allows to estimate the relative usefulness of each attribute in regard to all the others). The updated list of first line attributes was selected based on the first exercise, including performance attributes (sensitivity, false alarm rate, timeliness, bias, precision, positive predictive value, negative predictive value and repeatability) and economic attributes (cost and benefits). Technical impact and robustness were not included in this prioritisation since, as stated in section 2 of this report, their definition and role in the evaluation of surveillance systems was not clear enough to be included at that stage.

This second prioritisation workshop focussed on three case studies: 1) early detection of HPAI in poultry in the UK, 2) freedom from CSF in pigs in Germany and 3) monitoring salmonella prevalence in pigs in Vietnam. A description of each case study was available. Each case study was associated with a distinct surveillance objective. The evaluation questions used were also different between case studies.

The three proposed prioritisation techniques were tested on the three case studies. The discussions for prioritising the attributes were similar between prioritisation techniques, leading to similar classifications. It was therefore agreed that the simplest technique would be the most relevant. For the prioritisation of attributes, *simple categorisation* was therefore recommended asking a group of people to directly classify the attributes into three categories according to their level of usefulness (high, moderate or low usefulness).

As agreed by the consortium, the prioritisation needed to be performed according to the surveillance objective (4 levels: freedom from disease, early detection, prevalence monitoring, case finding) and the evaluation question (at least 8 different evaluation questions were identified; RISKSUR Deliverables 1.2 and 1.4). It was soon concluded that prioritising the attributes for each combination of surveillance objective and evaluation question (at least 40 scenarios) was hardly practical. A rationale was therefore developed to allow the selection of attributes given the surveillance objective and the evaluation question. This rationale is presented hereafter.

4.3 A suggested prioritisation of the usefulness of the evaluation attributes

Any evaluation question can always be related either to an effectiveness issue, to a cost-effectiveness issue or to a cost-benefit issue. Therefore, the evaluation question directly determines whether costs, benefits, effectiveness or any combination of these needs to be assessed. It also follows that the surveillance objective determines which effectiveness measure is useful to assess. In a nutshell, the evaluation question determines whether costs, benefits or effectiveness needs to be assessed and if the effectiveness is selected, the surveillance objective should determine which effectiveness measure is useful to evaluate. Using this approach and the outcome of the prioritisation exercises organised in Madrid and London, we adapted the results of Drewe et al. (2013) to prioritise evaluation attributes according to the surveillance objectives. As a result of this, three attribute selection matrices (one for each type of evaluation question) were produced to assist prioritisation of attributes for assessment of surveillance systems. Tables 3, 4 and 5 present the relevance of evaluation attributes as a function of the surveillance objective for evaluation questions related to effectiveness, cost-effectiveness and cost-benefit respectively. It has to be noted that the classification of the usefulness of the attributes is intended as a guide and may be modified to suit each evaluation as necessary. An ultimate prioritisation workshop was therefore no longer necessary.

Table 3: Attribute selection matrix to assist with prioritisation of attributes for the assessment of surveillance systems when the evaluation question is related to the effectiveness. The figures refer to the expected usefulness of the attribute from 1: very useful, to 3: not very useful.

	Surveillance objective			
Attribute	Monitor the prevalence of an infection	Case finding of infected animals to facilitate control	Early detection of new or re-emerging infection	Demonstrate freedom from infection
Sensitivity	2	1	1	1
False alarm rate/Specificity	2	1	2	1
Bias	1	2	3	2
Precision	1	2	3	2
Timeliness	1	1	1	2
Negative predictive value	3	1	3	1
Positive predictive value	3	1	3	3
Repeatability	2	3	3	3
Cost	3	3	3	3
Benefit	3	3	3	3

Table 4: Attribute selection matrix to assist with prioritisation of attributes for the assessment of surveillance systems when the evaluation question is related to the cost-effectiveness. The figures refer to the expected usefulness of the attribute from 1: very useful, to 3: not very useful.

	Surveillance objective			
Attribute	Monitor the prevalence of an infection	Case finding of infected animals to facilitate control	Early detection of new or re-emerging infection	Demonstrate freedom from infection
Sensitivity	2	1	1	1
False alarm rate	2	1	2	1
Bias	1	2	3	2
Precision	1	2	3	2
Timeliness	1	1	1	2
Negative predictive value	3	1	3	1
Positive predictive value	3	1	3	3
Repeatability	2	3	3	3
Cost	1	1	1	1
Benefit	3	3	3	3

Table 5: Attribute selection matrix to assist with prioritisation of attributes for the assessment of surveillance systems when the evaluation question is related to the cost-benefit. The figures refer to the expected usefulness of the attribute from 1: very useful, to 3: not very useful.

	Surveillance objective			
Attribute	Monitor the prevalence of an infection	Case finding of infected animals to facilitate control	Early detection of new or re-emerging infection	Demonstrate freedom from infection
Sensitivity	3	2	2	2
False alarm rate	3	2	3	2
Bias	2	3	3	3
Precision	2	3	3	3
Timeliness	2	2	2	3
Negative predictive value	3	2	3	2
Positive predictive value	3	2	3	3
Repeatability	3	3	3	3
Cost	1	1	1	1
Benefit	1	1	1	1

4.4 Limitation of such prioritisation

To date, the evaluation of the effectiveness of a surveillance system is generally done through the evaluation of various performance attributes, the most popular being the sensitivity and the timeliness (Drewe et al., 2012). The diversity of effectiveness measures is also highlighted in the RISKSUR Deliverable 1.2 which showed a lack of standardised approaches for evaluating the effectiveness of surveillance. This lack of a standardised approach was also recognised during the second prioritisation exercise where, although only four experts were involved, agreement on the usefulness of the performance attributes given a particular surveillance objective was almost impossible to reach. Indeed, as indicated by the network structures of evaluation attribute interconnectedness (Section 3 of this report) it appeared that most first line attributes were linked to each other. For example, sensitivity and specificity directly impact on bias, precision, negative and positive predictive values, and in some cases vice-versa, making it difficult to clearly define which performance attributes are the most useful to assess. These tight relationships made the prioritisation more difficult and highlighted the need for further characterisation of the attributes. It is therefore important to consider the suggested prioritisation matrices as guides for attribute selection rather than rigid frameworks.

These prioritisation exercises, the results of the systematic literature review of economic evaluations (RISKSUR deliverable 1.2) and the linking of the attributes (Section 3 of this report) confirmed the need to develop standardised measures of the effectiveness of a surveillance system and a standardised way of evaluating it. Consequently, we developed a unified rationale for defining and estimating the effectiveness of a surveillance system, which is presented in the next section.

5 A rationale for evaluating (cost-)effectiveness of surveillance systems

Highlights

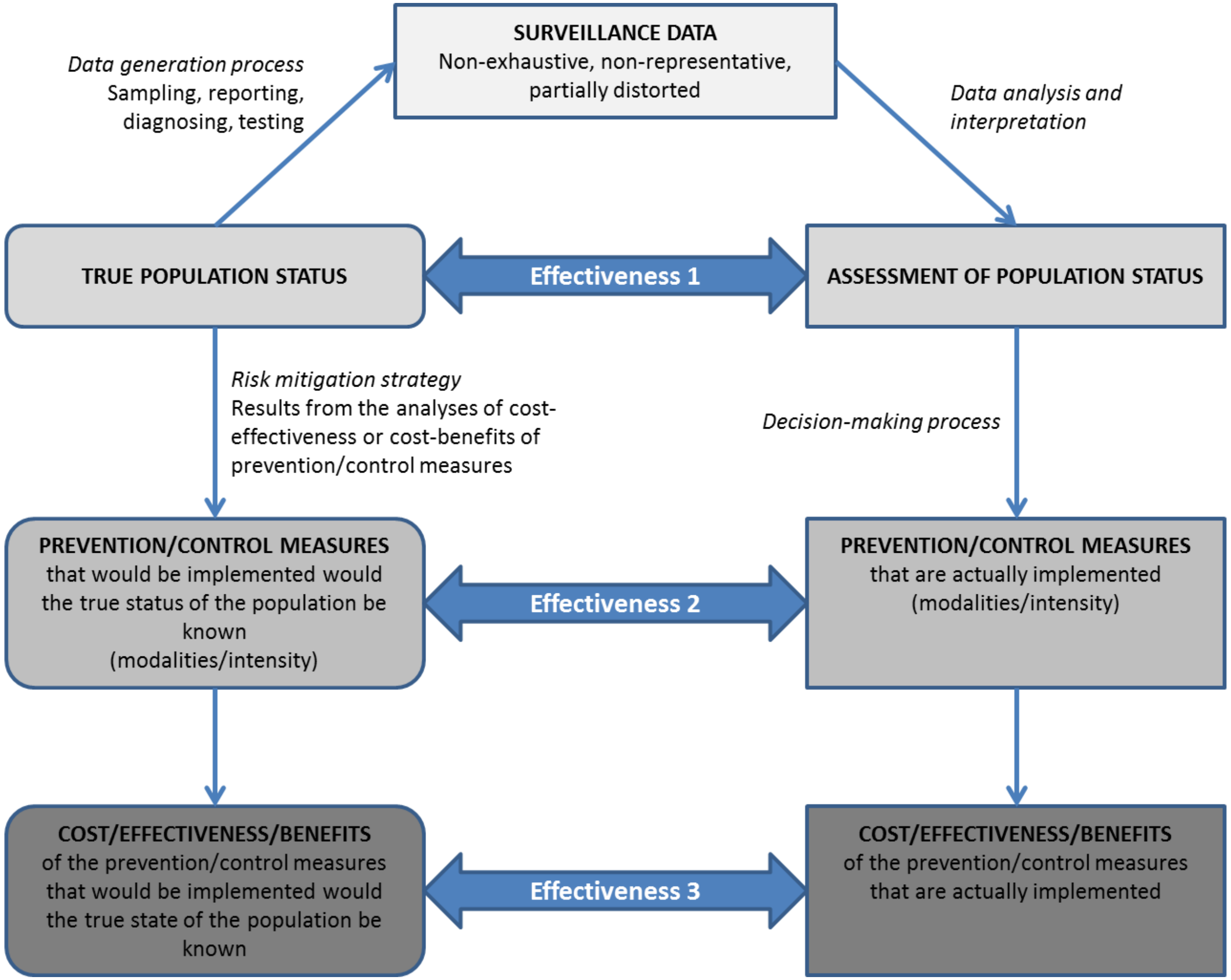
Surveillance systems produce data which, once analysed and interpreted, support decisions regarding disease management. While several effectiveness measures for surveillance are in use, no general theoretical framework has been developed yet for defining and estimating effectiveness of surveillance systems. An effective surveillance system is a system whose data collection, analysis and interpretation processes lead to decisions that are would be taken would the true state of the population be known. Accordingly, we developed a framework taking into account sampling, testing and data interpretation processes, which expresses in a probabilistic way the direction and magnitude of the discrepancy between "decisions that would be made would the true state of a population be known" and the "decisions that are actually made upon the analysis and interpretation of surveillance data". The framework provides a theoretical basis for standardised quantitative evaluation of the effectiveness of surveillance systems. We illustrate the application of the framework using a hypothetical surveillance system aimed at monitoring the prevalence of an endemic disease. Finally, we discuss the potential of this new approach for harmonising cost-effectiveness analyses of animal health surveillance systems.

5.1 Introduction

Surveillance has been defined as «the systematic measurement, collection, collation, analysis, interpretation, and timely dissemination of animal-health and -welfare data from defined populations. These data are essential for describing health-hazard occurrence and to contribute to the planning, implementation, and evaluation of risk-mitigation actions» (Hoinville et al., 2013). The data produced by a surveillance system are generated through reporting, diagnosing, sampling, and sample testing processes. Most of the time reporting and/or sampling is not exhaustive, and may also be non-representative of the underlying population at risk. Moreover diagnostic and/or sample testing procedures usually misclassify a fraction of the examined units or tested samples. The data generated by surveillance systems are thus most of the time non-exhaustive, partially distorted and may also be non-representative. Decisions regarding the implementation of prevention and control actions, and more generally regarding animal health management, nonetheless rely on the assessment of the epidemiological status of the population or of focal units or components of the population through the analysis and interpretation of such imperfect data (Häsler et al., 2011; Howe et al., 2013).

We propose a framework for evaluating the effectiveness of a surveillance system based on three different levels of evaluation. The first level assesses the discrepancy between the true epidemiological status of a population and the status derived based on the sampling, reporting, diagnosing, testing, analysis and interpretation processes of a surveillance system. Most of the evaluation frameworks that have been developed so far focused on this first type of evaluation with estimations of standard effectiveness measures such as the sensitivity or the timeliness of a surveillance system. Our aim is to extend this approach by including in the framework first the decision-making process and then the economics of the surveillance. Consequently, the second level assesses the difference between the decisions (mainly the modalities and intensity of prevention and control measures) that would be made would the true epidemiological status of a population be known and the decisions that are actually made based on the analysis and interpretation of the data produced by a surveillance system. The third level assesses the discrepancy between the cost, effectiveness and benefits of prevention and control measures that would be implemented if the true epidemiological status of a population was known and the cost, effectiveness and benefits of prevention and control measures that are actually implemented based on the analysis and interpretation of the data produced by a surveillance system. These approaches are illustrated in Figure 3, with "effectiveness 1", "effectiveness 2" and "effectiveness 3" being the three levels of evaluation of the effectiveness of a surveillance system.

Figure 3: Proposed approaches for the evaluation of the effectiveness of a surveillance system



5.2 General overview of the rationale

We describe below the steps involved in the evaluation of the effectiveness of a surveillance system based on the difference between the prevention and/or control measures that would be implemented if the true epidemiological status of a population and of its components was known and the modalities and intensity of prevention and/or control measures that are actually implemented based on the analysis and interpretation of the data produced by a surveillance system ("effectiveness 2" in figure 1).

5.2.1 Defining relevant state variables and epidemiological scales

In order to develop a framework for the assessment of the effectiveness of surveillance systems, we need to define the epidemiological scale and the state variable(s) describing the population status that are relevant with respect to the objectives of surveillance. The relevant state variable is the variable which determines the modalities and intensity of the prevention and/or control measures considered to be appropriate by stakeholders and decision makers. The relevant epidemiological scale is the scale at which prevention and/or control measures are implemented.

Consider a system whose objective is to provide information on the yearly prevalence of an endemic disease. Consider further that this information is used to decide whether or not systematic testing for this disease should be implemented at all the slaughterhouses of the country. In this example, the relevant scale is the country and the relevant state variable is yearly prevalence.

Consider a system whose objective is to detect all farms infected by an emerging pathogen. Consider further that this information is used to identify herds that have to be culled. In this example, the relevant scale is the farm and the relevant state variable is the infection status of a farm regarding the pathogen of interest.

5.2.2 Describing the mitigation strategy

Our objectives are restricted to the evaluation of surveillance and do not include the evaluation of prevention/control strategies. This is the reason why in the proposed framework uses as a reference the prevention and/or control measures that would be *implemented* if the true epidemiological status of a population and of its components was known (see Figure 3) rather than the prevention and/or control measures that would be *optimal* if the true epidemiological status of a population and of its components was known. One important step in the proposed approach is thus to describe the prevention and/or control measures that would be considered as appropriate by stakeholders and decision makers for any possible focal epidemiological state variable (see previous section). The relationship between the value of the epidemiological state variable and the prevention and/or control measures considered to be appropriate will hereafter be referred to as "mitigation strategy". It is worth distinguishing two types of mitigation strategies. In the first type the modalities and intensity of mitigation measures vary as a function of the value of the epidemiological state variable according to a step function, so that the possible values of the epidemiological state variable are categorized into subsets each of which is associated with a specific mitigation option. In the second type, the modalities and intensity of mitigation measures vary continuously as a function of the value of the epidemiological state variable. For the time being and for the sake of clarity we will only consider the first type of mitigation strategy.

In table 6, an example of a simple mitigation strategy is presented for each surveillance objective considered in RISKSUR. In these mitigation strategies, the possible values of the relevant state variable are classified according to two subsets referred to as S^+ and S^-. Each of these subsets is associated with a particular mitigation option (M^+ and M^-, respectively) that is considered to be appropriate by stakeholders and decision makers. In these simple examples we define S^+ as the subset of values of the state variable that requires the implementation of mitigation measures (*i.e.* mitigation option M^+) and S^- as the subset of values of the state variable that requires no mitigation measures (*i.e.* mitigation option M^-). For example, when considering a surveillance system aiming at monitoring prevalence, decision-makers may decide to do nothing if the true prevalence is below a pre-defined threshold and to implement an active surveillance in slaughterhouses if the prevalence is above that threshold.

Note that while at a given time, the epidemiological state of a population or one of its components is expected to fall only in one category, the states may change over time consistently with the evolution of disease occurrence and factors that determine the implementation of mitigation measures. In some situations, the implementation of mitigation measures is likely to be motivated by a change in the state of the population, so instead of S^- and S^+ it can be relevant to consider ΔS^- and ΔS^+. Note also that situations where possible values of epidemiological states are categorized according to more than two subsets can be considered, as will be illustrated below.

An important issue for the proposed framework to be applicable is the characterization of the mitigation strategy. There may exist a predefined mitigation strategy in which different possible scenarios regarding the status of the focal population and/or its components have been considered and the corresponding prevention/control measures have been determined (ideally through cost-effectiveness or cost-benefit analyses). Such a strategy should be communicated to the evaluator. However, for some sanitary situations, mitigation strategies will not have been formally defined prior to their occurrence and instead a reactive ad-hoc response will be implemented. Participatory approaches involving stakeholders and decision makers (for instance using companion modelling) could then be used to determine which prevention/control measures are considered to be appropriate for different scenarios depending on the status of the population of interest and of its components.

Table 6: Examples of simple mitigation strategies for various surveillance objectives

Surveillance objective	**Scale**	**State variable**	**S^-**	**M^-**	**S^+**	**M^+**
Monitoring prevalence	Country/ Region	Yearly prevalence of a disease (Prev)	Prev ≤ Threshold	Do nothing	Prev > Threshold	Implement systematic testing at slaughterhouses before products leave for market
Case detection of disease	Herd	Disease Status	No infected animal in the herd	Do nothing	≥1 infected animal in the herd	Cull the herd
Demonstrate freedom from disease	Country/ Region	Yearly prevalence of a disease (Prev)	Prev≤ Threshold	Allow exportations	Prev > Threshold	Ban exportations
Early detection of an emerging disease	Country/ Region	Daily Incidence Rate (DIR)	DIR=0	Do nothing	DIR>0	Launch massive testing, in depth case investigation, and limit movements

5.2.3 Describing the data generation process

The proposed framework also requires a thorough description of the processes through which surveillance data are generated. Information on reporting (e.g. underreporting rates, factors influencing reporting probability), diagnostic (case definition, probability of misclassification), sampling (coverage, stratification, intensity, frequency), sample testing (e.g. sensitivity and specificity of the tests used) are needed in order to assess the discrepancy between true epidemiological status (unknown) and the observed status derived from the analysis and interpretation of surveillance data (see figure 3). A preliminary study might be required to estimate the parameters that are necessary to describe the data generation process.

5.2.4 Describing the data analysis and interpretation processes leading to decisions regarding prevention/control measures

Our framework further requires a description of the process through which data analysis and interpretation leads to decisions regarding the implementation of mitigation measures (see figure 1). In most instances, this process involves the computation of selected statistics that provide an assessment of the epidemiological state and serve as a basis for informing decisions regarding the mitigation measures to implement. Considering the simple mitigation strategies presented in table 6 where two subsets of values of the focal state variable are considered, we can define two corresponding subsets of values for such a statistics. A^+ is the subset for which the focal epidemiological state variable is assessed as falling into the category requiring the implementation of mitigation measures (M^+). A^- is the subset for which the focal epidemiological state variable is assessed as falling into the category requiring no mitigation measure (M^-). Table 8 presents possible statistics and the A^- and A^+ subsets of their values for the mitigation strategies considered in Table7.

Table 7: Examples of decision making rules relying on the analysis and interpretation of surveillance data. The decision rules correspond to the mitigation strategies presented in table 6

Surveillance objective	**Scale**	**Statistics used to assess epidemiological status**	**A^-**	**Decision M^-**	**A^+**	**Decision M^+**
Monitoring prevalence	Country/ Region	Proportion of positive tests in the samples collected over a year P(+)	$P(+) \leq$ Threshold	Do nothing	$P(+) >$ Threshold	Implement systematic testing in slaughterhouses before products are put on the market
Case detection of disease	Herd	Result of a pooled test	Negative test result	Do nothing	Positive test result	Cull the herd
Demonstrate freedom from disease	Country/ Region	Proportion of positive tests in the samples collected over a year P(+)	$P(+)\leq$ Threshold	Allow exportations	$P(+) >$ Threshold	Ban exportations
Early detection of an emerging disease	Country/ Region	Case reporting	No case reported	Do nothing	$\geq$1 case reported	Launch massive testing and in depth case investigation. Limit movements

5.2.5 Effectiveness criteria

Once S^+, S^-, A^+, A^- have been determined, it is possible to define two types of errors, namely Type I and Type II errors analogous to the types of error used in statistical or diagnostic tests (Table 9).

Table 8: The two types of error used as effectiveness criteria

		True epidemiological status	
		S^+ mitigation required	**S^- mitigation not required**
Assessment of the epidemiological status resulting from the analysis and interpretation of surveillance data	**A^+ mitigation implemented**		Type I error
	A^- mitigation not implemented	Type II error	

Type I error occurs whenever a surveillance system generates outputs that result in the implementation of mitigation measures when in fact the true status of the population would not require it. Type I errors imply that costly mitigation measures are implemented unnecessarily. Type II error occurs whenever a surveillance system generates outputs that will not result in the implementation of any mitigation measure when in fact the true state of the population would require it. Type II errors result in an increased risk of failure to control a genuine disease threat.

The effectiveness of a surveillance system can be assessed by estimating the probability of a type I error (*i.e.* $P(A^+/S^-)$) and the probability of a type II error (*i.e.* $P(A^-/S^+)$) for that system. Importantly, $P(S^+/A^-)$ and $P(S^-/A^+)$ are also useful quantities that inform the decision makers about the level of risk they take when making a decision about implementation of mitigation measures and therefore constitute critical information in decision making. $P(S^+/A^-)$ informs the decision maker on the probability that the true epidemiological status would require the implementation of mitigation measures in situations where surveillance evidence suggests that no mitigation measures should be implemented (for example the probability of a disease being present with a prevalence higher than the threshold (below which the territory would be considered as free of the disease), when surveillance data suggest that prevalence is lower than this threshold and that the territory is therefore free of the disease). $P(S^-/A^+)$ informs the decision maker on the probability that the true epidemiological status would not require the implementation of mitigation measures in situations where surveillance evidence suggests that mitigation measures should be implemented (for example the probability that the true prevalence of a disease is lower than the threshold (below which the territory is considered as free of the disease) when surveillance data suggest that prevalence is higher than this threshold and that the territory is therefore not free of the disease). Note that $P(A^+/S^-)$, $P(A^-/S^+)$, $P(S^+/A^-)$ and (S^-/A^+) are closely related with standard surveillance effectiveness attributes (see Table 1 in section 2). $P(A^+/S^-)$ and $P(A^-/S^+)$ can be assimilated to the complements to 1 of respectively the specificity and sensitivity of a surveillance system. $P(S^+/A^-)$ and (S^-/A^+) can be assimilated to the complements to 1 of respectively the negative predictive value and the positive predictive value of a surveillance system. We believe however that the notations that we propose should be adopted because they are more explicit. Indeed, they make it clear that evaluating the effectiveness of surveillance requires the definition of sets of true epidemiological states (*i.e.* S^- and S^+) and sets of assessments of epidemiological states (*i.e.* A^- and A^+) associated with distinct mitigation options.

Provided that mitigation strategies, data generation processes and data analysis and interpretation processes leading to decisions regarding mitigation measures have been described, probabilities of type I and type II errors can be assessed either analytically using sampling and probability theories or through simulations.

It is worth noting that this approach is commonly used to determine sample sizes for surveillance system components aiming at demonstrating freedom of disease. Indeed, designers of such surveillance systems set sample sizes at values that are large enough for guarantying that if the prevalence in the population would exceed a usually very low threshold, the probability of not detecting any case would be lower than a desired value, usually set at 0.05 (Cameron and Baldock, 1998; Johnson et al., 2004). In such surveillance systems sample sizes are thus determined with respect to the probability of a type II error and therefore with respect to the decisions made (declare or not the territory free and therefore allow or not the exportation) based on the examination and interpretation of the data collected. We argue that this should be the case for any other types of surveillance systems.

In the next section we assess these effectiveness criteria for a hypothetical surveillance system aimed at monitoring the level of prevalence in a particular population.

5.3 From theory to empirical analysis

5.3.1 Surveillance objectives

We consider a hypothetical active surveillance system aimed at monitoring yearly prevalence of a cattle disease in a country to inform decision-makers on which vaccination strategy should be implemented at the national level. In this example the relevant state variable is thus the yearly prevalence and the relevant scale is the country.

5.3.2 Mitigation strategy

Three sets values of yearly prevalence (p) have been defined each requiring the implementation of a different type of vaccination strategy.

- If prevalence is below 0.1, then no vaccination (M^-) is implemented. S^- is thus $p \leq 0.1$
- If prevalence is higher than 0.1 but below 0.2 then vaccination is implemented only in high risk areas (M^+). S^+ is thus $0.1 < p \leq 0.2$

- If prevalence is higher than 0.2, then vaccination is implemented in the whole population (M^{++}). S^{++} is thus $p > 0.2$.

Note that in this mitigation strategy, three subsets of possible values of the epidemiological state have been defined.

5.3.3 Data generation process

Let us consider that N=100 units are randomly sampled over a one year period and that each sample is tested using a test with sensitivity $Se = 0.90$ and specificity $Sp = 0.95$.

5.3.4 Data analysis and interpretation processes

Let us assume that the statistic used by decision-makers to assess the value of the epidemiological state variable is the number (n_+) of samples that test positive. Let us furthermore assume that to inform decisions regarding the implementation of mitigation measures this number is compared with the expected number of samples testing positive for threshold values of the true prevalence given the sensitivity and specificity of the test used. Subsets of possible values of n_+ leading to different decisions regarding mitigation measures are then:

- A^-: if $n_+ \leq N*[0.1*Se+(1-0.1)*(1-Sp)]$, then no vaccination is implemented (M^-).
- A^+: if $N*[0.1*Se+(1-0.1)*(1-Sp)] < n_+ \leq N*[(0.2*Se+(1-0.2)*(1-Sp)]$, then targeted vaccination is implemented in high risk areas (M^+)
- A^{++}: if $n_+ > N*[0.2*Se+(1-0.2)*(1-Sp)]$, then large-scale vaccination is implemented (M^{++})

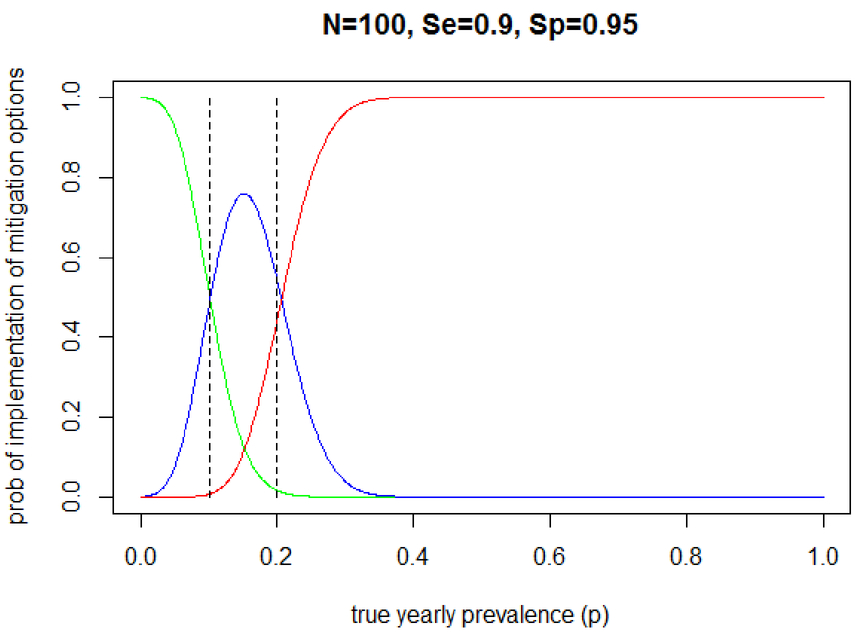
5.3.5 Computing effectiveness criteria

Evaluating the effectiveness of this surveillance system using the proposed rationale requires estimating the probabilities of implementing inappropriate mitigation measures such as:

- $Pr(A^-|S^+) = Pr(n_+ \leq N*[0.1*Se+(1-0.1)*(1-Sp)] \mid 0.1< p \leq 0.2)$: probability of a moderate type II error
- $Pr(A^- \mid S^{++}) = Pr(n_+ \leq N*[0.1*Se+(1-0.1)*(1-Sp)] \mid p > 0.2)$: probability of a severe type II error
- $Pr(A^+ \mid S^-) = Pr(N*[0.1*Se+(1-0.1)*(1-Sp)] < n_+ \leq N*[0.2*Se+(1-0.2)*(1-Sp)] \mid p \leq 0.1)$: probability of a moderate type I error
- $Pr(A^+ \mid S^{++}) = Pr(N*[0.1*Se+(1-0.1)*(1-Sp)] < n_+ \leq N*[0.2*Se+(1-0.2)*(1-Sp)] \mid p > 0.2)$: probability of a moderate type II error
- $Pr(A^{++} \mid S^-) = Pr(n_+ > N*[0.2*Se+(1-0.2)*(1-Sp)] \mid p \leq 0.1)$: probability of a severe type I error
- $Pr(A^{++} \mid S^+) = Pr(n_+ > N*[0.2*Se+(1-0.2)*(1-Sp)] \mid 0.1 < p \leq 0.2)$: probability of a moderate type I error

For this hypothetical situation, computing these quantities is straightforward because the probability distribution of the number of positive tests n_+ is known: it is a binomial distribution of parameter N (the sample size) and $p*Se+(1-p)*(1-Sp)$ (the apparent prevalence as a function of the true prevalence and the characteristics of test used). We can thus compute $\Pr(X < n_+ < Y \mid p, N, Se, Sp)$ for any value of X, Y, p, Se, Sp. This distribution has been used to describe for the example exposed above how the probabilities of implementing M^-, M^+ and M^{++} vary as a function of true prevalence in the population. This description is displayed on the Figure 4 below.

Figure 4: probability distribution that data generation, analysis and interpretation processes result in the implementation of different mitigation options as a function of the true epidemiological state.



On Figure 4,

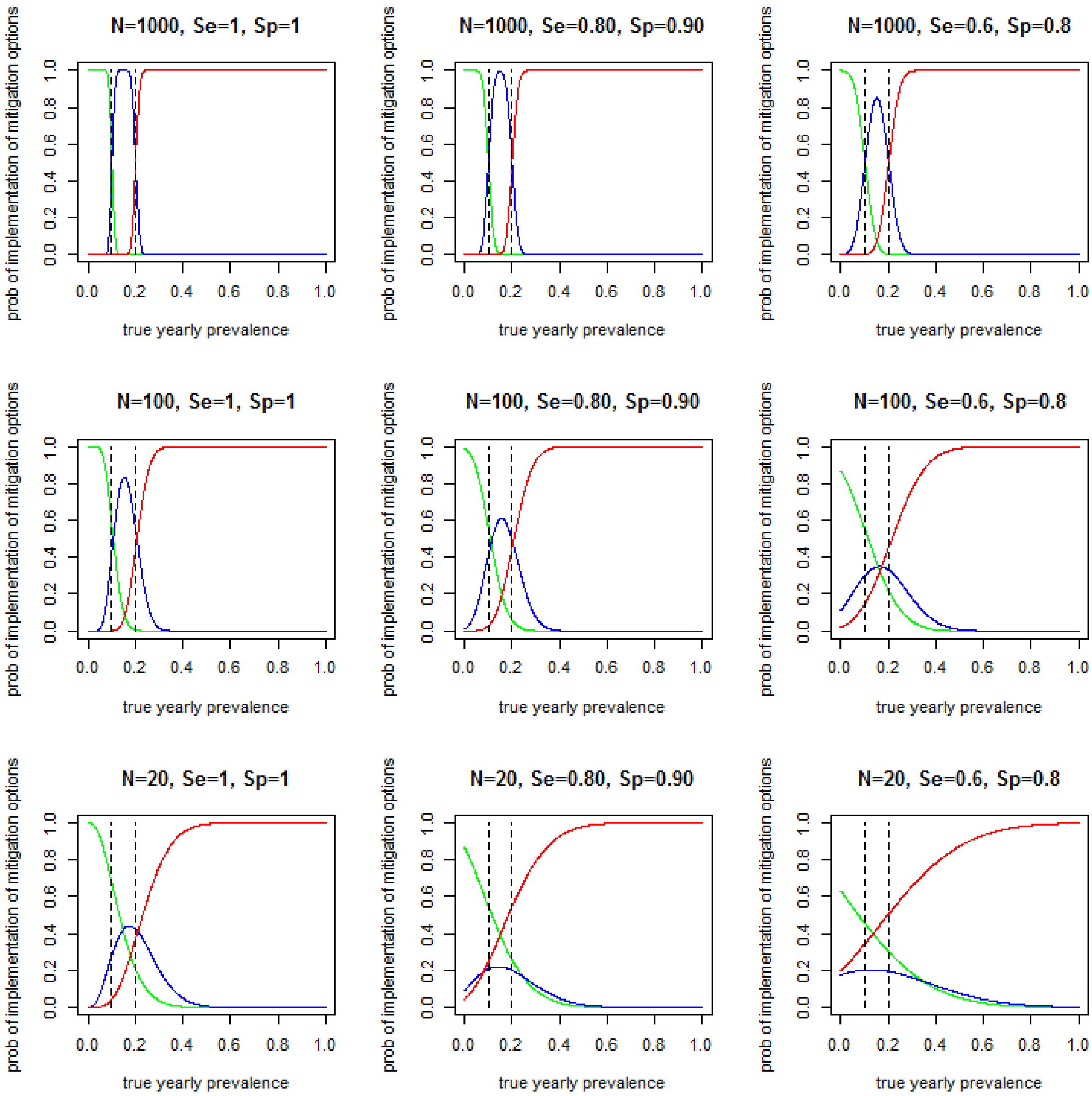
- The green curve represents $P(A^- \mid p)$: it shows for any given value of p (real yearly prevalence in the population) the probability that the analysis and the interpretation of the data produced by the surveillance system result in decision makers opting for no vaccination (mitigation option M^-),
- The blue curve represents $P(A^+ \mid p)$: it shows for any given value of p the probability that the analysis and the interpretation of the data produced by the surveillance system result in decision makers opting for targeted vaccination (mitigation option M^+),
- The red curve represents $P(A^{++} \mid p)$: it shows for any given value of p the probability that the analysis and the interpretation of the data produced by the surveillance system result in decision makers opting for mass vaccination (mitigation option M^{++})
- The vertical dashed black lines represent the mitigation strategy. They delimit the ranges for p which are considered by stakeholders and decision makers as requiring different mitigation measures to be implemented (left hand side region: no vaccination (M^-), middle region: targeted vaccination (M^+), right hand side region: mass vaccination (M^{++})).

Let us consider the values taken by the green, the blue and the red curve when the real prevalence in the population is just above 0.2 (thus mass vaccination would be required). The figure tells us that given the sample size, the diagnostic test characteristics, and the decision rule used, the probability of actually implementing mass vaccination is around 0.5, the probability of implementing only targeted vaccination (moderate type II error) is around 0.5 and the probability of not implementing vaccination (severe type II error) is close to 0. Let us now consider the situation where the real prevalence in the population is around 0.05 (thus no vaccination would be required). The figure tells us that given the sample size, the diagnostic test characteristics, and the decision rule used, the probability of nonetheless implementing mass vaccination (severe type I error) is 0, the probability of implementing targeted vaccination (moderate type I error) is around 0.1 and the probability of not implementing vaccination is around 0.9.

5.3.6 Sensitivity of effectiveness criteria to the characteristics of data generation process

Our approach allows us to assess how probabilities of type I and type II errors change when reporting or sampling or diagnostic or sample testing procedures are modified. On the different panels of Figure 5, we considered modifications in our hypothetical surveillance system in terms of sample size and performance of the diagnostic test used to detect the disease in each sampled unit. As expected, increasing sample sizes and improving test performances results in reducing the probabilities of type I and type II errors.

Figure 5: sensitivity of surveillance effectiveness to changes in sampling and sample testing procedures



5.3.7 Sensitivity of effectiveness criteria to the data analysis and interpretation processes underlying decisions regarding mitigation options

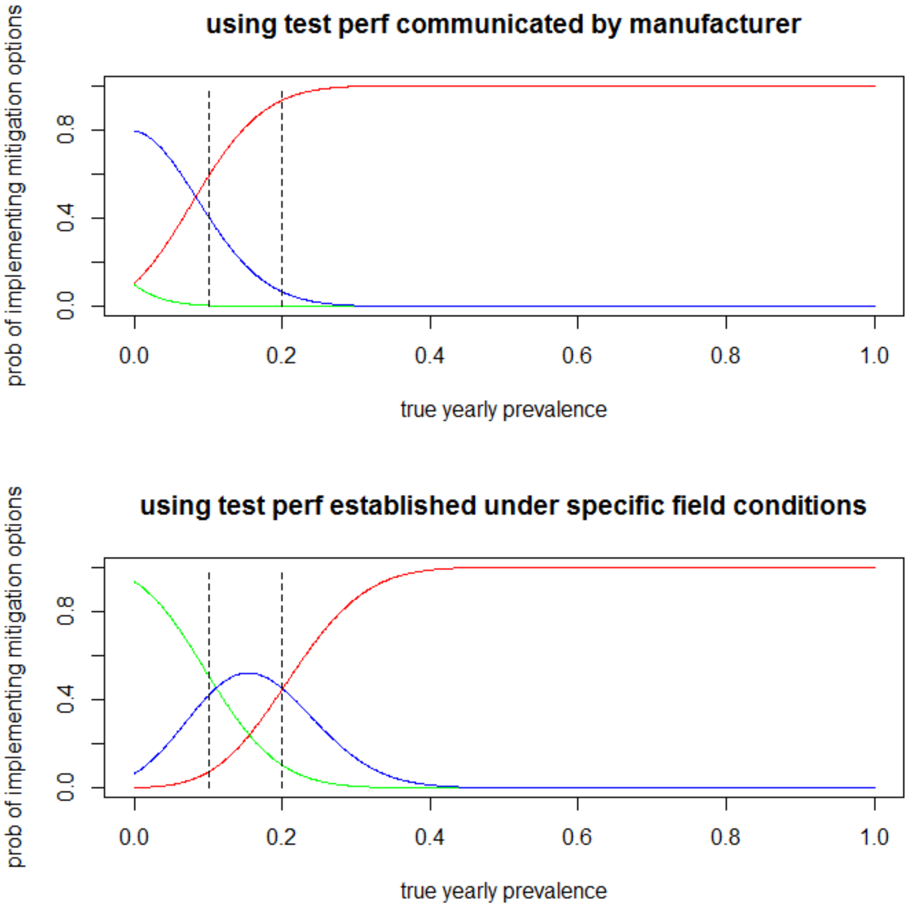
In the previous section we have shown how our rationale can be used to assess the sensitivity of surveillance effectiveness (expressed in terms of probabilities of Type I and type II errors) to changes in the attributes related to the data generation process. Importantly our rationale can also be used to address the sensitivity of surveillance effectiveness (expressed in terms of probabilities of Type I and type II errors) to changes in the way surveillance data are analysed and interpreted.

Let us consider the previous example in which *N*=100 units are randomly sampled and tested over a year period. Let us further consider that the test used has according to its manufacturer *Se* = 0.95 and *Sp* = 0.99. It is however likely that under specific field conditions, test sensitivity and specificity differ from the values provided by the manufacturer. Let us thus assume that under the specific conditions of the surveillance system and surveyed population the sensitivity and specificity of the test used are actually *Se*=0.7 and *Sp*=0.85. We can now consider two contrasted situations in terms of data interpretation:

- A situation where decision makers trust the information communicated by the manufacturer and thus consider that the *Se* and *Sp* of the test are 0.95 and 0.99, respectively.
- A situation where decision makers have evaluated the performance of the test they use under the specific conditions in which they are used in the surveillance process. They have thus figured out that *Se* and *Sp* are 0.7 and 0.85, respectively.

In Figure 6, the probabilities of implementing M^-, M^+ and M^{++} vary as a function of true prevalence in the population are displayed for these two scenarios (upper panel: the decision makers use the *Se* and *Sp* communicated by the manufacturer; lower panel: the decision makers have evaluated the performance of the test under the specific conditions of the surveillance system and surveyed population). It shows how misinterpretation of the data due to erroneous information about test sensitivity and specificity can increase dramatically the probabilities of type I and type II errors.. This is illustrated in Figure 6.

Figure 6: sensitivity of surveillance effectiveness to changes in data analysis and interpretation procedures



5.4 Further developments and discussion

So far, we have shown for a single context and surveillance objective how the effectiveness of a surveillance system can be evaluated in terms of discrepancy between the modalities and intensity of prevention and control measures that would be implemented would the true epidemiological status of a population be known and the modalities and intensity of prevention and control measures that are actually implemented based on the analysis and interpretation of the data produced by a surveillance system. We have also shown that our rationale allows analysing sensitivity of the effectiveness criteria to changes in the surveillance system that are likely to improve effectiveness. Importantly, we show that information on data generation processes alone does not allow thorough evaluations of surveillance effectiveness. Indeed, we claim that information on the process through which surveillance data are analysed and interpreted and on the decision-making process leading to the implementation of mitigation strategies are crucial.

Although we believe our rationale offers a sound theoretical basis for the assessment of the effectiveness of surveillance systems, further developments are needed for it to be applicable in real world situations. In particular, the data analysis and interpretation processes leading to the implementation of mitigation measures considered so far are over-simplistic. In future developments we will consider decisions regarding which mitigation measures to implement that account for uncertainty in estimations of focal epidemiological states. For the sake of clarity we have only considered so far mitigation strategies in which the possible values of the epidemiological state variable are categorized into subsets each of which is associated with a specific mitigation option. Mitigation strategies in which the modalities and intensity of mitigation measures vary continuously as a function of the value of the epidemiological state variable will also be considered in further developments.

As it stands, our approach allows evaluating the effectiveness of surveillance for any given true value of an epidemiological state variable. One possible limitation is that a true value of the epidemiological state variable has to be defined in order to get estimations of the probabilities of type I and type II errors. A question that will shortly be addressed is that of the integration of probabilities of type I and type II errors over the possible true values of the epidemiological state variable. One solution would be to consider that all possible values are equi-important (or equally likely to occur). Another solution would be to use a probability distribution of the true values of the epidemiological state variable to weight probabilities of type I and type II errors. Hence epidemiological models could be integrated in our framework to derive probability distribution of the true values of the epidemiological state variables.

Finally a major breakthrough would be to assess surveillance effectiveness through the discrepancy between the cost, effectiveness and benefits of prevention and control measures that would be implemented would the true epidemiological status of a population and of its components be known and the cost, effectiveness and benefits of prevention and control measures that are actually implemented based on the analysis and interpretation of the data produced by a surveillance system ("effectiveness 3" in Figure 3). To do so, we need to assess the economic consequences of alternative mitigation measures.

We also have so far developed our methodology only for surveillance systems aiming at monitoring the prevalence of an endemic disease. One important next step is to adapt it to surveillance systems with other objectives (demonstrating freedom from disease, early detecting the emergence of exotic diseases, finding all cases of an emerging disease).

Finally, we need to apply our methodology to different real surveillance case studies, for each of the four surveillance objectives.

6 The use of attributes by decision makers

Highlights

As part of the RISKSUR task 1.1, a survey was conducted among decision-makers and technical advisors to investigate which attributes they use when deciding whether animal health surveillance should be funded and how surveillance should be carried out. International and national legal requirements as well as negative impact on public health or animal production appeared to be the main reasons deciding whether or not to allocate resources to surveillance. Surveillance attributes such as effectiveness attributes and economic criteria were more often used for deciding how to do the surveillance.

Apart from understanding the links between evaluation attributes and working on the usefulness of evaluation attributes from a conceptual point of view, we investigated whether decision-makers use evaluation attributes in the context of animal health surveillance, and how they use them.

The data collected as part of the RISKSUR task 1.1 was used to investigate how decision-makers use decision-criteria when deciding ***whether*** animal health surveillance should be funded and ***how*** animal health surveillance should be implemented. It has to be noted that a major aim of this activity was to investigate what decision-criteria were used most frequently by decision-makers when allocating resources to surveillance. Therefore, 30 decision criteria were evaluated, encompassing both evaluation attributes such as sensitivity, coverage or cost, and other more general factors such as international legal requirements or negative impact of the disease on animal production. The list of decision-criteria provided for the interviews was based on the list of attributes used in RISKSUR, criteria described in the scientific literature, and practical experience of members of the RISKSUR consortium. Respondents further had the possibility to add other decision-criteria they commonly used. Key information on the use of decision criteria with a special focus on evaluation attributes is summarised and highlighted hereafter.

Thirty four interviews were conducted with 14 decision-makers and 20 technical advisors in France, Germany, Great Britain, the Netherlands, Spain, Sweden, and Switzerland. The interviewees were asked about their role in the surveillance, how animal health surveillance was carried out, how resources in the private or public sector were allocated to animal health surveillance, and what the processes were to allocate resources to surveillance. The questionnaire further included questions about public-private partnerships, cost-sharing, decision-criteria used, constraints perceived, and further information desired for decision-making on surveillance.

Twenty seven respondents ranked the five most important decision-criteria from 1 to 5 in the list provided both for the questions related to ***whether*** animal health surveillance should be funded and ***how*** animal health surveillance should be funded. The results indicate some basic trends and highlight areas of particular relevance for the development of the evaluation tool. The relevance of decision-criteria was investigated for various specific surveillance programmes, and was analysed according to the four following categories of surveillance programmes: (i) all, (ii) endemic; (iii) exotic, emerging disease, freedom from disease and (iv) zoonoses.

6.1 The most popular decision-criteria used by decision-makers overall

Figure 7 gives an overview of the counts of the decision-criteria mentioned across all categories, independent of the ranking of the criteria, illustrated for the questions ***whether*** animal health surveillance should be funded and ***how*** animal health surveillance should be funded and as total. The frequent mentioning of international legal requirements shows the importance EU and other international legislation has in resource allocation to surveillance. Therefore, taking into account the international perspective and how international legislation is created is paramount when evaluating surveillance. Other context specific criteria are also relevant in decision-making, in particular the disease situation in the country and the negative impact on public health and, to a lesser extent, on animal production.

The evaluation attributes cited most often were cost-effectiveness, cost-benefit, effectiveness, and practicality of the surveillance[2]. Almost all the other evaluation attributes were also mentioned, but less frequently. The frequent mentioning of cost-benefit and cost-effectiveness demonstrate that the value of surveillance is important when it comes to making an investment.

[2] Note that here we considered cost-effectiveness and cost-benefit as evaluation attributes although, strictly, they have been defined as evaluation criteria combining evaluation attributes such as cost, benefit and measures of effectiveness

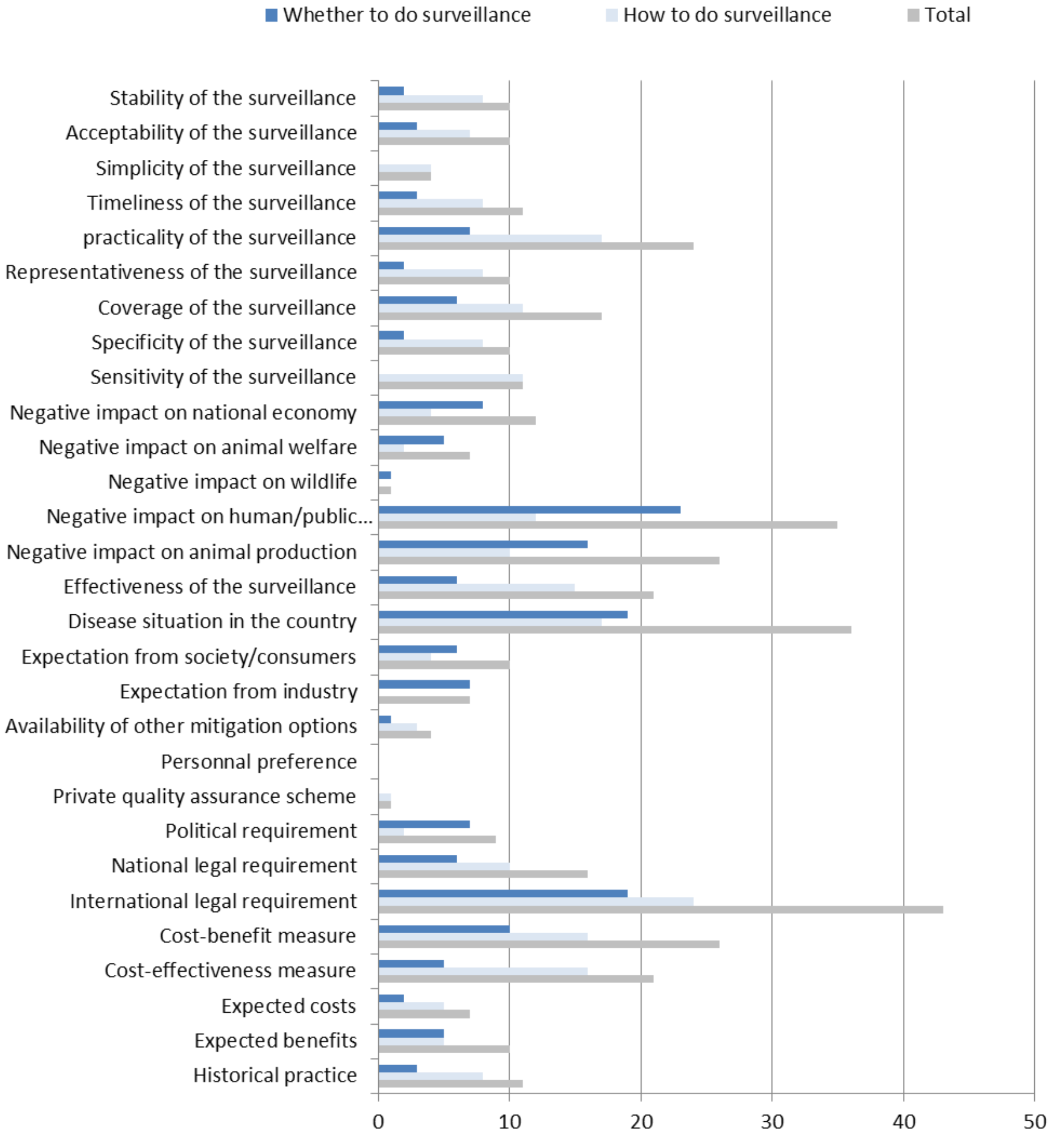


Figure 7: Number of time each criterion was considered to be an useful criterion by the interviewees for making decisions related to whether to fund surveillance and how to fund surveillance (the four categories of surveillance programmes have been aggregated).

When deciding ***whether*** to commit money to surveillance, international and national legal requirements were identified as important criteria by both types of respondent. So too were the impacts of disease on human/public health, animal production and the national economy. Decision maker respondents also mentioned the disease situation in the country, expected costs and benefits and the political requirement.

When deciding ***how*** to do surveillance more technical aspects were identified. Indeed practicality, representativeness, timeliness, sensitivity or effectiveness of the surveillance system were identified by the respondents, in addition to all the criteria already identified for the decision whether to do surveillance.

6.2 Decision-criteria used for exotic and emerging diseases or freedom from disease

When deciding ***whether*** to commit money to surveillance for exotic diseases, the decision maker respondents identified a wide range of criteria, but international legal requirement stood above the rest. The technical advisors were more in agreement on the (potential) impacts on human/public health and animal production. International legal requirement and the disease situation in the country were also cited but less frequently (38% of the answers).

When deciding ***how*** to do surveillance for exotic diseases more technical aspects were identified, such as practicality, timeliness or effectiveness of the system were identified by the respondents. Some decision maker respondents also identified cost-benefit and cost-effectiveness criteria but less frequently.

6.3 Decision-criteria used for endemic diseases

When deciding ***whether*** to commit money to surveillance for endemic diseases, the decision maker respondents again identified a wide range of criteria, with international legal requirement still above the rest. The technical advisors were more in agreement on a few criteria: the impacts on human/public health and animal production, plus the disease situation in the country.

When deciding ***how*** to do surveillance for endemic diseases, more decision maker respondents identified overall effectiveness and more technical advisors identified stability and sustainability of the surveillance system, both criteria related to the long-term performance. Other more technical aspects such as sensitivity and specificity were identified by the technical advisors while the decision makers identified cost-benefit and cost-effectiveness criteria.

6.4 Decision-criteria used for zoonotic diseases (exotic or endemic)

When deciding ***whether*** to commit money to surveillance for zoonotic diseases, not surprisingly, both types of respondent identified the negative impact on human/public health as the most important criteria. The decision makers also identified expectation from society/consumers as an important criterion, while technical advisors also identified the disease situation in the country as important. International legal requirement was also identified by many of both types of respondent.

When deciding ***how*** to do surveillance for zoonotic diseases, most decision maker respondents identified overall effectiveness, cost-benefit, international legal requirement and impact on human/public health as important criteria. Most technical advisors identified cost-effectiveness, practicality and coverage of the surveillance system, and international legal requirement.

6.5 Conclusion

The decision-making processes used for deciding whether and how to conduct surveillance is likely to involve a complex combination of different decision-criteria.

The most common criteria used for deciding whether to implement surveillance are extrinsic criteria that do not depend on the surveillance itself, and therefore cannot serve as a basis for surveillance evaluation. These are mainly international and national legal requirements, and negative impact of the disease on production or on human/public health, that still have to be clearly understood for defining the context of the evaluation. However, a few intrinsic criteria relating to evaluation attributes were also cited by the respondents (although to a much lower extent) such as practicality and cost-benefit or cost-effectiveness measures.

For deciding how to conduct surveillance, intrinsic criteria were much more often used, especially economic criteria such as cost-effectiveness and cost-benefit measures. This highlights the fact that, to date, surveillance evaluation might have more utility for identifying the best system configuration amongst a set of alternatives when the implementation of a surveillance system *per se* has already been decided, than to decide whether or not to conduct surveillance. This finding is in agreement with the findings from the systematic literature review of economic evaluation of animal health surveillance systems (Deliverable 1.2) which showed that 81% of the economic evaluations included in the review focused on questions relating to “how to perform surveillance?”, and only 19% on questions related to “do we perform surveillance?”

Further in-depth analyses can be found in the Deliverable 1.1.

7 General conclusions

In this deliverable, we reviewed the attributes that are suitable for evaluating disease surveillance systems and defined them in the context of animal health. Network analysis was used to try and formalise the links between the attributes of a surveillance system. In-degree and out-degree measures informed which attributes have a strong influence on others and which are strongly affected by others. Based on expert opinion elicitation, we proposed a ranking of the usefulness of common effectiveness attributes and economic criteria depending on the evaluation question and the surveillance objectives. Similar to other evaluation frameworks, these rankings should be used as a *guide* for the evaluation, especially because it does not appear to be possible to achieve consensus amongst experts when characterising the usefulness of the effectiveness attributes. In order to investigate alternative solutions to this problem, we developed a new methodology for evaluating the effectiveness of surveillance systems, and introduced two effectiveness measures that can be applied to any surveillance system. We propose to integrate these two new measures within the framework of the EVA tool in addition to other effectiveness measures, so as to allow standardisation of effectiveness evaluation and the comparability between evaluations.

8 References

Cameron, A.R., Baldock, F.C., 1998. A new probability formula for surveys to substantiate freedom from disease. Preventive veterinary medicine 34, 1-17.

Drewe, J.A., Hoinville, L.J., Cook, A.J., Floyd, T., Gunn, G., Stark, K.D., 2013. SERVAL: A New Framework for the Evaluation of Animal Health Surveillance. Transbound Emerg Dis.

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

German, R.R., 2000. Sensitivity and predictive value positive measurements for public health surveillance systems. Epidemiology (Cambridge, Mass 11, 720-727.

German, R.R., Lee, L.M., Horan, J.M., Milstein, R.L., Pertowski, C.A., Waller, M.N., 2001. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep 50, 1-35; quiz CE31-37.

Häsler, B., Howe, K.S., Stärk, K.D., 2011. Conceptualising the technical relationship of animal disease surveillance to intervention and mitigation as a basis for economic analysis. BMC health services research 11, 225.

Hendrikx, P., Gay, E., Chazel, M., Moutou, F., Danan, C., Richomme, C., Boue, F., Souillard, R., Gauchard, F., Dufour, B., 2011. OASIS: an assessment tool of epidemiological surveillance systems in animal health and food safety. Epidemiology and infection, 1-11.

Hoinville, L., 2012. Animal Health surveillance terminology final report from pre-ICAHS workshop. Available online at: http://www.defra.gov.uk/ahvla-en/files/icahs-workshop-report.pdf.

Hoinville, L.J., Alban, L., Drewe, J.A., Gibbens, J.C., Gustafson, L., Hasler, B., Saegerman, C., Salman, M., Stark, K.D., 2013. Proposed terms and concepts for describing and evaluating animal-health surveillance systems. Preventive veterinary medicine 112, 1-12.

Howe, K.S., Häsler, B., Stärk, K.D.C., 2013. Economic principles for resource allocation decisions at national level to mitigate the effects of disease in farm animal populations. Epidemiology and Infection 141, 91-101.

Jackson, M.O., 2008. Social and economic networks. Princeton University Press Princeton, NJ.

Johnson, W.O., Su, C.L., Gardner, I.A., Christensen, R., 2004. Sample size calculations for surveys to substantiate freedom of populations from infectious agents. Biometrics 60, 165-171.

R-Development-Core-Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.r-project.org.

9 Appendix 1: Definition of organisational attributes

Attributes aimed at evaluating the management processes

Organisation and management	An assessment of organisational structures of the surveillance including whether the objectives are relevant and clearly defined and the existence of steering and technical committees whose members are representative of the surveillance stakeholders. The members of these committees should have appropriate expertise, clearly defined roles and responsibilities and should hold meetings regularly to oversee the function of the system at which minutes are kept.
Training provision	Provision of adequate initial training and an ongoing program of training for those implementing the surveillance system, particularly those collecting the data
Performance indicators and evaluation	Whether performance indicators are routinely used to monitor system performance and whether periodic external evaluations are used to assess the system outputs in relation to its objectives
Resource availability	An assessment of the financial and human resources available for implementing the surveillance activity including the expertise and capability of personnel

Attributes aimed at evaluating the technical processes

Data collection	The use of appropriate data sources and collection methods including automation of data collection where appropriate and the existence of a case definition and data collection protocol including an appropriate sampling strategy
Sampling strategy	The use of appropriate sampling strategies including the use of risk-based approaches and pooled sampling where appropriate. This could include a risk-based requirement calculations or risk-based sampling. The basis of the risks used in the design of the risk-based sampling strategy should be assessed.
Data storage and management	Appropriate use and documentation of data management systems for processing information, including data processing protocols, and effective use of data verification procedures and data storage and back-up procedures
Internal communication	An assessment of the methods used and ease of information exchange between all those involved in providing, managing, analysing and disseminating information for the surveillance system. The methods used to provide feedback to data providers and to increase their awareness about hazards and surveillance activities should also be assessed.
External communication and dissemination	An assessment of the data and information provided to those outside the surveillance system including the timeliness and types of output produced. The efforts made to disseminate these outputs including the use of web-based systems should also be assessed.
Laboratory testing and analyses	Whether testing is carried out using appropriate methods, including an assessment of diagnostic test sensitivity and specificity, with quality assurance scheme and timely and accurate delivery of results.
Data analysis	Whether appropriate methods are used for the analysis and interpretation of data at an appropriate frequency
Quality assurance	Whether the laboratory or other surveillance processes are quality assured or accredited

10 Appendix 2: Definition of functional attributes

Attributes aimed at evaluating the system function

Stability and sustainability	The ability to function without failure (reliability), to be operational when needed (availability) and the robustness and ability of system to be ongoing in the long term (sustainability).
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 include an assessment of stakeholder awareness of the system and their understanding of it. 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.
Simplicity	Refers to the surveillance system structure, ease of operation and flow of data through the system.
Flexibility	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 case definitions or technology; and variations in funding sources or reporting methods should be assessed.
Portability	Evaluating the possible use of the system in other circumstances or at a different location
Interoperability	Compatibility with and ability to integrate data from other sources and surveillance components

Attributes aimed at evaluating the quality of the data collected

Data completeness and correctness	The proportion of data that was intended to be collected that actually was and the proportion of data entries that correctly reflect the true value of the data collected
Historical data	Quality and accessibility of archived data

11 Appendix 3: Definition of attributes related to surveillance effectiveness

Attributes aimed at evaluating inclusion

Coverage	The proportion of the population of interest (target population) that is included in the surveillance activity.
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 of sampling (important for some systems e.g. for vector borne disease)
Multiple utility	Whether the system captures information about more than one hazard

Attributes aimed at evaluating the quality of the evidence provided

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.
Bias	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
Precision	How closely defined a numerical estimate is. A precise estimate has a narrow confidence interval. Precision is influenced by the prevalence of disease as well as the nature of the surveillance approach that is used (e.g. sensitivity of diagnostic test and sample size)
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 definitions for the RISKSUR project are; **For early detection** Measured using time - Time between introduction of infection and detection of outbreak Measured using case numbers - Number of animals/farms infected when outbreak detected **For demonstrating freedom** Measured using time - Time between introduction of infection and detection of presence by surveillance system Measured using case numbers – Number of animals/farms infected when 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.
Sensitivity	Sensitivity of a surveillance system can be considered on three levels. • **Surveillance sensitivity (case detection)** refers to the proportion of individual animals or herds in the population of interest that have the health-related condition of interest that the surveillance system is able to detect • **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.
PPV	Probability that health event is present given that health event is detected
NPV	The probability that no health event is present given that no health event is detected
Repeatability	How consistently the surveillance component performance can be maintained over time.
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.

12 Appendix 4: Definition of attributes assessing surveillance value

Cost	The evaluation should list and quantify each of the resources required to operate the surveillance system and identify who provides this resource. These resources could include: time and personnel (labour), services (e.g. laboratory tests, postage), travel, consumables, and equipment.
Technical impact	This indicates the changes that have been based on the results of the surveillance providing a measure of the usefulness of the surveillance system in relation to its aims. This should include details of actions taken as a result of the information provided by the surveillance system e.g. changes in protocols or behaviour and changes in mitigation measures and particularly changes in disease occurrence
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, 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.

13 Appendix 5: Definition of economic efficiency criteria

Optimal economic efficiency	The net benefit to society shall be maximised. Achieved where the marginal costs of least-cost combinations of surveillance and intervention resources equal the marginal benefits of mitigation (=loss avoidance).
Economic acceptability	Ensuring that the benefits (=loss avoidance) generated by a mitigation policy at least cover the costs for surveillance and intervention.
Least-cost choice	Ensuring that a technical target for disease mitigation (e.g. time to detection) is achieved at minimum cost without quantifying the benefit.