RECOMMENDATIONS FOR METRICS TO EVALUATE SURVEILLANCE PERFORMANCE

JRP6 – NOVA D-JRP6-5.1 and 5.2

Responsible Partner: SVA

Contributing partners: SVA, DTU, RIVM, APHA

GENERAL INFORMATION

European Joint Programme full title	Promoting One Health in Europe through joint actions on foodborne zoonoses, antimicrobial resistance and emerging microbiological hazards
European Joint Programme acronym	One Health EJP
Funding	This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.
Grant Agreement	Grant agreement n° 773830
Start Date	01/01/2018
Duration	60 Months

DOCUMENT MANAGEMENT

Project deliverable	EJP-NOVA-D-JRP6-5.1 and 5.2
Project Acronym	JRP6 – NOVA – FBZ1 – 1st Call
Author	Thomas Rosendal (SVA)
Other contributors	Håkan Vigre, Ofosuhene Okofrobour Apenteng, Eric Evers, Mark Arnold
Due month of the report	M45
Actual submission month	M45
Type *R: Document, report DEC: Websites, patent filings, videos, etc.; OTHER*	R
Dissemination level *PU: Public (default) CO: confidential, only for members of the consortium (including the Commission Services)*	PU
Dissemination *Author's suggestion to inform the following possible interested parties.*	No specific recipients

RECOMMENDATIONS FOR METRICS TO EVALUATE SURVEILLANCE PERFORMANCE

Recommendations for metrics to evaluate surveillance performance

The objective of this work package was to develop modelling approaches to predict the impact of surveillance programs applied in the food production chain on human exposure to food borne pathogens and ultimately public health. All models were developed as case-studies focusing on specific pathogen-food combinations or animal health problems that are of primary importance to the public and animal health within the EU. Each case study used national and region-specific data with an approach of addressing part of or the complete food production chain from farm to consumer. These models were used to develop frameworks for surveillance evaluation that can be applied to other animal systems, pathogens and food items. The establishment of key performance metrics that should be evaluated when measuring surveillance performance was an underlying theme in each of these case studies. Surveillance performance has previously been evaluated within the context of protecting a system from introduction of disease. This evaluation has focused on measurement of the sensitivity of a surveillance strategy to detect the disease and in turn the probability that a population is free from a given disease. Probability of freedom from a disease is clearly not relevant in an endemic state and the sensitivity of a surveillance system does not capture the complexity of features that are relevant when comparing candidate surveillance strategies. To identify alternate and useful approaches, several metrics of performance were considered within the case studies completed in this work package such as the standard sensitivity of a strategy, the expected time to the first detection, extent of an outbreak at the time of detection, which population strata are affected prior to detection, the burden of human disease measured by disability-adjusted life years (DALYs) prior to detection and the cumulative hazard of disease detection over time to capture both detection rate and sensitivity in one metric. The utility of these metrics in evaluation of surveillance are discussed below in the 5 case studies performed in OHEJP - NOVA work package 5.

1. Case study 1 – *Salmonella* in poultry model

Since 2018, the EU commission has declared the Danish broiler industry to be *Salmonella* free. However, there is continuous *Salmonella* pressure from the environment which results in several parent and broiler flocks becoming infected annually. When a parent flock becomes infected, the infection can be transmitted vertically to recipient broiler flocks before the parent flock is detected and destroyed, including the eggs at the hatchery. To address this issue, a stochastic dynamic model of the transmission of *Salmonella* in parent flocks was developed. This model included the relationship between flock prevalence and test sensitivity for environmental samples in the flock. When an artificial infection was started in the model by seeding 10 and 100 infected hens, the likelihood of detecting an infected parent flock within the first three weeks post-infection was strongly improved by the taking of five boot swabs (95% CI 70-100) instead of two (95% CI 40-100) or by supplementing of the two boot swabs by a dust sample (95% CI 43-100). These results suggest that the likelihood of detecting a *Salmonella* infection in a broiler flock in the programme was at least 99% in broiler flock even if only one chicken was initially infected. These findings are of relevance for managing parent flocks and eggs at the hatchery in case of *Salmonella* infection in parent flocks in Danish poultry. This also illustrates the utility of considering timeliness of detection as a metric for surveillance evaluation since a fast detection can result in less downstream impact on recipient broiler flocks and in turn human exposure to *Salmonella*.

This meeting is part of the European Joint Programme One Health EJP.
This project has received funding from the European Union's Horizon 2020
research and innovation programme under Grant Agreement No 773830.

2. Case study 2 – *Salmonella* in pigs model

A transmission model was built to evaluate a national control programme for *Salmonella* in pigs at national level in Great Britain. *Salmonella* is an endemic disease of pigs in Great Britain which has a significant health impact on humans, so the aim of a control programme is to identify infected herds and implement control measures in those herds to reduce the national slaughter prevalence and hence exposure of humans to *Salmonella*. Therefore, the metric used in this study was the national prevalence of infected pigs at slaughter over time. The benefit of this metric is that it is directly related to the aim of the surveillance programme i.e., reduction in prevalence of *Salmonella* at slaughter. The model considered the sensitivity and specificity of the tests used and could therefore be used to evaluate the effectiveness of the surveillance scheme using alternative diagnostic tests and sampling designs. It could also be used to evaluate how well any surveillance scheme would monitor progress towards other targets by evaluating the precision with which the herd-level prevalence and annual reduction is estimated for different surveillance designs.

A method to estimate DALYS (Disability-adjusted life years) from the slaughter prevalence was also developed as an extension of this work. This was performed for a simpler surveillance approach, but the method is readily adaptable to relate the impact of any surveillance programme on the actual number of DALYS attributed to the food chain. This could enable a cost-benefit analyses to be carried out for the above model evaluating a national control programme and is arguably the metric of a surveillance programme that is closest to the ultimate target of improving public health.

3. Case study 3 – *Mycobacterium avium* subsp. *paratuberculosis* in cattle model

To monitor a state of disease freedom and to ensure a timely detection of new introductions of disease, surveillance programmes need be evaluated prior to implementation. A strategy was developed to evaluate surveillance of *Mycobacterium avium* subsp. *paratuberculosis* (MAP) using simulated testing of bulk milk in an infectious disease spread model. MAP is a globally distributed, chronic infectious disease with substantial animal health impact and has been hypothesized to be associated with human illness. Designing surveillance for this disease poses specific challenges because methods for surveillance evaluation have focused on estimating surveillance system sensitivity and probability of freedom from disease and do not account for spread of disease or complex and changing population structure over long periods. The aims of the study were to 1. define a model that describes the spread of MAP within and between Swedish herds; 2. define a method for simulation of imperfect diagnostic testing in this framework; 3. to compare surveillance strategies to support surveillance design choices. The results illustrate how this approach can be used to identify differences between the probability of detecting disease in the population based on choices of the number of herds sampled and the use of risk based or random selection of these herds. The approach was also used to assess surveillance to detect introduction of disease and to detect a very low prevalence endemic state. The use of bulk milk sampling was determined to be an effective method to detect MAP in the population with as few as 500 herds tested per year if the herd-level prevalence was 0.2 %. However, detection of point introductions in the population was unlikely in the 13-year simulation period even if as many as 2000 herds were tested per year. Interestingly, the use of a risk-based selection strategy was found to be a disadvantage to detect MAP given the modelled disease dynamics. This case study illustrated the utility of surveillance sensitivity, time to detection and size of an outbreak at the time of detection in evaluating the surveillance of a very slow spreading disease. The results indicate that although the metrics were useful, the study period for the surveillance evaluation was too short to measure performance of the surveillance strategies to detect point introductions of disease.

4. Case study 4 – Risk assessment model

4.1. Risk-based sampling in the retail phase

A risk-based sampling approach focuses on the question of how to best subdivide sampling capacity of a surveillance activity most effectively monitor DALYs in the human population. Sampling is performed preferentially in those pathogen-product combinations for which the surveillance activity has the lowest cost per DALYs monitored. This case study focused on *Salmonella* in pig meat in the retail phase and the costs per sample in the surveillance programme refer to the microbiological analysis costs. The number of samples was set using a criterion of a constant relative uncertainty of estimated pathogen prevalence. The number of DALYs was estimated using an overall pathogen-food animal value, which was subdivided over products using exposure assessment estimates. Risk based sampling methodology was applied to France, United Kingdom, Denmark, Sweden and The Netherlands and this led to two improvements in the methodology, so that the number of food product categories and the number of inhabitants of a country could be considered. Ranking of food-product groups in terms of costs / DALYs monitored was performed for the respective countries both separately and in combination.

4.2. Cost-effectiveness of sampling in the retail phase

A method to evaluate the cost-effectiveness of product testing in the pre-retail and retail phase, by food industry and the government, respectively, related to food product withdrawal was also explored. The metric used to evaluate the surveillance effort was costs / DALYs evaded in the context of *Salmonella* in pig meat. Costs of surveillance included the microbiological analysis and the number of samples collected. The effect included a positive effect on public health, expressed in DALYs, as batches with a positive test would not reach the consumer or would be heat-treated first. Cost-effectiveness can be compared with a criterion set by WHO. This comparison indicated that the short-term effect of product testing was not cost-effective. However, it is suspected that the present calculations may have been an underestimate of the disease burden and uncertainty of the estimates were large. The long-term effects of product testing on hygiene in the food industry were not considered in this work because they were considered too difficult to quantify.

4.3. Cost-effectiveness of sampling in the farm phase

A method to evaluate the cost-effectiveness of surveillance in the farm phase, focusing on *Salmonella* in pigs, and applying it to The Netherlands and Great Britain was explored. The metric used to evaluate the surveillance strategy was, again, costs / DALYs evaded. The costs considered included the number of samples taken and the costs of microbiological analysis per sample. To calculate DALYs evaded, it was assumed that pig meat from *Salmonella* test-positive farms did not reach the consumer. It was calculated as a decrease in *Salmonella* prevalence at animal-level as a result of surveillance, which was carried through the abattoir- and consumer phases to result in a reduction of DALYs. It was shown that cost-effectiveness was independent of the number of farms sampled but decreased with the number of samples taken per farm. Cost-effectiveness was more favourable (lower) for Great Britain than for The Netherlands, 4.78×10^3 and 8.70×10^3 EUR/DALY evaded, respectively, for comparable scenarios. Both the costs and the DALYs evaded were higher for Great Britain, being a larger country, but DALYs evaded with a higher factor than costs. However, the calculations had a high uncertainty so it is not clear if a conclusion can be drawn on the relative effectiveness of this surveillance in the two countries.
Cost-effectiveness of sampling in the farm phase was much more favourable than in the retail or abattoir phases, being about 8.5×10^3 EUR/DALY vs 6.08×10^4 or 2.26×10^5 EUR/DALY (depending on sampling system) (Evers et al., 2020). The value of 8.5×10^3 EUR/DALY is also lower than 4.6×10^4 EUR/DALY, the standard value set by WHO for cost-effectiveness of an intervention (WHO Europe, 2014).
These case studies using the risk assessment methodology and considering the efficiency of a surveillance system in terms of the relative resources invested to reduction in human disease are very inclusive, but the estimates are highly variable. These metrics allow for the direct comparison of surveillance initiatives in different phases of the food-chain on a common monetary scale but because of the difficultly in estimating costs and the variability in human disease the estimates are uncertain.

This meeting is part of the European Joint Programme One Health EJP.
This project has received funding from the European Union's Horizon 2020
research and innovation programme under Grant Agreement No 773830.

5. Case study 5 – Use of gene-sequencing technology for surveillance of antimicrobial resistance in pig production to detect changes and forecast.

The overall aim of this work was to develop a modelling framework for a holistic assessment of how a new laboratory technology could be used to improve the performance of surveillance systems before the method is actually implemented. This case study simulated the use of metagenomics in the monitoring programme of antimicrobial resistance (AMR) in the Danish pig population. The assessment of the performance was both on detecting changes in the occurrence of existing AMR and to detect the emergence of previously undetected genes coding for specific AMR.

The ability to take action on the increased occurrence of AMR is dependent on a monitoring system of AMR in humans, animals and in the environment. In Denmark, the current surveillance programme DANMAP is based on isolation of indicator bacteria from samples in humans, animals and food. These samples are tested phenotypically for AMR using minimum inhibitory concentration (MIC) panels. Currently, only *E.* coli and *Enterobacteriaceae* are used for monitoring phenotypic occurrence of AMR in the microbiome in animal populations. This limits the detection of changes in AMR to only those genes that are expressed in these indicator bacteria. AMR is expected to increase in all bacteria species, and this may not be immediately detectable with the current strategy.

Due to the rapid development of gene sequencing technology, it is now feasible to investigate the presence of all known AMR-genes in all bacteria in samples directly (metagenomics). Therefore, using metagenomics, it may be possible to detect emerging AMR in a population earlier than with phenotypic test of indicator bacteria.

The model framework developed for this case study has two dimensions: 1) the dynamic change in the occurrence of AMR genes in the population over time and 2) the monitoring procedure (sampling schedule), laboratory method, interpretation and statistical analysis of laboratory results. The simulation of the monitoring procedure included modules describing stochastic processes along the whole chain from sampling including: random selection of herds and animals, pooling of samples and thereby dilution of the genes of interest, randomness in the sequencing procedure related to how much genetic material that is sequenced. Using this detailed modelling approach, it was possible to assess how much of the uncertainty in the surveillance programme that originated from the sampling procedure and how much originated from the diagnostic procedure. This information can support surveillance design that reduced uncertainty of the surveillance results.

The data obtained from the metagenomic procedure has a quantitative format in the form of the number of genes in a sample. It is therefore possible to utilise more computationally intensive methods in the analysis of the laboratory results. In this case study, we explored different statistical analyses to objectively detect changes in the observed data, and to forecast AMR based on the trend in the historical data. Forecasting is important for the control of AMR because precautionary actions need to be implemented early before the problem becomes a burden on human health.

In the design of many surveillance programmes, decision makers must make a subjective eye-ball forecast based on graphs illustrating the historical occurrence over time. The ability for risk managers to make evidence-based decisions about how to control risks prospectively, depends on the information available. In this context objective forecasts based on the observed data are highly informative for supporting decisions. The importance of forecasting is of specific relevance for health issues as AMR, because the time between implementation of control actions and effect on AMR is relatively long, on the scale of months or years.

This case study on illustrates the use of novel surveillance approaches for AMR and uses a model-based strategy to evaluate the performance of the new approach considering the uncertainty of the model estimates and the ability of the method to forecast future changes in AMR. The ability of a surveillance programme to support forecasting is important to allow decision makers to make timely decisions about intervention to reduce AMR.

6. Conclusions

The metrics for evaluation of a surveillance programme need to be informative to the decision makers that are responsible for the design and investment choices when enacting a new surveillance programmes or changes are proposed existing initiatives. Therefore, a close dialog is needed between assessors and managers to determine the utility of measured parameters such as DALYs, cost per DALY, time to detection or the extent of an outbreak at the time of detection. It is also important to consider the dynamic nature of hazards in a population as the prevalence of disease changes over time in an outbreak or may vary in an endemic phase of disease. These changes may have a strong effect on the estimated performance of a surveillance system and must therefore be considered by using methods such as infectious disease modelling as a foundation for surveillance evaluation. A Surveillance programme also needs to be assessed holistically, considering any sources of variability that could affect surveillance performance. This should be considered to avoid drawing conclusions about the efficacy of a proposed surveillance programme that fulfils only one criterion among several competing interests or ignores uncertainty in input parameters to give a false confidence. In some cases, where the timeliness of interventions as a result of surveillance results is important, a surveillance system should be evaluated on its ability to forecast future changes in the occurrence of the hazard. Forecasting is not commonly used surveillance but improvements in computationally intensive statistical methods allows forecast with many types of monitoring data.