Deliverable

D1.2 Risk based surveillance plans, sampling algorithms and protocols for surveillance of emerging and foodborne diseases and pathogens not covered by existing surveillance.

Version: 4

Due: Month 52

Completed: Month 52

Authors: Reina Sikkemma, Lineke Begeman, Thijs Kuiken, Marion Koopmans (ErasmusMC), Lidewij Wiersma (FAO), Stephanie Lim (Artemis), Sofia Duarte, Hakan Vigre (DTU), Emma Snary (APHA).

Contents

Deliverable Description

Workpackage 1 within COMPARE focuses on Risk Assessment and Risk-Based strategies for Sample and Data Collection. Once a pathogen, disease, or outbreak is detected, population-level analyses of next generation sequencing (NGS) data can provide important insights into the evolution of outbreaks, sources of infection, modes of transmission, changes in virulence and transmission dynamics, and effects of control measures. Hence, for an optimal outbreak risk assessment, the probability of occurrence is predicted, surveillance is designed based on these predictions, and methods are developed to prepare for outbreak investigations to determine its size, the affected population, the potential for further spread, and the (potential) modes of transmission. As part of this work, sampling strategies are required that capture the essential specimens and metadata.

Deliverable report 1.2 provides a summary of activities relating to the following tasks, which aimed to develop risk-based sampling and data collection strategies for early detection and investigation of unusual patterns of infectious disease outbreaks:

Task 2.1: To develop risk-based sampling for unusual clinical symptoms in humans and domestic animals

Task 2.2: Targeted sampling for early detection of emerging and re-emerging infections coming from wild or feral animals

Task 2.3: To develop risk-based sampling algorithms and protocols for detection of human pathogen circulation in the absence of recognized illness

Task 2.4: To develop food-level sampling strategies for surveillance as well as foodborne outbreak investigation

Due to synergies between Tasks 2.1-2.3, it was possible to develop a harmonised sampling strategy for outbreak situations in humans, domestic animals and free-ranging animals. Tools have been developed to assist the sampling and information collection process in the case of a suspicion / knowledge of the specific pathogen or not (i.e. a syndrome-based sampling protocol). Such tools can be used to guide healthcare workers during an outbreak; ensuring that, wherever possible, sampling is both science based and cost effective. The syndrome-based sampling protocol is currently being used in the SEVTRAV Study, which is a GEoSentinel collaboration investigating severe, undiagnosed infections in returning travellers. In addition, the sampling tools have been shared with the GOARN network for use during the Ebola outbreak. There is also interest from the FAO to use the tools as part of an app for sampling in domestic animals.

In order to help focus food sampling in the early stages of a foodborne outbreak and to facilitate international collaboration, a searchable data catalogue of past foodborne outbreaks has been created and is publicly available in a virtual research environment named Global Foodsource Identifier (GFI). Such a platform will encourage harmonised, international data sharing among laboratories. The data catalogue in GFI has so far been populated with more than 100foodborne outbreaks, including most of the frequent foodborne pathogens and a broad range of typing methods.

Taking this work forward, at least four publications are planned and further interest from international organisations/consortiums is anticipated.

Harmonised sampling for the detection of emerging infectious diseases

Overview

A large part of emerging infectious diseases in humans is of zoonotic origin. Early warning and surveillance of (re-)emerging infectious diseases in humans and animals is vital for detection and timely control of outbreaks. Correct sampling is essential for proper laboratory testing and epidemiological analyses in an outbreak investigation.

We attempt to make a harmonised sampling strategy for outbreak situations in humans, domestic animals and free-ranging animals that could guide healthcare workers during an outbreak to make sampling more cost-effective and science-based.

Tools

Several tools were developed to assist the sampling and information collection process in case of (increased risk of) infectious disease outbreaks in humans, domestic animals and/or free-ranging animals.

First, a **sampling flowchart** (**ANNEX 1**) was developed, covering the appropriate steps for harmonised sampling for the detection of emerging infectious diseases in case of four different scenarios:

- Scenario 1: Increased morbidity (in humans, domestic animals and free-ranging animals)
- Scenario 2: Increased mortality (in humans, domestic animals and free-ranging animals)
- Scenario 3: Sudden decrease in population size (in free-ranging animals)
- Scenario 4: Increased risk of disease emergence or introduction (in humans, domestic animals and free-ranging animals)

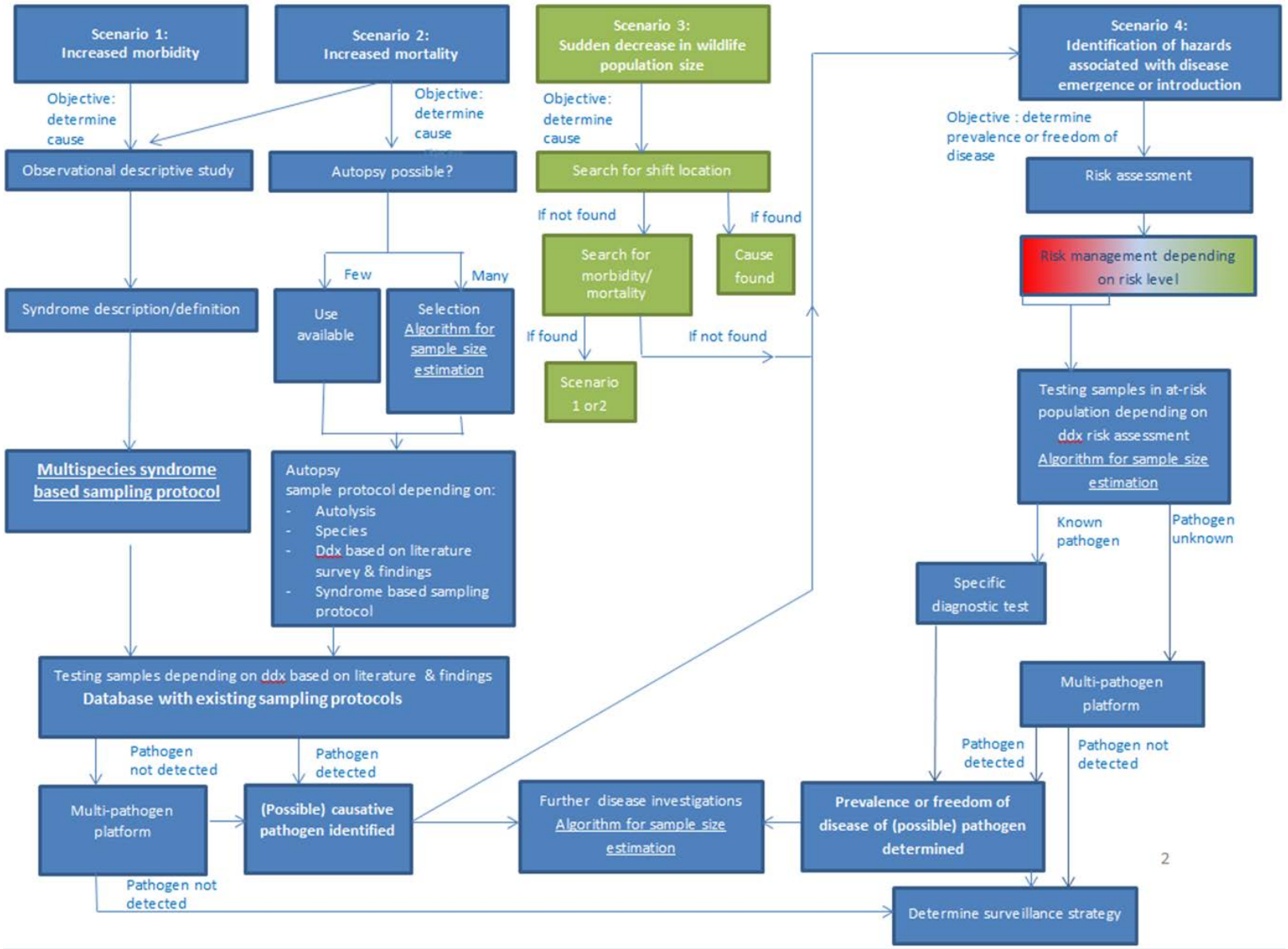
When there is a suspicion or clear indication of infection with specific pathogens, our **database with existing sampling protocols** can be used. It includes diseases of veterinary and public health importance:

https://www.compare-europe.eu/Library/Epidemiological-Datasets

When there is no suspicion or clear indication for a specific pathogen, we propose sampling could be standardized by using our **multispecies syndrome-based sampling protocol** (**ANNEX 2**) that is based on syndromes. We have tried to cover all possible disease outbreaks in 13 syndromes. These syndromes match with the database with existing sampling protocols. When the researcher arrives at the stage of sample collection, our **list of metadata** (**ANNEX 3**) can be used to assist and standardise the data that should accompany the sample so that it can be used for later analyses.

A background document (**ANNEX 4**) provides additional explanation on the use of the sampling flowchart.

ANNEX 1: Sampling flowchart for harmonised sampling for the detection of emerging infectious diseases.



ANNEX 2: Multispecies syndrome-based sampling protocol for infectious disease outbreaks of unknown origin.

Each sample set should include the following:

1. Informed consent (owner) or permission to sample free-ranging animals
2. Completed **list of metadata**
3. Full blood (EDTA tube) and serum (serum tube) at admission/ capture (if possible in the acute phase of disease) and > 2 weeks later (or at discharge/ release)
4. Biopsies of organs are very useful for later analyses. These can be stored partly in formalin for histopathology and partly in freezer for NGS.

	Nasopharyngeal swab [d]	Bronchoalveolar lavage fluid (BALF) [e]	Stool [e]/ rectal swab / cloacal swab [a,d]	CSF [b,e]	Urine [e]	Amniotic fluid, placenta, fetal tissue, ora [e]	Ocular swab [d]	Vesicular fluid [e], crust, lesion swab (vesicular exudate) [d]	Milk [e]
Gastrointestinal	+		++						
Respiratory	++	+	+		+				
Neurological	++		++	++	+		+		
Hemorrhagic	+		+		+				
Weight loss			++		++				
Cutaneous and mucosal membrane	++		++					++	+
Icterus	+		++		++				
Musculoskeletal									
Reproductive and teratogenic						++			
Urinary					++				
Ocular							++		
Multisystemic	++	+	++	+	+				
Mortality [c]	++		++	+	++				

++=minimum required; += preferable

a. Stool is the preferred sample. If this is not possible, a swab can be collected.
b. If the CSF sample is not taken on the day of admission, an additional serum sample should be taken at the same date as the lumbar puncture to obtain CSF.
c. Sample set depends on pathology results.

d. Three swabs with plastic shaft and polyester/ dacron tip, to be stored in three cryovials filled with 500 µL of VTM, 500 µL of Trizol and 500 µL of a 15% glycerol solution, respectively.
e. Three cryovials with 500 µL of fluid (in case of feces: pea size/200 mg) with 500 µL TRIzol, 500 µL of VTM and without medium (500µL), respectively.

ANNEX 3: List of metadata to be collected at the time of sampling[a].

Priority[b]	Event data	
++	Reason for sampling: Diagnostic/ Surveillance/ Outbreak/ Other	
+	Outbreak: - Related to event (event identifier) - Number of humans/animals involved - Site description	
Data associated to the sample		
++	Human/animal ID number / Food item ID / Water ID / Other ID	
++	Animal species (if applicable)	
++	Sample ID	
++	Sampled organ and tissue (if applicable)	
++	Date of sample collection (Day-Month-Year)	
++	Location of sampling: as specific as possible, minimum is province level - Site latitude and longitude (decimal degrees, example: 53.18333, 6.13333)	
+	Specimen medium or preparation: - No medium/ TRIzol/ VTM incl. reference to recipe/ 10% buffered formalin / Lysis buffer/ PBS/ Antibiotics/ Other - None / EDTA/ Sodium Citrate/ CTAD/ Lithium/Sodium Heparin / Sodium Fluoride / Acid Citrate Dextrose / other	
+	Specimen container: - 2ml cryovial/ Plastic container/ Vacutainer/ Filter paper/ Other	
+	Sample volume	
+	Transport temperature: - Room temperature/ +4°C/ On ice/ -20°C/ Dry ice/ Liquid Nitrogen/ -80°C	
+	Transport time	
+	Storage temperature: - Room temperature/ +4°C/ On ice/ -20°C/ Dry ice/ Liquid Nitrogen/ -80°C	
+	Storage time	
+	Reference to the standardized protocol (SOP) used for sampling, if available	
+	Freshness/quality of sample at the moment the sample was taken	
	Other, namely:	

Contextual metadata covering "clinical, microbiological, epidemiological, and other data"		
	Clinical data:	
+	Healthcare setting (e.g. General Practitioner, Outpatient department, Hospital, Intensive Care Unit); Domestic animals: species, production sector, husbandry, number of animals on farm, age categories present on farm etc.; Free ranging animals: all observations that lead to the investigation, species involved, population density, recent changes in population density, possibility of contacts with people and domestic animals.	
+	Age (category) and gender of donor of the sample	
++	Clinical syndrome(s): cutaneous and mucosal membrane, ocular, musculoskeletal, weight loss, respiratory, gastrointestinal, neurological, reproductive and teratogenic, urinary, haemorrhagic, icterus, multisystemic, mortality Report ranked clinical signs in order of frequency of appearance and provide details on the number of humans/animals affected and dead at the time of sampling, split by age category if possible (along with the total population size)	
+	Comorbidities; (unknown/no/yes; if yes, please list)	
++	Date of onset syndrome (Day-Month-Year)	
+	Outcome of disease (cure without sequelae, cure with sequelae, death, unknown)	
+	History of travel; international movements of humans, animals or germplasm (unknown/no/yes; if yes, please list country/countries)	
++	History of vaccination; (unknown/no/yes; if yes, name vaccination(s)	
++	Antimicrobial treatment, immunosuppressive drugs or other drugs at time of sample collection (unknown/no/yes; if yes, please list)	
+	History of contacts with other species (domestic and wild) (unknown/no/yes; if yes, name species)	
	Provider related data:	
++	Identifier and contact details of provider;	
	Other, namely:	

a. The list is compatible with GSC minimum information standards (http://gensc.org/mixs/) and EBI-standards (ftp://ftp.sra.ebi.ac.uk/meta/xml/checklist.xml)
b. ++=minimum required; += preferable

ANNEX 4: Harmonised sampling for the detection of emerging infectious diseases: definitions, explanatory document.

1. Introduction

This document serves as explanatory document for the use of the sampling flowchart for outbreaks of unknown disease and increased risk of such outbreaks. The flowchart was developed by the COMPARE consortium. This document provides an explanation of the terms used in the flowchart and links to existing tools and protocols to aid the user in the application of appropriate sampling strategies under different circumstances.

2. Scenario 1 and 2: Increased morbidity and/or increased mortality

In case of mortality, both autopsies and an observational descriptive study assisting in carcass selection, including a syndrome description and sampling according to the multispecies syndrome based sampling protocol, should be executed.

Observational descriptive study
An observational descriptive study[1] should include a description of the disease, as well as a description of the affected humans or animals, place and time of disease occurrence. This can be summarised as the "what", "who", "where" and "when" .

What (outcome):
- Morbidity, mortality, or both
- In case of morbidity: disease signs

Who (affected humans/animals)
- Number of cases, percentage of individuals affected
- Species
- Age, sex
- Immune status/comorbidities
- Use of medication
- Vaccination status
- Activities, such as occupation, leisure activities, use of tobacco/drugs (humans)
- Husbandry system (domestic animals and captive wildlife)
- Stage of breeding or migratory cycle (domestic animals and wildlife)

Where (place)
- Location of cases
- Unusual environmental conditions

When (time)
- Dates onset disease, duration, disease progression

Syndrome description/definition
A case definition is important to focus the investigation. It should include information on the clinical features and key characteristics that all (human or animal) patients have in common. It can also include time and place, to limit the case definition to the outbreak that is to be investigated.[2]

To be able to use the syndrome based sample protocol, symptoms should be grouped under the following syndromes:
- Cutaneous and mucosal membrane

- Ocular
- Musculoskeletal
- Weight loss
- Respiratory
- Gastrointestinal
- Neurological
- Reproductive and teratogenic
- Urinary
- Haemorrhagic
- Icterus
- Multisystemic
- Mortality

Multispecies syndrome-based sampling protocol
The sampling protocol can be used for infectious disease outbreaks of unknown origin in humans, domestic animals and wildlife. For each syndrome the minimum and desirable dataset is outlined. The appropriate metadata should be collected at the time of sample collection (see COMPARE metadata form)

Autopsy
Autopsies should be executed according to protocols of the institute involved, with the expertise of people involved. Recommended necropsy protocols for different species are included in the **database with existing sampling protocols.**

Testing samples depending on differential diagnosis (ddx)
The COMPARE Consortium developed a database with sampling protocols for a large number of infectious diseases of humans, domestic animals and free-ranging animals. They include human infectious reportable diseases for which surveillance exists in the European Union (ECDC regulation 851/2004 and decision 1082/2013/EU), OIE listed notifiable diseases (resolutions 72 GS/FR 2004), infectious diseases specifically covered by ECDC/CDC/WHO, other diseases of importance to animal health and important zoonotic diseases and diseases that are not zoonotic but have a common source that are not covered by EU-, OIE-, CDC- and IHR regulations (e.g. Histoplasma from the environment). The database is searchable for host species, disease name, primary syndromes, other syndromes, pathogen type and pathogen species. The inventory does not include non-infectious diseases or toxins. The COMPARE multispecies syndrome based sampling protocol for infectious disease outbreaks of unknown origin can be used for diseases of unknown origin.

Multi-pathogen platform
Diagnostic tests or platforms that can detect multiple pathogens simultaneously. Examples are: gross autopsy, histopathological analysis, multiplex PCR systems, multiplexed antibody array technologies, pan-viral microarrays and metagenomics technologies.

Further disease investigations after pathogen has been identified
Knowledge on the identified pathogen, combined with the data collected during the initial observational descriptive analysis, risk assessment and sampling can be used to generate further hypotheses on prevalence, risk factors and mechanisms of disease. Studies to test these hypotheses include experimental and observational studies.[3] Also, sequence information of the causative pathogen can be further studied to learn more about its characteristics.

If the identified pathogen is zoonotic, or if other animal groups could be at risk, scenario 4 should be followed for the human and/or animal population at-risk.

3. Scenario 3: Sudden decrease in wildlife population size

Shift location free ranging animals

Any geographical change of the population, which can cause a decrease in a population at a certain location, without any (abnormal) mortality having occurred here.

4. Scenario 4: Identification of hazards associated with disease emergence or introduction

Scenario 4 comprises several scenarios that differ per sector

General[4]	Proximate drivers (direct determinants of changes in human, animal reservoir, and vector population dynamics) o Movement/migration o Habitat change o Food and water change Ultimate drivers (occurring at broader (regional or global) geographic scales temporally precede and govern changes in proximate drivers) o Climate o Land use change o Changes in animal management.
Humans	o Increase or introduction of vectors o Increased contact with possible reservoir species (Land use change, food, etc.) o Increased travel/trade with endemic area/area with infectious diseases prevalent that are not (yet) introduced o Suspicion due to morbidity / mortality in wildlife/ domestic animals
Domestic animals	o Increase population size or introduction new species o Increase or introduction of vectors o Change in husbandry or housing o Increased live animal trade/direct contact with animals from risk countries* o Suspicion due to morbidity / mortality in humans / wildlife *countries with endemic infectious diseases that are not introduced in another country yet, but could become endemic
Free-ranging animals	o Increase population size or introduction new species o Increase or introduction of vectors o Suspicion due to morbidity / mortality in humans/ domestic animals

The cut-off point for action is determined by the national responsible institutes and is not included in this flowchart.

Risk assessment

The WHO defines risk assessment as a systematic process for gathering, assessing and documenting information to assign a level of risk.[5] The risk assessment following scenario 4 specifically assesses the likelihood of entry, establishment or spread of a pathogenic agent within the territory of an importing country.[6] Following a risk

assessment, an impact or consequence assessment is usually also performed. Guidance on risk assessment and risk analysis by WHO and FAO is available here:

- FAO Good Emergency Management Practices: The Essentials
- WHO Rapid Risk Assessment of Acute Public Health Events
- WHO Risk assessment for human exposure to foodborne hazards

Algorithm for sample size estimation

If you need assistance in sample size estimations for sampling domestic animals and humans, we recommend using existing online tools such as Epitools and sample-size.net. Such tools are less suitable for wildlife sampling because wildlife distribution, population size and health status are usually largely unknown. We aim to design a sampling strategy that can be applied for two scenarios. The first is a morbidity or mortality event in wildlife. The second scenario is that wildlife are identified as a potential hazard associated with observed or predicted disease emergence in humans and/ or domestic animals. For each of these scenarios we aim to offer steps and guidance for developing a sampling strategy. [in preparation]

5. Syndrome-based sampling

Sampling techniques

Sampling should be executed as soon as possible after onset of symptoms. A convalescent serum should be collected > 2 weeks later. It is important that sterile dedicated plastic-ware is used for sample collection and gloves are worn during all stages of sample handling.

Appropriate sampling techniques for humans and animals have been described previously by WHO and OIE:

- WHO Guidelines for the collection of clinical specimens during field investigation of outbreaks, 2000
- OIE Manual of Diagnostic Tests and Vaccines for Terrestrial Animals 2013: Chapter 1.1.2. Collection, submission and storage of diagnostic specimens
- U.S. Geological Survey, U.S. Fish and Wildlife Service, and National Park Service, Wildlife Specimen Collection, Preservation, and Shipment

Storage and Transport

Note that samples in VTM have the most stringent cold chain requirements; these should be put in liquid nitrogen, a dry shipper or a -80°C freezer as quickly as possible following sampling. Failure to do so will limit the possibility of virus isolation and also decrease the quality of the sample for PCR. Serum samples can be kept at +4°C for up to a week and samples in TRIzol for up to 48 hours. Stool specimens can be kept at +4°C for 1-2 days.

Transport of specimens within national borders should comply with applicable national regulations. International transport should follow applicable international regulations, as has been described by WHO and OIE guidance. Correct labelling and appropriate shipping documents are vital. Always contact the receiving laboratory before shipping samples.

- WHO Guidance on regulations for the Transport of Infectious Substances for category A infectious substances
- OIE Transport of specimens of animal origin

- U.S. Geological Survey, U.S. Fish and Wildlife Service, and National Park Service, Wildlife Specimen Collection, Preservation, and Shipment

Environmental sampling

When it is not feasible to capture animals for sampling, collection of environmental samples, such as water or feces, may also provide insight into the causative agent of disease. For humans, it may also prove to be difficult to collect the appropriate patient samples. Environmental samples can provide a solution in some cases. Sewage or wastewater sampling can be very informative, as well as air sampling. Additionally, evidence of a pathogen in environmental specimens, food or on fomites may provide evidence of the source of infection.

Ethics

Sampling humans and animals should comply with national and international regulations such as the obligation for ethical approval and data sharing, animal welfare laws and privacy legislation. Moreover, many ethical considerations. Guidance from the WHO on ethical issues in infectious disease outbreaks is available here:

- WHO Guidance for Managing Ethical Issues in Infectious Disease Outbreaks

6. References

(1) https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section6.html, 01-02-2018
(2) https://www.cdc.gov/urdo/downloads/casedefinitions.pdf, 01-02-2018
(3) https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section7.htm l, 01-02-2018
(4) Gortazar et al., Crossing the interspecies barrier: opening the door to zoonotic pathogens., PLoS Pathog., 2014, doi: 10.1371/journal.ppat.1004129
(5) WHO, Rapid Risk Assessment of Acute Public Health Events, 2012, http://apps.who.int/iris/bitstream/handle/10665/70810/WHO_HSE_GAR_ARO_2012.1_eng.pdf?sequence=1, 5-4-2018
(6) Import Risk Analysis, Part 1 Section 1.4 of the International Animal Health Code Articles 1.4.2.3. and 1.4.2.4., http://www.fao.org/docrep/003/x7354e/x7354e12.htm, 5-4-2018

Global Foodsource Identifier: A database for foodborne outbreaks to assist the development of sampling strategies in food.

Introduction

An open, searchable data catalogue of past foodborne outbreaks (FBO) has been created, integrated in a virtual research environment named Global Foodsource Identifier (GFI). The objectives of this resource are to help focus food sampling in the initial investigation of future foodborne outbreaks based on information of previous outbreaks with a similar causative agent, to host harmonised, international data sharing among laboratories and to facilitate international collaboration by providing a channel for communication.

Before the creation of the data catalogue, a dataset was built that collated records from previous foodborne outbreaks.

Based on scientific literature on outbreak investigation and discussions within a working group, we decided to include variables in the dataset that are relevant for FBO investigators in an outbreak situation and/or are usually presented in reports of FBO. The number of fields in the dataset are as parsimonious as possible, and fields are distributed under the categories *Causative agent*, *Epidata*, *Food source* and *Report details*.

In the category *causative agent*, several considerations were taken, to ensure transparency and harmonisation of reporting, and subsequent comparability of records:

- A dropdown menu to choose among the most common typing methods used for a particular causative agent.
- For each typing method, a dropdown menu for the most common protocols used.
- Whenever possible, a default nomenclature/syntax format to report the result of a specific typing method. For PFGE and WGS there is no default format, but the user can report whether the outbreak isolate/strain was typed using the method.

Population of the data catalogue

The data catalogue was initially populated with outbreak data described in the annual Danish Zoonoses Report from years 2005 to 2016. Additionally, data were extracted from the National Database of Foodborne Outbreaks (Fødevareudbrudsdatabase - ***FUD***) and specific scientific literature describing the outbreaks in more detail.

We intended to populate the catalogue with approximately 100 outbreaks covering the most frequent pathogens and a broad range of typing methods. This was done to make sure that it encompasses options for as many typing methods as possible, and especially those used for common foodborne pathogens.

Food sources were categorized based on a food categorization scheme from the Interagency Food Safety Analytics Collaboration (IFSAC), with 234 food categories, which takes into account the implication of different processing, preparation and consumption types on the microbiological environment in the food.

For several of the fields, drop-down menus were created to ease the reporting process and to avoid erroneous entries. When reporting data, food sources shall be assigned to the most specific hierarchical food category possible.

So far, the dataset has been populated with 102 foodborne outbreaks from Denmark (Table 1), covering the most frequent pathogens and a broad range of typing methods. *Salmonella* enterica, Norovirus and ETEC were the most common causative agents. Serotype, MLVA profile and genogroup were the most frequently used typing methods (Table 2).

Table 1: Descriptive statistics of outbreaks, outbreak strains and cases per year.

Year	Outbreaks	Outbreak strains	Cases
2005	10	10	1263
2006	5	6	533
2007	4	23	429
2008	7	13	1825
2009	6	6	786
2010	25	54	1091
2011	5	5	4024
2012	9	10	546
2013	8	14	729
2014	7	7	525
2015	6	6	296
2016	10	12	1419
Total	102	166	13466

Table 2: Descriptive statistics of causative agents and typing results.

Causative agent		Sero-type	PFGE typed	Phage-type	MLVA profile	MLST	AMR profile	Viru-lence profile	Geno-group	Was the isolate sequenced?
Campylobacter	3									1
C. jenuni	6					2	1			3
C. perfringens	2									
C. hominis	1									1
EPEC	3	3						1		
ETEC	**22**	21								
Other pathogenic										
E. coli	2						1	1		
VTEC/STEC	8	6	1			3		5		3
E. histolytica	1									
G. intestinalis/										
lamblia/duodenalis	1									
Hepatitis A virus	2									
L. monocytogenes	4	1	2			2				2
Norovirus	42								27	22

Causative agent		Sero-type	PFGE typed	Phage-type	MLVA profile	MLST	AMR profile	Viru-lence profile	Geno-group	Was the isolate sequenced?
Other parasitic agents	3									
Rotavirus	2									
S. enterica	54	54	7	25	37	1	30			15
Sapovirus	1									
S. sonnei	4		1				1			
S. aureus	1									
Streptococcus	1		1							
Y. enterocolitica	3	1								
Total	166	86	12	25	37	8	33	7	27	47

The pilot project

Partners in COMPARE WP1, -4 and -7 were invited in January 2018 to contribute with foodborne outbreak data from their own countries and to provide feedback on the datatset's structure. For this, a description of the variables and a template to report the data were provided.

The Dutch National Institute for Public Health and the Environment (RIVM) and the French Research Institute for the Exploitation of the Sea (Ifremer) provided comments and proposals for improvement, listed below. Additionally, Ifremer provided data on three outbreaks of Norovirus in molluscs.

- Provide a definition of a foodborne outbreak*; (this comment has been addressed in the wiki page)*
- There is no case for number of people exposed*; (this comment has not been addressed)*
- Consider synchronizing the value list for types of food with the RASFF list*; (this comment has not been addressed)*
- Only very few countries are able to report number of cases or many of the other epi data; *(this comment has been addressed – countries who can report it, should have the opportunity to do so)*
- For norovirus and hepatitis A (HAV) virus the typing information does not necessarily relate to the food type, as those are considered contaminants *(this comment has been addressed – the data catalogue does not imply any causality between food source and agent, and we expect the users to be experts in outbreak investigation and therefore to know that norovirus and HAV are contaminants)*
- The information whether a person is infected with one or more than one genotype/sub-type could be relevant information for norovirus: Infection with multiple strains is an indication of contact with food or water with sewage contamination*. (this comment has been addressed – multiple agents for the same outbreak can be reported in separate records)*
- For norovirus and HAV a database of only confirmed foodborne cases would not be informative as often foodborne outbreak are not recognized/reported as such. We use NoroNet and HAVNet to find out whether infections with an unknown source might be linked to others using the sequences. Next step is to find out whether transmission could have been person to person or perhaps via a common (imported) food source. So we could not use a GFI database for that. Perhaps this approach works better for bacteria? The main difference is in the epidemiology: HAV and norovirus are human pathogens; the food is mainly a vehicle, directly or indirectly contaminated by an infected person, whereas many bacteria are zoonotic agents. Thus, outbreak tracing needs a completely different

approach. *(this comment has been addressed – the data catalogue does not imply any causality between food source and agent, and we expect the users to be experts in outbreak investigation and therefore to know that norovirus and HAV are contaminants; the records can be linked to other information sources, such as WG sequences of the outbreak agent(s))*

Global Foodsource Identifier (GFI) – a Virtual Research Environment for Outbreak Investigators.

A virtual research environment (VRE) to host the data catalogue was created with the technical support of the project AGINFRA+ (http://plus.aginfra.eu/), which focuses on supporting the development of cloud-based infrastructures for communities around agriculture and food.

We migrated the data from the spreadsheet format into the catalogue format. In the final version of GFI, users will report outbreak data into the data catalogue. The full content of the catalogue will be regularly and automatically exported into a spreadsheet located in GFI's shared workspace. Data searches can be performed both inside of the catalogue and on the spreadsheet with the full list of all records. GFI incorporates different statistical and graphical analysis tools, such as R studio; making it possible to the users to conduct and share exploratory data analysis directly within the platform using the spreadsheet. Additionally, it includes a wiki page with detailed descriptions of the variables present in the data catalogue and instructions to the users on how to report data. Users can post comments on a shared environment and communicate with each other within GFI.

Currently, GFI's data catalogue is accessible in "read only" mode for guest users at https://aginfra.d4science.org/web/foodborneoutbreak.

Users who wish to become members of the VRE, with access to all its features and the opportunity to report data into the catalogue, must follow the following steps:

1. visit https://aginfra.d4science.org/group/foodborneoutbreak
2. create an account to register as user
3. request GFI's Admin to join as a member

The policy currently governing GFI membership is "Restricted", i.e. users having an account can request to join and the VRE Admin can approve or reject. Note that the policy can be easily switched to "Open", i.e. any user willing to access the VRE is allowed to.

A scientific article describing GFI has been drafted and will be ready for submission in the second half of 2019, along with the public launch of GFI.